

Understanding Complex Systems

Springer :
COMPLEXITY

Charles L. Webber, Jr.
Norbert Marwan *Editors*

Recurrence Quantification Analysis

Theory and Best Practices



Springer

Springer Complexity

Springer Complexity is an interdisciplinary program publishing the best research and academic-level teaching on both fundamental and applied aspects of complex systems-cutting across all traditional disciplines of the natural and life sciences, engineering, economics, medicine, neuroscience, social and computer science.

Complex Systems are systems that comprise many interacting parts with the ability to generate a new quality of macroscopic collective behavior the manifestations of which are the spontaneous formation of distinctive temporal, spatial or functional structures. Models of such systems can be successfully mapped onto quite diverse “real-life” situations like the climate, the coherent emission of light from lasers, chemical reaction-diffusion systems, biological cellular networks, the dynamics of stock markets and of the internet, earthquake statistics and prediction, freeway traffic, the human brain, or the formation of opinions in social systems, to name just some of the popular applications.

Although their scope and methodologies overlap somewhat, one can distinguish the following main concepts and tools: self-organization, nonlinear dynamics, synergetics, turbulence, dynamical systems, catastrophes, instabilities, stochastic processes, chaos, graphs and networks, cellular automata, adaptive systems, genetic algorithms and computational intelligence.

The three major book publication platforms of the Springer Complexity program are the monograph series “Understanding Complex Systems” focusing on the various applications of complexity, the “Springer Series in Synergetics”, which is devoted to the quantitative theoretical and methodological foundations, and the “Springer Briefs in Complexity” which are concise and topical working reports, case-studies, surveys, essays and lecture notes of relevance to the field. In addition to the books in these two core series, the program also incorporates individual titles ranging from textbooks to major reference works.

Editorial and Programme Advisory Board

Henry Abarbanel, Institute for Nonlinear Science, University of California, San Diego, USA

Dan Braha, New England Complex Systems Institute and University of Massachusetts Dartmouth, USA

Péter Érdi, Center for Complex Systems Studies, Kalamazoo College, USA and Hungarian Academy of Sciences, Budapest, Hungary

Karl Friston, Institute of Cognitive Neuroscience, University College London, London, UK

Hermann Haken, Center of Synergetics, University of Stuttgart, Stuttgart, Germany

Viktor Jirsa, Centre National de la Recherche Scientifique (CNRS), Université de la Méditerranée, Marseille, France

Janusz Kacprzyk, System Research, Polish Academy of Sciences, Warsaw, Poland

Kunihiko Kaneko, Research Center for Complex Systems Biology, The University of Tokyo, Tokyo, Japan

Scott Kelso, Center for Complex Systems and Brain Sciences, Florida Atlantic University, Boca Raton, USA

Markus Kirkilionis, Mathematics Institute and Centre for Complex Systems, University of Warwick, Coventry, UK

Jürgen Kurths, Nonlinear Dynamics Group, University of Potsdam, Potsdam, Germany

Andrzej Nowak, Department of Psychology, Warsaw University, Poland

Linda Reichl, Center for Complex Quantum Systems, University of Texas, Austin, USA

Peter Schuster, Theoretical Chemistry and Structural Biology, University of Vienna, Vienna, Austria

Frank Schweitzer, System Design, ETH Zurich, Zurich, Switzerland

Didier Sornette, Entrepreneurial Risk, ETH Zurich, Zurich, Switzerland

Stefan Thurner, Section for Science of Complex Systems, Medical University of Vienna, Vienna, Austria

Understanding Complex Systems

Founding Editor: S. Kelso

Future scientific and technological developments in many fields will necessarily depend upon coming to grips with complex systems. Such systems are complex in both their composition – typically many different kinds of components interacting simultaneously and nonlinearly with each other and their environments on multiple levels – and in the rich diversity of behavior of which they are capable.

The Springer Series in Understanding Complex Systems series (UCS) promotes new strategies and paradigms for understanding and realizing applications of complex systems research in a wide variety of fields and endeavors. UCS is explicitly transdisciplinary. It has three main goals: First, to elaborate the concepts, methods and tools of complex systems at all levels of description and in all scientific fields, especially newly emerging areas within the life, social, behavioral, economic, neuro- and cognitive sciences (and derivatives thereof); second, to encourage novel applications of these ideas in various fields of engineering and computation such as robotics, nano-technology and informatics; third, to provide a single forum within which commonalities and differences in the workings of complex systems may be discerned, hence leading to deeper insight and understanding.

UCS will publish monographs, lecture notes and selected edited contributions aimed at communicating new findings to a large multidisciplinary audience.

For further volumes:

<http://www.springer.com/series/5394>

Charles L. Webber, Jr. • Norbert Marwan
Editors

Recurrence Quantification Analysis

Theory and Best Practices

Editors

Charles L. Webber, Jr.
Department of Cell and Molecular
Physiology
Loyola University
Maywood, Illinois, USA

Norbert Marwan
Transdisciplinary Concepts and Methods
Potsdam Institute for Climate Research
Research Domain
Potsdam, Germany

ISSN 1860-0832

ISBN 978-3-319-07154-1

DOI 10.1007/978-3-319-07155-8

Springer Cham Heidelberg New York Dordrecht London

ISSN 1860-0840 (electronic)

ISBN 978-3-319-07155-8 (eBook)

Library of Congress Control Number: 2014946073

© Springer International Publishing Switzerland 2015

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

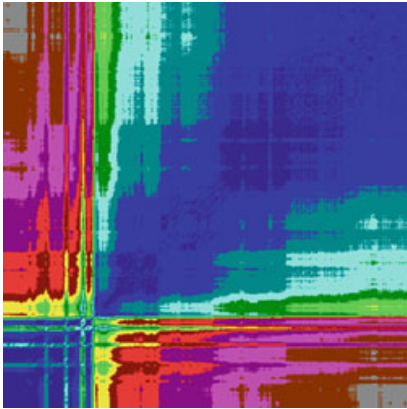
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

*In memory of Joseph P. Zbilut, esteemed
colleague, honest critic, and faithful friend
whose insights into the recurrence properties
of nonlinear dynamical systems have inspired
fruitful contributions across scientific
disciplines as testified to herein.*

Preface



This book has been almost one decade in preparation from original thoughts to finalized compilations (hardcover and e-book formats). Back in 2007 Marwan et al. published *Recurrence Plots for the Analysis of Complex Systems* in Physics Reports. Preceding that in 2005 Joe Zbilut and Chuck Webber published a comprehensive review *Recurrence Quantification Analysis of Nonlinear Dynamical Systems* as part of an e-book for the National Science Foundation (NSF) edited by Michael Riley and Guy Van Orden. Our contribution already had the feel of a book within a book for it captured the collaborative work of Zbilut and Webber over the previous 15 years. Figure 1 shows an early photo of the complimentary researcher pair from days gone by.

Shortly after their NSF chapter appeared Zbilut and Webber started thinking more seriously about writing their own book that went beyond what they had written for NSF. Drafts of several chapters were generated, but attempts to tie the whole package together were slowed by two facts: the first exciting; the second tragic. At the turn of the new millennium it was becoming more and more obvious that recurrence plot strategies and recurrence quantifications were growing in popularity and utility across wider and more numerous scientific disciplines. Substantial progress with new methodological procedures and theoretical advances was being made. Norbert Marwan started counting these references and posting them on his website which are captured in graphic form in Fig. 2. It was very satisfying to actually see and experience the spread of recurrences (can it be called viral?) across broad fields of inquiry. But this situation introduced a new and serious problem. How could any two authors capture the breadth and depth of this dynamically changing field and do any justice to it?



Fig. 1 Chuck Webber (*left*) and Joe Zbilut (*right*) in Chuck’s office at Loyola University Chicago, Stritch School of Medicine (January 7, 1992)

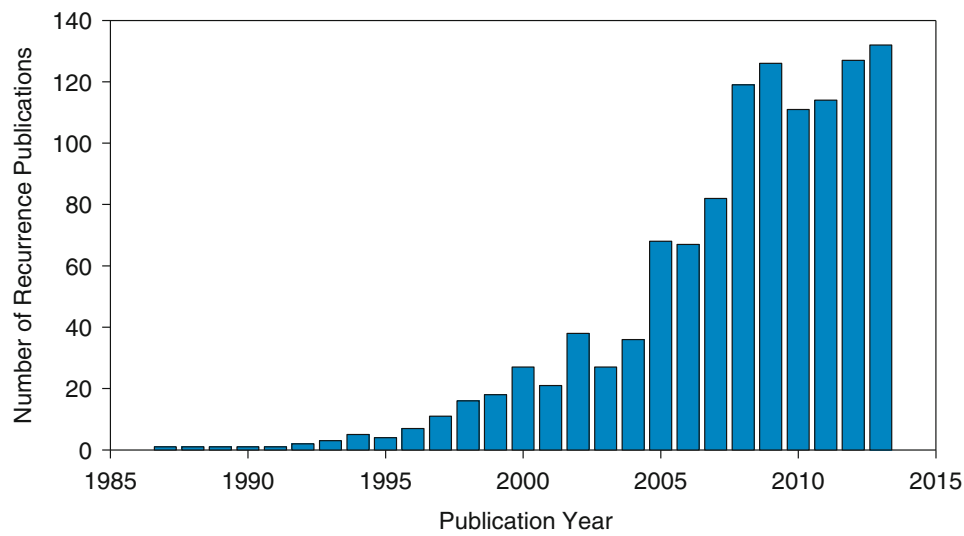


Fig. 2 Growth of professional publications applying recurrence plots (RP) or recurrence quantifications (RQA) to specific scientific problems, theoretical and practical. Note the accelerated growth in 1995 and even steeper increase in 2005

On the sad side of life, one cold and snowy Saturday afternoon in Chicago Joe Zbilut suffered a fatal heart attack at his home after shoveling snow for the second time that day (January 10, 2009). Just the day before Chuck had received an e-mail message from Joe, “How’s the book coming?” to which the honest reply was given, “slowly but surely.” This tragic death of a trusted friend and faithful colleague at the age of 60 years came as such a shock that thoughts of writing a new RQA book literally vaporized.

In the years to come, the memory of Joe was honored in the literature as well as at biannual International Symposia on Recurrence Plots (Montreal, 2009; Hong Kong, 2011; Chicago, 2013). Joe had faithfully attended the two previous symposia (Potsdam, 2005; Siena, 2007) and had served on the organizing committees for both. So more recently with hope restored Webber reached out to Marwan and suggested that we co-edit a book on recurrences which described both theoretical matters and practical. We put our heads together and invited several accomplished authors to join us in this massive project. Our responses were affirmative and confidence building. This modern book on recurrences was indeed possible! Hope was renewed.

The book that readers now hold in hand represents the combined efforts of 39 different authors writing 14 distinct and unique chapters. The trajectory in completing this project has been very nonlinear with unexpected twists and turns along the way. Stuttering pauses (singularities?) with dynamical noise have been endured, escaped, and tamed. But in the end the finished project is written with so much depth and breadth that no one person on the planet could have amassed such recurrence knowledge in one place! So to all the authors who contributed to this book, Chuck Webber and Norbert Marwan give a big “Thank you” with deep gratitude and sincere appreciation. Any success this work will bring to the field of nonlinear dynamics is totally and truly the success of many experts. As a corporate group, we are all very proud.

To distinguish and honor the various contributing authors, each chapter has been tagged with a unique identifying icon of a favorite recurrence plot from each research group. To represent the unity of all authors, these icons have been arranged into a super recurrence-plot quilt as shown in Fig. 3. The $[6, 6]$ symmetrical matrix has 16 different icons (one per chapter, preface, and ubiquitous line of identity). For fun, which one recurrence plot icon is actually a cross-recurrence plot?

As promised in the title, this book is divided into two parts. Part I describes the theoretical details of recurrence plots and quantifications including: mathematical and computational underpinnings (Chap. 1); Kolmogorov entropy estimations (Chap. 2); identification of coupling directions (Chap. 3); and complex network analyses (Chap. 4). Part II of the book is much more practical and covers the application of recurrence quantifications to systems such as: biological and physiological systems (Chaps. 5–8); vibrations on small scales (materials science) and large scales (earthquakes) (Chaps. 9 and 10); climatic dynamics, biodiversity profiles, and global photosynthetic activities (Chaps. 11 and 12); and interactive patterns in complex road traffic situations as well as couplings between interacting humans (psychology) (Chaps. 13 and 14).

We hope that the reader finds this book not only useful but inspiring as well. Remember, any system that “wiggles” in time or space is amenable to recurrence analysis. We trust that “any system” is extended to include new systems not yet introduced to recurrence plots and quantifications. As specific follow-up questions arise, please contact the various authors who are all willing to assist the reader

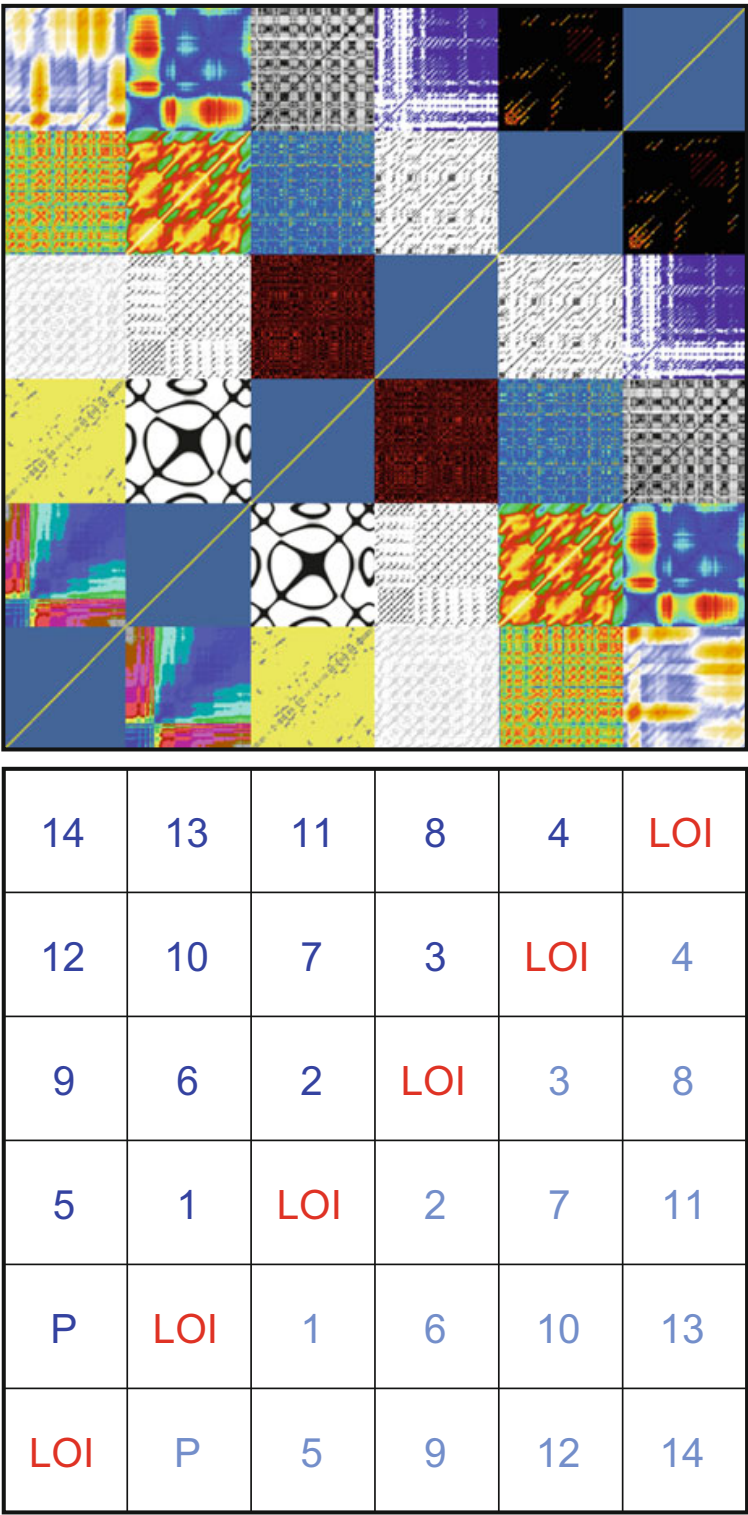


Fig. 3 Recurrence plot quilt formed from unique recurrence plot icons from each chapter (1–14), preface (P), and line-of-identity (LOI)

get a good foundation in recurrence analyses. Never forget that these are powerful nonlinear tools stemming from the ground-breaking work of giants like Poincaré and Eckmann and Ruelle to whom we are all indebted. Since this book is not the final conclusion on the topics of either recurrence theory or recurrence applications, how might your research contribute to new future developments?

Chicago, USA
Potsdam, Germany

Charles L. Webber, Jr.
Norbert Marwan

Contents

Part I RQA Theory

1	Mathematical and Computational Foundations of Recurrence Quantifications	3
	Norbert Marwan and Charles L. Webber, Jr.	
2	Estimating Kolmogorov Entropy from Recurrence Plots	45
	Philippe Faure and Annick Lesne	
3	Identifying Coupling Directions by Recurrences	65
	Yong Zou, M. Carmen Romano, Marco Thiel, and Jürgen Kurths	
4	Complex Network Analysis of Recurrences	101
	Reik V. Donner, Jonathan F. Donges, Yong Zou, and Jan H. Feldhoff	

Part II RQA Best Practices

5	From Time to Space Recurrences in Biopolymers	167
	Alfredo Colosimo and Alessandro Giuliani	
6	Dynamic Coupling Between Respiratory and Cardiovascular System	195
	Federica Censi, Giovanni Calcagnini, and Sergio Cerutti	
7	Analysis of Brain Recurrence	213
	Clifton Frilot II, Paul Y. Kim, Simona Carrubba, David E. McCarty, Andrew L. Chesson Jr., and Andrew A. Marino	
8	Recurrence Analysis of Otoacoustic Emissions	253
	Giovanna Zimatore and Marta Cavagnaro	

9 Vibration Analysis in Cutting Materials 279
Arkadiusz Syta and Grzegorz Litak

10 Dynamical Patterns in Seismology 291
Tamaz Chelidze and Teimuraz Matcharashvili

**11 Long Time-Scale Recurrences in Ecology: Detecting
Relationships Between Climate Dynamics and Biodiversity
Along a Latitudinal Gradient** 335
Raphaël Proulx, Lael Parrott, Lenore Fahrig,
and David J. Currie

**12 Recurrence Quantification and Recurrence Network
Analysis of Global Photosynthetic Activity** 349
Holger Lange and Sven Boese

**13 Recurrence Analysis Applications to Short-Term
Macroscopic and Microscopic Road Traffic** 375
Eleni I. Vlahogianni, Matthew G. Karlaftis, and John C. Golias

14 Interpersonal Couplings in Human Interactions..... 399
Kevin Shockley and Michael A. Riley

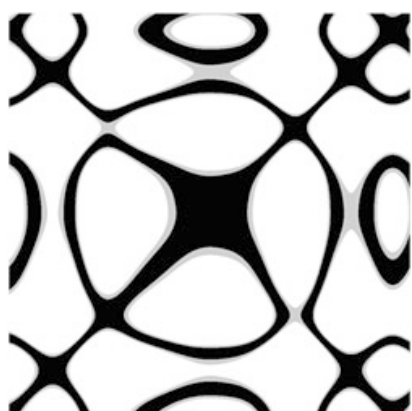
Part I

RQA Theory

Chapter 1

Mathematical and Computational Foundations of Recurrence Quantifications

Norbert Marwan and Charles L. Webber, Jr.



Abstract Real-world systems possess deterministic trajectories, phase singularities and noise. Dynamic trajectories have been studied in temporal and frequency domains, but these are linear approaches. Basic to the field of nonlinear dynamics is the representation of trajectories in phase space. A variety of nonlinear tools such as the Lyapunov exponent, Kolmogorov–Sinai entropy, correlation dimension, etc. have successfully characterized trajectories in phase space, provided the systems studied were stationary in time. Ubiquitous in nature, however, are systems that are nonlinear and nonstationary, existing in noisy environments all of which are assumption breaking to otherwise powerful linear tools. What has been unfolding over the last quarter of a century, however, is the timely discovery and practical demonstration that the recurrences of system trajectories in phase space can provide important clues to the system designs from which they derive. In this chapter we will introduce the basics of recurrence plots (RP) and their quantification analysis (RQA). We will begin by summarizing the concept of phase space reconstructions. Then we will provide the mathematical underpinnings of recurrence plots followed by the details of recurrence quantifications. Finally, we will discuss computational approaches that have been implemented to make recurrence strategies feasible and

N. Marwan (✉)

Potsdam Institute for Climate Impact Research, Potsdam, Germany

e-mail: marwan@pik-potsdam.de

C.L. Webber, Jr.

Loyola University Chicago, Chicago, IL, USA

e-mail: cwebber@luc.edu

useful. As computers become faster and computer languages advance, younger generations of researchers will be stimulated and encouraged to capture nonlinear recurrence patterns and quantification in even better formats. This particular branch of nonlinear dynamics remains wide open for the definition of new recurrence variables and new applications untouched to date.

1.1 Phase Space Trajectories

Systems in nature or engineering typically exist in either quasi-stationary states or in non-stationary states as they move or transition between states. These complicated processes derive mostly from complex systems (nonlinear, many coupled variables, polluted by noise, etc.) and defy meaningful analysis. Still, approximate investigations of these processes remain an important focus among numerous scientific disciplines (e.g. meteorology). To the extent that systems are deterministic (rule-driven) there still remains the hope and challenge of describing dynamical system changes to such a degree rendering it possible to predict future states of the system (e.g. make forecasts). Practically, the usual aim is to find mathematical models which can be adapted to the real processes (mimicry) and then used for solving given problems. The measuring of a state (which leads to observations of the state but not to the state itself) and subsequent data analysis are the first steps toward the understanding of a process. Well known and approved methods for data analysis are those based on linear concepts as estimations of moments, correlations, power spectra, or principal components analyses etc. In the last two decades this zoo of analytical methods has been enriched with methods of the theory of nonlinear dynamics. Some of these new methods are rooted in the topological analysis of the phase space of the underlying dynamics or on an appropriate reconstruction of it [1, 2]

The state of a system can be described by its d state variables

$$x_1(t), x_2(t), \dots, x_d(t), \quad (1.1)$$

for example the two state variables temperature and pressure in a thermodynamic system. The d state variables at time t form a vector $\mathbf{x}(t)$ in a d -dimensional space which is called phase space. This vector moves in time and in the direction that is specified by its velocity vector

$$\dot{\mathbf{x}}(t) = \partial_t \mathbf{x}(t) = \mathbf{F}(\mathbf{x}). \quad (1.2)$$

The temporary succession of the phase space vectors forms a trajectory (phase space trajectory, orbit). The velocity field $\mathbf{F}(\mathbf{x})$ is tangent to this trajectory. For autonomous systems the trajectory must not cross itself. The time evolution of the trajectory explains the dynamics of the system, i.e., the attractor of the system. If $\mathbf{F}(\mathbf{x})$ is known, the state at a given time can be determined by integrating the equation system [Eq. (1.2)]. However, a graphical visualization of the trajectory enables the determination of a state without integrating the equations. The shape

of the trajectory gives hints about the system; periodic or chaotic systems have characteristic phase space portraits.

The observation of a real process usually does not yield all possible state variables. Either not all state variables are known or not all of them can be measured. Most often only one observation $u(t)$ is available. Since measurements result in discrete time series, the observations will be written in the following as u_i , where $t = i \Delta t$ and Δt is the sampling rate of the measurement. [Henceforth, variables with a subscribed index are in this work time discrete (e.g. \mathbf{x}_i , $\mathbf{R}_{i,j}$), whereas a braced t denotes continuous variables (e.g. $\mathbf{x}(t)$, $\mathbf{R}(t_1, t_2)$).]

Couplings between the system's components imply that each single component contains essential information about the dynamics of the whole system. Therefore, an equivalent phase space trajectory, which preserves the topological structures of the original phase space trajectory, can be reconstructed by using only one observation or time series, respectively [2, 3]. A method frequently used for reconstructing such a trajectory $\hat{\mathbf{x}}(t)$ is the time delay method: $\hat{\mathbf{x}}_i = (u_i, u_{i+\tau}, \dots, u_{i+(m-1)\tau})^T$, where m is the embedding dimension and τ is the time delay (index based; the real time delay is $\tau \Delta t$). The preservation of the topological structures of the original trajectory is guaranteed if $m \geq 2d + 1$, where d is the dimension of the attractor [2].

Both embedding parameters, the dimension m and the delay τ , have to be chosen appropriately. Different approaches are applicable for the determination of the smallest sufficient embedding dimension [1, 4]. Here we focus on an approach which uses the number of false nearest neighbours.

There are various methods that use false nearest neighbours in order to determine the embedding dimension. The basic idea is that by decreasing the dimension an increasing amount of phase space points will be projected into the neighbourhood of any phase space point, even if they are not real neighbours. Such points are called *false nearest neighbours* (FNNs). The simplest method uses the amount of these FNNs as a function of the embedding dimension in order to find the minimal embedding dimension [1]. Such a dimension has to be taken where the FNNs vanish. Other methods use the ratios of the distances between the same neighbouring points for different dimensions [4, 5].

Random errors and low measurement precision can lead to a linear dependence between the subsequent vectors \mathbf{x}_i . Hence, the delay has to be chosen in such a way that such dependences vanishes. One possible means of determining the delay is by using the *autocovariance function* $C(\tau) = \langle u_i u_{i-\tau} \rangle$ (using the assumption $\langle u_i \rangle = 0$).

A delay may be appropriate when the autocovariance approaches zero. This minimizes the linear correlation between the components but does not have to mean they are independent. However, the converse is true: if two variables are independent they will be uncorrelated. Therefore, another well established possibility for determining the delay is the *mutual information* [6]

$$I(\tau) = - \sum_{\varphi, \psi} p_{\varphi, \psi}(\tau) \log \frac{p_{\varphi, \psi}(\tau)}{p_{\varphi} p_{\psi}}. \quad (1.3)$$

Here $p_{\varphi, \psi}(\tau)$ is the joint probability that $u_i = \varphi$ and $u_{i+\tau} = \psi$. p_φ and p_ψ are the probabilities that u_i has the value φ and ψ , respectively. In order to simplify the notations, we use $p_{u_i} = p_\varphi$, $p_{u_{i+\tau}} = p_\psi$ and $p_{u_i, u_{i+\tau}} = p_{\varphi, \psi}(\tau)$. The mutual information is not a function of the variables φ and ψ but of the joint probability $p_{\varphi, \psi}(\tau)$. It is the average of the information about a value after a delay τ , which can be yielded from the knowledge of the current value. The best choice for the delay is where $I(\tau)$ has its smallest local minimum. The advantage of the mutual information vs. the autocovariance function is that it finds the nonlinear interrelations and, hence, determines such a delay which fulfils the criterion of independence. The experience has shown that the delay is sometimes overestimated by auto-correlation and mutual information.

An alternative approach for finding optimal embedding parameters is using recurrence plots [7]. First we create a recurrence plot (RP) with a high embedding dimension ($m = 20 \dots 25$). Then we decrease progressively the dimension until a significant change in the RP results. In particular, we are interested in a RP that is cleaned from single points and where linear structures dominate [8]. Since this change is due to a topological change of the phase space trajectory caused by the occurrence of FNNs, the current dimension plus a few dimensions should be sufficient for the embedding. However, this criterion has to be considered with the utmost caution because with high embedding dimensions (e.g., $m = 10$ would be enough) we can get spurious recurrences which can create an RP with a large amount of diagonal lines even for stochastic data [9]. Non-optimal embedding parameters can cause many interruptions of diagonal lines in the RP, small blocks, or even diagonal lines perpendicular to the LOI (this corresponds to parallel trajectory segments running in opposite time direction).

A phase space reconstruction can be used in order to estimate characteristic properties of the dynamical system. For reviews on corresponding methods see for example [10, 11] or [12]. Besides, the phase space reconstruction is the starting point for the construction of a recurrence plot.

1.2 Recurrence Plots

1.2.1 Definition of Recurrence Plots

Natural processes can have a distinct recurrent behaviour, e.g., periodicities (as seasonal or Milanković cycles), but also irregular cyclicities (as El Niño/Southern Oscillation). Moreover, the recurrence of states, in the meaning that states are arbitrary close after some time, is a fundamental property of deterministic dynamical systems and is typical for nonlinear or chaotic systems [12–14].

Recurrences in the dynamics of a dynamical system can be visualised by the recurrence plot (RP), introduced by Eckmann et al. in 1987 [15]. The RP represents the times at which states \mathbf{x}_i in a phase space recur.

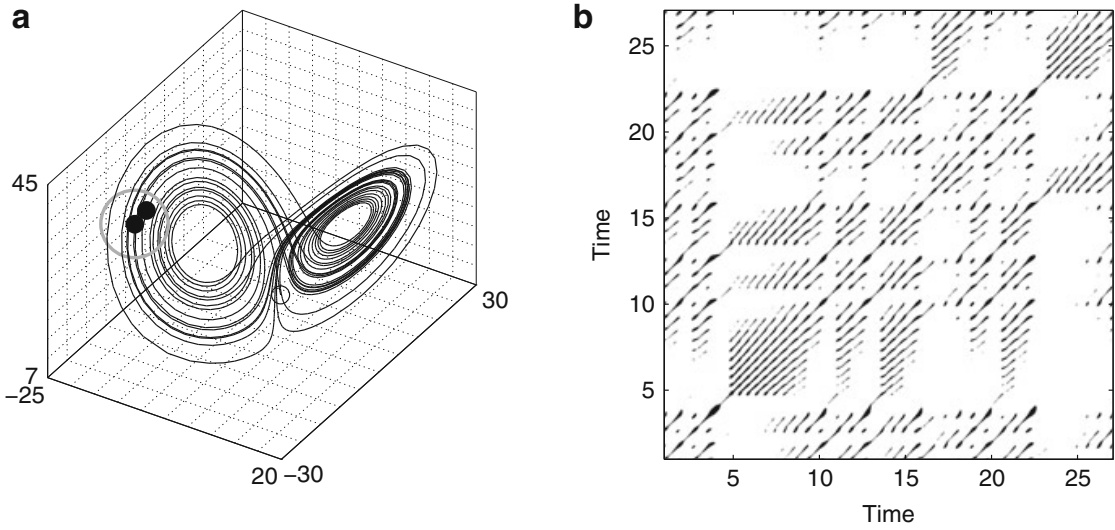


Fig. 1.1 (a) Segment of the phase space trajectory of the Lorenz system, Eq. (1.34) (parameters $r = 28$, $\sigma = 10$, $b = \frac{8}{3}$) [17] by using its three components and (b) its corresponding recurrence plot. A point of the trajectory at j which falls into the neighbourhood [gray circle in (a)] of a given point at i is considered as a recurrence point [black point on the trajectory in (a)]. This is marked with a black point in the RP at the location (i, j) . A point outside the neighbourhood [small circle in (a)] causes a white point in the RP. The radius of the neighbourhood for the RP is $\varepsilon = 5$

The original intention was to provide a tool which can easily provide insights into even high-dimensional dynamical systems, those phase space trajectories are otherwise very difficult to visualise [15, 16]. A RP enables us to investigate the m -dimensional phase space trajectory through a two-dimensional representation of its recurrences (Fig. 1.1). Such recurrence of a state at time i at a different time j is pictured within a two-dimensional squared matrix \mathbf{R} with dots, where both axes are time axes [9]:

$$R_{i,j}^{m,\varepsilon_i} = \Theta(\varepsilon_i - \|\mathbf{x}_i - \mathbf{x}_j\|), \quad \mathbf{x}_i \in \mathbb{R}^m, \quad i, j = 1 \dots N, \quad (1.4)$$

where N is the number of considered states x_i ; ε_i is a threshold distance, $\|\cdot\|$ a norm, and $\Theta(\cdot)$ the Heaviside function.

Since $R_{i,i} = 1$ ($i = 1 \dots N$) by definition, the RP has a black main diagonal line, the *line of identity* (LOI), with an angle of $\pi/4$. It has to be noted that a single recurrence point at (i, j) does not contain any information about the current states at the times i and j . However, from the totality of all recurrence points it is possible to reconstruct the phase space trajectory [9, 18, 19].

In practice it is not useful and largely impossible to find complete recurrences in the sense $\mathbf{x}_i \equiv \mathbf{x}_j$ (e. g. the state of a chaotic system would not recur exactly to the initial state but approaches the initial state arbitrarily close). Therefore, a recurrence is defined as a state \mathbf{x}_j is sufficiently close to \mathbf{x}_i . This means that those states \mathbf{x}_j that fall into an m -dimensional neighbourhood of size ε_i centered at \mathbf{x}_i are recurrent. These \mathbf{x}_j are called *recurrence points*. In Eq. (1.4), this is simply expressed by the Heaviside function and its argument ε_i .

In the original definition of the RPs, the neighbourhood is a ball (i.e., L_2 -norm is used) and its radius is chosen in such a way that it contains a fixed number of closest states \mathbf{x}_j [15]. With such a neighbourhood, the radius ε_i changes for each \mathbf{x}_i ($i = 1 \dots N$) and $R_{i,j} \neq R_{j,i}$ because the neighbourhood of \mathbf{x}_i does not have to be the same as that of \mathbf{x}_j . This property leads to an asymmetric RP, but all columns of the RP have the same recurrence density. We denote this neighbourhood as *fixed amount of nearest neighbours (FAN)*. However, the most commonly used neighbourhood is that using a metric and a fixed radius $\varepsilon_i = \varepsilon, \forall i$. A metric and a fixed radius ensures that $R_{i,j} = R_{j,i}$, i.e., a symmetric RP. The type of neighbourhood that should be used depends on the application [9, 20]. For example, the FAN neighbourhood is useful for nonstationary data, for bivariate recurrence investigations using cross recurrence plots, or for the comparison of RPs of different systems, because it is not necessary to normalise the time series beforehand and it allows an investigation on the basis of comparable recurrence structures [9].

The most commonly used norms are the L_2 -norm (Euclidean norm) and the L_∞ -norm (Maximum or Supremum norm). The L_∞ -norm is often used because it is independent of the phase space dimension, easier to calculate, and allows some analytical expressions [21–23]. However, this latter choice is more prone to noise or outliers, because it depends on a single point that if it is an extreme, it will from a statistical point of view, thus, result in an outlying distance and results in a wrong measure.

The recurrence threshold ε is a crucial parameter in the RP analysis. Although several works have contributed to this discussion [9, 22, 24, 25], a general and systematic study on the recurrence threshold selection remains an open task for future work. Nevertheless, recurrence threshold selection is a trade-off of to have a threshold as small as possible but at the same time a sufficient number of recurrences and recurrence structures. In general, the optimal choice of ε depends on the application and the experimental conditions (e.g., noise), but in all cases it is desirable that the smallest threshold possible is chosen.

A “rule of thumb” for the choice of the threshold ε is to select it as a few per cent (not larger than 10 %) of the maximum phase space diameter [26–28]. For classification purpose and signal detection, a better choice is to select ε between 20 and 40 % of the signal’s standard deviation σ [25].

However, the influence of noise can necessitate a larger threshold, because noise would distort any existing structure in the RP. Higher threshold may preserve these structures [9]. Suggestions from literature show that this threshold should be a few per cent of the maximum phase space diameter [26] and should not exceed 10 % of the mean or the maximum phase space diameter [27, 28]. Using the recurrence point density of the RP, the threshold can be chosen from the analysis of this measure in respect to a changing threshold [7]. The threshold can then be found by looking for a scaling region in the recurrence point density. However, this may not work for nonstationary data. For this case Zbilut et al. [7] have suggested to choose ε so that the recurrence point density is approximately 1 %. For noisy periodic processes, [24] have suggested to use the diagonal structures within the RP in order to determine an optimal threshold. Their criterion minimizes the fragmentation and thickness of

the diagonal lines in respect to the threshold. Recent studies about RPs in our group have revealed a more exact criterion for choosing this threshold. This criterion takes into account that a measurement of a process is a composition of the real signal and some observational noise with standard deviation. In order to get similar results by using RPs, a threshold has to be chosen which is five times larger than the standard deviation of the observational noise [22]. This criterion holds for a wide class of processes.

For specific purposes (e.g., quantification of recurrences), it can be useful to exclude the LOI from the RP, as the trivial recurrence of a state with itself might not be of interest. Moreover, due to the finite threshold value ε , further long diagonal lines can occur directly below and above the LOI for smooth or high resolution data (highly sampled data). Therefore, the diagonal lines in a small corridor around the LOI correspond to the *tangential motion* of the phase space trajectory, but not to different orbits. Thus, for quantification purposes it is better to exclude this entire predefined corridor and not only the LOI. This step corresponds with suggestions to exclude the tangential motion as it is done for the computation of the correlation dimension (known as Theiler correction or Theiler window) [29] or for the alternative estimators of Lyapunov exponents [30] in which only those phase space points are considered that fulfil the constraint $|j - i| \geq w$. Theiler has suggested using the autocorrelation time as an appropriate value for w [29], and Gao et al. state that $w = (m - 1)\tau$ would be a sufficient approach [30]. However, in a visual representation of an RP it is better to keep the LOI.

For the definition of a recurrence, other metrics or criteria can be used. Some extensions of recurrence definitions have been proposed in order to improve the representation and quantification of the characteristic recurrence structure [9].

For example, the *perpendicular recurrence plot* has been suggested in order to reduce the effects of the tangential motion. The perpendicular RP is defined as

$$R_{i,j}^{m,\varepsilon} = \Theta(\varepsilon - \|\mathbf{x}_i - \mathbf{x}_j\|) \cdot \delta(\dot{\mathbf{x}}_i \cdot (\mathbf{x}_i - \mathbf{x}_j)), \quad (1.5)$$

where δ is the Delta function. This recurrence plot contains only those points \mathbf{x}_j that fall into the neighbourhood of \mathbf{x}_i and lie in the $(m - 1)$ -dimensional subspace of \mathbb{R}^m that is perpendicular to the phasespace trajectory at \mathbf{x}_i . These points correspond locally to those lying on a Poincaré section.

1.2.2 Structures in Recurrence Plots

The fundamental purpose of RPs is the visualization of higher dimensional phase space trajectories. Structural patterns in RPs reveals hints about the time evolution of these trajectories. Other distinct advantages of RPs is that not only can they operate on noisy data, they can also be applied to rather nonstationary data as well as rather short data sets. The RPs exhibit characteristic large scale and small scale patterns.

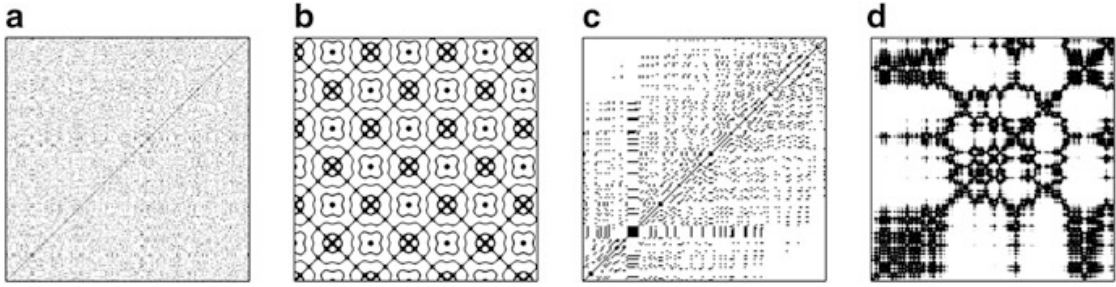


Fig. 1.2 Characteristic typology of recurrence plots: (a) homogeneous (uniformly distributed noise), (b) periodic (super-positioned harmonic oscillations), (c) drift (logistic map $x_{i+1} = 4x_i(1 - x_i)$ corrupted with a linearly increasing term) and (d) disrupted (Brownian motion). These examples illustrate how different RPs can be. The used data have the length 400 (a, b, d) and 150 (c), respectively; no embeddings are used; the thresholds are $\varepsilon = 0.2\sigma$ (a, c, d) and $\varepsilon = 0.4\sigma$ (b)

The large scale appearance of RPs, their *typology*, can be classified as *homogeneous*, *periodic*, *drift*, and *disrupted* [9, 15]:

- *Homogeneous* RPs typify stationary and autonomous systems in which relaxation times are short in comparison with the time spanned by the RP. An example of such an RP is that of a random time series (Fig. 1.2a).
- RPs with diagonal oriented, *periodic* recurrent structures (diagonal lines, checkerboard structures) are hallmarks for oscillating systems. The illustration in Fig. 1.2b is a rather clear periodic system with two frequencies and a frequency ratio of four (the main diagonal lines are divided by four crossing short lines; irrational frequency ratios cause more complex periodic recurrent structures). However, even for those oscillating systems whose oscillations are not easily recognizable, the RPs can be used to find their oscillations [15].
- Paling or darkening of recurrent points away from the LOI (*drift*) is caused by drifting systems with slowly varying parameters. Thus slow (adiabatic) changes in the dynamic over time brightens the RPs upper-left and lower-right corners (Fig. 1.2c).
- *White areas or bands* in RPs indicate abrupt changes in the dynamic as well as extreme events (Fig. 1.2d). In these cases, RPs can be used to find and assess extreme and rare events by scoring the frequency of their repeats

Close inspection of the RPs reveals small scale structures (the texture) which consists as a combination of *isolated dots* (chance recurrences), dots forming *diagonal lines* (deterministic structures) as well as *vertical/horizontal lines* or dots clustering to inscribe *rectangular regions* (laminar states, singularities). These small scale structures are the basis for quantitative analysis of the RPs.

- *Single, isolated recurrence points* can occur if states are rare, if they do not persist for any time or if they fluctuate heavily. However, they are not a unique sign of chance or noise (for example in maps).
- A *diagonal line* $R_{i+k,j+k} = 1$ (for $k = 1 \dots l$, where l is the length of the diagonal line) occurs when a segment of the trajectory runs parallel to

another segment, i.e., the trajectory visits the same region of the phase space at different times. The length of this diagonal line is determined by the duration of such similar local evolution of the trajectory segments. The direction of these diagonal structures can differ. Diagonal lines parallel to the LOI (angle $\pi/4$) represent the parallel running of trajectories for the same time evolution. The diagonal structures perpendicular to the LOI represent the parallel running with contrary times (mirrored segments; this is often a hint for an inappropriate embedding). The lengths of diagonal lines in an RP are directly related to the ratio of determinism or predictability immanent to the dynamics. If a system is predictable, similar situations (states), i.e., $R_{i,j} = 1$, will lead to a similar future, i.e., the probability to have $R_{i+1,j+1} = 1$ is high. Perfectly predictable systems, thus, would have infinitely long diagonal lines in the RP. In contrast, the probability for $R_{i+1,j+1} = 1$ is very low for stochastic systems, i.e., we find only single points or short lines. If the system is chaotic, close states will diverge exponentially in the future. The faster the divergence, i.e., the higher the Lyapunov exponent, the shorter the diagonals.

- A vertical (horizontal) line $R_{i,j+k} = 1$ (for $k = 1 \dots v$, with v the length of the vertical line) marks a time length in which a state does not change or changes very slowly. It seems, that the state is trapped for some time. This is a typical behaviour of laminar states (intermittency) or systems that paused at singularities. Such structures can reveal discontinuities in the signal which portends special states of the system.

These small scale structures are the base of a quantitative analysis of the RPs which is introduced in Sect. 1.3.

The examples in Fig. 1.2 illustrate the appearance of RPs for sundry dynamics. A large number of single points with a vanishing line structures are caused by heavy fluctuations in the dynamic such as seen in uncorrelated noise (Fig. 1.2a). The relationship between periodically recurrent structures and oscillators is obvious as the exact recurrent dynamics score as long diagonal lines separated by a fixed distance (Fig. 1.2b). The non-regular appearance of both short and long diagonal lines is characteristic for chaotic processes (Fig. 1.2c). The uneven occurrence of extended black clusters and extended white areas corresponds to non-regular behavior in the system such as found with correlated (red) noise (Fig. 1.2d).

The structures in a RP can also be used to estimate the recurrence times that have been also used to characterize the dynamics of a system [31–34]. The distance between recurrence points in a column of the RP corresponds to the duration until a state recurs. There are several possibilities of the recurrence time estimation: We can either measure the distance between the starting point of a recurrence structure and the starting point of the next recurrence structure [35] (Fig. 1.3a), or we measure the distance between the end point of a recurrence structure and the starting point of the next recurrence structure, i.e., we measure the length of the white vertical lines in a RP [34] (Fig. 1.3b). The more recurrence points are formed by the tangential motion (Fig. 1.4), the more extended the vertical recurrence structures are, leading to stronger differences in these two estimators. For such cases, recurrence times would

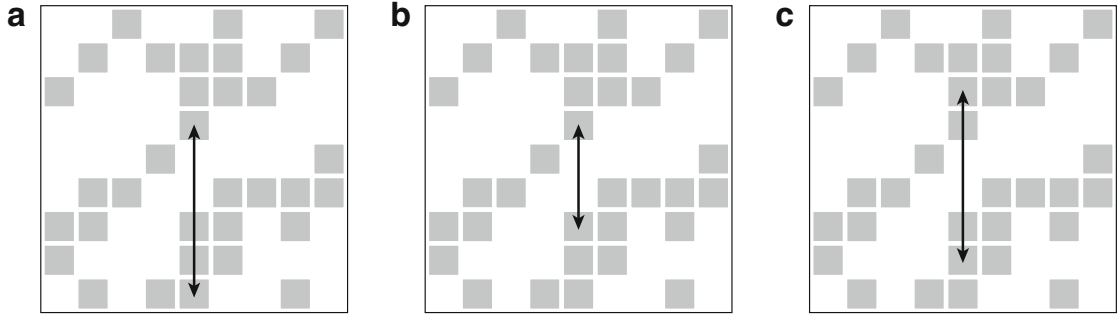


Fig. 1.3 RP illustrating the estimators of recurrence times (*black arrows*), based on (a) distance between starting points, (b) *white vertical lines*, and (c) distance between the midpoints of the recurrence structures

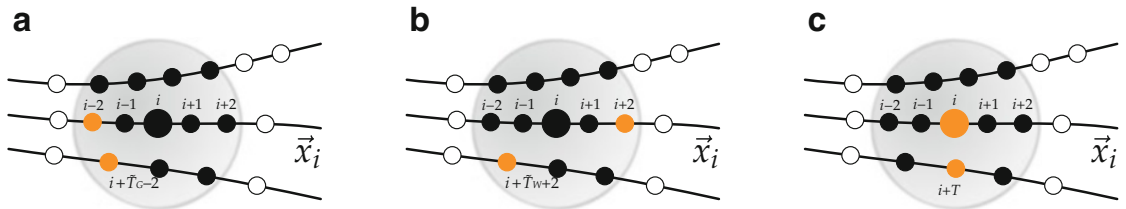


Fig. 1.4 Recurrence points of a state \mathbf{x}_i illustrating the differences in the estimators for the recurrence time T in case of tangential motion (the preceding and subsequent recurrence points along the trajectory). (a) The estimator as given in Fig. 1.3a corresponds to the time distance between the first points falling into the neighborhood of state i (here $i - 2$ and $i + T - 2$). (b) The estimator as given in Fig. 1.3b corresponds to the time distance between the last and first points falling into the neighborhood of state i (here $i + 2$ and $i + T + 2$). (c) The Poincaré recurrence time T for the illustrated case

be better estimated by using the (vertical) midpoints of the recurrence structures (Figs. 1.3c and 1.4c). Recurrence time definition based on distance between starting points is an upper limit and the estimator based on white vertical lines is a lower limit of the recurrence times estimation.

Summarizing the aforementioned points we can establish the following list of recurrent pattern structures and their corresponding qualitative interpretations:

1. Homogeneity \rightarrow the process is stationary
2. Fading or darkening to the upper left and lower right corners \rightarrow the process is nonstationary or on a transient between states
3. White banding disruptions \rightarrow the process is nonstationarity in which some states are rare or far from the normal or transitions may have occurred
4. Periodic patterns \rightarrow the process has characteristic cyclicities with periods corresponding to the time distance between periodic structures
5. Single isolated points \rightarrow the process is dominated with by heavy fluctuations or may even be stochastic
6. Diagonal lines (parallel to the LOI) \rightarrow the process is deterministic in the periodic sense (long diagonals) or chaotic sense (short diagonals)
7. Diagonal lines (orthogonal to the LOI) \rightarrow the process is characterized by the evolution of palindromic states (time reversed)

8. Vertical and horizontal lines forming rectangles \rightarrow some states do not change or change slowly for some time (laminar states) or the process is halted at a singularity in which the dynamic is stuck in paused states

The visual interpretation of RPs requires some experience. The study of RPs from paradigmatic systems gives a good introduction into characteristic typology and texture. However, the quantification of RPs offers a more objective way for evaluating the system under investigation. As we will see, quantitative recurrence variables are defined and extracted from recurrence plots in a non-biased fashion.

1.3 Recurrence Quantifications

Visually RPs can provide some useful insights in the dynamics of dynamical systems. However, graphical displays with insufficient resolution to display RPs come with the disadvantage that users are forced to subjectively intuit and interpret patterns and structures presented within the recurrence plot. And different observers see things differently. To overcome the subjectivity of the methodology, in the early 1990s Zbilut and Webber introduced definitions and procedures to quantify RP structures [28, 36, 37]. They defined a set of five recurrence variables that functioned as complexity measures based on diagonal line structuring in RPs and coined the name recurrence quantification analysis (RQA).

1.3.1 Classical Recurrence Quantification Analysis

The first variable in RQA is *percent recurrence (REC)* or *recurrence rate (RR)*

$$RR(\varepsilon, N) = \frac{1}{N^2 - N} \sum_{i \neq j=1}^N R_{i,j}^{m,\varepsilon}, \quad (1.6)$$

which simply enumerates or counts the black dots in the RP excluding the LOI. It is a measure of the relative density of recurrence points in the sparse matrix and is related to the definition of the correlation sum [38]. Large segments of data are required when RR is used as an estimator of the correlation sum. In the limit of long time series

$$P = \lim_{N \rightarrow \infty} RR(\varepsilon, N), \quad (1.7)$$

is the probability of finding a recurrence point within the RP and is the probability that states will recur. With the knowledge of the probability $\rho(x)$ of a state x of a stochastic process and for dimension $m = 1$ and using the maximum norm, RR can be analytically computed by using the convolution [23]

$$P_o = \rho(x) * \rho(x). \quad (1.8)$$

This probability P_o can be used to analytically describe the RQA measures for some systems [22, 23].

The following measures are based on line structures in the RP. First, we consider the histogram of the lengths of the diagonal structures in the RP,

$$H_D(l) = \sum_{i,j=1}^N (1 - R_{i-1,j-1})(1 - R_{i+l,j+l}) \prod_{k=0}^{l-1} R_{i+k,j+k}. \quad (1.9)$$

The second variable in RQA is the *percent determinism (DET)*, defined as the fraction of recurrence points that form diagonal lines

$$DET = \frac{\sum_{l=d_{\min}}^N l H_D(l)}{\sum_{i,j=1}^N R_{i,j}}. \quad (1.10)$$

Systems possessing deterministic (rule-obeying) dynamics are characterized by diagonal lines indicating repeating recurrences within a state. For periodic signals the diagonal lines are long. For chaotic signals the diagonal lines are short. For stochastic signals the diagonal lines are absent, save for chance recurrences forming very short lines. *DET* can be interpreted as the predictability of the system more so for periodic behaviors than chaotic processes. But it must be realized that *DET* does not have the same strict meaning of the determinism of a process. The d_{\min} parameter sets the lower bounds on the definition of lines in the RP. The threshold d_{\min} excludes the diagonal lines which are formed by the tangential motion of the phase space trajectory. Typically, d_{\min} is set to 2. If d_{\min} is set to 1, *DET* and *RR* are identical. For $d_{\min} > 2$, this parameter serves as a filter, excluding the shorter lines and decreasing *DET* which is practically useful for the study of some dynamical systems. The choice of d_{\min} could be made in a similar way as the choice of the size for the Theiler window, but we have to take into account that a too large l_{\min} can worsen the histogram $H_D(l)$ and thus the reliability of the measure *DET*.

The derived variable *ratio (RATIO)*, has been defined as the ratio of *DET* to *RR* [36] and can be computed from the frequency distributions of the lengths of the diagonal lines

$$RATIO = N^2 \frac{\sum_{l=l_{\min}}^N l H_D(l)}{\left(\sum_{l=1}^N l H_D(l)\right)^2}. \quad (1.11)$$

A heuristic study of physiological systems has revealed that *RATIO* is useful in discovering dynamic transitions, since during certain types of transitions the *RR* decreases while *DET* remains unchanged, thereby increasing *RATIO* [36].

The third variable in RQA is the *maximal line length in the diagonal direction* (D_{\max})

$$D_{\max} = \arg \max_l H_D(l) \quad (1.12)$$

which is simply the length of the single longest diagonal within the entire RP. Since diagonal structures show the range in which a segment of the trajectory is rather close to another segment of the trajectory at a different time, these lines give a hint about the divergence of the trajectory segments. The smaller D_{\max} the more divergent the trajectories. Based on this idea it is obvious that there is a relationship between the largest positive Lyapunov exponent (if there is one in the considered system) and D_{\max} . Indeed, the relationship can be found by considering the (cumulative) frequency distribution of the lengths of the diagonal lines and the K_2 entropy that is a lower limit of the sum of the positive Lyapunov exponents [9].

Related to D_{\max} is the *average diagonal line length*

$$\langle D \rangle = \frac{\sum_{l=d_{\min}}^N l H_D(l)}{\sum_{l=d_{\min}}^N H_D(l)} \quad (1.13)$$

which is the average time two segments of the trajectory are close to each other. In this case, $\langle D \rangle$ can be interpreted as the mean prediction time.

The fourth variable in RQA is the *Shannon entropy of the frequency distribution of the diagonal line lengths (ENT)*

$$ENT = - \sum_{l=d_{\min}}^N p(l) \ln p(l) \quad \text{with} \quad p(l) = \frac{H_D(l)}{\sum_{l=d_{\min}}^N H_D(l)} \quad (1.14)$$

which reflects the complexity of the deterministic structure in the system. Calibrated in bits/bin, the higher the ENT the more complex the dynamics, e.g., for uncorrelated noise or oscillations the value of ENT is rather small, indicating its low complexity. It can be noted that ENT depends sensitively on the bin number and, thus, may differ for different parameter choices of the same process (e.g. different thresholds ε , different l_{\min} values, etc.), as well as for different data sets of course.

The measures introduced up to now, RR , DET , D_{\max} , etc. can also be computed separately for each diagonal parallel to the LOI and with distance k to the LOI. For example, the recurrence point density along a diagonal of distance k from the LOI is

$$RR_k = \frac{1}{N-k} \sum_{k,j=1}^{N-k} R_{k,j}. \quad (1.15)$$

We denote such diagonalwise computed measures with a subscribed index or, in general, an asterisk, e.g., RR_* or RR_i . Diagonalwise calculated RQA measures are useful for studying the periodicity of a signal [39], to indicate periodic orbits [26,40,41], or to investigate the interrelationship between complex systems [42]. Moreover, RR_k can be interpreted as the probability that a system occurs after k time steps. This measure can be used to study phase-synchronization [43].

The fifth RQA measure is the *trend* (TND), which is a linear regression coefficient over the recurrence point density RR_* of the diagonals parallel to the LOI, Eq. (1.15), as a function of the time distance between these diagonals and the LOI

$$TND = \frac{\sum_{i=1}^{\tilde{N}} (i - \tilde{N}/2)(RR_i - \langle RR_i \rangle)}{\sum_{i=1}^{\tilde{N}} (i - \tilde{N}/2)^2}. \quad (1.16)$$

The trend gives information about the stationarity versus nonstationarity in the process. Quasi-stationary dynamics will have TND values that hover near 0. Nonstationary dynamics will have TND values far from 0 revealing drift in the dynamics possibly indicating that the system is en route between more stationary states. The TND computation excludes the edges of the RP ($\tilde{N} < N$) because the statistic lacks by reason of less recurrence points. The choice of \tilde{N} depends on the studied process. Whereas $N - \tilde{N} > 10$ should be sufficient for noise, this difference should be much larger for a process with some autocorrelation (ten times the order of magnitude of the autocorrelation time should always be enough). It should be noted that if the time dependent RQA (measures are computed in shifted windows, see Sect. 1.3.6) is used, TND will depend strongly on the size of the windows and may reveal contrary results for different window sizes [20].

1.3.2 Extended Recurrence Quantification Analysis

The five RQA variables defined above are based largely on the lengths, number and distributions of diagonal lines in RPs. That is, they are sensitive to parallel trajectories along different segments of the time series. Additional information about other geometrical structures is not included. But RPs have not only diagonal lines, but also vertical and horizontal elements as well. From these vertical lines, additional recurrence quantifications have been posited by Marwan et al. [44].

From these considerations, the sixth RQA measure is the *laminarity*

$$LAM = \frac{\sum_{l=v_{min}}^N l H_V(l)}{\sum_{i,j=1}^N R_{i,j}}, \quad (1.17)$$

with

$$H_V(l) = \sum_{i,j=1}^N (1 - R_{i,j-1})(1 - R_{i,j+l}) \prod_{k=0}^{l-1} R_{i,j+k} \quad (1.18)$$

as the histogram of lengths of vertical lines. LAM carries a definition analogous to that of definition of DET , that is LAM reports the percentage of recurrent points

in vertical structures whereas *DET* reports the percentage of recurrent points in diagonal structures.

The computation of *LAM* is realized for those l that exceed a minimal length v_{min} in order to decrease the influence of sojourn points. For iterated maps (as opposed to continuous flows) we typically set $v_{min} = 2$. Since *LAM* quantifies the relative amount of vertical structuring over the entire RP, it also represents the frequency of occurrence of laminar states within the system. The length of the laminar phases in time is ignored, but *LAM* will decrease if the RP contains recurrent points that are more isolated than in vertical or diagonal structures.

Consequently, we define the seventh RQA measure as the average length of vertical structures

$$TT = \frac{\sum_{l=v_{min}}^N v H_V(l)}{\sum_{l=v_{min}}^N H_V^\varepsilon(l)}, \quad (1.19)$$

which we call *trapping time* TT . The computation also uses the minimal length v_{min} as in *LAM*, Eq. (1.17). TT contains information about the amount and the length of the vertical structures in the RP by reporting the mean time the system will abide at a specific state (how long the state is trapped).

The eighth and last RQA measure is the *maximal length of the vertical structures* V_{max} measuring the longest vertical line in the RP

$$V_{max} = \arg \max_l H_V(l) \quad (1.20)$$

and is analogous to the diagonal measure D_{max} Eq. (1.12). The dynamical interpretation of V_{max} has not been clearly delineated, however, but it can be related to singular states in which the system is stuck in a holding pattern inscribing rectangles in the RP.

In contrast to the five basic RQA measures, these new measures are able to find chaos-chaos transitions [44]. Hence, these measures make the investigation of intermittency possible, even if they are only occurring in rather short and nonstationary time series. Since these measures are zero for periodic dynamics, chaos-order transitions can also be identified.

1.3.3 Recurrence Time Based Measures

We can also use recurrence times for the definition of measures of complexity. Most of these measures make use of the probability distribution of the recurrence times T , i.e., $p(T)$. There are measures like *recurrence period density entropy* (entropy of $p(T)$) [33], the *mean recurrence time*, or the *number of the most probable recurrence time* [45]. These measures allow to distinguish between different types of

dynamics (like periodic, chaotic, or stochastic) or the onset of dynamical transitions (e.g., from chaos to strange non-chaotic attractors).

1.3.4 Complex Network Based Quantification

A RP can be considered as the adjacency matrix of an undirected, unweighted complex network [46]. This allows the application of measures from complex networks statistics, like *clustering coefficient*, *betweenness coefficient*, or *average shortest path length* [47–49]. In the last years this new field has been developed fast and has shown that these new measures add an additional and complement view to the recurrence quantification analysis. In Chap. 4, a comprehensive overview on this topic will be given.

1.3.5 Advanced Quantification

Diagonal lines in a RP allow for the calculation of dynamical invariants, like the Rényi entropy of second order (correlation entropy). In the definition of the second order Rényi entropy an attractor is covered by partitions of boxes of size ε , $1, 2, \dots, M(\varepsilon)$. Then, we measure the joint probability $p(i_1, \dots, i_l)$ that $x(i)$ is in box i_1 , $x(i+1)$ in box i_2 , \dots , and $(i + l - 1)$ finally in box i_l , and get the Rényi entropy [50] as

$$K_2 = - \lim_{\Delta t \rightarrow 0} \lim_{\varepsilon \rightarrow 0} \lim_{l \rightarrow \infty} \frac{1}{l \Delta t} \sum_{i_1, \dots, i_l} p^2(i_1, \dots, i_l), \quad (1.21)$$

with Δt the sampling time step. This measure is directly related to the number of possible trajectories that the system can take for l time steps in the future. If the system is perfectly deterministic in the classical sense, there will be only one possibility for the trajectory to evolve and, hence, $K_2 = 0$. In contrast, for purely stochastic systems the number of possible future trajectories increases to infinity so fast, that $K_2 \rightarrow \infty$. Chaotic systems are characterised by a finite value of K_2 , as they belong to a category between pure deterministic and pure stochastic systems. The inverse of K_2 has units of time and can be interpreted as the mean prediction time of the system.

The sum of the probabilities $p(i_1, \dots, i_l)$ can be approximated by the probability $p_i(l)$ of finding a sequence of l points in boxes of size ε centred at the points $x(i), \dots, x(i + (l - 1))$, that can be estimated by the RP:

$$p_i(l) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{s=1}^N \prod_{k=0}^{l-1} \mathbf{R}_{i+k, s+k}. \quad (1.22)$$

Using this approximation, the second order Rényi entropy is [21, 23]

$$K_2(l) = -\frac{1}{l \Delta t} \ln(p^c(l)) = -\frac{1}{l \Delta t} \ln \left(\frac{1}{N^2} \sum_{t,s=1}^N \prod_{k=0}^{l-1} \mathbf{R}_{t+k,s+k} \right), \quad (1.23)$$

where $p^c(l)$ is the probability to find a diagonal of *at least* length l in the RP.

On the other hand, the l -dimensional correlation sum can be used to define K_2 [51]. This definition of K_2 can also be expressed by means of RPs and yields the relationship [23]

$$\ln p^c(l) \sim \varepsilon^{D_2} e^{-K_2(\varepsilon)\tau}. \quad (1.24)$$

D_2 is the correlation dimension of the system under consideration [52]. K_2 can be estimated from the slope of a logarithmic plot of $p^c(l)$ which corresponds to $-K_2 \Delta t$ for large l .

From the explanations above we have seen that the RP can be used to estimate probabilities, where the neighbourhood around a state \mathbf{x}_i corresponds to a certain binning. Based on this considerations, we can pursue in this direction and also calculate further measures which use probabilities, as, e.g., the generalised mutual information.

The mutual information quantifies the amount of information that we obtain from the measurement of one variable on another. It has become a widely applied measure to quantify dependencies within or between time series (auto and cross mutual information). The time delayed generalised mutual information (redundancy) $I_q(\tau)$ of a system \mathbf{x}_i is defined by [50]

$$I_q^{\mathbf{x}}(\tau) = 2H_q - H_q(\tau). \quad (1.25)$$

H_q is the q th-order Rényi entropy of \mathbf{x}_i and $H_q(\tau)$ is the q th-order joint Rényi entropy of \mathbf{x}_i and $\mathbf{x}_i + \tau$:

$$H_q = -\ln \sum_i p^q(i), \quad H_q(\tau) = -\ln \sum_{i,j} p^q(i, j; \tau), \quad (1.26)$$

where $p(i)$ is the probability that the trajectory visits the i th box and $p(i, j; \tau)$ is the joint probability that the trajectory is first in box i and after some time τ in box j . Hence, for the case $q = 2$ we can use the RP to estimate H_2

$$H_2 = -\ln \left(\frac{1}{N^2} \sum_{i,j=1}^N \mathbf{R}_{i,j} \right) \quad (1.27)$$

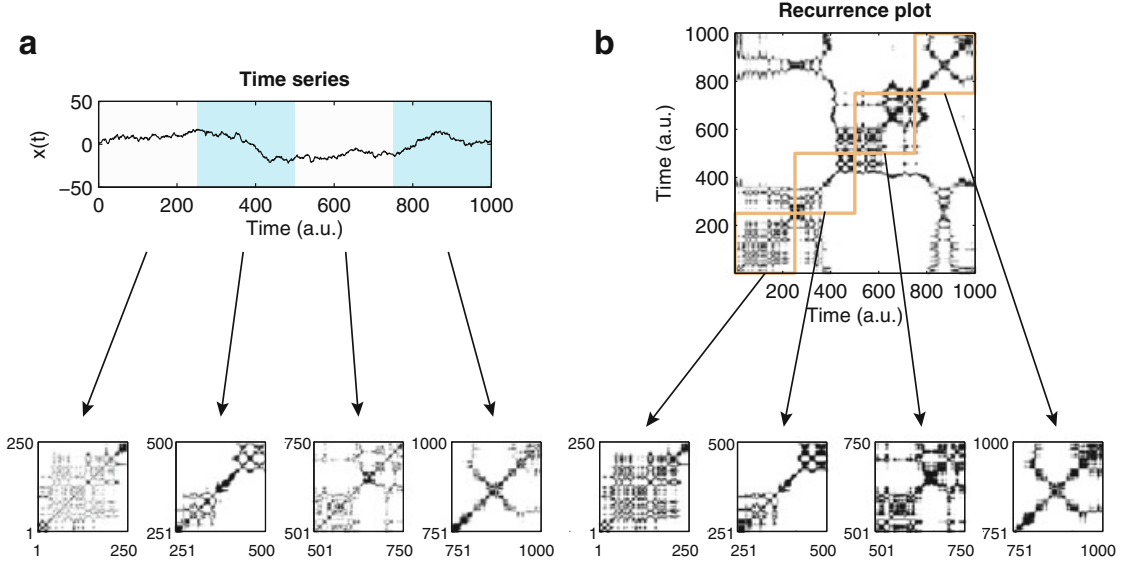


Fig. 1.5 Two possibilities of windowed RQA: **(a)** windowing of time series and **(b)** windowing of RP. The example is an auto-regressive process, Eq. (1.33), the RP is calculated using a constant number of neighbours (10 % of all points) and without embedding. The sub-RPs at the *bottom* clearly demonstrate the differences between the two approaches

and $H_q(\tau)$

$$H_2(\tau) = -\ln \left(\frac{1}{N^2} \sum_{i,j=1}^N \mathbf{R}_{i,j} \mathbf{R}_{i+\tau,j+\tau} \right) = -\ln \left(\frac{1}{N^2} \sum_{i,j=1}^N \mathbf{J}\mathbf{R}_{i,j}^{\mathbf{x},\mathbf{x}}(\tau) \right), \quad (1.28)$$

where $\mathbf{J}\mathbf{R}_{i,j}(\tau)$ denotes the delayed joint recurrence matrix (see Sect. 1.4.2). Then, the second order generalised mutual information can be estimated from a RP by [23]

$$I_2^{\mathbf{x}}(\tau) = \ln \left(\frac{1}{N^2} \sum_{i,j=1}^N \mathbf{J}\mathbf{R}_{i,j}^{\mathbf{x},\mathbf{x}}(\tau) \right) - 2 \ln \left(\frac{1}{N^2} \sum_{i,j=1}^N \mathbf{R}_{i,j} \right). \quad (1.29)$$

A comprehensive description and explanation on estimating entropies from recurrence plots is given in Chap. 2.

1.3.6 Windowing Techniques

RQA is powerful for the analysis of slight changes and transitions in the dynamics of a complex system. For this purpose we need a time-dependent RQA (a RQA series) what can be realised in two ways (Fig. 1.5):

1. The RP is covered with small overlapping windows of size w spreading along the LOI and in which the RQA will be calculated, $R_{i,j}$ ($i, j = k, \dots, k + w - 1$).

2. The time series (or phase space trajectory) is divided into overlapping segments x_i ($i = k, \dots, k+w-1$) from which RPs and subsequent RQA will be calculated separately.

Such time dependent approach can be used to analyse the stationarity of the dynamical system or dynamical transitions, like period-chaos or chaos-chaos transitions.

Here we should note the following important points. The time scale of the RQA values depends on the choice which point in the window should be considered as the corresponding time point. Selecting the first point k of the window as the time point of the RQA measures allows to directly transfer the time scale of the time series to the RQA series. However, the window reaches into the future of the current time point and, thus, the RQA measures represent a state which lies in the future. Variations in the RQA measures can be misinterpreted as early signs of later state transitions (like a prediction). A better choice is therefore to select the centre of the window as the current time point of the RQA. Then the RQA considers states in the past and in the future. If strict causality is required (crucial when attempting to detect subtle changes in the dynamics just prior the onset of dramatic state changes), it might be even useful to select the end point of the window as the current time point of the RQA (using embedding we have to add $(m-1)\tau-1$). For most applications the centre point should be appropriate.

The different windowing methods (1) or (2) are only equivalent when we do not normalise the time series (or its pieces) from which the RP is calculated and when we chose a fixed threshold recurrence criterion. Both approaches (1) and (2) can be useful and depend on the given question. If we know that the time series shows some nonstationarities or trends which are not of interest, then approach (2) can help to find transitions neglecting these nonstationarities. But, if we are interested in the detection of the overall changes (e.g., to test for nonstationarity), we should keep the numerical conditions for the entire available time constant and chose approach (1). Anyway, for each RQA we should explicitly state how the windowing procedure has been performed.

For the choice of the window size we have to consider the following fact. Because the RQA measures are statistical measures derived from histograms, the window should be large enough to cover a sufficient number of recurrence lines or orbits. A too small window can pretend strong fluctuations in the RQA measures just by weak statistical significance (e.g., the RQA measure *TND* is very sensitive to the window size and can reveal even contrary results. Therefore, chosen window size has to be checked carefully and conclusions about nonstationarity or dynamical transitions have to be validated by significance tests [20, 53].

1.3.7 Remark on Significance

Variations in these measures of complexity are mostly of relative nature. In order to not draw wrong conclusions from non-significant variations, a statistical test is

strongly suggested. Two approaches have been recently suggested, both basing on bootstrapping recurrence features (e.g., diagonal lines). The first approach estimates the local variation of a RQA measure within one window by bootstrapping the recurrence structures in the current window and calculating an empirical test distribution for the RQA measures for each sliding window separately [25]. This approach results in a confidence band around the varying RQA measure and can help when comparing the RQA variation of several studied time series. The second approach merges the local distributions, e.g., of the diagonal lines $H_D(i)$, Eq. (1.9), of the sliding windows i by $\hat{H}_D = \sum_i H_D(i)$, and then bootstrapping from this “global” distribution the recurrence structures [53]. The bootstrapped recurrence structures are used to calculate the RQA measures, finally providing one empirical test distribution of the current RQA measure for the entire time series. This can be used to estimate a confidence level for the variation of the RQA measure in order to check whether a transition is significant or not (see the confidence levels in Fig. 1.6).

1.3.8 Example: Rössler System with Regime Transitions

To illustrate the potential of recurrence quantification we consider the Rössler system, Eq. (1.35), [54], with $a = b = 0.25$ and continuously changing parameter c . We integrate the equations with a Runge–Kutta of fourth order of a time period of 2,200 s and with a sampling time of $\Delta t = 0.05$. We remove the first 1,000 values in order to remove transients, i.e., the resulting time series have a length of $N = 43,001$. With each time step, c increases by 0.004, resulting in a range of $c \in [35.2, 43.8]$. Within this interval, the system exhibits transitions from chaotic to periodic and back to chaotic states, e.g., a periodic window around $c \in [36.5, 37.5]$ (Fig. 1.6a), and an unstable periodic orbit (UPO) at $c \approx 41$.

The RQA measures are calculated from the x -component of the system, using an embedding with dimension $m = 3$ and delay $\tau = 6$, a fixed recurrence rate of $RR = 0.05$, and a minimal line length of $d_{\min} = 10$.

In order to calculate the RQA measures within each window (of window size 1,500 and 20 % overlap), we had to calculate the local distributions of recurrence structures, e.g., of the diagonal lines $H_D(i)$. These local distributions are merged together and used for a bootstrapping based estimation of a test distribution of the RQA measures, using 1,000 samplings and a confidence level of 99 % [53].

From the x -component of the Rössler system we calculate the RQA measures DET , $\langle D \rangle$, ENT , as well as the mean recurrence time T_2 [35], and the transitivity coefficient \mathcal{T} [47]. The deterministic character of the system is well reflected by high values of DET for the complete time interval (Fig. 1.6b), also revealing that the RP mainly consists of diagonal lines. The confidence interval clearly indicates that the variation of DET is not significantly differing over the time. Without such statistical test we might be inveigled to interpret the variations as real and draw, therefore, wrong conclusions about the transitions in the system. During the periodic regime the diagonal lines are spanning the whole RP, leading to an increase of

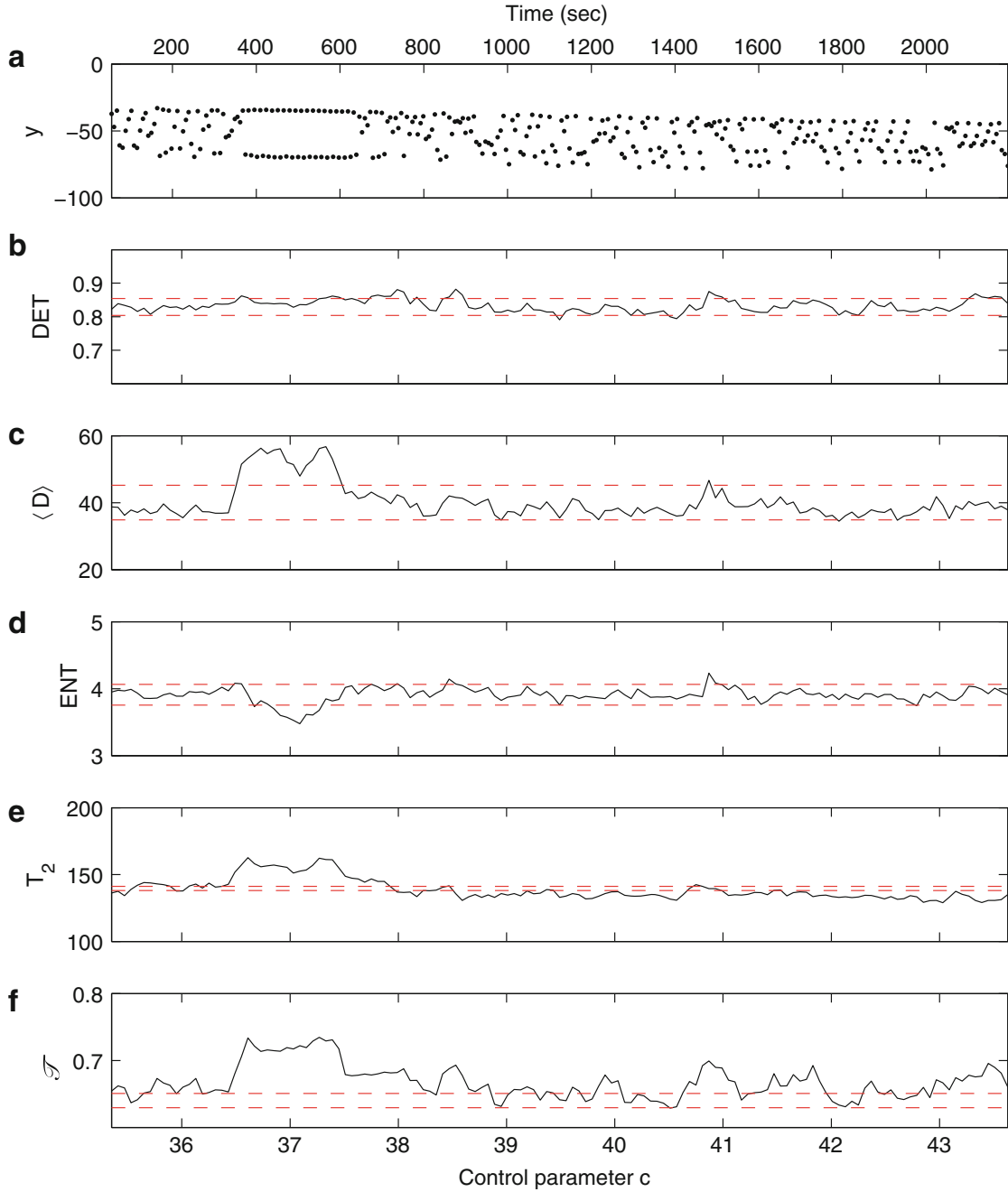


Fig. 1.6 (a) Poincaré section of the y -component of the Rössler system at $x = 0$. A periodic window at time $t \approx 400, \dots, 600$, corresponding to the control parameter range around $c \in [36.5, 37.5]$ is clearly visible. An unstable periodic period (UPO) appears at $c \approx 41$. The RQA measures are calculated within moving windows of size $w = 1,500$ with an overlap of 20%. (b) The RQA measure DET reveals very high values for the entire time period due to the deterministic nature of the system. In the periodic window, (c) the mean diagonal line length $\langle D \rangle$ has high values, whereas (d) the measure ENT decreases. Both, $\langle D \rangle$ and ENT have increased values at the UPO. (e) T_2 reveals the longer recurrence times during periodic dynamics and (f) the transitivity coefficient \mathcal{T} measures regular dynamics, as present in the periodic window and UPO. A surrogate test was applied to mark the 99 % confidence interval

the mean diagonal line length $\langle D \rangle$ during this period (Fig. 1.6c). An UPO also causes many diagonal lines with increased length and a reduction of short lines, yielding increased $\langle D \rangle$. The distribution of the line lengths in the periodic window is less complex than during chaotic dynamics. Therefore, the entropy of the line length distributions ENT decreases in the periodic window, but increases at the UPO, because the UPO duration is smaller than the window length (Fig. 1.6d). The mean recurrence time also increases during the periodic regime and the UPO (Fig. 1.6e). The transitivity coefficient \mathcal{T} can distinguish regular from irregular dynamics [46, 55], and thus, reveals the periodic window and the UPO due to the more regular dynamics than during the other chaotic regimes.

1.4 Bivariate Extensions of Recurrence Analysis

Bivariate recurrence analysis allow the study of correlations, couplings, coupling directions, or synchronization between dynamical systems. Depending on the purpose and application, there are two major directions for such bivariate extension: the *cross recurrence plot* and the *joint recurrence plot*.

1.4.1 Cross Recurrence Plot

The *cross recurrence plot (CRP)* is a bivariate extension of the RP and was introduced for the investigation of the simultaneous evolution of two different phase space trajectories, allowing the study of dependencies between two different systems [56, 57]. Suppose we have two dynamical systems, each one represented by the trajectories \mathbf{x}_i and \mathbf{y}_i in the same d -dimensional phase space (Fig. 1.7a). We find the corresponding cross recurrence matrix (Fig. 1.7b) by computing the pairwise mutual distances between the phase vectors of the two systems:

$$CR_{i,j}^{\mathbf{x},\mathbf{y}}(\varepsilon) = \Theta(\varepsilon - \|\mathbf{x}_i - \mathbf{y}_j\|), \quad i = 1, \dots, N, \quad j = 1, \dots, M, \quad (1.30)$$

where the length of the trajectories of \mathbf{x} and \mathbf{y} is not required to be identical, hence, the matrix \mathbf{CR} is not necessarily square. Note that both systems are represented in the same phase space, because a CRP looks for those times when a state of the first system recurs to one of the other system. Using experimental data, it is sometimes difficult to reconstruct the phase space. If the embedding parameters are estimated from both time series, but are not equal, the higher embedding should be chosen. However, the data under consideration should be from the same (or a very comparable) process and, actually, should represent the same observable. Therefore, the reconstructed phase space should be the same. The components of \mathbf{x}_i and \mathbf{y}_i are usually normalised before computing the cross recurrence matrix, in order to make both systems comparable.

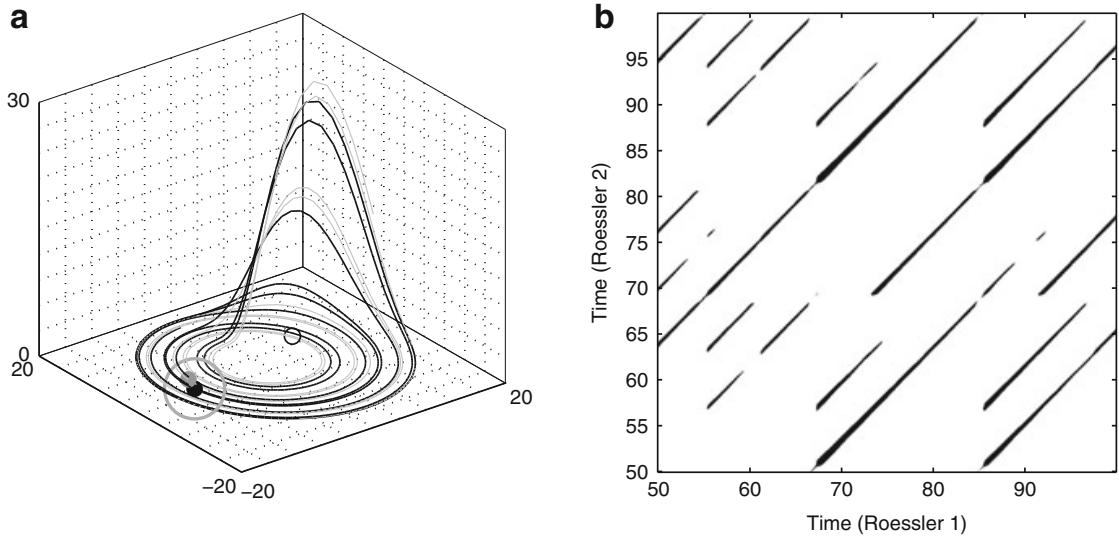


Fig. 1.7 (a) Phase space trajectories of two coupled Rössler systems, Eqs. (1.36) and (1.37), with $a = 0.15$, $b = 0.20$, $c = 10$, $\nu = 0.015$ and $\mu = 0.01$ by using their three components (black and grey line correspond to the first and second oscillator). In (b) the corresponding CRP is shown (Euclidean norm and $\varepsilon = 3$ is used). (a) If a phase space vector of the second Rössler system at j (grey point on the grey line) falls into the neighbourhood (grey circle) of a phase space vector of the first Rössler system at i , in the CRP (b) at the location (i, j) a black point will occur

Since the values of the main diagonal $CR_{i,i}$ ($i = 1, \dots, N$) are not necessarily one, there is usually not a black main diagonal (Fig. 1.7b). Apart from that, the statements given in the subsection about all the structures in RPs (Sect. 1.2.2) hold also for CRPs. The lines which are diagonally oriented are here of major interest too. They represent segments on both trajectories, which run parallel for some time. The frequency and length of these lines are obviously related to a certain similarity between the dynamics of both systems. A measure based on the lengths of such lines can be used to find nonlinear interrelations between two systems, which cannot be detected by the common cross-correlation function [57].

An important advantage of CRPs is that they reveal the local difference of the dynamical evolution of close trajectory segments, represented by bowed lines. A time dilatation or time compression of one of the trajectories causes a distortion of the diagonal lines [58]. A time shift between the trajectories causes a dislocation of the LOS. Hence, the LOS may lie rather far from the main diagonal of the CRP.

1.4.2 Joint Recurrence Plot

If we ask whether two systems have a similar recurrence structure, i.e., their states recur in a simultaneous way, we will use the *joint recurrence plot* [59, 60]. Here we consider the *recurrences* of the trajectories of the two systems in their respective phase spaces separately and look for the times when both of them recur simultaneously, i.e. when a *joint recurrence* occurs. By means of this approach,

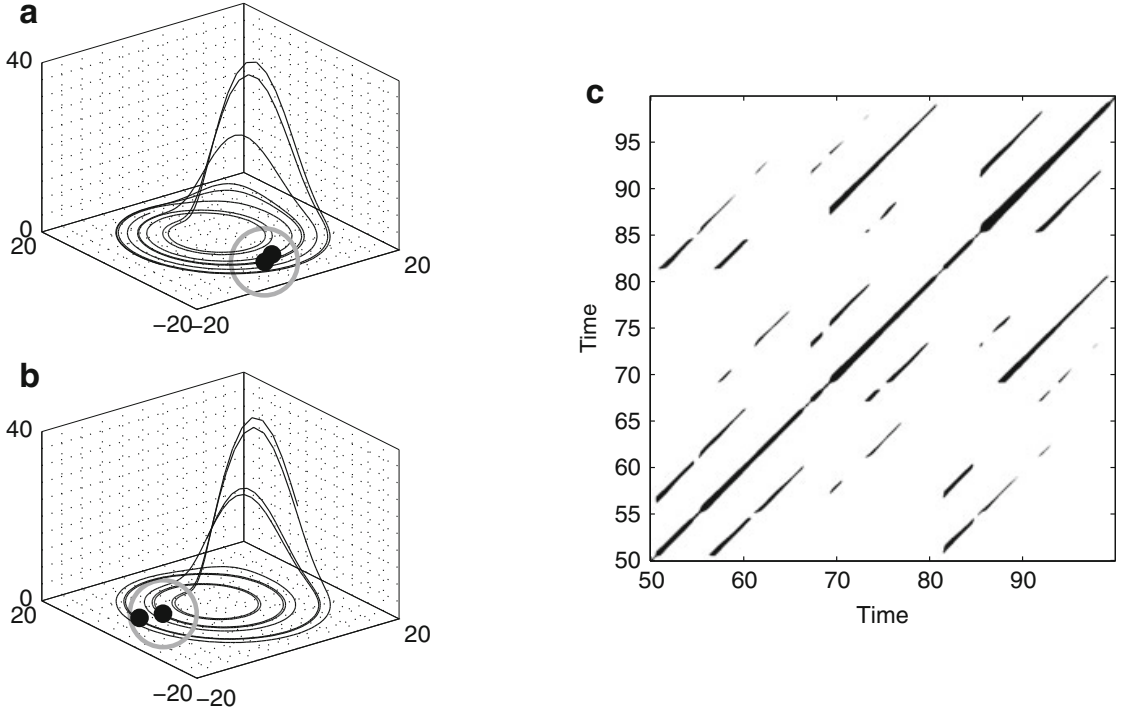


Fig. 1.8 (a, b) Phase space trajectories of two coupled Rössler systems, Eqs. (1.36) and (1.37), with $a = 0.15$, $b = 0.20$, $c = 10$, $\nu = 0.015$ and $\mu = 0.01$. In (c) the corresponding JRP is shown (L_2 norm and $\varepsilon = 5$ is used for both systems). If two phase space vectors of the second Rössler system at i and j are neighbours [black points in (b)] and if two phase space vectors of the first Rössler system at same i and j are also neighbours [black points in (a)], a black point in the JRP at the location (i, j) will occur

the individual phase spaces of both systems can be used that could have even different embedding dimension. Furthermore, two different thresholds for each system, ε^x and ε^y , can be considered, so that the criteria for choosing the threshold (Sect. 1.2.1) can be applied separately, respecting the natural measure of both systems. The *joint recurrence matrix* (Fig. 1.8) for two systems \mathbf{x} and \mathbf{y} is then the element wise product of the single RPs

$$JR_{i,j}^{x,y}(\varepsilon^x, \varepsilon^y) = \Theta(\varepsilon^x - \|\mathbf{x}_i - \mathbf{x}_j\|) \Theta(\varepsilon^y - \|\mathbf{y}_i - \mathbf{y}_j\|), \quad i, j = 1, \dots, N. \quad (1.31)$$

In this approach, a recurrence will take place if a point \mathbf{x}_j on the first trajectory returns to the neighbourhood of a former point \mathbf{x}_i , and *simultaneously* a point \mathbf{y}_j on the second trajectory returns to the neighbourhood of a former point \mathbf{y}_i . That means, that the joint probability that both recurrences (or n recurrences, in the multidimensional case) happen simultaneously in their respective phase spaces are studied. In such a definition of a recurrence it is not necessary that the recurrence occurs at same states of the considered systems.

The JRP is invariant under permutation of the coordinates in one or both of the considered systems.

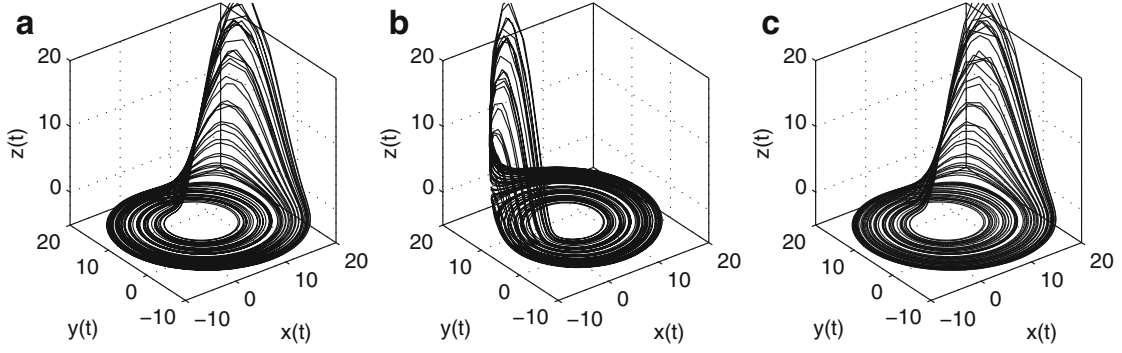


Fig. 1.9 (a) Phase space trajectory of the Rössler system [Eqs. (1.35)], with $a = 0.15$, $b = 0.2$ and $c = 10$. (b) Same trajectory as in (a) but rotated on the z -axis by $\frac{3}{5}\pi$. (c) Same trajectory as in (a) but time scale transformed by $\tilde{t} = t^2$

Moreover, a delayed version of the joint recurrence matrix can be introduced

$$JR_{i,j}^{x,y}(\varepsilon^x, \varepsilon^y, \tau) = R_{i,j}^x(\varepsilon^x) R_{i+\tau, j+\tau}^y(\varepsilon^y), \quad i, j = 1, \dots, N - \tau, \quad (1.32)$$

which is very useful for the analysis of interacting delayed systems (e. g. for lag synchronisation).

The JRP can be used to estimate joint recurrence probabilities and even conditional recurrence probabilities [61, 62], what is useful for the study of coupling directions (Chap. 3).

1.4.3 Comparison Between CRPs and JRPs

In order to illustrate the difference between CRPs and JRPs, we consider the phase space trajectory of the Rössler system, Eq. (1.35), in three different situations: the original trajectory (Fig. 1.9a), the trajectory rotated on the z -axis (Fig. 1.9b) and the trajectory under a parabolic stretching/compression of the time scale (Fig. 1.9c). These three trajectories look very similar; one of them is rotated and the other one contains another time parametrisation (but looks identical to the original trajectory in phase space).

At first, let us look at the RPs of these three trajectories. The RP of the original trajectory is identical to the RP of the rotated one, as expected (Fig. 1.10a, b). The RP of the stretched/compressed trajectory looks different than the RP of the original trajectory (Fig. 1.10c): it contains bowed lines, as the recurrent structures are shifted and stretched in time with respect to the original RP.

Now we calculate the CRP between the original trajectory and the rotated one (Fig. 1.10d) and observe, that it is rather different from the RP of the original trajectory (Fig. 1.10a). This is because in the CRP the difference between each pair of vectors is computed, and this difference is not invariant under rotation of

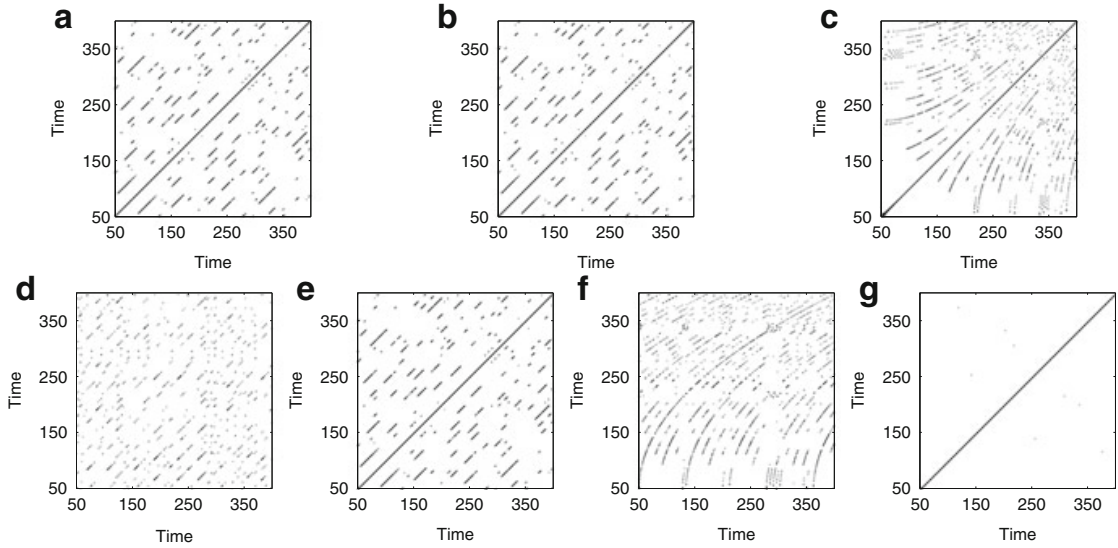


Fig. 1.10 RPs of the (a) original trajectory of the Rössler system, (b) of the rotated trajectory and (c) of the stretched/compressed trajectory. (d) CRP and (e) JRP of the original and rotated trajectories and (f) CRP and (g) JRP of the original and stretched/compressed trajectories. The threshold for recurrence is $\varepsilon = 1$

one of the systems. Hence, a rotation of the reference system of one trajectory changes the CRP. Therefore, the CRP cannot detect that both trajectories are identical up to a rotation. In contrast, the JRP of the original trajectory and the rotated one (Fig. 1.10e) is identical to the RP of the original trajectory (Fig. 1.10a). This is because the JRP considers joint recurrences, i. e. recurrences which occur simultaneously in both systems, and they are invariant under affine transformations.

The CRP between the original trajectory and the stretched/compressed one contains the bowed LOS, which reveals the functional shape of the parabolic transformation of the time scale (Fig. 1.10f). Note that the CRP represents the times at which both trajectories visit the same region of the phase space. On the other hand, the JRP of these trajectories is almost empty (Fig. 1.10g) because the recurrence structure of both systems is now different. Both trajectories have different time scales, and hence, there are almost no joint recurrences. Therefore, the JRP is not able to detect the time transformation applied to the trajectory, even though the shape of the phase space trajectories is very similar.

To conclude we can state that CRPs are more appropriate to investigate relationships between the parts of the same system which have been subjected to different physical or mechanical processes, e. g., two borehole cores in a lake subjected to different compression rates. On the other hand, JRPs are more appropriate for the investigation of two interacting systems which influence each other, and hence, adapt to each other, e. g., in the framework of phase and generalised synchronisation or causal couplings (see Chap. 3).

1.5 Computational Foundations of Recurrence Quantification Analysis

1.5.1 *Brief Historical Background*

The mathematical concept of recurrences traces back to Feller (1950) [63] and Poincaré (1890) [13] and has direct application to dynamical systems. Eckmann et al. (1987) [15] incorporated these ideas into the qualitative tool, the recurrence plot. A few years later Zbilut and Webber [28, 36] quantified the recurrence plot and introduced the concept of recurrence quantifications by defining five recurrence variables: *recurrence rate*, *determinism*, *max diagonal line*, *line entropy*, and *trend*. A decade later Marwan [44] added three new recurrence variables: *laminarity*, *max vertical line*, and *trapping time*. The details of these measures of complexity have already been discussed in depth in Sect. 1.3 of this chapter. To make long story short, from these three foundational papers has grown up a large and vast literature across fields and five international symposia on recurrence plots conducted every other year from 2005 through 2013 respectively in Potsdam, Siena, Montreal, Hong Kong, and Chicago, also reflected by the impressive growing list of recurrence papers at the webpage <http://www.recurrence-plot.tk/>.

The generation of recurrence plots and recurrence quantifications requires high capacity and high speed computers. As the speed of machines has increased dramatically over the last four decades [64], so too has the ease of computation of distance and recurrence matrices. The authors are very familiar with various programming languages and have implemented computational strategies in their doctoral dissertations [65, 66]. One of the authors (Webber) has programmed recurrence algorithms using the C language [67] for the disk operating system (DOS) [68] and the other (Marwan) has programmed the algorithms as a MATLAB Toolbox [69]. A third popular format has been devised for MS Windows-based operating systems by Kononov [70].

1.5.2 *Computational Strategies*

The purpose of this section is to provide a generalized flow chart for RQA calculations covering three principle categories: recurrence plots (RQD, KRQD, JRQD), recurrence quantifications (RQS, KRQS, JRQS), and recurrence windows (RQE, KRQE, JRQE). Each class of programs will be covered separately. The ubiquitous flow chart for all these approaches is shown in Fig. 1.11.

First, the program starts (S) by receiving a single input data vector for auto-recurrences or dual input data vectors for cross-recurrences. These data are in the standard ASCII format (numeric codes devoid of any alphabetic codes). Second, RQA parameters are entered either manually or from automatically from a

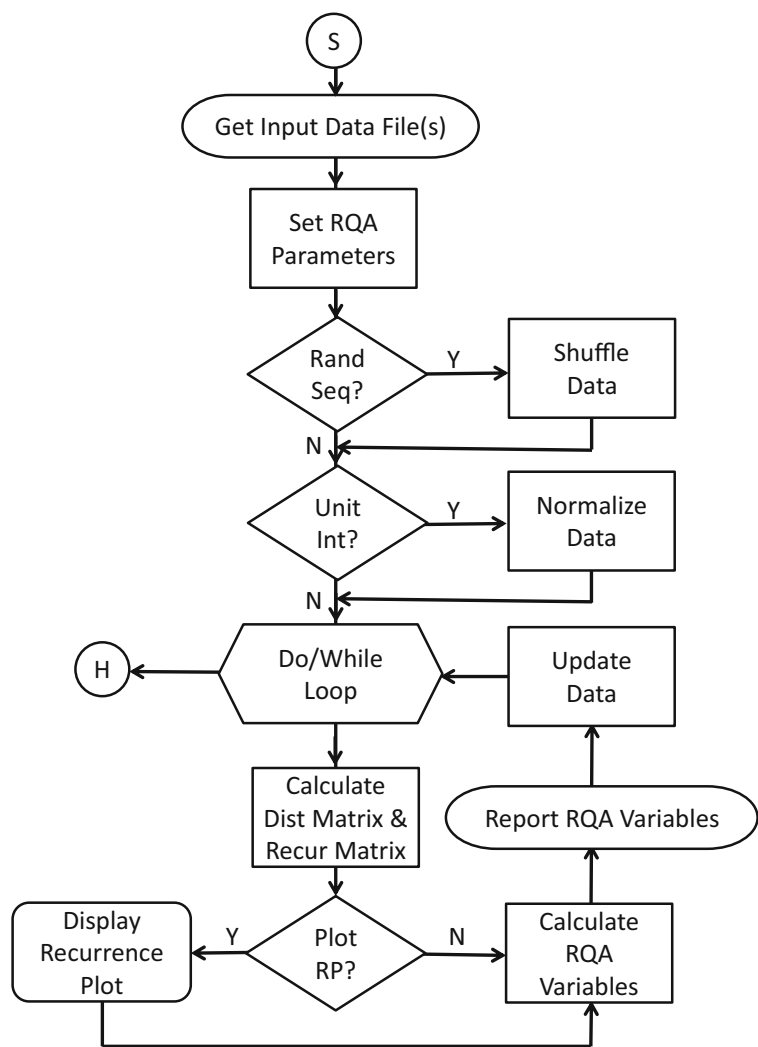


Fig. 1.11 Generalized flow chart for all recurrence computations

parameter file. Third, the ordering of the input data set within the recurrence window can either be retained or shuffled to destroy correlative coupling within the data stream (randomized control). Fourth, the data set can be left alone or normalized over the unit interval. Fifth, the computation enters a do/while loop in which the distance and the thresholded recurrence matrices are computed. Here the distance matrix can be rescaled to the maximum distance or mean distance of the matrix or left in absolute distance units. Sixth, the recurrence plot can be displayed or skipped prior to the calculation of RQA variables. Seventh, the RQA variables are reported to the screen and/or output file in ASCII format. Eighth, the data can be updated by the selection of new parameters or by shifting the recurrence window to a new segment of the input data. Ninth, the looping halts (H) either by manual interruption or automatic exiting depending upon the input data length.

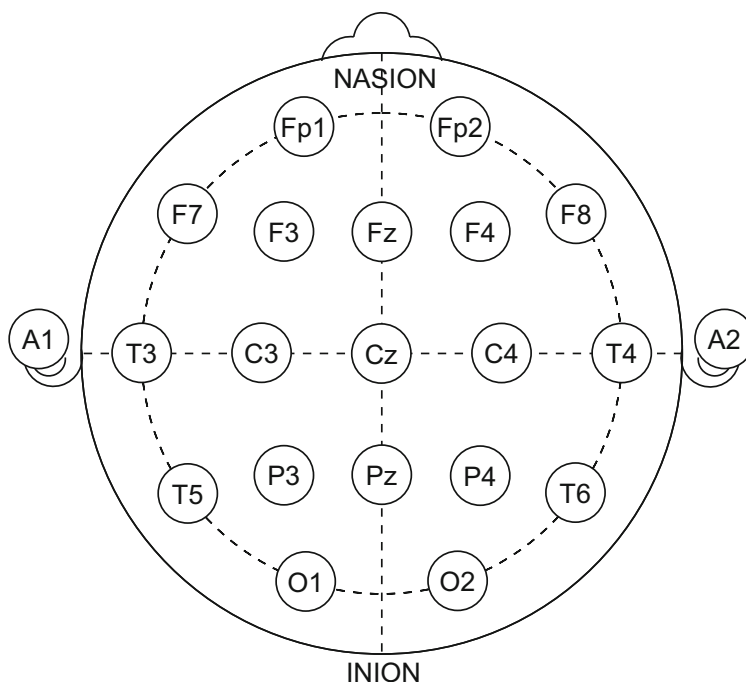


Fig. 1.12 Electrode placements in the 10–20 EEG system. Figure reproduced from Wikipedia (public source)

1.5.3 Example Program Runs

For demonstration purposes three brainwave signals from a single subject will be used to illustrate the operation of recurrence programs found within a larger suite of RQA programs [68].

Electroencephalographic (EEG) signals were digitized at 1,000 Hz in a patient using the standard 10–20 electrode system as diagrammed in Fig. 1.12 (e.g., see [71]). One-second traces of the waveforms from three sites (Fz, F3 and F7) are displayed in Fig. 1.13. Note that electrode FZ is closer to electrode F3 than electrode F7. These physical separations become important when comparing electrical activations in pairs (kross and joint recurrences) (e.g. near pair KZ-F3 versus distant pair KZ-F7).

1.5.3.1 Programs RQD, KRQD and JRQD

Individual RPs of the three EEG signals are shown in Fig. 1.14. Parameter settings for each time series were identical: window of 500 points (500 ms), embedding dimension of 10, delay of 50 (50 ms), Euclidean norm, maximum distance rescaling, radius of 30 % maximum distance and line of 2. The distribution of recurrent points in each plot are not homogeneous, showing a non-stationarity in the signals. The recurrence density and deterministic structuring for each signal are unique: FZ (7.428 and 96.903 %); F3 (2.833 and 86.333 %); F7 (4.336 and 89.832 %).

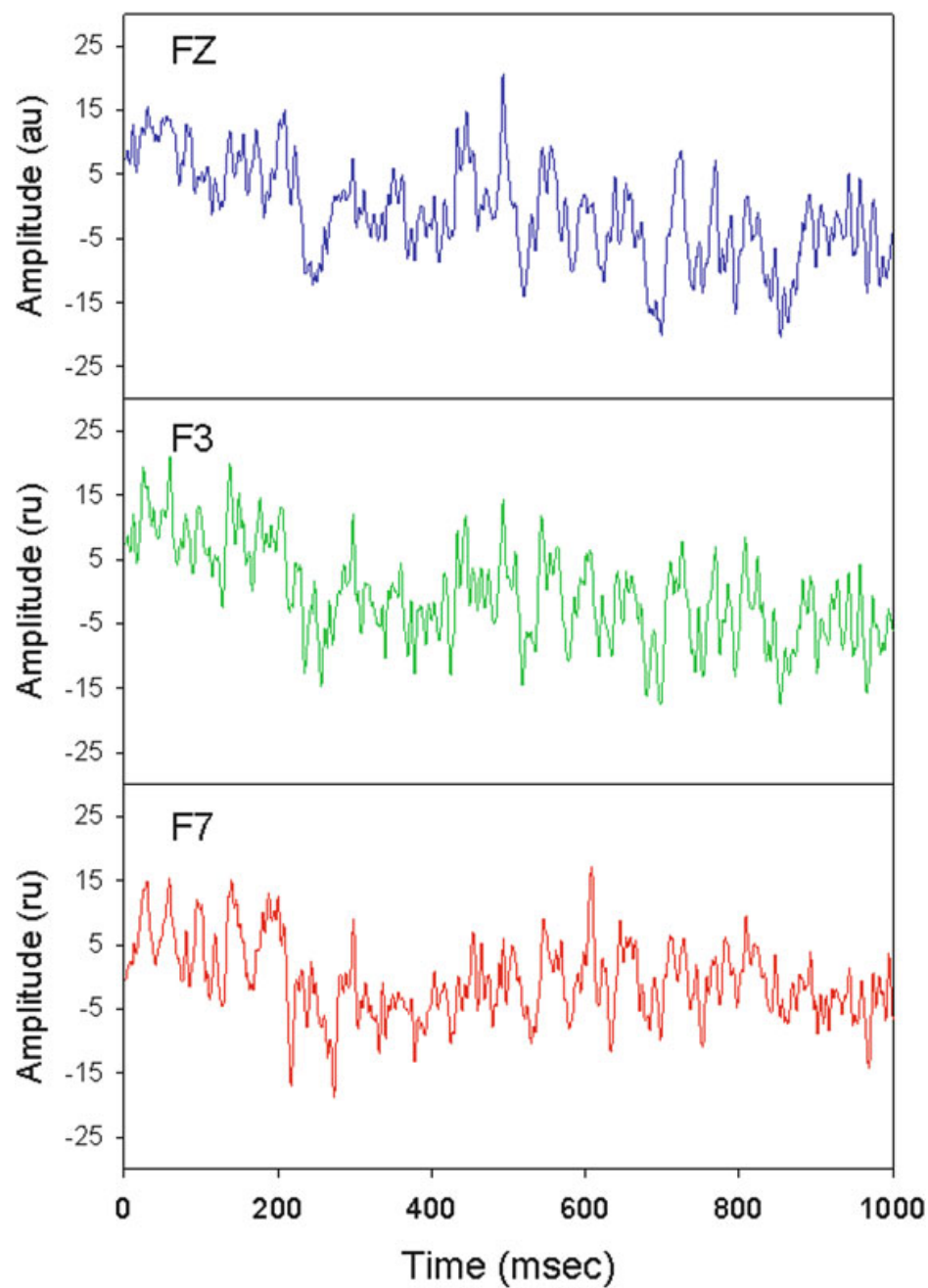
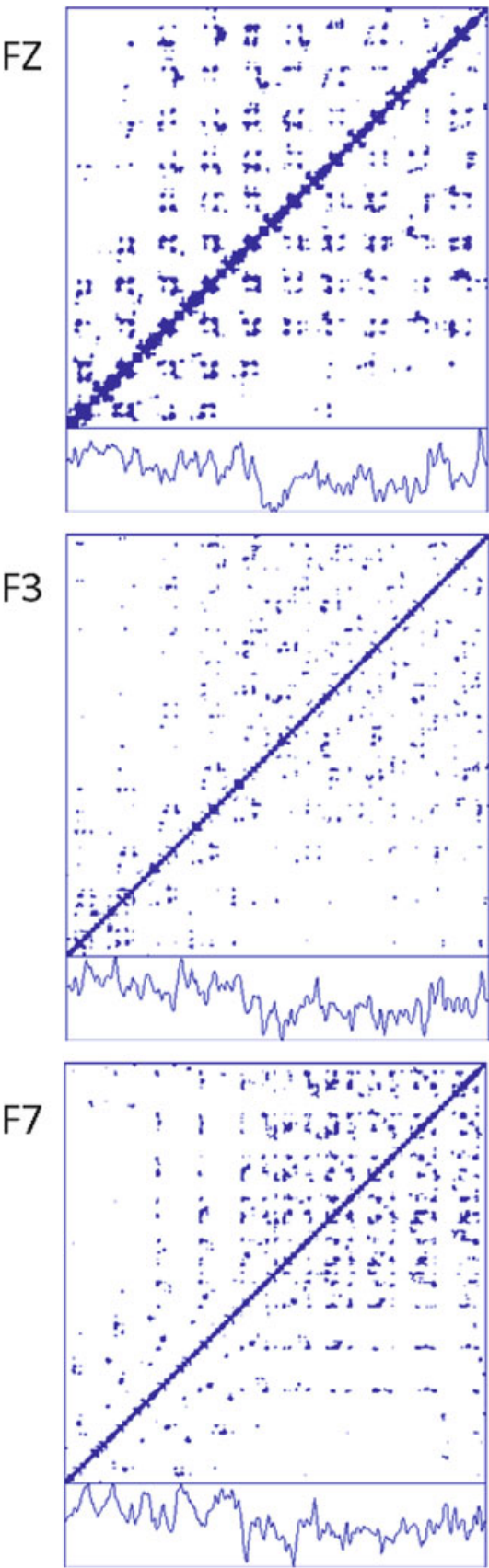


Fig. 1.13 One-second recordings from EEG electrode positions FZ, F3 and F7 in a normal, healthy and resting human subject

To study how two signals are correlated in time, CRPs are shown in Fig. 1.15 (left). Program KRQD was run on paired signals using the same parameters settings as for the RPs. As can be seen the recurrences and determinisms for the two pairs are rather similar for this single window in time: FZ:L3 (4.769 and 92.997 %); FZ:F7 (4.252 and 91.552 %). This means that the location of the signals on the skull, near (FZ to F3) or distant (FZ to F7), cannot be discriminated.

Lastly, to study how two signals have a similar recurring behavior, JRPs are shown in Fig. 1.15 (right). Program JRQD was run on paired signals using the same parameter settings as for the auto and cross recurrences. Recurrent points in JRPs

Fig. 1.14 Recurrence plots of EEG signals at three brain sites, FZ, F3 and F7



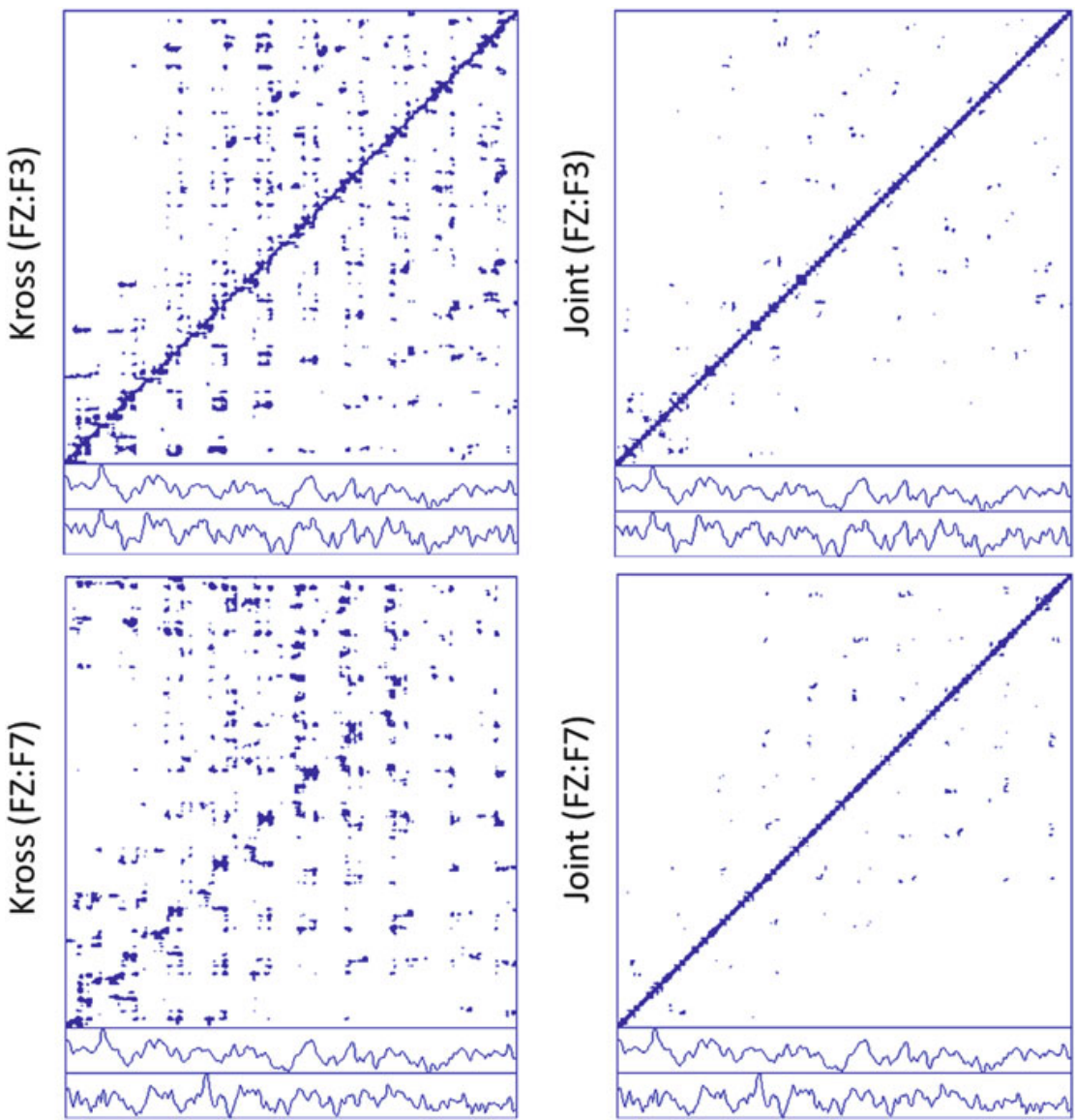


Fig. 1.15 Cross recurrence plots (*left*) and joint recurrence plots (*right*) for paired EEG signals FZ:F3 and FZ:F7

signify shared recurrent points in the RPs of the individual signals. For this reason, JRPs are necessarily symmetrical about the LOI as are the RPs. For the double pairs, the recurrence densities are lower than the auto recurrences and kross recurrences, but the deterministic structuring is still high: FZ:F3 (1.699 and 89.901 %); FZ:F7 (1.525 and 90.168 %). Again, the data at this time window do not discriminate on signal location.

1.5.3.2 Programs RQS, KRQS and JRQS

Implementation of RPs and RQA requires a familiarity with some characteristics of the dynamical system one is exploring. Some systems are best studied as flows

(continuous smooth fluctuations of time) whereas other systems are best studied as maps (discontinuous jumps in time). For example, the electrocardiogram (EEG) is a flow. But the series of time intervals from one zero-crossing the next is a map (literally a Poincare section of the EEG flow). But whether flows or maps, it is critical to set the RQA parameters appropriate for each signal. There are seven key parameters: window size, embedding dimension, delay between embedded points, type of norm (min, max or Euclidean), method of rescaling the distance matrix (absolute, mean distance, maximum distance), radius threshold, and line parameter (defining d_{\min} and v_{\min}). The proper selection of these parameters are describe earlier in this chapter as well as elsewhere [72].

The three programs of interest (RQS, KRQS and JRQS) are all scaling programs that conveniently increment four key RQA parameters and generate large matrices of recurrence quantifications. The user can select the range of points to be studied, the range of delays, range of embedding dimensions and the range of thresholds each of which is systematically incremented. The delay can be found using either the autocovariance function or minimal mutual information and held constant when running these scaling programs. For example, program RQS was run on EEG variable FZ on a single window of 500 points with a delay of 50 points while incrementing the embedding dimension from 1 to 40 and the radius threshold from 0 to 50. The output matrix contained sufficient data to plot three-dimensional surfaces for each of the 8 RQA variables as functions of embedding and radius. We show only the topology of the entropy variable in Fig. 1.16 over the embedding and radius parameter space. Three-dimensional graphs such as these for determinism, for example, can be used for visual selection the radius threshold and embedding space [68].

1.5.3.3 Programs RQE, KRQE and JRQE

One of the most useful applications of recurrence quantifications is to examine long time series of data using a small moving window traversing the data. For example, in retrospective studies it is possible to study subtle shifts in dynamical properties just before a large event occurs. One can ask questions like [44, 72]: What RQA quantifications change just prior to a brain seizure or heart fibrillation? To show the dynamical richness of EEG signals, we used program RQE to compute recurrence variables with 838 sliding windows. Each window was 500 points line 500 ms) with starting times offset by only 5 points (5 ms) giving exactly 99 % overlap for adjacent windows (Fig. 1.17). The other RQA parameters were selected as before for these same EEG signals.

This windowed process is also fully applicable to quantifications derived from CRPs and JRP as shown in Figs. 1.18 and 1.19. Two things are to be noted from these plots. First, running determinism values in cross and joint recurrences do not remain fixed or constant over time, by show complex rhythms. For example, the cross recurrence picks up a 1 Hz rhythm with repeating (recurring) nadirs or dips in determinism every 1,000 ms. Running determinism values in joint recurrences,

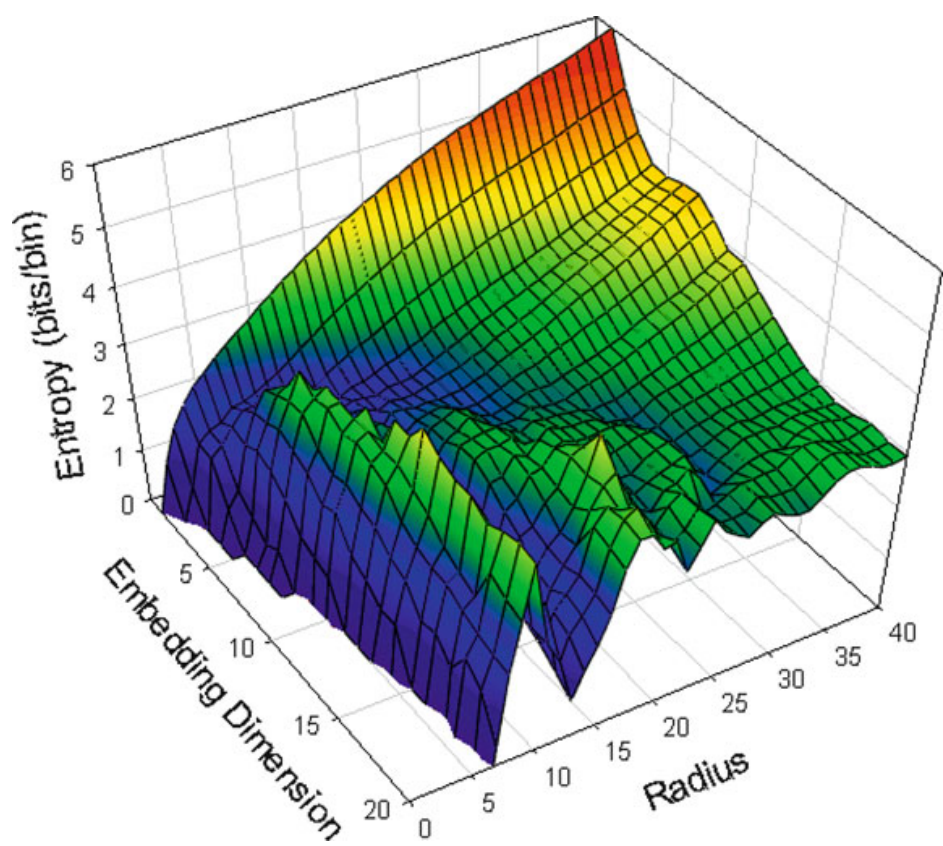


Fig. 1.16 Topology of RQA variable entropy in parameter space of radius and embedding dimension

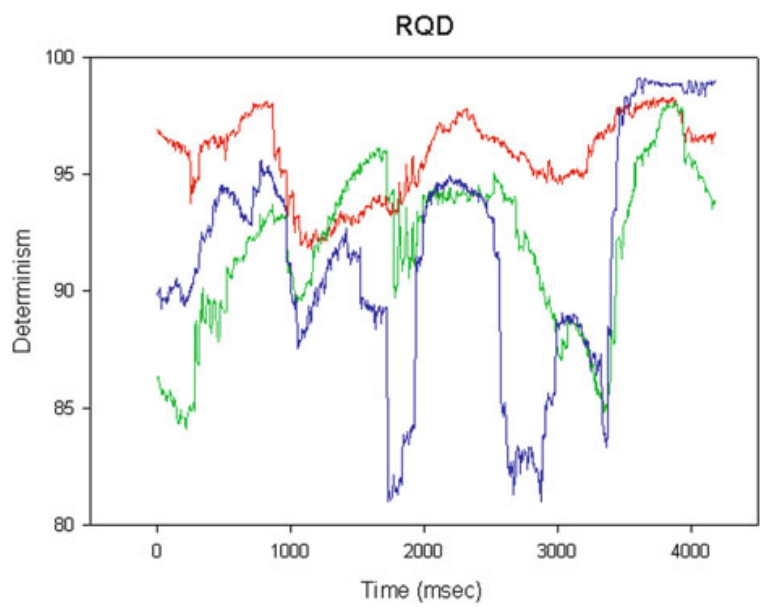


Fig. 1.17 Sensitivity of RQA variable determinism within a 500-point (500 ms) sliding window for EEG signals recorded at sites FZ (*red*), F3 (*green*) and F7 (*blue*). The dynamical complexities shown correspond to unknown processes in place for this resting human subject

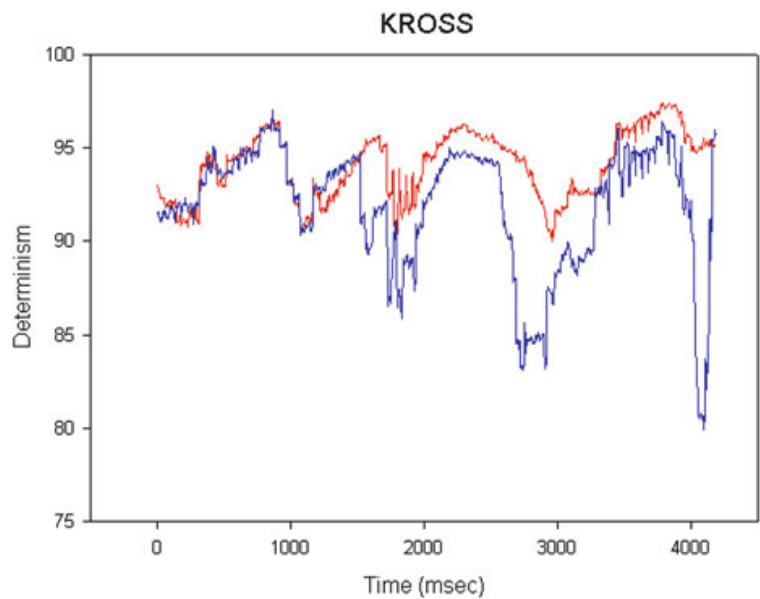


Fig. 1.18 Sensitivity of RQA variable determinism within a 500-point (500 ms) sliding window for paired EEG signals recorded at sites FZ, F3 and F7. The RQA cross correlations are performed for the FZ:F3 close pair (*red*) and Fz:F7 distant pair (*blue*). Both pairs show similar dynamical correlations within the first 2,000 ms, but soon after there is a departure or bifurcation in the dynamics in which the distant pairing loses deterministic coupling as compared to the close pairing of electrodes

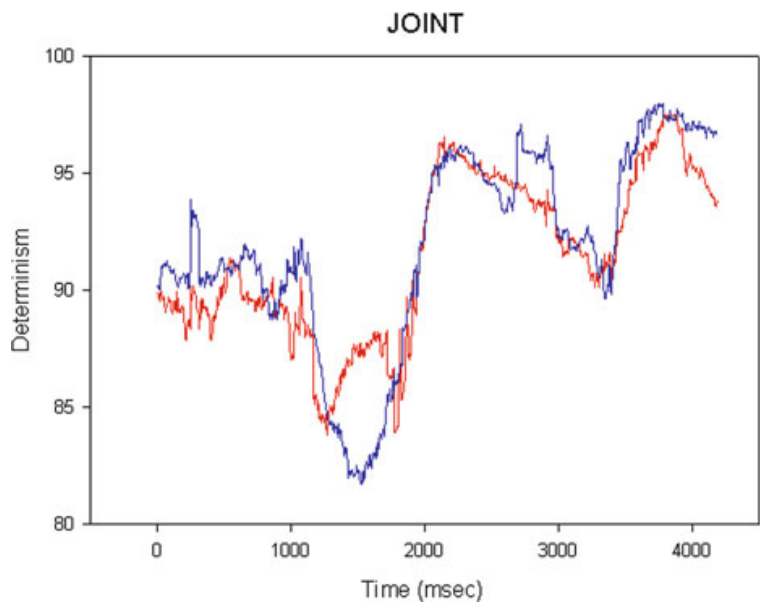


Fig. 1.19 Sensitivity of RQA variable determinism within a 500-point (500 ms) sliding window for paired EEG signals recorded at sites FZ, F3 and F7. The RQA joint correlations are performed for the FZ:F3 close pair (*red*) and FX:F7 distant pair (*blue*). Both electrode pairings show similar dynamical correlations for most of the epochs save in the windows around 1,250–1,750 ms and 2,750–3,000 ms. These departures reveal dynamical bifurcations of the complex system of resting brainwaves recorded at specific sites

however, reveal a slower rhythm of about 0.5 Hz with repeating (recurring) nadirs or dips in determinism every 2,000 ms. Second, running determinism values for the different pairs of electrodes representing close sites (FZ:F3) versus distant (FZ:F7) sites sometimes track together and sometimes diverge from one another. These results give important hints regarding diverging (and converging) dynamics occurring over time. Remember that these examples come from a resting human subject. Maybe changes in attention or other state changes in this free-run mode are responsible for the dynamical bifurcations. The point is that windowed cross and joint recurrences gives the investigator powerful tools to study complex brain activities. Thus one might make similar EEG recordings and recurrence analyses during the performance of specific tasks. One wonders if this type of perspective could be allied to robotics in which human EEG patterns are used to control artificial limbs [73].

1.5.4 *Advanced Topics*

There are several advanced recurrence topics that will just be mentioned here because they are not well explored in terms of dynamical performance. Just as there are frequency spectra (linear) so too there are recurrence spectra (nonlinear) [39, 74] that can be studied as auto-spectra, cross spectra and joint spectra using programs RQF, KRQF and JRQF respectively. Also, histogram distribution of recurrence intervals (inverse of frequencies) can be studied with program RQI, KRQI and JRQI [72]. When comparing linear spectra with nonlinear spectra, the latter methodologies have a higher resolution and sensitivity that pick up subtleties missed by the former.

Windowed recurrences, whether auto, cross or joint all generate large matrices of data. The typical approach has been to examine those RQA variables which best diagnosis the system under study. Recurrence rate and determinism are two favorite variables, for example. However, instead of insisting that the investigator make the judgment call in this respect, the full matrix of data $[N, 8]$ where N is the number of epoch rows and eight is the number of RQA variables can be submitted to principal component analysis. Typically, the first three principal components account for 95 % or better of the variability in the data. So if one has 20 subjects in the study, 20 sets of three principal components will be produced. PC1, PC2 and PC3 can then be plotted as three-dimensional scatter plots to see if any bunching of points occurs. Cluster analysis can formalize the grouping of points which may be diagnostic for different patient types.

1.6 **Summary**

Recurrence plot (RP) and recurrence quantification analysis (RQA) are easily accessible tools for investigating the dynamics of complex systems. Since its introduction in 1987 as a visualization tool for revealing hidden rhythms, RP methodology

has not only been enriched by a heuristic quantification approach (RQA), but also advanced by a growing number of add-on applications. For example, RPs produce meaningful graphical displays relating theoretically founded measures of complexity (e.g., K2 entropy), complex network relationships, and even synchronized systems replete with coupling directions. Simplicity of implementation and wide applicability of RP and RQA technologies across diverse systems continue to attract and expand the utilization of these measures in a growing number of scientific fields. As more researchers utilize recurrence strategies on their particular systems of interest, the limitations and pitfalls of these nonlinear techniques are being better appreciated. Better understood also are the forthcoming results, their assessment and interpretation, and how they infer dynamical structuring or topological properties of complex systems. In short, recurrence analysis is a statistical tool which works remarkably well on non-linear, non-deterministic, non-stationary, noisy dynamical systems of short duration. Twenty seven years have passed since the foundational paper of Eckmann, Kamphorst, and Ruelle [15], yet recurrence analysis remains an active field with open questions and promising new directions which the following chapters in this book will illustrate remarkably well.

Appendix: Mathematical Models

To illustrate our statements we are using some prototypical model systems, which are listed below:

Auto-Regressive Process of 1st Order

$$x_i = 0.95x_{i-1} + 0.05x_{i-2} + 0.9\xi \quad \text{where } \xi \text{ is white Gaussian noise.} \quad (1.33)$$

Lorenz System [17]

$$\begin{aligned} \dot{x} &= -\sigma(x - y), \\ \dot{y} &= -xz + rx - y, \\ \dot{z} &= xy - bz. \end{aligned} \quad (1.34)$$

Rössler System [54]

$$\begin{aligned} \dot{x} &= -y - z, \\ \dot{y} &= x + ay, \\ \dot{z} &= b + z(x - c). \end{aligned} \quad (1.35)$$

Mutually Coupled Rössler Systems

$$\begin{aligned}\dot{x}_1 &= -(1 + \nu)x_2 - x_3 + \mu(y_1 - x_1), \\ \dot{x}_2 &= (1 + \nu)x_1 + a x_2,\end{aligned}\tag{1.36}$$

$$\begin{aligned}\dot{x}_3 &= b + x_3 (x_1 - c), \\ \dot{y}_1 &= -(1 - \nu)y_2 - y_3 + \mu(x_1 - y_1), \\ \dot{y}_2 &= (1 - \nu)y_1 + a y_2,\end{aligned}\tag{1.37}$$

$$\dot{y}_3 = b + y_3 (y_1 - c).$$

References

1. H. Kantz, T. Schreiber, *Nonlinear Time Series Analysis* (University Press, Cambridge, 1997)
2. F. Takens, Detecting strange attractors in turbulence, in *Dynamical Systems and Turbulence*, ed. by D. Rand, L.-S. Young. Lecture Notes in Mathematics, vol. 898 (Springer, Berlin, 1981), pp. 366–381
3. N.H. Packard, J.P. Crutchfield, J.D. Farmer, R.S. Shaw, Geometry from a time series. *Phys. Rev. Lett.* **45**(9), 712–716 (1980)
4. L. Cao, Practical method for determining the minimum embedding dimension of a scalar time series. *Physica D* **110**(1–2), 43–50 (1997)
5. M.B. Kennel, R. Brown, H.D.I. Abarbanel, Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys. Rev. A* **45**(6), 3403–3411 (1992)
6. A.M. Fraser, H.L. Swinney, Independent coordinates for strange attractors from mutual information. *Phys. Rev. A* **33**(2), 1134–1140 (1986)
7. J.P. Zbilut, J.-M. Zaldívar-Comenges, F. Strozzi, Recurrence quantification based Liapunov exponents for monitoring divergence in experimental data. *Phys. Lett. A* **297**(3–4), 173–181 (2002)
8. F.M. Atay, Y. Altıntaş, Recovering smooth dynamics from time series with the aid of recurrence plots. *Phys. Rev. E* **59**(6), 6593–6598 (1999)
9. N. Marwan, M.C. Romano, M. Thiel, J. Kurths, Recurrence plots for the analysis of complex systems. *Phys. Rep.* **438**(5–6), 237–329 (2007)
10. J.-P. Eckmann, D. Ruelle, Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys.* **57**(3), 617–656 (1985)
11. H.D.I. Abarbanel, R. Brown, J.J. Sidorowich, L.S. Tsimring, The analysis of observed chaotic data in physical systems. *Rev. Mod. Phys.* **65**(4), 1331–1392 (1993)
12. E. Ott, *Chaos in Dynamical Systems* (University Press, Cambridge, 1993)
13. H. Poincaré, Sur la probleme des trois corps et les équations de la dynamique. *Acta Math.* **13**, 1–271 (1890)
14. J.H. Argyris, G. Faust, M. Haase, *An Exploration of Chaos* (North Holland, Amsterdam, 1994)
15. J.-P. Eckmann, S. Oliffson Kamphorst, D. Ruelle, Recurrence plots of dynamical systems. *Europhys. Lett.* **4**, 973–977 (1987)
16. N. Marwan, A historical review of recurrence plots. *Eur. Phys. J. Spec. Top.* **164**(1), 3–12 (2008)
17. E.N. Lorenz, Deterministic nonperiodic flow. *J. Atmos. Sci.* **20**, 120–141 (1963)
18. G. Robinson, M. Thiel, Recurrences determine the dynamics. *Chaos* **19**, 023104 (2009)

19. Y. Hirata, S. Horai, K. Aihara, Reproduction of distance matrices from recurrence plots and its applications. *Eur. Phys. J. Spec. Top.* **164**(1), 13–22 (2008)
20. N. Marwan, How to avoid potential pitfalls in recurrence plot based data analysis. *Int. J. Bifurcat. Chaos* **21**(4), 1003–1017 (2011)
21. P. Faure, H. Korn, A new method to estimate the Kolmogorov entropy from recurrence plots: its application to neuronal signals. *Physica D* **122**(1–4), 265–279 (1998)
22. M. Thiel, M.C. Romano, J. Kurths, R. Meucci, E. Allaria, F.T. Arecchi, Influence of observational noise on the recurrence quantification analysis. *Physica D* **171**(3), 138–152 (2002)
23. M. Thiel, M.C. Romano, J. Kurths, Analytical Description of Recurrence Plots of white noise and chaotic processes. *Appl. Nonlinear Dyn.* **11**(3), 20–30 (2003)
24. L. Matassini, H. Kantz, J.A. Holyst, R. Hegger, Optimizing of recurrence plots for noise reduction. *Phys. Rev. E* **65**(2), 021102 (2002)
25. S. Schinkel, O. Dimigen, N. Marwan, Selection of recurrence threshold for signal detection. *Eur. Phys. J. Spec. Top.* **164**(1), 45–53 (2008)
26. G.M. Mindlin, R. Gilmore, Topological analysis and synthesis of chaotic time series. *Physica D* **58**(1–4), 229–242 (1992)
27. M. Koebbe, G. Mayer-Kress, Use of recurrence plots in the analysis of time-series data, in *Proceedings of SFI Studies in the Science of Complexity*, vol. XXI, ed. by M. Casdagli, S. Eubank (Addison-Wesley, Redwood City, 1992), pp. 361–378
28. J.P. Zbilut, C.L. Webber Jr., Embeddings and delays as derived from quantification of recurrence plots. *Phys. Lett. A* **171**(3–4), 199–203 (1992)
29. J. Theiler, Spurious dimension from correlation algorithms applied to limited time-series data. *Phys. Rev. A* **34**(3), 2427–2432 (1986)
30. J. Gao, Z. Zheng, Direct dynamical test for deterministic chaos and optimal embedding of a chaotic time series. *Phys. Rev. E* **49**, 3807–3814 (1994)
31. V. Balakrishnan, G. Nicolis, C. Nicolis, Recurrence time statistics in deterministic and stochastic dynamical systems in continuous time: A comparison. *Phys. Rev. E* **61**(3), 2490–2499 (2000)
32. E.G. Altmann, E.C. da Silva, I.L. Caldas, Recurrence time statistics for finite size intervals. *Chaos* **14**(4), 975–981 (2004)
33. L.M. Little, P. McSharry, S.J. Roberts, D.A.E. Costello, I.M. Moroz, Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. *BioMed. Eng. OnLine* **6**(23), 1–19 (2007)
34. E.J. Ngamga, D.V. Senthilkumar, A. Prasad, P. Parmananda, N. Marwan, J. Kurths, Distinguishing dynamics using recurrence-time statistics. *Phys. Rev. E* **85**(2), 026217 (2012)
35. J.B. Gao, H.Q. Cai, On the structures and quantification of recurrence plots. *Phys. Lett. A* **270**(1–2), 75–87 (2000)
36. C.L. Webber Jr., J.P. Zbilut, Dynamical assessment of physiological systems and states using recurrence plot strategies. *J. Appl. Physiol.* **76**(2), 965–973 (1994)
37. J.P. Zbilut, C.L. Webber Jr., Recurrence quantification analysis: Introduction and historical context. *Int. J. Bifurcat. Chaos* **17**(10), 3477–3481 (2007)
38. P. Grassberger, I. Procaccia, Characterization of strange attractors. *Phys. Rev. Lett.* **50**(5), 346–349 (1983)
39. J.P. Zbilut, N. Marwan, The Wiener-Khinchin theorem and recurrence quantification. *Phys. Lett. A* **372**(44), 6622–6626 (2008)
40. D.P. Lathrop, E.J. Kostelich, Characterization of an experimental strange attractor by periodic orbits. *Phys. Rev. A* **40**(7), 4028–4031 (1989)
41. R. Gilmore, Topological analysis of chaotic dynamical systems. *Rev. Mod. Phys.* **70**(4), 1455–1529 (1998)
42. N. Marwan, M. Thiel, N.R. Nowaczyk, Cross recurrence plot based synchronization of time series. *Nonlinear Process. Geophys.* **9**(3/4), 325–331 (2002)
43. M.C. Romano, M. Thiel, J. Kurths, I.Z. Kiss, J. Hudson, Detection of synchronization for non-phase-coherent and non-stationary data. *Europhys. Lett.* **71**(3), 466–472 (2005)

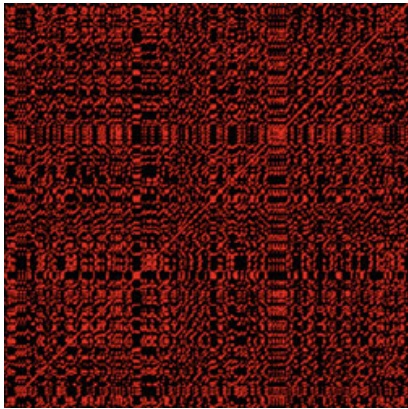
44. N. Marwan, N. Wessel, U. Meyerfeldt, A. Schirdewan, J. Kurths, Recurrence plot based measures of complexity and its application to heart rate variability data. *Phys. Rev. E* **66**(2), 026702 (2002)
45. E.J. Ngamga, A. Nandi, R. Ramaswamy, M.C. Romano, M. Thiel, J. Kurths, Recurrence analysis of strange nonchaotic dynamics. *Phys. Rev. E* **75**(3), 036222 (2007)
46. N. Marwan, J.F. Donges, Y. Zou, R.V. Donner, J. Kurths, Complex network approach for recurrence analysis of time series. *Phys. Lett. A* **373**(46), 4246–4254 (2009)
47. R.V. Donner, Y. Zou, J.F. Donges, N. Marwan, J. Kurths, Recurrence networks – A novel paradigm for nonlinear time series analysis. *New J. Phys.* **12**(3), 033025 (2010)
48. R.V. Donner, M. Small, J.F. Donges, N. Marwan, Y. Zou, R. Xiang, J. Kurths, Recurrence-based time series analysis by means of complex network methods. *Int. J. Bifurcat. Chaos* **21**(4), 1019–1046 (2011)
49. R.V. Donner, J. Heitzig, J.F. Donges, Y. Zou, N. Marwan, J. Kurths, The geometry of chaotic dynamics – a complex network perspective. *Eur. Phys. J. B* **84**, 653–672 (2011)
50. A. Rényi, *Probability Theory* (North-Holland, Amsterdam, 1970)
51. P. Grassberger, I. Procaccia, Estimation of the Kolmogorov entropy from a chaotic signal. *Phys. Rev. A* **9**(1–2), 2591–2593 (1983)
52. P. Grassberger, I. Procaccia, Measuring the strangeness of strange attractors. *Physica D* **9**(1–2), 189–208 (1983)
53. N. Marwan, S. Schinkel, J. Kurths, Recurrence plots 25 years later – gaining confidence in dynamical transitions. *Europhys. Lett.* **101**, 20007 (2013)
54. O.E. Rössler, An equation for continuous chaos. *Phys. Lett. A* **57**(5), 397–398 (1976)
55. Y. Zou, R.V. Donner, J.F. Donges, N. Marwan, J. Kurths, Identifying complex periodic windows in continuous-time dynamical systems using recurrence-based methods. *Chaos* **20**(4), 043130 (2010)
56. J.P. Zbilut, A. Giuliani, C.L. Webber Jr., Detecting deterministic signals in exceptionally noisy environments using cross-recurrence quantification. *Phys. Lett. A* **246**(1–2), 122–128 (1998)
57. N. Marwan, J. Kurths, Nonlinear analysis of bivariate data with cross recurrence plots. *Phys. Lett. A* **302**(5–6), 299–307 (2002)
58. N. Marwan, J. Kurths, Line structures in recurrence plots. *Phys. Lett. A* **336**(4–5), 349–357 (2005)
59. A. Porta, G. Baselli, N. Montano, T. Gneccchi-Ruscone, F. Lombardi, A. Malliani, S. Cerutti, Classification of coupling patterns among spontaneous rhythms and ventilation in the sympathetic discharge of decerebrate cats. *Biol. Cybern.* **75**(2), 163–172 (1996)
60. M.C. Romano, M. Thiel, J. Kurths, W. von Bloh, Multivariate recurrence plots. *Phys. Lett. A* **330**(3–4), 214–223 (2004)
61. Y. Zou, M.C. Romano, M. Thiel, N. Marwan, J. Kurths, Inferring indirect coupling by means of recurrences. *Int. J. Bifurcat. Chaos* **21**(4), 1099–1111 (2011)
62. N. Marwan, Y. Zou, N. Wessel, M. Riedl, J. Kurths, Estimating coupling directions in the cardio-respiratory system using recurrence properties. *Philos. Trans. R. Soc. A* **371**(1997), 20110624 (2013)
63. W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. 1 (Wiley, New York, 1950)
64. D.C. Brock, *Understanding Moore's Law: Four Decades of Innovation* (Chemical Heritage Foundation, Philadelphia, 2006)
65. C.L. Webber Jr., *Quantitative Analysis of Respiratory Cell Activity*. PhD Dissertation, Loyola University Chicago, 1974
66. N. Marwan, *Encounters With Neighbours – Current Developments Of Concepts Based On Recurrence Plots And Their Applications*. PhD thesis, University of Potsdam, 2003
67. B. Kernighan, D. Ritchie, *The C Programming Language* (Prentice Hall, Englewood Cliffs, 1978)
68. C.L. Webber, Jr., Introduction to recurrence quantification analysis. RQA version 14.1 README.PDF. 2012
69. N. Marwan. CRP Toolbox 5.17, 2013, platform independent (for Matlab)

- 70. E. Kononov, Visual Recurrence Analysis 4.9, 2009, only for Windows
- 71. N. Thomasson, T.J. Hoeppner, C.L. Webber Jr., J.P. Zbilut, Recurrence quantification in epileptic EEGs. *Phys. Lett. A* **279**(1–2), 94–101 (2001)
- 72. C.L. Webber Jr., J.P. Zbilut, *Recurrence Quantification Analysis of Nonlinear Dynamical Systems* (National Science Foundation, Arlington, 2005), pp. 26–94
- 73. D.J. McFarland, W.A. Sarnacki, J.R. Wolpaw, Electroencephalographic (EEG) control of three-dimensional movement. *J. Neural Eng.* **7**(3), 036007 (2010)
- 74. K. Shockley, M. Butwill, J.P. Zbilut, C.L. Webber Jr., Cross recurrence quantification of coupled oscillators. *Phys. Lett. A* **305**(1–2), 59–69 (2002)

Chapter 2

Estimating Kolmogorov Entropy from Recurrence Plots

Philippe Faure and Annick Lesne



Abstract Kolmogorov entropy, actually an entropy rate h , has been introduced in chaos theory to characterize quantitatively the overall temporal organization of a dynamics. Several methods have been devised to turn the mathematical definition into an operational quantity that can be estimated from experimental time series. The method based on recurrence quantitative analysis (RQA) is one of the most successful. Indeed, recurrence plots (RPs) offer a trajectory-centered viewpoint circumventing the need of a complete phase space reconstruction and estimation of the invariant measure. RP-based entropy estimation methods have been developed for either discrete-state or continuous-state systems. They rely on the statistical analysis of the length of diagonal lines in the RP. For continuous-state systems,

P. Faure (✉)

CNRS UMR 7102, Neurobiologie des Processus Adaptatifs, Université Pierre et Marie Curie-Paris 6, 4 place Jussieu, 75252 Paris Cedex 05, France
e-mail: philippe.faure@snv.jussieu.fr

A. Lesne

CNRS UMR 7600, Laboratoire de Physique Théorique de la Matière Condensée, Université Pierre et Marie Curie-Paris 6, 4 place Jussieu, 75252 Paris Cedex 05, France
e-mail: lesne@lptmc.jussieu.fr

only a lower bound K_2 can be estimated. The dependence of the estimated quantity on the tunable neighborhood radius ϵ involved in constructing the RP, termed the ϵ -entropy, gives a qualitative information on the regular, chaotic or stochastic nature of the underlying dynamics. Although some caveats have to be raised about its interpretation, Kolmogorov entropy estimated from RPs offers a simple, reliable and quantitative index, all the more if it is supplemented with other characteristics of the dynamics.

2.1 Introduction

2.1.1 Entropy

The quantity today termed “Kolmogorov entropy” has been introduced for deterministic dynamical systems by Kolmogorov [1] and 1 year later by Sinai [2] hence it is sometimes called Kolmogorov–Sinai entropy or metric entropy. This quantity is similar to the entropy rate introduced in a different context by Shannon [3, 4] for symbolic sequences. They should not be confused with Kolmogorov complexity, which is another name for algorithmic complexity [5], nor with thermodynamic entropy, which is a different concept, only slightly and indirectly related through statistical mechanics and Boltzmann entropy [6].

Kolmogorov entropy, henceforth denoted h , is a global quantity providing a quantitative measure of the overall temporal organization of the dynamics. The initial motivation of Kolmogorov was to investigate whether some dynamical systems were isomorphic or not. He demonstrated that entropy is an invariant of the dynamics, i.e. it is preserved upon any isomorphism. To assess that two dynamical systems are non isomorphic, it is then sufficient to show that their entropies differ. Kolmogorov entropy is in fact an entropy rate. For a discrete-state source, Sect. 2.2, a direct and simple interpretation is provided by the asymptotic equipartition property. This property states that n -words, i.e. sequences of length n produced by the source, asymptotically separate in two classes: typical and non-typical n -words. Typical n -words have asymptotically the same probability e^{-nh} (hence the name “equipartition property”) and their number scales as e^{nh} , while all the other n -words have a vanishing probability and in this respect no chance to be observed. Entropy has a meaning for both deterministic and stochastic dynamics. We will see, Sect. 2.4.1, that in the case of continuous-state dynamical systems, the intermediary steps of the computation at finite resolution ϵ in the phase space provide an auxiliary ϵ -entropy $h(\epsilon)$, whose dependence with respect to ϵ reflects the deterministic or stochastic nature of the dynamics. To date, h has been used to characterize linguistic data [7], DNA sequences [8, 9], behavioral sequences [10, 11], speech analysis [12, 13], or spike emission in neurons [14, 15]. Estimating Kolmogorov entropy from experimental data is not an easy task. We will see that using recurrence plots yields a simple and reliable estimation of this value, or at least a lower bound.

2.1.2 Recurrence Plots

Several methods have been devised to turn the mathematical definition of Kolmogorov entropy into an operational quantity that can be estimated from experimental time series. The method based on recurrence plots (RPs) has proven to be an efficient one. RPs have been introduced as a graphical representation of a dynamical system well-suited for data analysis [16]. They consist in square binary matrices, depicting how the trajectory repeats itself: a black dot (i, j) is drawn in the plot when a recurrence is observed in the trajectory for times i and j . Patterns of activity can be identified in the organization of the recurrences along vertical or diagonal lines [17]. RPs thus provide a useful framework for visualizing the time evolution of a system and for discerning subtle transitions or drifts in the dynamics. Beyond being a graphical tool, RPs offer a means for the quantitative analysis of the underlying dynamics, what is summarized by the words “recurrence quantification analysis” (RQA) [18–20]. In particular, the graphical power of RPs will not be the point in computing Kolmogorov entropy, except if used for prior detection of non-stationarity in the data prompting to windowing the series. Recurrence quantification analysis is particularly powerful because it relies on a trajectory-centered exploration of the phase space. In contrast to methods involving a blind partition of the phase space, containing many regions of negligible weight that considerably slow down the computation for no significant gain in accuracy, the recurrence pattern naturally samples important regions of the phase space, thus reducing computational complexity and time. This is analogous in spirit to Monte-Carlo-Markov-Chain sampling of the integration domain for computing an integral, instead of using a Riemann discretization of the integration domain.

As in a number of nonlinear analysis methods, it is necessary for building a RP to first reconstruct a multidimensional signal from a single observed variable. The reconstruction amounts to consider a series $(\mathbf{x}_i)_{i \geq 0}$ of embedded vectors \mathbf{x}_i instead of the original one-dimensional time series $(u_t)_{t \geq 0}$. Their definition involves an embedding dimension m and a time delay τ , according to:

$$\mathbf{x}_i = (u_{i\tau}, u_{(i+1)\tau}, \dots, u_{(i+m-1)\tau}) \quad (2.1)$$

The embedding of a one-dimensional signal in a multidimensional phase space and the choice of the involved time delay are standard procedures presented in Chap. 1 (see also [21]). The choice of m and τ is critical for any subsequent analysis since an inappropriate choice can either “insufficiently unfold” the high-dimensional dynamic or lead to false positive indications of chaos. An additional parameter in the RP is the radius ϵ of the neighborhoods defining recurrence. We will see that the dimension m is involved jointly with the length l of the diagonal lines, namely the relevant quantity is $m + l$, hence there is no need to consider m as an independent parameter.

RPs portray the dynamics of the embedded signals in the form of dots interspersed in a square matrix. Let \mathbf{x}_i be the i -th point on the reconstructed trajectory,

of length N , describing the system in an m -dimensional space. A recurrence plot is the $N \times N$ matrix in which a dot is placed at (i, j) whenever \mathbf{x}_j is close to \mathbf{x}_i , i.e. whenever the distance $d(\mathbf{x}_i, \mathbf{x}_j)$ is lower than a given cutoff value ϵ . Different metrics can be used, for instance Euclidian distance or Maximum norm. The RP then contains N^2 black or white dots, see Fig. 2.1a (left). The black dots represent the recurrence of the dynamical process determined with a given resolution ϵ , and their organization characterizes the recurrence properties of the dynamics. A vertical line of length l (i.e. comprising l dots) starting from a dot (i, j) means that the trajectory starting from \mathbf{x}_j remains close to \mathbf{x}_i during $l - 1$ time steps. A diagonal black line of length l starting from a dot (i, j) means that trajectories starting from \mathbf{x}_i and \mathbf{x}_j remain close during $l - 1$ time steps, Fig. 2.1a (right). Diagonal segments in the RP (excluding the main diagonal) can be counted and plotted as an histogram according to their length. We will see below that this histogram, and in particular the slope α for the linear region in the log-log plot is the basis for estimating Kolmogorov entropy. It should be noted that changing the embedding dimension m amounts to shift by a few units the length of the diagonal lines, which does not affect the slope α . Hence the estimated value for Kolmogorov entropy will be independent of the embedding dimension, as shown in [22].

2.2 Discrete-State Signals

2.2.1 Messages, Symbolic Sequences and Discrete Sources

Information theory is concerned with the analysis of messages written with letters from a given alphabet [3]. This symbolic setting relates to continuous-state dynamical systems through encoding continuous states into discrete ones. Such encoding is an acknowledged approach allowing to prune irrelevant information, to improve statistics by reducing the dimension of the sequence and overall to simplify the system description without altering its essential dynamical properties [23]. For discrete-time dynamics in a continuous phase space, symbolic sequences can be obtained from the discretization of continuous-valued trajectories using a partition of the phase space in subsets A_w [24]. Each trajectory $(z_i)_{i \geq 0}$ is associated with a symbolic sequence $(w_i)_{i \geq 0}$ describing the array of visited subsets according to $z_i \in A_{w_i}$. The partition is said to be generating when the knowledge of the semi-infinite symbolic sequence $(w_i)_{i \geq 0}$ fully determines a unique initial condition $z_0 \in A_{w_0}$ in the continuous phase space. In this special case, the symbolic encoding is asymptotically faithful, with no loss of information compared to the continuous-valued trajectory (a loss of information nevertheless occurs when considering trajectories of finite length). However, generating partitions are very rare, existing only for sufficiently chaotic dynamical systems [25]. Even if generating partitions exist, a constructive method to determine them may not be available [26]. Discretization

has then to be done using an a priori chosen partition of the phase space, with a main issue being to make the proper choice [27, 28].

Often in practice the system phase space is not fully formalized, think for instance of behavioral sequences recorded with a CCD camera. Encoding is then achieved in an heuristic way. Typically, discrete states are defined by partitioning the values of a few relevant observables. In the example of behavioral sequences, when considering the velocity V of a moving individual to be the relevant observable, time steps where $V < V_c$ will be coded 0 and time steps where $V \geq V_c$ will be coded 1, transforming the video recording of the individual into a binary sequence. Another situation is the case where data are intrinsically discrete, e.g. language or DNA sequences [29, 30]. In this case, and more generally in the information-theoretic terminology, one speaks of a symbolic source generating messages, instead of discrete-state dynamics generating trajectories.

2.2.2 Entropy Rate of Symbolic Sequences

In information theory, entropy rate has been basically introduced to characterize languages modeled with Markov chains of increasing orders [3]. It measures the time-average information per symbol needed to transmit messages. It has been later demonstrated to coincide with Kolmogorov entropy (up to a factor relating \ln and \log_2) in the case where the symbolic sequences originate from the phase-space discretization of a dynamical system according to a generating partition [31, 32]. Formally, it is defined as the limit of normalized block-entropies or block-entropy differences (Theorem 5 in [3]). Block entropy H_n is defined as the Shannon entropy of the n -word distribution $p_n(\cdot)$, namely $H_n = - \sum_{\bar{w}_n} p_n(\bar{w}_n) \log_2 p_n(\bar{w}_n)$ where the sum runs over the set of n -words \bar{w}_n . H_n increases with n , while the sequence of increments $h_n = H_{n+1} - H_n$ is a decreasing positive sequence. The increments h_n and the normalized quantities $\tilde{h}_n = H_n/n$ have the same limit $h = \lim_{n \rightarrow \infty} h_n = \lim_{n \rightarrow \infty} \tilde{h}_n$ (if it exists), which defines the entropy rate h . The definition of entropy h as a rate is thus far more than a mere normalization by a duration: it involves a time integration, by considering words of increasing length. The above information-theoretic definition can be reformulated in terms of temporal correlations, which reduce the amount of information required to retrieve a message. In other words, h reflects the temporal organization of the dynamics, taking small values when the dynamics has a strongly correlated structure. h thus provides an integrated measure of the overall temporal correlations present in the dynamics, and $1/h$ can roughly be seen as a correlation time. It is to note that h_n corresponds to the entropy rate of the $(n - 1)$ -th order Markov approximation of the source, involving only n -point joint probabilities [3]. In the above definition, we used the binary logarithm \log_2 to match information-theoretic usage and Shannon definition; h is then expressed in bits per time unit. It is straightforward to replace \log_2 by the Neperian logarithm \ln so as to exactly recover Kolmogorov entropy.

Estimation of h is currently based on the above definition [10, 33, 34], with h_n appearing as a better estimator than \tilde{h}_n (although both are unbiased). In practice, due to the finite size N of the data sequence, the estimated value \hat{H}_n plotted as a function of the block size n saturates to $\log_2 N$. If h has a non trivial value, this plot displays a linear region of slope h , and the crossover to the asymptotic plateau occurs around $n^* = (\log_2 N)/h$ [34, 35]. Another method uses the identity for ergodic sources between the entropy rate h and the Lempel–Ziv complexity, defined for a single sequence and computed using compression algorithms [36–38]. The latter method may perform better, in particular for short sequences [35]. Actually a proper implementation requires a two-step estimation. The first step is to obtain a rough estimate of h , delineating the validity and the performance of these alternative methods, giving in particular a lower bound on the sequence length required for the block-entropy method to be reliable. Then a refined estimate of h is obtained using the best of the two methods [35]. A third alternative is provided by RP representation of the symbolic sequences, as we will see below, Sect. 2.2.4.

2.2.3 Shannon–McMillan–Breiman Theorem

Both the interpretation of the above-defined quantity h and its estimation via RP, Sect. 2.2.4, rely on the Shannon–McMillan–Breiman theorem [39]. This theorem, established as Theorem 3 in [3] and further improved by McMillan then Breiman [4, 40, 41], states that the number of typical n -words (i.e. n -words that have the same statistical properties corresponding to the almost sure behavior) scales like e^{nh} as $n \rightarrow \infty$, where the exponent h is the entropy rate of the source. A corollary of this theorem is the asymptotic equipartition property, stating that the probability $p_n(\bar{w}_n)$ of every typical n -word \bar{w}_n takes asymptotically the same value e^{-nh} . Although containing the core idea of the asymptotic equipartition property, this finite-size statement has been made more rigorous on mathematical grounds. Its formulation requires to introduce random variables $\hat{\mathcal{P}}_n$ depending on the realization \bar{w} of the whole symbolic sequence according to $\hat{\mathcal{P}}_n(\bar{w}) = p_n(w_0, \dots, w_{n-1})$. The asymptotic equipartition property then writes

$$\lim_{n \rightarrow \infty} (-1/n) \ln \hat{\mathcal{P}}_n \rightarrow h \quad \text{in probability} \quad (2.2)$$

This means that for any arbitrary small $\delta > 0$ and $\epsilon > 0$, there exists a word threshold size $n^*(\delta, \epsilon)$ such that $\text{Prob}(\{\bar{w}, p_n(w_0, \dots, w_{n-1}) > e^{n(-h+\delta)}\}) < \epsilon$ and $\text{Prob}(\{\bar{w}, p_n(w_0, \dots, w_{n-1}) < e^{n(-h-\delta)}\}) < \epsilon$ for any $n \geq n^*(\delta, \epsilon)$, or equivalently in terms of n -word subsets, $p_n(\{\bar{w}_n, p_n(\bar{w}_n) > e^{n(-h+\delta)}\}) < \epsilon$ and $p_n(\{\bar{w}_n, p_n(\bar{w}_n) < e^{n(-h-\delta)}\}) < \epsilon$. As a corollary of this result, the number of typical n -words \bar{w}_n for which the asymptotic equipartition property $p_n(\bar{w}_n) \sim e^{-nh}$ holds scales as e^{nh} for n large enough, providing another interpretation of h .

Shannon–McMillan–Breiman theorem will be the basis of the entropy estimation method from RPs. It justifies that all the observed n -words belong to the set of typical words since non-typical ones are too rare to be observed. An important caveat is the asymptotic nature of this theorem, making its application to finite words and finite sequences a questionable extrapolation. However, numerical experiments show that the asymptotic regime is reached rapidly and the theorem yields the correct dominant behavior even for moderate values of n (lower than 10).

2.2.4 RP-Based Estimation of the Entropy (Per Unit Time)

Our starting point will be the representation of the data as a m -RP describing the recurrence of m -words: a black dot (i, j) means that $\bar{w}_m(i) = \bar{w}_m(j)$ where $\bar{w}_m(i) = (w_i, \dots, w_{i+m-1})$ denotes the m -words starting at time i in the original sequence. The integer m thus appears as an embedding dimension. Observing a diagonal line of length l starting in (i, j) in the m -RP means that the two $(m+l-1)$ -words starting at times i and j coincide: $\bar{w}_{m+l-1}(i) = \bar{w}_{m+l-1}(j)$. Such a line corresponds to a diagonal line of length $l+1$ in the $(m-1)$ -RP and a single dot in the $(m+l-1)$ -RP. In fact, all the quantities that may be introduced regarding the statistics of diagonal lines are relative to a given realization of the m -RP and they depend not only on m but also on the sequence length N and its realization \bar{w} , which will be skipped for simplicity. We will assume that the size of the m -RP, or equivalently the sequence length N , is large enough to identify quantities computed in one realization of the m -RP and their statistical average, based on the assumed ergodicity of the dynamics. We will also assume (and numerically check) that the asymptotic probability estimate given by Shannon–McMillan–Breiman theorem, that centrally involves the entropy rate h of the source, is valid at the leading order for the considered words.

We consider a length l large enough, so that a $(m+l-1)$ -word occurs at most twice and Shannon–McMillan–Breiman theorem approximately holds. The probability of double occurrence of a typical $(m+l-1)$ -word, $(N-l-m+1)e^{-h(m+l-1)}$, multiplied by the number $e^{h(m+l-1)}$ of these non-identical typical words yields the number $\nu_m^{(N)}(l)$ of diagonal (and possibly overlapping) segments of length l in the upper triangle of the m -RP, not counting the main diagonal line:

$$\nu_m^{(N)}(l) = \frac{(N-l-m+2)(N-m-l+1)}{2} e^{-h(m+l-1)} \quad (2.3)$$

Note that this histogram of the line-lengths l is currently denoted $H_D(l)$ (see e.g. Chap. 1); we adopt the notation $\nu(l)$ to avoid any confusion with block entropies H_n .

Since $m + l \ll N$ we may identify $N - l - m + 2$ and $N - l - m + 1$ with N , getting:

$$v_m^{(N)}(l) \sim (N^2/2) e^{-h(m+l-1)} \quad (2.4)$$

A semi-log plot of $v_m^{(N)}(l)$ with respect to l will exhibit a slope $-h$ in its linear region. While the theoretical derivation of the scaling behavior is done for $v_m^{(N)}(l)$, numerical implementation is more easily done in practice using the number $\eta_m^{(N)}(l)$ of diagonal lines of length exactly equal to l in the m -RP, or the cumulative number $\phi_m^{(N)}(l)$ of diagonal lines of total length larger or equal than l (Fig. 2.1b). After normalization by the total number of diagonal lines, $\phi_m^{(N)}(l)$ coincides with the cumulative probability $p^c(l)$ introduced in Chap. 1. However, $\phi_m^{(N)}(l)$ and $p^c(l)$ satisfy the same scaling laws, which allows to circumvent the normalization issue. We will henceforth work with the raw number $\phi_m^{(N)}(l)$. A diagonal line of total length $l + r$ yields $r + 1$ (partly overlapping) diagonal stretches of length l contributing to $v_m^{(N)}(l)$. It follows that $v_m^{(N)}(l) = \sum_{r \geq 0} (r + 1) \eta_m^{(N)}(l + r)$ [39]. The scaling behavior of $\phi_m^{(N)}(l)$ is then derived using $v_m^{(N)}(l) - v_m^{(N)}(l + 1) \approx -dv_m^{(N)}(l)/dl = hv_m^{(N)}(l)$:

$$\phi_m^{(N)}(l) \equiv \sum_{r \geq 0} \eta_m^{(N)}(l + r) \approx -\frac{dv_m^{(N)}(l)}{dl} \sim (hN^2/2) e^{-h(m+l-1)} \quad (2.5)$$

A semi-log representation of the number $\phi_m^{(N)}(l)$ of diagonal lines of total length larger than or equal to l as a function of the length l would also have a slope $-h$ in its linear region, which provides a direct way to estimate h from RP, as presented in Fig. 2.1. Note finally that the average length of the diagonal lines in the m -RP expresses:

$$\langle D \rangle = \frac{\sum_{l \geq 1} l \eta_m^{(N)}(l)}{\sum_{l \geq 1} \eta_m^{(N)}(l)} = \frac{v_m^{(N)}(1)}{v_m^{(N)}(1) - v_m^{(N)}(2)} \quad (2.6)$$

The second equality is obtained using the change of variable $l = r + 1$ in $v_m^{(N)}(1)$ and $l = r + 2$ in $v_m^{(N)}(2)$ where the expressions for $v_m^{(N)}(1)$ and $v_m^{(N)}(2)$ have been obtained by plugging $l = 1$ and $l = 2$ in the identity $v_m^{(N)}(l) = \sum_{r \geq 0} (r + 1) \eta_m^{(N)}(l + r)$. At the leading order, the scaling $\sum_{r \geq 0} \eta_m^{(N)}(l + r) \sim (hN^2/2) e^{-h(m+l-1)}$ yields $\langle D \rangle \sim e^h/(e^h - 1)$ and even $\langle D \rangle \sim 1/h$ if $h \ll 1$, which gives an intuitive interpretation of $1/h$ as a characteristic time (correlation time) of the source. Note that the sums include single dots ($l = 1$), which is not always the case in RP. Indeed, as noted before, the relevant quantity in a m -RP is $m + l$. A line of length 1 in a m -dimensional space is a line of length 2 in a $(m - 1)$ -dimensional space, hence there is no obvious reason to discriminate single dots in a m -RP.

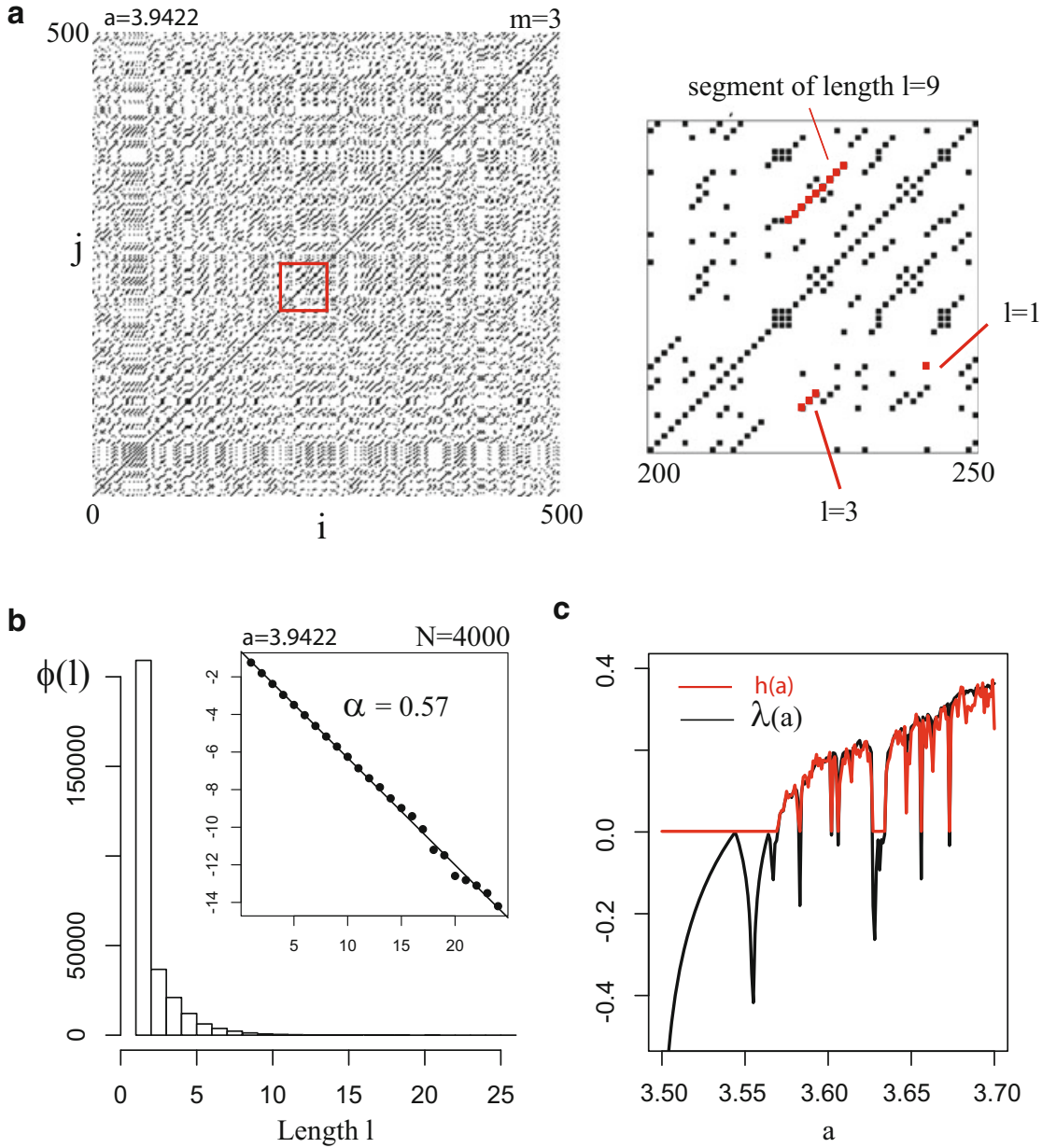


Fig. 2.1 Method for constructing length histograms of diagonal line segments. **(a, left)** RP obtained from a logistic map $g_a(z) = az(1 - z)$ in $[0, 1]$ with $a = 3.9422$ and state-discretized using a symbolic encoding based on the simple rule: if $z_i > 0.5$ then $w_i = 1$ else $w_i = 0$. The embedding dimension is $m = 3$, hence a dot (i, j) means that the three-words (w_i, w_{i+1}, w_{i+2}) and (w_j, w_{j+1}, w_{j+2}) are identical. **(a, right)** Detail of the RP with diagonal line segments of various length l . **(b)** Number $\phi_m^{(N)}(l)$ of diagonal lines of total length larger than or equal to l , counted in the RP obtained from $(w_i)_{i=1, \dots, 4,000}$. *(Inset)* Semi-log representation of $\phi_m^{(N)}(l)$; the absolute value — of the slope of the fitting line yields an estimation of h . **(c)** Comparison of the RP-estimated value of $h(a)$ (red line) with Lyapunov exponent value $\lambda(a)$ (black line). Trajectories used to estimate $h(a)$ have a length $N = 2,000$ with $3.5 \leq a \leq 3.7$. A negative value of $\lambda(a)$ corresponds to an entropy value $h(a) = 0$, whereas Pesin equality ensures $h(a) = \lambda(a)$ for positive values of $\lambda(a)$.

2.3 Continuous-State Dynamics

2.3.1 Kolmogorov Entropy for a Continuous State System

For dissipative dynamical systems, attractors provide a global picture of the long-term behavior. A more quantitative representation of the latter is given by the invariant measure on the attractor, that is, the probability measure invariant upon the action of the dynamics. It describes how frequently a given trajectory visits any particular region of the state space. The state space can be divided into a finite number of intervals (one dimension) or boxes (two or more dimensions) which defines a finite partition $\mathcal{A} = \{A_i, i = 1, \dots, m\}$. The frequency at which a trajectory visits these specific boxes thus gives a partial insight into this invariant measure.

Let us first consider discrete-time dynamics, i.e. dynamics generated by maps. As explained in Sect. 2.2.1, a trajectory $(z_i)_{i \geq 1}$ in the continuous phase space can be encoded by a n -word $\bar{w}_n = (w_1, \dots, w_n)$, meaning that the trajectory successively visits the regions A_{w_1}, \dots, A_{w_n} , with $z_i \in A_{w_i}$. Denoting $P_n(\bar{w}_n)$ the probability of a n -word \bar{w}_n , equal to the measure of the set of points whose n -step trajectory is encoded by \bar{w}_n , the n -block entropy for this partition \mathcal{A} is

$$H_n(\mathcal{A}) = - \sum_{\bar{w}_n} P_n(\bar{w}_n) \ln P_n(\bar{w}_n) \quad (2.7)$$

where the sum runs over the set of all possible n -words. The relationship to Shannon block-entropies is obvious. As for Shannon entropy rate, Kolmogorov entropy (also termed Kolmogorov–Sinai entropy [1, 2] or metric entropy) is defined as the limit

$$h = \sup_{\mathcal{A}} \lim_{n \rightarrow \infty} h_n(\mathcal{A}) \quad (2.8)$$

The supremum is reached for special partitions, called “generating partitions”, when they exist (Sect. 2.2.1). However, this powerful theorem is rarely applicable in practice as we do not know how to construct generating partitions except for unimodal maps and certain maps of the planes. In general, the supremum over the partitions \mathcal{A} is reached by infinitely refining the partitions. It is enough to consider a sequence of partitions \mathcal{A}_ϵ whose boxes have a typical linear size ϵ . Denoting $h_n(\epsilon) \equiv h_n(\mathcal{A}_\epsilon)$ and $h(\epsilon) = \lim_{n \rightarrow \infty} h_n(\epsilon)$, it comes $h = \lim_{\epsilon \rightarrow 0} h(\epsilon)$.

In the case of continuous-time dynamics, an additional parameter is the time delay τ involved in the reconstruction of the dynamics. As explained in the introduction, the first step is the reconstruction of a m -dimensional trajectory $(\mathbf{x}_i)_{i=1, \dots, N}$ of length N from the continuous-state experimental trajectory $u(t)$, according to $\mathbf{x}_i = [u(i\tau), u((i+1)\tau), \dots, u((i+m-1)\tau)]$. We denote $H_{n,\tau}(\epsilon)$ the n -block entropy corresponding to the sequence $(\mathbf{x}_i)_{i=1, \dots, N}$. The difference $H_{n+1,\tau}(\epsilon) - H_{n,\tau}(\epsilon)$ is the average information needed to predict which box of

the partition \mathcal{A}_ϵ will be visited at time $(n + 1)\tau$, given the n boxes visited up to $n\tau$. Definition of Kolmogorov entropy is then similar to the discrete-time case, except for an additional limit $\tau \rightarrow 0$

$$h = \lim_{\tau \rightarrow 0} \frac{1}{\tau} \lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow \infty} (h_{n,\tau}(\epsilon)) \quad (2.9)$$

Generalized entropies $h^{(q)}$ can be defined according to [42, 43]

$$h^{(q)} = - \lim_{\tau \rightarrow 0} \frac{1}{\tau} \lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{q-1} \ln \sum_{\bar{w}_n} [P_n(\bar{w}_n)]^q \quad (2.10)$$

For $q = 1$, we have $h = h^{(1)}$, and it can be demonstrated that $h^{(q)} \geq h^{(q')}$ for any $q' > q$. The inequality is strict as soon as the attractor has a nontrivial multifractal structure. In particular, $h^{(2)}$, currently denoted K_2 , is a lower bound for Kolmogorov entropy. This property is centrally used in the estimation of h . Indeed, the mathematical definition of Kolmogorov entropy for a continuous-state dynamics involves several non-commuting limits, which prevent any direct implementation. Actually, only $K_2 = h^{(2)}$ can be extracted from experimental time series, either by a method based on phase space reconstruction and the computation of a correlation integral [42, 44], or by a method based on RQA [45], both presented below.

2.3.2 Grassberger and Procaccia Method for Computing K_2

A first method for estimating K_2 from experimental data has been developed by Grassberger and Procaccia [42]. The estimation method is based on the computation of the correlation integral of the reconstructed trajectory in embedding dimension m

$$C^m(\epsilon) = \frac{1}{N} \sum_{i=1}^N C_i^m(\epsilon) \quad (2.11)$$

where $C_i^m(\epsilon)$ is the number of time indices j ($1 \leq j \leq N$) for which $d(\mathbf{x}_i, \mathbf{x}_j) \leq \epsilon$. Note that the quantity $C^m(\epsilon)$ can be viewed as the average probability that two trajectories visit jointly a given sequence of m boxes of the partition \mathcal{A}_ϵ . This quantity provides a direct access to the entropy bound $K_2 \leq h$ according to:

$$\tau K_2 = \lim_{\epsilon \rightarrow 0} \lim_{m \rightarrow \infty} \ln \frac{C^m(\epsilon)}{C^{m+1}(\epsilon)} \quad (2.12)$$

Since $C_i^m(\epsilon)$ is the number of black dots in line i when the RP is defined at resolution ϵ , $C^m(\epsilon)/N$ is also the recurrence density at resolution ϵ and embedding

dimension m . This establishes a simple link between RPs and correlation integrals. However a method based on an estimation of the density would ignore the spatial organization of dots in the RP, which contains important information about the dynamics. It would moreover require to explicitly construct RPs for any embedding dimension m .

2.3.3 *RP-Based Method for Computing K_2*

RQA, here based on diagonal line statistics, offers an alternative method for computing K_2 , highly beneficial since it will be enough to consider the reconstructed sequence $(\mathbf{x}_i)_{i=1,\dots,N}$ for a single embedding dimension m [45]. Indeed, as in the discrete-state case, changing the embedding dimension only shifts the diagonal line statistics, since a diagonal line of length l in dimension m corresponds to a diagonal line of length $l - 1$ in dimension $m + 1$. Let us denote $\nu_\epsilon(l)$ the number of diagonal segments of length l , possibly included in longer segments. A diagonal line thus contributes by all l -segments that can be delineated in it. For instance, a diagonal line of four dots contributes to $\nu_\epsilon(l = 2)$ by three segments of two dots, to $\nu_\epsilon(l = 3)$ by two segments of three dots and to $\nu_\epsilon(l = 4)$ by a single segment (itself). As in the discrete case, it can be shown that

$$\nu_\epsilon(l) = \text{const.} e^{-l\alpha(\epsilon)} \quad (2.13)$$

hence the exponent $\alpha(\epsilon)$ can be obtained by fitting the linear part of the log histogram of $\nu_\epsilon(l)$. The expected limiting behavior is:

$$\lim_{\epsilon \rightarrow 0} \alpha(\epsilon) = \tau K_2 \quad (2.14)$$

Actually, as presented in Fig.2.1 in the discrete-state case, the cumulative number $\phi_\epsilon(l)$ of diagonal lines of total length equal to or larger than l displays the same scaling behavior, and could be easier to extract from the RP. A noticeable point is that for a discrete-state system, it follows from Shannon–McMillan–Breiman theorem and the ensuing asymptotic estimate of n -word probability $p_n(\cdot)$ that all generalized entropies $h^{(q)}$ coincide with h , in particular $K_2 = h$, so that the results of the present Sect. 2.3.3 are fully consistent with those of Sect. 2.2.4. The main difference between this approach and the Grassberger–Procaccia method is that the convergence of $\alpha(\epsilon)$ is now studied in terms of the $\epsilon \rightarrow 0$ limit, and not by increasing the embedding dimension $m \rightarrow \infty$. The resulting advantage is the fact that the distribution of the distances between points need to be calculated only once, for constructing a single RP, instead of for each new embedding dimension.

2.4 Some Factors Influencing Computation of Entropy

2.4.1 The Notion of ϵ -Entropy for Analyzing Noisy Dynamics

The presence of the limits $\epsilon \rightarrow 0$ (resolution ϵ in the phase space) and $n \rightarrow \infty$ (length n of words, i.e. trajectory segments) in the entropy definition has the consequence that the estimated entropy is, at best, an approximation. In particular, for small resolutions ϵ , the number of neighboring points become too small to get reliable statistics. This limitation becomes more critical as the word-length n increases or the length N of the time series decreases. Overall, the major constraint imposed by the limit $\epsilon \rightarrow 0$ is due to the noise inherent to experimental data. Noise prevents any analysis below a given resolution at which deterministic structures are destroyed. This resolution threshold is directly related to the variance of the noise. The choice of an optimal threshold ϵ a priori depends on the considered time series but Thiel et al. [46] suggested that a value of $\epsilon = 5\sigma$ (where σ is the fluctuation level in the signal) is appropriate for a wide class of processes. Letellier suggested the value $\epsilon = \sigma\sqrt{m}/10$ where m is the dimension of the embedding [47]. Effect of noise is illustrated in Fig. 2.2. where the convergence of the slope $\alpha(\epsilon)$ towards h is depicted for a Hénon map system with additive Gaussian noise. For a noise value equal to 0 there is a clear logarithmic convergence towards K_2 . As soon as noise increases, a threshold of divergence $\epsilon_{\text{div}}(\zeta)$ appears, with a related upward swing of the curves. Such a result is consistent with the fact that $\alpha(\epsilon)$ converges towards K_2 in a chaotic map, while it diverges as $\epsilon \rightarrow 0$ in a stochastic dynamics [5, 32, 48]. As a consequence, a plot like that of Fig. 2.2a, provides complete information on the system, with a good estimate of K_2 in the range $\epsilon > \epsilon_{\text{div}}(\zeta)$. In contrast, it only gives information on the noise component if $\epsilon < \epsilon_{\text{div}}(\zeta)$.

2.4.2 Non-stationarity

A key assumption in entropy estimation is the statistical stationarity of the source. When dealing with experimental data, people are confronted to a trade-off between the requirement of recording long time series and the non-stationarity of real systems. Non-stationarity can produce spurious identification of chaos [21], hence it has triggered the development of statistical methods (such as surrogates) to test the obtained results. Development of methods allowing to analyze non-stationary time series is thus important [49]. RPs allow to get a visual assessment of the assumption of stationarity. Homogeneity of the RP gives a support of sequence stationarity. Else, non-stationary features like the presence of a drift in the evolution law (reflecting in inhomogeneous lower right and upper left corners, compared to the RP core), the occurrence of transitions (reflecting in disruptions within the RP) or periodicities (reflecting in periodic patterns in the RP) can be easily detected. This is illustrated in Fig. 2.3b where we plotted the recurrence pattern of a time series obtained

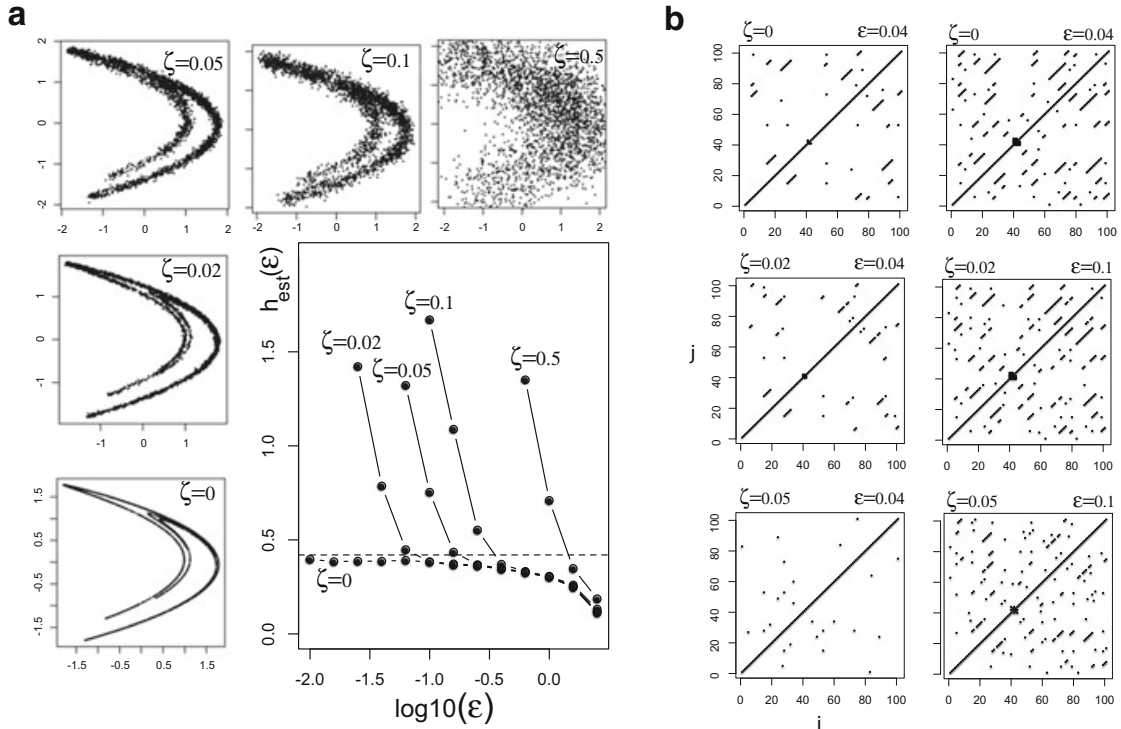


Fig. 2.2 Plot of the RP-estimated ϵ -entropy $h_{est}(\epsilon) = \alpha(\epsilon)$ as a function of the resolution ϵ (neighborhood radius) for different noise levels ζ . **(a)** Time series of $N = 2,000$ points obtained from a Hénon map (with $a = 1.4$, $b = 0.3$) for which the exact entropy rate is $h = 0.42$. Gaussian white noise with standard deviation ζ was added, with ζ varying from 0 to 0.5. The RP-estimated value $h_{est}(\epsilon) = \alpha(\epsilon)$ is represented as a function of ϵ for the five times series. As soon as the noise level increases, the convergence of $h_{est}(\epsilon)$ to the exact value of K_2 (horizontal dashed line) as ϵ tends to 0, which indicates chaos, is interrupted by an upward swing at some value $\epsilon_{div}(\zeta)$, reflecting the stochastic component of the dynamics. For a given level of noise, the convergence can only be inferred from the inflexion point in the computed curve. **(b)** RP for different values of the noise level ζ and resolution ϵ

by concatenating two time series from logistic maps with different values of the parameter a . The RP shows a clearly heterogeneous organization in four quadrants illustrating the modification of temporal properties when passing from the first sequence to the second one. In such non-stationary situations, statistical analysis and in particular entropy estimation should be restricted to time windows where the RP is statistically homogeneous. As the transition points are usually not known in advance, a sliding window is used in practice. The optimal size of the sliding window has to be determined in a preliminary step, either by visual inspection of the RP, or by more quantitative image analysis techniques to determine the typical size of statistically homogeneous regions in the RP, or by invoking additional knowledge about the system. We have numerically implemented the sliding window procedure on a binary discretized trajectory, initially generated by a logistic map whose control parameter a slowly increases by small steps in the course of evolution (Fig. 2.3a). This increase is slow enough for a quasi-stationary approximation to make sense. It allows to consider an entropy rate $h(a)$ corresponding to the instantaneous value

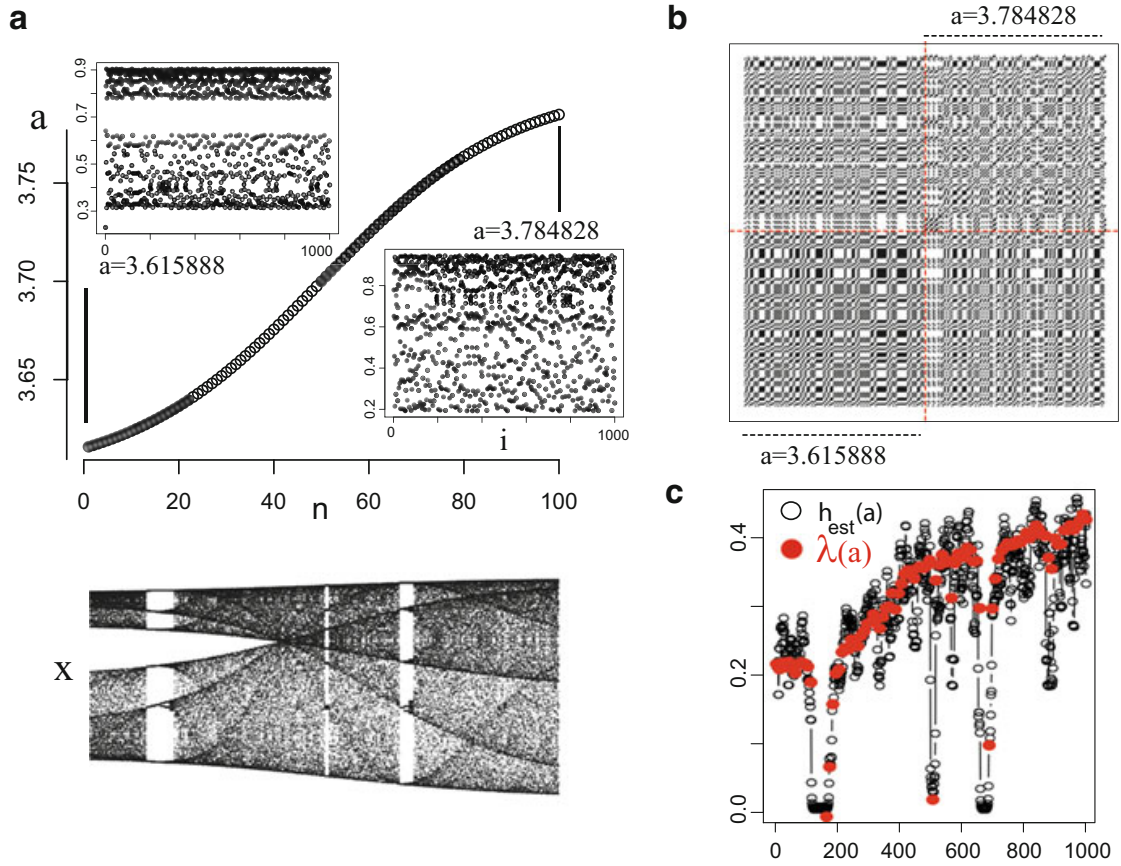


Fig. 2.3 Entropy estimation for non-stationary dynamics. (a, top left) Variation of a from 3.6 to 3.8, according to a sigmoidal function. The value of a in the n -th part of the simulation is given by $a_n = 3.6 + \frac{2}{10 \cdot (1 + \exp(-0.5 \cdot n))}$ with $n = 1, \dots, 100$. (a, bottom left) Time series obtained from concatenation of 100 time series $z_{a_n}(i)_{i=1, \dots, 1,000}$ for $n = 1, \dots, 100$ from logistic map $g_a(z) = az(1 - z)$, with a varying from 3.6 to 3.8, as can be read on the above curve. The last point of a time series $z_{a_n}(i)$ is used as the initial point to calculate the following $z_{a_{n+1}}(i)$ series. (a, right) Two of these time series for $a = 3.615888$ and $a = 3.784828$. (b) Recurrence plot obtained by concatenating the two time series illustrated in (a, right). (c) Variation of h along the non-stationary dynamics. $h(a)$ is estimated on successive overlapping windows of length $N = 1,000$ with a shift of 100 time steps. For each value of a , estimated value $h_{est}(a)$ of $h(a)$ (black points) is superimposed with the value of Lyapunov exponent $\lambda(a)$ (red points)

of a and characterizing the non-stationary dynamics during the associated transient stage. Figure 2.3c shows that the evolution of the entropy rate $h(a)$ as a varies, although very irregular, can be faithfully captured by entropy RP-estimation in a sliding window.

2.5 Discussion

For continuous-state dynamics, whose attractor has a non trivial multifractal structure, only a lower bound K_2 of the Kolmogorov entropy h can be estimated from experimental data. Discrete-state dynamics has a simpler structure, for which

the different generalized entropies $h^{(q)}$ coincide, hence it is possible to estimate $h = K_2$. However, discretizing the continuous states does not solve the issue, since what will be estimated is the Kolmogorov entropy of an approximation of the continuous-state dynamics, missing the multifractal structure of the attractor, and providing anyhow a lower bound on the actual Kolmogorov entropy of the original dynamical system. More generally, discretization is associated with a loss of information about the states, and all the discretization procedures presented above are quite sensitive to the contamination by an additive noise. However, quantitative analysis, and specifically entropy estimation, is expected to be statistically more reliable when performed on symbolic sequences.

Kolmogorov entropy rate h should not be confused with the Shannon entropy of the length distribution of black diagonal lines [17] or white diagonal lines [47]. Black diagonal lines correspond to the recurrence of a segment of trajectory, that is, two stretches of trajectory remaining close one to the other during l time steps if the diagonal length is l . Intuitively, it is thus expected in case of a chaotic dynamics that their average length scales as the inverse of the maximal Lyapunov exponent. Numerical evidences, for some 1D maps, seem to suggest that these Shannon entropies display the same behavior than the maximal Lyapunov exponent, coinciding for hyperbolic 1D maps with the Kolmogorov entropy rate h . However, to our knowledge, an analytical and general link between Kolmogorov entropy rate and the Shannon entropy of the length distribution of black or white diagonal lines is still lacking.

We have seen that h roughly measures the range of temporal correlations. In other words, the time during which the behavior of the system can be predicted is proportional to $1/h$. More rigorously, $h = \lambda$ when there is a single strictly positive Lyapunov exponent λ . In the case where two or more Lyapunov exponents are strictly positive, what is termed hyperchaos, Kolmogorov entropy determines only a lower bound on the sum of positive Lyapunov exponents according to $h \leq \sum_{\lambda \geq 0} \lambda$. This inequality, known as the Pesin inequality, turns into an equality under the condition of uniform hyperbolicity (typically Anosov and Axiom A systems). As a consequence, if h approaches zero, the system becomes fully predictable (for example the case of periodic dynamics). On the other hand, a finite positive h indicates the presence of chaos while h diverges for a stochastic dynamics. Entropy is also closely related to recurrence times [39, 50]. In the discrete case, Wyner–Ziv theorem states that the minimal recurrence time at the level of m -words (i.e. the smallest time t such that $\bar{w}_m(0) = \bar{w}_m(t)$) behaves asymptotically as e^{mh} [51]. This theorem can be exploited either to give an interpretation of h in terms of recurrence times, or conversely to estimate h from the recurrence times. Note that recurrence times involved in this theorem are minimal recurrence times given by the vertical distance to the main diagonal (first bisector), that is, the length of the white vertical line separating a point (i, i) from its first recurrence, associated with some dot (i, j) [39, 52]

Overall the use of a RP-based, finite-size estimation of the Kolmogorov entropy for characterizing an experimental dynamics is clearly limited. Two types of

limiting factors can be distinguished. The first type refers to the dynamics of the investigated system and the very nature of Kolmogorov entropy. Although entropy is an absolute dynamical invariant, it characterizes only stationary regimes, and the superimposition of slow and fast components in the dynamics [32] does not straightforwardly reflect in the entropy. Another set of limiting factors is related to the nature of the data and estimation issues. Experimental data are generally associated with short time series, high-dimensional underlying dynamics, noise and non-stationarity. These factors (some of which have been discussed above) all limit the accuracy and reliability of the entropy estimation. Furthermore, except in the case of symbolic dynamics, only a lower bound K_2 can be obtained. Estimating Kolmogorov entropy is however appropriated for comparison purposes. In data analysis, entropy is often used for comparing the real system with null models, through a comparison of their entropies. Moreover, entropy is a unifying concept insofar as it applies to both deterministic and stochastic dynamics. Such a feature alleviates data analysis from the need of assessing the deterministic nature of the dynamics and makes the same estimation procedure valid in both cases. Finally, entropy provides an overall quantification of the complexity of the dynamics but it cannot answer questions about a specific moment, nor about a specific region of the phase space. This gap can be filled with further analysis of the RPs, which possibly provides information about what happens in a localized region of the phase space [50]. While RPs are suitable for statistical analysis of sequences (extraction of average or integrated features like the entropy or the average recurrence time) they also allow to keep track of the temporal location of specific events, hence allowing to visually evidence and locate dynamic transitions and more generally non-stationary features of the evolution.

References

1. A.N. Kolmogorov, A new metric invariant of transient dynamical systems and automorphisms in lebesgue spaces (in Russian). *Doklady Akademii Nauk SSSR* **119**, 768–771 (1958)
2. Ya.G. Sinai, On the concept of entropy for a dynamic system. *Doklady Akademii Nauk SSSR* **124**, 768–771 (1959)
3. C.E. Shannon, A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 479–423, 623–656 (1948)
4. T.M. Cover, J.A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991)
5. G. Nicolis, P. Gaspard, Toward a probabilistic approach to complex systems. *Chaos Solitons Fractals* **4**, 41–57 (1994)
6. A. Lesne, Shannon entropy: a rigorous mathematical notion at the crossroads between probability, information theory, dynamical systems and statistical physics. *Math. Struct. Comput. Sci.* **24**, e240311 (2014)
7. W. Ebeling, T. Pöschel, Entropy and long-range correlations in literary English. *Europhys. Lett.* **26**, 241–246 (1994)
8. H. Herzel, W. Ebeling, A.O. Schmitt, Entropies of biosequences: the role of repeats. *Phys. Rev. E* **50**, 5061–5071 (1994)
9. C.K. Peng, S.V. Buldyrev, A.L. Goldberger, S. Havlin, M. Simons, H.E. Stanley, Finite-size effects on long-range correlations: implications for analyzing DNA sequences. *Phys. Rev. E* **47**, 3730–3733 (1993)

10. M.P Paulus, M.A. Geyer, L.H. Gold, A.J. Mandell, Application of entropy measures derived from ergodic theory of dynamical systems to rat locomotor behavior. *Proc. Natl. Acad. Sci. USA* **87**, 723–727 (1990)
11. P. Faure, H. Neumeister, D.S. Faber, H. Korn, Symbolic analysis of swimming trajectories reveals scale invariance and provides model for fish locomotion. *Fractals* **11**, 233–243 (2003)
12. K. Doba, L. Pezard, A. Lesne, V. Christophe, J.L. Nandrino, Dynamics of emotional expression in autobiographical speech of patients with anorexia nervosa. *Psychol. Rep.* **101**, 237–249 (2007)
13. K. Doba, J.L. Nandrino, A. Lesne, J. Vignau, L. Pezard, Organization of the narrative components in the autobiographical speech of anorexic patients: a statistical and non-linear dynamical analysis. *New Ideas Psychol.* **26**, 295–308 (2008)
14. S.P. Strong, R.B. Koberle, R.R. de Ruyter van Steveninck, W. Bialek, Entropy and information in neural spike trains. *Phys. Rev. Lett.* **80**, 197–200 (1998)
15. J.M. Amigo, J. Szczepanski, E. Wajnryb, M.V. Sanchez-Vives, Estimating the entropy rate of spike trains via Lempel-Ziv complexity. *Neural Comput.* **16**, 717–736 (2004)
16. J.P. Eckmann, S. Kamphorst, D. Ruelle, Recurrence plots of dynamical systems. *Europhys. Lett.* **5**, 973–977 (1987)
17. L.L. Trulla, A. Giuliani, J.P. Zbilut, C.L. Webber, Jr., Recurrence quantification analysis of the logistic equation with transients. *Phys. Lett. A* **223**, 255–260 (1996)
18. C.L. Webber, J.P. Zbilut, Dynamical assessment of physiological systems and states using recurrence plot strategies. *J. Appl. Physiol.* **76**, 965–973 (1994)
19. C.L. Webber, J.P. Zbilut, Recurrence quantifications: feature extractions from recurrence plots. *Int. J. Bifurcat. Chaos* **17**, 3467–3475 (2007)
20. N. Marwan, M.C. Romano, M. Thiel, J. Kurths, Recurrence plots for the analysis of complex systems. *Phys. Rep.* **438**, 237–329 (2007)
21. H. Kantz, T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, 1997)
22. M. Thiel, M.C. Romano, P.L. Read, J. Kurths, Estimation of dynamical invariants without embedding by recurrence plots. *Chaos* **14**, 234–243 (2004)
23. C.S. Daw, C.E.A. Finney, E.R. Tracy, A review of symbolic analysis of experimental data. *Rev. Sci. Instrum.* **74**, 916–930 (2003)
24. D. Lind, B. Marcus, *An Introduction to Symbolic Dynamics and Coding*. (Cambridge University Press, Cambridge, 1995)
25. J. Guckenheimer, P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*. (Springer, Berlin, 1983)
26. R.L. Davidchack, Y.C. Lai, E.M. Bollt, M. Dhamala, Estimating generating partitions of chaotic systems by unstable periodic orbits. *Phys. Rev. E* **61**, 1353–1356 (2000)
27. E.M. Bollt, T. Stanford, Y.C. Lai, K. Zyczkowski, Validity of threshold-crossing analysis of symbolic dynamics from chaotic time series. *Phys. Rev. Lett.* **85**, 3524–3527 (2000)
28. E.M. Bollt, T. Stanford, Y.C. Lai, K. Zyczkowski, What symbolic dynamics do we get with a misplaced partition? On the validity of threshold crossings analysis of chaotic time-series. *Physica D*, **154**, 259–286 (2001)
29. W. Ebeling, G. Nicolis, Word frequency and entropy of symbolic sequences: a dynamical perspective. *Chaos, Solitons Fractals* **2**, 635–650 (1992)
30. H. Herzel, I. Grosse, Measuring correlations in symbol sequences. *Physica A* **216**, 518–542 (1995)
31. R. Badii, A. Politi, Thermodynamics and complexity of cellular automata. *Phys. Rev. Lett.* **78**, 444–447 (1997)
32. P. Castiglione, M. Falcioni, A. Lesne, A. Vulpiani, *Chaos and Coarse-Graining in Statistical Mechanics* (Cambridge University Press, Cambridge, 2008)
33. J. Kurths, A. Voss, A. Witt, P. Saparin, H.J. Kleiner, N. Wessel, Quantitative analysis of heart rate variability. *Chaos* **5**, 88–94 (1995)
34. T. Schürmann, P. Grassberger, Entropy estimation of symbol sequences. *Chaos* **6**, 414–427 (1996)

35. A. Lesne, J.L. Blanc, L. Pezard, Entropy estimation of very short symbolic sequences. *Physical Review E* **79**, 046208 (2009)
36. A. Lempel, J. Ziv, On the complexity of finite sequences. *IEEE Trans. Inf. Theory* **22**, 75–81 (1976)
37. J. Ziv, A. Lempel, A universal algorithm for sequential data compression. *IEEE Trans. Inf. Theory* **23**, 337–343 (1977)
38. J. Ziv, A. Lempel, Compression of individual sequences by variable rate coding. *IEEE Trans. Inf. Theory* **24**, 530–536 (1978)
39. P. Faure, A. Lesne, Recurrence plots for symbolic sequences. *Int. J. Bifurcat. Chaos* **20**, 1731–1749 (2010)
40. L. Breiman, The individual ergodic theorem of information theory. *Ann. Math. Stat.* **28**, 809–811 (1957)
41. B. McMillan, The basic theorems of information theory. *Ann. Math. Stat.* **24**, 196–219 (1953)
42. P. Grassberger, I. Procaccia, Computing the Kolmogorov entropy from a chaotic signal. *Phys. Rev. A* **28**, 2591–2593 (1983)
43. A. Cohen, I. Procaccia, Computing the Kolmogorov entropy from time signals of dissipative and conservative dynamical systems. *Phys. Rev. A* **31**, 1872–1882 (1985)
44. J.P. Eckmann, D. Ruelle, Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys.* **57**, 617–656 (1985)
45. P. Faure, H. Korn, A new method to estimate the Kolmogorov entropy from recurrence plots: its application to neuronal signals. *Physica D* **122**, 265–279 (1998)
46. M. Thiel, M.C. Romano, J. Kurths, R. Meucci, E. Allaria, T. Arecchi, Influence of observational noise on the recurrence quantification analysis. *Physica D* **171**, 138 (2002)
47. C. Letellier, Estimating the Shannon entropy: recurrence plots versus symbolic dynamics. *Phys. Rev. Lett.* **96**, 254102 (2006)
48. P. Gaspard, X.J. Wang, Noise, chaos, and (ϵ, τ) -entropy per unit time. *Phys. Rep.* **235**, 291 (1993)
49. M.C. Casdagli, Recurrence plots revisited. *Physica D* **108**, 12–44 (1997)
50. M.S. Baptista, E.J. Ngamga, P.R.F. Pinto, M. Brito, J. Kurths, Kolmogorov-Sinai entropy from recurrence times. *Phys. Lett. A* **374**, 1135–1140 (2010)
51. A.D. Wyner, J. Ziv, Some asymptotic properties of the entropy of a stationary ergodic data source with applications to data compression. *IEEE Trans. Inf. Theory* **35**, 1250–1258 (1989)
52. E.J. Ngamga, D.V. Senthilkumar, A. Prasad, P. Parmananda, N. Marwan, J. Kurths, Distinguishing dynamics using recurrence-time statistics. *Phys. Rev. E* **85**, 026217 (2012)

Chapter 3

Identifying Coupling Directions by Recurrences

Yong Zou, M. Carmen Romano, Marco Thiel, and Jürgen Kurths



Abstract The identification of the coupling direction from measured time series taking place in a group of interacting components is an important challenge for many experimental studies. In *Part I* of this chapter, we introduce a method to detect and quantify the asymmetry of the coupling between two interacting systems based on their recurrence properties. This method can detect the direction of the coupling in weakly as well as strongly coupled systems. It even allows detecting the asymmetry

The main part of this chapter is compiled with permission from publications [33, 56], respectively, Copyright ©2007 American Physical Society, and Copyright ©2011 World Scientific Publishing Company.

Y. Zou (✉)

Department of Physics, East China Normal University, Shanghai 200062, China

Potsdam Institute for Climate Impact Research, P.O. Box 601203, 14412 Potsdam, Germany

e-mail: yzou@phy.ecnu.edu.cn

M.C. Romano • M. Thiel

Institute for Complex Systems and Mathematical Biology, University of Aberdeen, Aberdeen AB243UE, UK

J. Kurths

Potsdam Institute for Climate Impact Research, P.O. Box 601203, 14412 Potsdam, Germany

Department of Physics, Humboldt University Berlin, Newtonstr. 15, 12489 Berlin, Germany

Institute for Complex Systems and Mathematical Biology, University of Aberdeen, Aberdeen AB243UE, UK

of the coupling in the more challenging case of structurally different systems and it is very robust against noise. We also address the problem of detecting the asymmetry of the coupling in passive experiments, i.e., when the strength of the coupling cannot be systematically changed, which is of great relevance for the analysis of experimental time series. *Part II* of this chapter hinges on a generalisation of conditional probability of recurrence to the case of multivariate time series where indirect interactions might be present. We test our method by an example of three coupled Lorenz systems. Our results confirm that the proposed method has much potential to identify indirect coupling.

3.1 Part I: Estimation of the Direction of the Coupling by Conditional Probabilities of Recurrence

3.1.1 Introduction: Part I

The interplay among different complex dynamical systems is a central issue in nonlinear dynamics as well as in nonlinear time series analysis. Under certain assumptions different types of synchronisation can occur between the interacting systems. This topic has been intensively studied in the last years and has been applied to various fields, such as physics, engineering and biology [2, 6, 8, 9, 29, 52, 53, 55]. In such systems it is important not only to analyse the synchronisation but also to identify causal (drive-response) or mutual relationships. There are four major approaches to address this problem: state-space based methods [4, 31, 32, 42], information theory based methods [26, 41, 54], methods based on the interrelations between the phases of the systems under consideration [37, 38], and recurrence-based methods [12, 13, 18, 24, 33].

In the state-space based approach, the state vectors are usually reconstructed by means of delay embedding [19]. The direction of the coupling is then assessed by considering the correspondence between neighbours in the phase spaces of the driver and response. If there exists a functional relationship between the driver X and the response system Y , i.e., $\mathbf{y}(t) = \Psi(\mathbf{x}(t))$, they are said to be generalised synchronised [20, 39]. If Ψ exists and is smooth, it follows that close states of the driver will be mapped to close states of the response. However, if Ψ is bijective, also close states of the response will be mapped to close states of the driver. Therefore, if X and Y are generalised synchronised it is in general impossible to assess the direction of the coupling reliably and the state-space based methods are only applicable in the non-synchronised regime [25, 43].

The understanding of a driver-response relationship was firstly evaluated in a linear framework by bivariate autoregressive models, by means of Granger causality [16]. This has been mainly applied to economy and neurosciences [7]. From the nonlinear perspective, there are several methods based on information theory to determine the direction of the coupling [5, 14, 26, 41, 47]. They are usually applied to systems which are strongly coupled. In order to treat also weakly coupled

systems, the phases of the signals are determined beforehand, and then information theory based indices are applied to the phases [27, 28]. In [38], a technique based on the fitting of the functional relationship between the phases of the two interacting systems has been proposed to detect and quantify the asymmetry in the coupling. A systematic comparison between the phase-dynamics and the state-space approach in the case of weak directional coupling has been done in [43], where the authors concluded that neither one of the approaches is generally superior and that both approaches have difficulties in assessing the direction of the coupling in systems which are structurally different.

In Part I of this chapter, we summarize the recently proposed method to uncover directional coupling. This approach is based on the recurrence properties of both interacting systems. The concept of recurrence has been used to detect relationships between interacting systems in [46], where the so-called synchronisation likelihood has been introduced. This method allows for a multivariate analysis of generalised synchronisation. Moreover, in [34] the concept of recurrence has been used to quantify a weaker form of synchronisation, namely phase synchronisation. Here, we extend these measures in order to detect the direction of the coupling. The proposed method is rather straightforward to compute, in contrast to the more complicated information theory approaches. Furthermore, it has the advantage that it is applicable to both weak and strong directional coupling, as well as to structurally different systems.

The outline of Part I of this chapter is as follows: in Sect. 3.1.2 we introduce measures for the analysis of directional coupling based on recurrences. In Sect. 3.1.3 we demonstrate the proposed measures in some numerical examples and discuss the choice of the parameters of the method in Sect. 3.1.4. In Sect. 3.1.5 we discuss the dependence of these measures on observational noise. We consider in Sect. 3.1.6 the problem of passive experiments, where the coupling strength between the two interacting systems cannot be varied systematically. In Sect. 3.1.7 we compare the proposed method with other existing techniques and, finally, we give some conclusions.

3.1.2 Detection of the Coupling Direction by Recurrences

Recurrence is a fundamental property of dynamical systems. The concept of recurrence was introduced by Poincaré [30], where he showed that the trajectory of a dynamical system with a measure preserving flow recurs infinitely many times to some neighbourhood of a former visited state on an invariant set in phase space. There are many different techniques in nonlinear dynamics which exploit the concept of recurrence [1, 15, 40]. We concentrate on the method of Recurrence Plots (RPs), introduced by Eckmann et al. to visualise the behaviour of dynamical systems in the phase space [10]. They are defined by means of the recurrence matrix

$$R_{i,j}^X = \Theta(\varepsilon - \|\mathbf{x}_i - \mathbf{x}_j\|), \quad i, j = 1, \dots, N, \quad (3.1)$$

where \mathbf{x}_i denotes the state of the system X at time $i \Delta t$ with Δt being the sampling rate, ε is a predefined threshold, $\Theta(\cdot)$ is the Heaviside function and N is the length of the trajectory considered. The RP is obtained plotting a dot at the coordinates (i, j) if $R_{i,j} = 1$. By looking at the patterns of the RP, one gets at the outset a visual impression about the dynamics of the system under consideration. In order to go beyond the visual impression, several measures have been proposed to quantify the patterns in the RP. They have found numerous applications in very different kinds of systems [21, 22]. Moreover, somehow more formal relationships between the patterns obtained in RPs and main dynamical invariants, such as K_2 and D_2 , have been found [11, 51] (cf. Chap. 2). It has also been shown that the RP contains all necessary information to reconstruct the underlying trajectory, at least topologically [49].

The method of RPs has been extended to Joint Recurrence Plots (JRPs) to analyse the interplay of two or more dynamical systems [36, 46]. The JRP of X and Y is defined as

$$J_{R_{i,j}}^{X,Y} = \Theta(\varepsilon_X - \|\mathbf{x}_i - \mathbf{x}_j\|) \Theta(\varepsilon_Y - \|\mathbf{y}_i - \mathbf{y}_j\|), \quad (3.2)$$

i.e., a joint recurrence occurs if the system X recurs in its own phase space and simultaneously, the system Y recurs also in its own phase space. Based on JRPs it is possible to analyse different kinds of synchronisation of coupled complex systems [34, 45, 46]. In order to illustrate this, we consider two rather different chaotic oscillators, namely the Rössler system

$$\begin{aligned} \dot{x}_1 &= 2 + x_1(x_2 - 4), \\ \dot{x}_2 &= -x_1 - x_3, \\ \dot{x}_3 &= x_2 + 0.45x_3, \end{aligned} \quad (3.3)$$

which drives the Lorenz system

$$\begin{aligned} \dot{y}_1 &= -10(y_1 - y_2), \\ \dot{y}_2 &= 28u - y_2 - uy_3, \\ \dot{y}_3 &= uy_2 - 8/3y_3, \end{aligned} \quad (3.4)$$

by means of the variable $u = x_1 + x_2 + x_3$. In [20] it has been shown that the driven Lorenz system is asymptotically stable and that both systems are in generalised synchronisation. Hence, two close neighbours in the phase space of the driver system correspond to two close neighbours in the phase space of the driven system [39]. This relationship is reflected very clearly in the RPs of both systems. In Fig. 3.1a, c we plot the trajectories in phase space of the Rössler Eq. (3.3) and of the Lorenz system Eq. (3.4), respectively. To calculate their corresponding RPs, we have used the third component of each system and reconstructed the respective trajectories in phase space using delay embedding [19] with embedding dimension

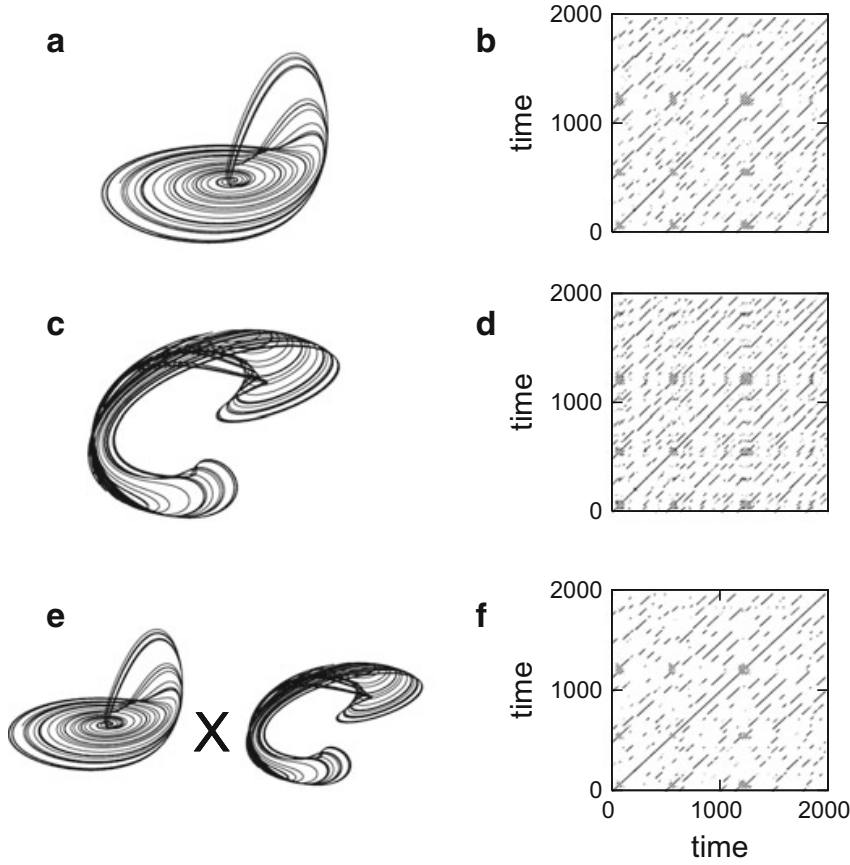


Fig. 3.1 (a) Rössler driving system, (b) the RP of the Rössler system ($m = 3, \tau = 5$), (c) the driven Lorenz system, (d) the RP of the Lorenz system ($m = 7, \tau = 5$), (e) representation of the “joint” system, and (f) the joint recurrence plot of both systems. The threshold for the computation of the RPs has been chosen so that the recurrence rate (number of recurrence points divided by N^2) is equal for both systems. In this case the recurrence rate was 0.005. The equations were integrated using fourth order Runge–Kutta of and the sampling time was 0.2

$m = 7$ and time delay $\tau = 5$ (the time step between two consecutive points being 0.2), since dealing with experimental time series, usually only one observable of the system is available. Even though the shapes of both attractors in the phase space look rather different (Fig. 3.1a, c), both RPs are very similar (Fig. 3.1b, d). Therefore, the joint recurrence plot (Fig. 3.1f) resembles the very similar recurrence patterns as the RPs of the single systems.

This property of joint recurrence plots has been treated in detail in [34], where it has been used for the detection of generalised synchronisation, also in more difficult cases where other methods, such as the mutual false nearest neighbours, are not appropriate any longer. In [46] the authors have introduced the synchronisation likelihood, which is a multivariate measure for generalised synchronisation. This measure is based on a very similar concept to the joint recurrence matrix of Eq. (3.2). However, the thresholds ε_X and ε_Y are not fixed for the whole trajectories, but are dependent on time.

Only considering the concept of joint recurrence is not sufficient to identify which system is the driver and which one is the response. In order to accomplish that, it is necessary to assess conditional probabilities of recurrence. Therefore, we propose the mean conditional probabilities of recurrence (MCR) between two systems X and Y , which are defined as follows

$$M_{CR}(Y|X) = \frac{1}{N} \sum_{i=1}^N p(\mathbf{y}_i|\mathbf{x}_i) = \frac{1}{N} \sum_{i=1}^N \frac{\sum_{j=1}^N J_{Ri,j}^{X,Y}}{\sum_{j=1}^N R_{i,j}^X}, \quad (3.5)$$

and

$$M_{CR}(X|Y) = \frac{1}{N} \sum_{i=1}^N p(\mathbf{x}_i|\mathbf{y}_i) = \frac{1}{N} \sum_{i=1}^N \frac{\sum_{j=1}^N J_{Ri,j}^{X,Y}}{\sum_{j=1}^N R_{i,j}^Y}, \quad (3.6)$$

where $p(\mathbf{y}_i|\mathbf{x}_i)$ is an estimate of the probability that the trajectory of Y recurs to the neighbourhood of \mathbf{y}_i under the condition that the trajectory of X recurs to the neighbourhood of \mathbf{x}_i ($p(\mathbf{x}_i|\mathbf{y}_i)$ is defined analogously). One can consider these measures as an extension of the methods presented in [34, 46].

The criterion that we use for detecting the asymmetry of the coupling is the following

$$\text{If } X \text{ drives } Y, \quad M_{CR}(Y|X) < M_{CR}(X|Y). \quad (3.7a)$$

$$\text{If } Y \text{ drives } X, \quad M_{CR}(X|Y) < M_{CR}(Y|X). \quad (3.7b)$$

If the coupling is symmetric, then $M_{CR}(X|Y) = M_{CR}(Y|X)$.

This criterion might appear counterintuitive at first, because if X is the driver, one could think that the probability of recurrence of a state \mathbf{y}_i given that the state \mathbf{x}_i recurs is larger than vice versa, since X is independent of Y .

A heuristic argumentation for this criterion is the following: if X drives Y , the dimension of Y will be in general larger than the dimension of X , because the evolution of Y is determined by both the states of X and Y . Moreover, the higher the complexity of Y , the smaller is the probability of recurrence of $\mathbf{y}_i \quad \forall i$. Hence, by increasing the coupling strength from X to Y , the probability $p(\mathbf{y}_i)$ that the trajectory of Y recurs to the neighbourhood of \mathbf{y}_i will decrease. In contrast, the complexity of X remains constant with increasing coupling strength, because the evolution of X depends only on the states of X . Hence, the probability $p(\mathbf{x}_i)$ that the trajectory of X recurs to the neighbourhood of \mathbf{x}_i does not change with the coupling strength. We choose the thresholds ε_X and ε_Y in such a way, that if the coupling strength is equal to zero, $\langle p(\mathbf{x}_i) \rangle = \langle p(\mathbf{y}_i) \rangle$. Therefore, if the coupling strength from X to Y is larger than zero, in general $p(\mathbf{y}_i) < p(\mathbf{x}_i)$. That implies $p(\mathbf{x}_i, \mathbf{y}_i)/p(\mathbf{x}_i) < p(\mathbf{x}_i, \mathbf{y}_i)/p(\mathbf{y}_i)$ and hence, $M_{CR}(Y|X) < M_{CR}(X|Y)$.

3.1.3 Numerical Examples

In this section we illustrate the performance of the proposed measures for the direction of the coupling by three kinds of examples: strongly coupled systems (close to the onset of complete synchronisation), weakly coupled systems (close to the onset of phase synchronisation), and structurally different systems. The number of data points of the trajectories used in each case, if not stated otherwise, is equal to 10,000 throughout the chapter.

3.1.3.1 Strongly Coupled Systems

We consider two unidirectionally coupled Hénon maps, given by the following equations

$$\begin{aligned}x_1(i + 1) &= 1.4 - x_1(i)^2 + b_1 x_2(i), \\x_2(i + 1) &= x_1(i)\end{aligned}\tag{3.8}$$

for the driving system X , and

$$\begin{aligned}y_1(i + 1) &= 1.4 - (\mu x_1(i) y_1(i) + (1 - \mu) y_1^2(i)) + b_2 y_2(i), \\y_2(i + 1) &= y_1(i)\end{aligned}\tag{3.9}$$

for the response system Y [32], where μ is the coupling strength. We analyse both the case of identical systems ($b_1 = b_2 = 0.3$) and non-identical systems ($b_1 = 0.1, b_2 = 0.3$). To mimic this problem for data analysis, we assume that we have observed the two scalar time series $\{x_1(i)\}_{i=1}^N$ and $\{y_1(i)\}_{i=1}^N$. Hence, we have to reconstruct the trajectories of X and Y in phase space [19]; this will be done by delay embedding. We choose embedding dimension $m = 3$ and time delay $\tau = 1$, but we note that the results are qualitatively the same with other reasonable choices. The values of the thresholds ε_X and ε_Y have been chosen such that for no coupling both mean probabilities of recurrences $\langle p(\mathbf{x}(i)) \rangle$ and $\langle p(\mathbf{y}(i)) \rangle$ are equal to 0.01. We use 10,000 data points and compute the indices $M_{CR}(X|Y)$ and $M_{CR}(Y|X)$ in dependence on the coupling strength μ . The results are shown in Fig. 3.2.

For two identical Hénon maps (Fig. 3.2a), the onset to identical synchronisation occurs at approximately $\mu = 0.65$, as reported in [26, 32]. As expected from this, we yield for $\mu > 0.65$, $M_{CR}(X|Y) = M_{CR}(X|Y)$. Before the onset of synchronisation, we get $M_{CR}(X|Y) > M_{CR}(Y|X)$, indicating correctly the direction of the coupling.

On the other hand, for the non-identical Hénon maps, the onset to generalised synchronisation occurs at approximately $\mu = 0.4$ [32]. Note that in general the detection of the directionality is only possible before the onset of synchronisation. In the case of identical synchronisation, the series $\{x_i\}$ and $\{y_i\}$ are identical and hence there is no possibility of establishing the causal relationship between X and Y just from the data. This argument can be also extended to the case

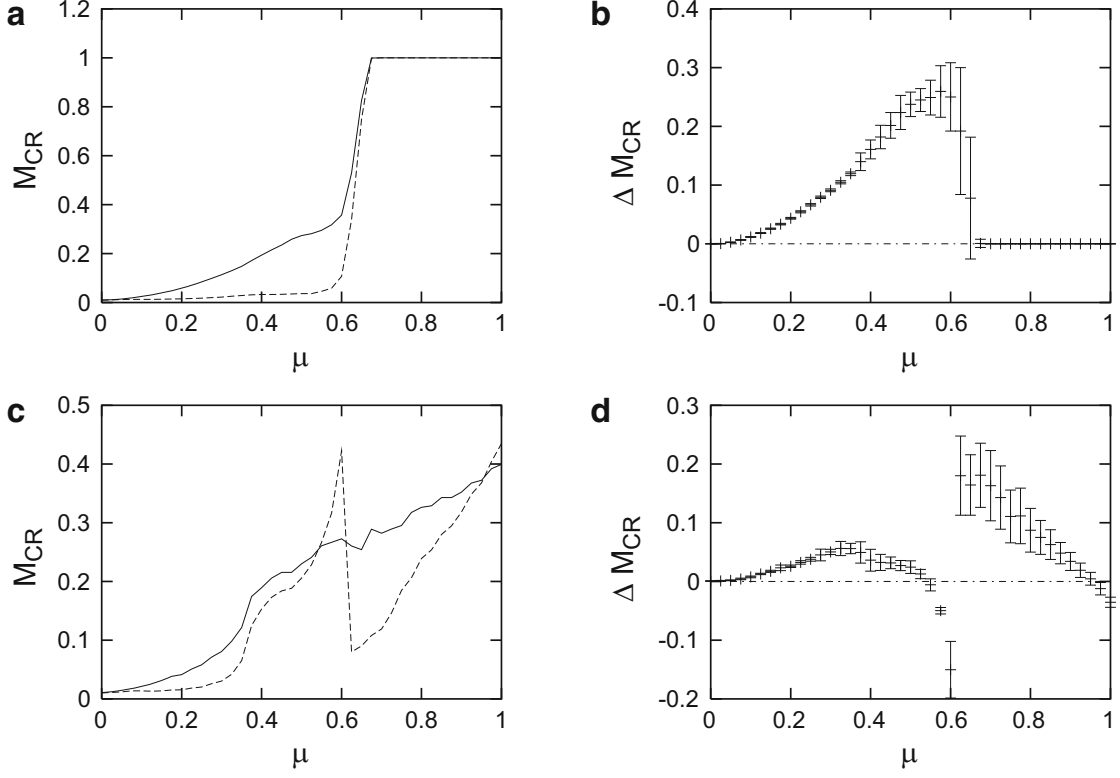


Fig. 3.2 Mean conditional probabilities of recurrence $M_{CR}(X|Y)$ (solid) and $M_{CR}(Y|X)$ (dashed) for two unidirectionally coupled identical (a) and non-identical (c) Hénon maps. The system X is in both cases the driver, and hence, $M_{CR}(X|Y) > M_{CR}(Y|X)$. For each value of the coupling strength μ , the mean value over 100 trajectories for uniformly distributed initial conditions has been computed. In b and d we have plotted the mean value of the difference $\Delta M_{CR} = M_{CR}(X|Y) - M_{CR}(Y|X)$ over 100 trajectories and the corresponding standard deviation for the identical and non-identical Hénon systems, respectively. The zero line is also plotted for orientation (dotted-dashed)

of generalised synchronisation, where the systems are related by a one-to-one function [4]. Therefore, in the case of the two non-identical Hénon maps the directionality parameters are reliable for $0 < \mu < 0.4$. The sharp drop of $M_{CR}(Y|X)$ (dashed curve in Fig. 3.2c) at approximately $\mu = 0.6$ is due to the non-monotonic dependence of the maximum Lyapunov exponent of the response system on the coupling strength [32].

3.1.3.2 Weakly Coupled Systems

Now we study two non-identical unidirectionally coupled Lorenz systems, given by the equations

$$\begin{aligned}\dot{x}_1 &= 10(x_1 - x_2), \\ \dot{x}_2 &= 40x_1 - x_2 - x_1x_3, \\ \dot{x}_3 &= x_1x_2 - 8/3x_3,\end{aligned}\tag{3.10}$$

for the driver system X and

$$\begin{aligned}\dot{y}_1 &= 10(y_2 - y_1) + \mu(x_1 - y_1), \\ \dot{y}_2 &= 35y_1 - y_2 - y_1y_2, \\ \dot{y}_3 &= y_1y_2 - 8/3y_3,\end{aligned}\tag{3.11}$$

for the response system Y . The equations have been integrated by a fourth-order Runge–Kutta algorithm and the time step between two consecutive points is equal to 0.03. We use 10,000 data points and assume that only the scalar variables x_3 and y_3 have been observed. The embedding parameters used for the reconstruction are $m = 10$ and $\tau = 12$. As in the former case, the results do not depend on the details of this choice. We have not used the optimal embedding parameters which can be estimated by, e.g., the methods of false nearest neighbours and the autocorrelation function, in order to show that the results are robust with respect to different embedding parameters [43]. We compute the directionality parameters MCR in dependence on the coupling strength μ between 0 and 10, which is before the onset of phase synchronisation [43]. The results are shown in Fig. 3.3a. We clearly see that $M_{CR}(X|Y) > M_{CR}(Y|X)$ for all computed values of the coupling strength μ , i.e., the recurrence based indices detect the direction of the coupling correctly. The values of the thresholds ε_X and ε_Y have been chosen such that for no coupling both mean probabilities of recurrences $\langle p(\mathbf{x}(i)) \rangle$ and $\langle p(\mathbf{y}(i)) \rangle$ are equal to 0.01. However, note that for $\mu = 0$, the values of $M_{CR}(X|Y)$ and $M_{CR}(Y|X)$ are larger and not equal to 0.01, as one would expect. This is because the estimated joint probability of recurrence is larger than $(0.01)^2$, due to the limited number of data used for the computation. Nevertheless, the expected qualitative behaviour, i.e. $M_{CR}(X|Y) > M_{CR}(Y|X)$ still holds, which is the important fact for our analysis.

The next example that we consider is a bidirectionally coupled system, namely, two stochastic Van der Pol oscillators with slightly different mean frequencies ω_x and ω_y

$$\begin{aligned}\ddot{x} &= 0.2(1 - x^2)\dot{x} - \omega_x^2x + \xi_x + 0.03(y - x), \\ \ddot{y} &= 0.2(1 - y^2)\dot{y} - \omega_y^2y + \xi_y + \mu(x - y),\end{aligned}\tag{3.12}$$

where $\omega_x = 1.02$ and $\omega_y = 0.98$, ξ_x and ξ_y are independent Gaussian white noise with standard deviation 0.04. This example has been considered in [38, 43]. The equations have been integrated with the Euler scheme and the sampling time was 0.1π . The variables x and y have been used to reconstruct the phase space with embedding dimension 10 and delay 12, as in the former case. The thresholds ε_X and ε_Y have been chosen such that for symmetrical coupling $\langle p(\mathbf{x}_i) \rangle = \langle p(\mathbf{y}_i) \rangle = 0.1$. The results for the indices MCR are shown in Fig. 3.3b in dependence on the coupling strength μ . For $\mu < 0.03$, $M_{CR}(X|Y) < M_{CR}(Y|X)$, since the coupling is stronger from Y to X than vice versa. At the coupling strength 0.03, we obtain $M_{CR}(X|Y) = M_{CR}(Y|X)$, because the coupling is symmetrical, and for $\mu > 0.03$,

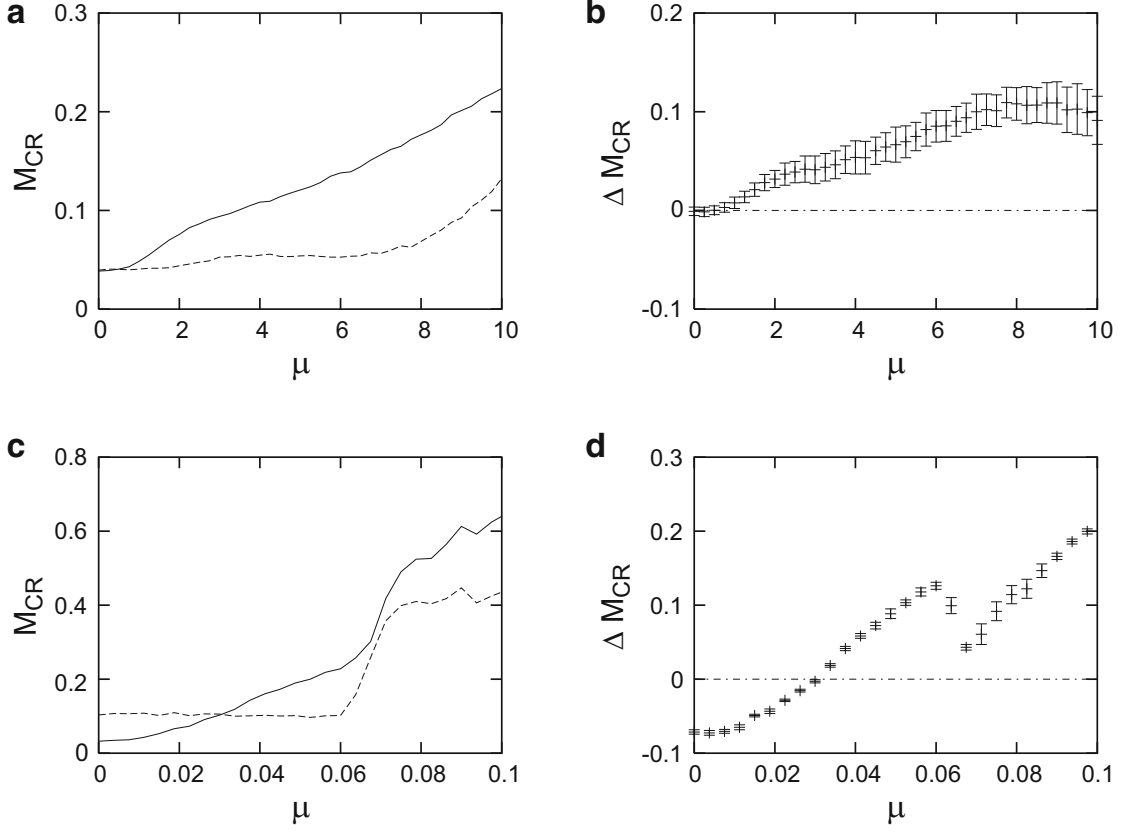


Fig. 3.3 Mean conditional probabilities of recurrence $M_{CR}(X|Y)$ (solid) and $M_{CR}(Y|X)$ (dashed) for (a): two weakly unidirectionally coupled non-identical Lorenz systems. For each value of the coupling strength μ , the mean value over 100 trajectories for uniformly distributed initial conditions has been computed. The system X is the driver, and hence, $M_{CR}(X|Y) > M_{CR}(Y|X)$. (c): Two weakly bidirectionally coupled stochastic Van der Pol oscillators. The coupling strength from X to Y is fixed and equal to 0.3. In b and d we have plotted the mean value of $\Delta M_{CR} = M_{CR}(X|Y) - M_{CR}(Y|X)$ and the corresponding standard deviation over 100 trajectories for each system, respectively. The zero line is also plotted for orientation (dotted-dashed)

we observe that $M_{CR}(X|Y) > M_{CR}(Y|X)$, because the coupling from X to Y is stronger than vice versa. Note that at $\mu \approx 0.06$ both oscillators become phase synchronised and the value of $M_{CR}(Y|X)$ increases much faster.

3.1.3.3 Structurally Different Systems

Next, we study the more challenging case of two structurally different systems, namely a stochastic Van der Pol system which drives a Rössler system. The equation of the driving system X is

$$\ddot{x} = 0.1(1 - x^2)\dot{x} - \omega_x^2 x + \xi_x, \quad (3.13)$$

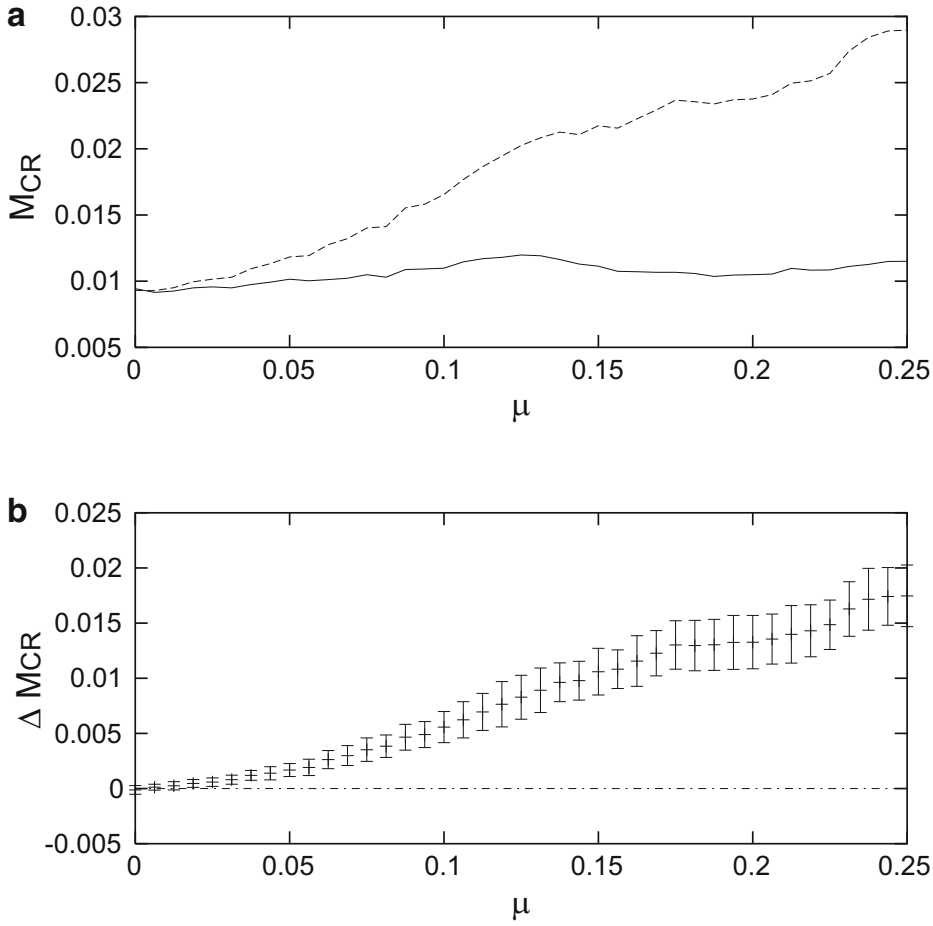


Fig. 3.4 (a) Mean conditional probabilities of recurrence $M_{CR}(X|Y)$ (solid) and $M_{CR}(Y|X)$ (dashed) for the chaotic Rössler system driven by the stochastic Van der Pol system. For each value of the coupling strength μ , the mean value over 100 trajectories for uniformly distributed initial conditions has been computed. The system X is the driver, and hence, we find $M_{CR}(X|Y) > M_{CR}(Y|X)$. (b) Mean value of the difference $\Delta M_{CR} = M_{CR}(X|Y) - M_{CR}(Y|X)$ and corresponding standard deviation over 100 trajectories. The zero line is also plotted for orientation (dotted-dashed)

where $\omega_x = 0.98$ and ξ_x is Gaussian white noise with standard deviation 0.05. The equations of the response system Y are given by

$$\begin{aligned}\dot{y}_1 &= -y_2 - y_3, \\ \dot{y}_2 &= y_1 + 0.15y_2 + \mu x, \\ \dot{y}_3 &= (y_1 - 10)y_3 + 0.2.\end{aligned}\tag{3.14}$$

The equations have been integrated with a Euler scheme and the sampling time was 0.1π . The phase space has been reconstructed using the variables x and y_1 and embedding dimension 10 and delay 12. The values of the thresholds ε_X and ε_Y have been chosen as in the former cases. The curves for MCR are shown in dependence on the coupling strength μ in Fig. 3.4. In this interval of values of the coupling

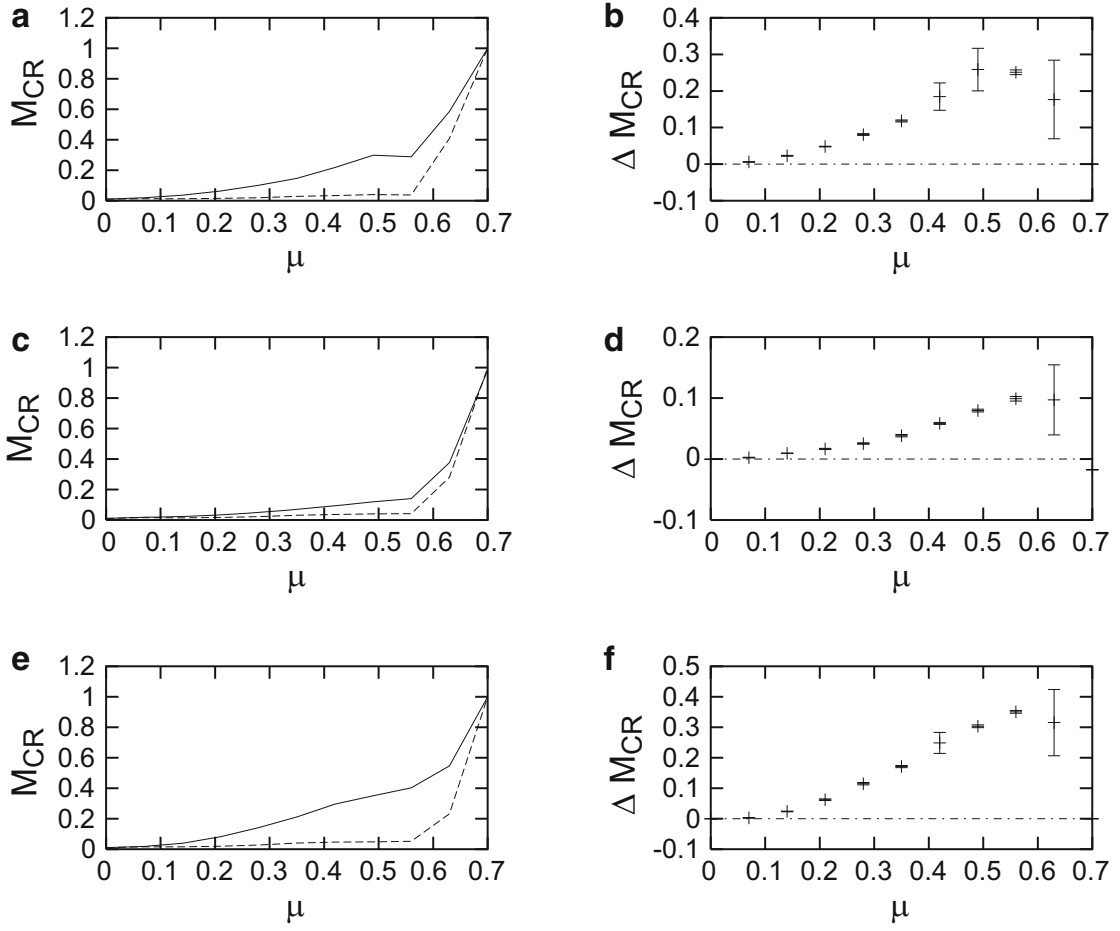


Fig. 3.5 Mean conditional probabilities of recurrence $M_{CR}(X|Y)$ (*solid*) and $M_{CR}(Y|X)$ (*dashed*) for two identical unidirectionally coupled Hénon maps for different choices of the embedding parameters: (a) $m = 3$, $\tau = 1$, (c) $m = 2$, $\tau = 3$, (e) $m = 5$, $\tau = 1$. For each value of the coupling strength μ , the mean value over ten trajectories for uniformly distributed initial conditions has been computed. In b, d and f the mean value and standard deviation over ten trajectories of the corresponding ΔM_{CR} are represented. The zero line is also plotted for orientation (*dotted-dashed*)

strength both systems are before the onset of phase synchronisation [43]. For all values of the coupling strength we obtain $M_{CR}(X|Y) > M_{CR}(Y|X)$, i.e., we are able to detect the direction of the coupling also in this case.

3.1.4 Choice of the Parameters

In order to compute the indices MCR, we need to fix four parameters: the embedding dimension m and the delay τ for the reconstruction of the phase space, and the thresholds ε_X and ε_Y for the computation of the recurrence matrices. As we have mentioned in the previous section, the special choice of the embedding parameters does not influence the results. In Fig. 3.5 we show the results for the

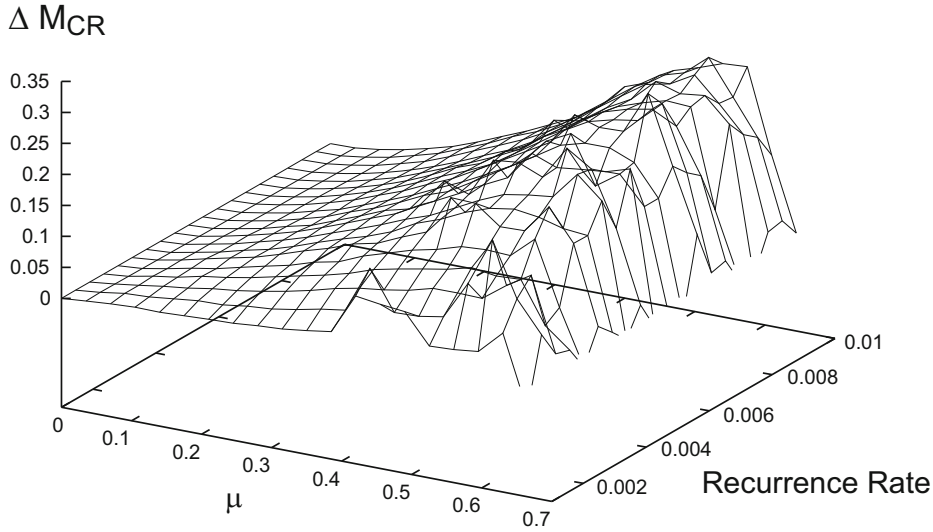


Fig. 3.6 $\Delta M_{CR} = M_{CR}(X|Y) - M_{CR}(Y|X)$ for two identical unidirectionally coupled Hénon maps in dependence on the coupling strength μ and the mean probability of recurrence or recurrence rate. For each value of the coupling strength μ , the mean value over ten trajectories for uniformly distributed initial conditions has been computed

direction parameters MCR for different choices of m and τ for the two identical unidirectionally coupled Hénon systems [Eqs. (3.8) and (3.9)]. We get, regardless of the choice of the embedding parameters, $M_{CR}(X|Y) > M_{CR}(Y|X)$ for all values of the coupling strength μ before the onset of synchronisation. This is the correct behaviour, since the system X is the driver and Y is the response.

With regard to the choice of the thresholds ε_X and ε_Y , we have mentioned in the previous section that they were chosen such that the mean probabilities of recurrence for both systems at coupling strength $\mu = 0$ are equal. In this way, it is not necessary to normalise the data x_i and y_i beforehand. In the numerical examples considered in Sect. 3.1.3, we chose the mean probability of recurrence to be equal to 0.01. In order to demonstrate how the results depend on this choice, we show in Fig. 3.6 $\Delta M_{CR} = M_{CR}(X|Y) - M_{CR}(Y|X)$ in dependence on the coupling strength and on the mean probability of recurrence (labelled as “Recurrence Rate” in the plot) for the Hénon systems [Eqs. (3.8) and (3.9)]. As system X is the driver, we expect that the surface ΔM_{CR} takes only positive values, which is the case in fact. Hence, we see that the estimation of MCR does not depend crucially on the choice of the thresholds ε_X and ε_Y . Hence, for a rather broad range of values of the thresholds, the direction of the coupling can be estimated correctly.

3.1.5 Influence of Noise

We now study the influence of observational noise on the MCR measures [Eqs. (3.5) and (3.6)]. Therefore, we add different levels of noise to the scalar time series

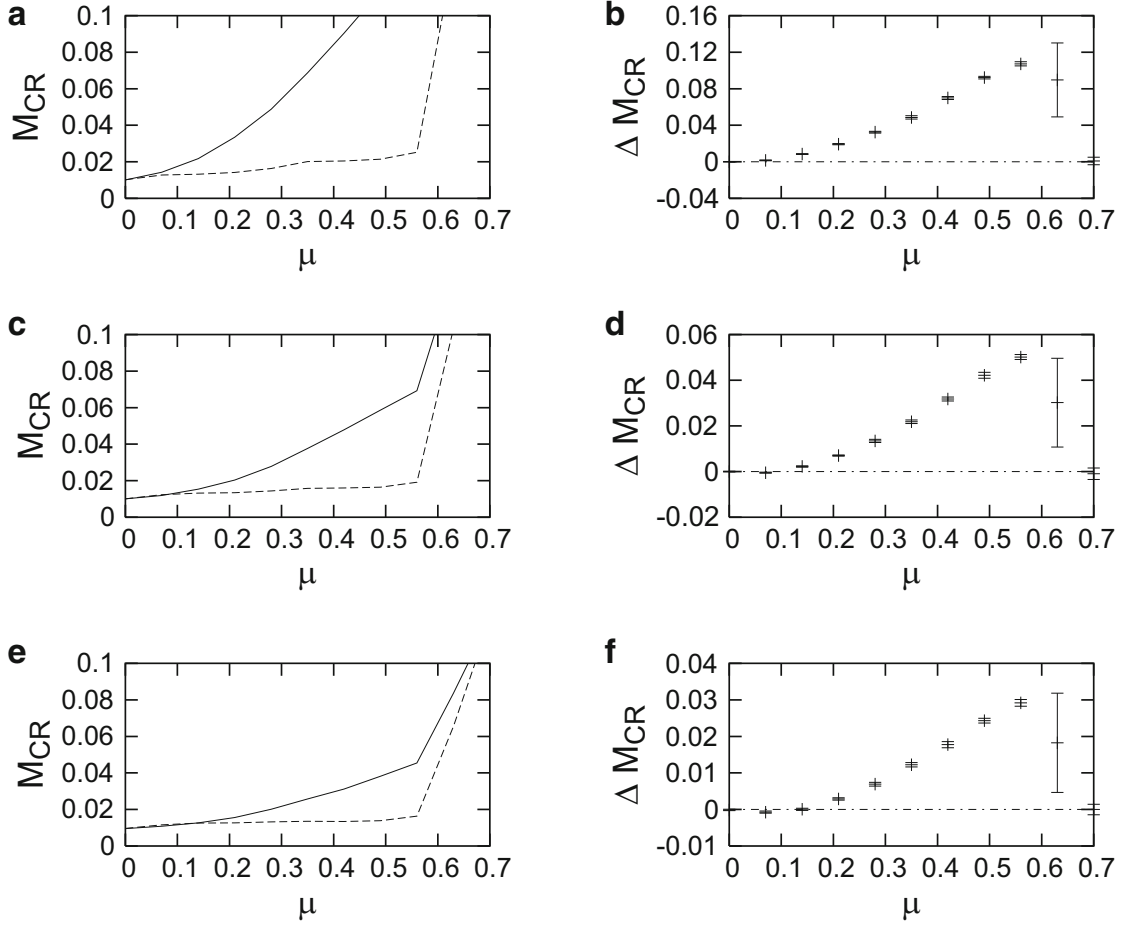


Fig. 3.7 $M_{CR}(X|Y)$ and $M_{CR}(Y|X)$ for two identical unidirectionally coupled Hénon maps contaminated by uniformly distributed noise in dependence on the coupling strength μ . (a) $\gamma = 0.2$, (c) $\gamma = 0.4$, (e) $\gamma = 0.6$. For each value of the coupling strength μ , the mean value over ten trajectories for uniformly distributed initial conditions has been computed. In b, d and (f) the corresponding mean value of the difference ΔM_{CR} and the standard deviation over ten trajectories are plotted. The zero line is also plotted for orientation (*dotted-dashed*)

$\{x_i\}_{i=1}^N$ and $\{y_i\}_{i=1}^N$, so that we compute the MCR indices for the series $x'_i = x_i + \gamma\sigma_x\eta_i^x$ and $y'_i = y_i + \gamma\sigma_y\eta_i^y$, where γ denotes the level of noise, σ_x and σ_y are the standard deviation of x_i and y_i , respectively, and η_x and η_y are two independent realisations of uniformly distributed random noise between -0.5 and 0.5 . Figure 3.7 shows the results obtained for the MCR indices for three different values of γ corresponding to 20, 40 and 60 % of observational noise.

We observe that when the level of noise γ increases, it becomes more difficult to detect the asymmetry of the coupling for very small values of the coupling strength μ , because both curves $M_{CR}(X|Y)$ and $M_{CR}(Y|X)$ are almost equal. The larger the level of noise, the stronger must be the coupling strength in order to detect the asymmetry. Nevertheless, even with such high levels of observational noise, the asymmetry of the coupling can still be correctly detected for relatively small values of the coupling strength. Hence, we conclude that the MCR indices are a rather robust measure for the detection of the asymmetry of the coupling, also in the presence of relative high levels of observational noise.

3.1.6 *Passive Experiments*

One crucial problem of all measures for the detection of asymmetry of the coupling is the assessment of the significance of the results for passive experiments, i.e., when the coupling strength between both systems X and Y cannot be varied systematically in experiments. In such cases, we usually just have one scalar measurement sequence for each system $\{x_i\}_{i=1}^N$ and $\{y_i\}_{i=1}^N$ for a fixed coupling strength μ . This is the case in numerous situations. For example, the experimental data used in [23, 41] to illustrate the applicability of the method for the detection of the asymmetry of the coupling, are time series of breath rate and instantaneous heart rate of a sleeping human. It is very hard, if not impossible, to change the coupling strength between the respiratory and cardiological system of a person in a systematic way. The authors in [4] apply their proposed method for the detection of asymmetry of the coupling to intracranially recorded EEG data. In this case, it is also obvious that it is not possible to change the coupling strength between different areas of the brain, in a controlled manner.

In all these cases we obtain just one value for the directionality indices and then it is not trivial to decide whether the computed values have been obtained just by chance or whether they are significant. In order to address this question, we propose the following statistical test. Our null hypothesis is that the two systems X and Y are independent. To test this null hypothesis, we generate the so called *natural or twin surrogates* [50]. Suppose that we have one time series for each system $\{x_i\}_{i=1}^N$ and $\{y_i\}_{i=1}^N$ for a fixed coupling strength μ . The natural surrogates are trajectories from the same underlying dynamical systems X and Y with identical coupling strength μ between both of them, but starting at different initial conditions. We denote them by $\{x_i^s\}_{i=1}^N$ and $\{y_i^s\}_{i=1}^N$. If we have computed the directionality indices $M_{CR}(x|y)$ and $M_{CR}(y|x)$ for the measured time series, we can compare the obtained values with the distribution of $M_{CR}(x|y^s)$ and $M_{CR}(y^s|x)$,¹ respectively, generated from a large number of surrogates. If both systems X and Y are independent, the value $M_{CR}(x|y)$ will not differ significantly from the distribution of the values $M_{CR}(x|y^s)$. Otherwise, we can reject the null hypothesis, indicating that the obtained values for the directionality indices are significant.

At this point, the following question arises naturally: dealing with experimental data, one usually does not have a model for the governing dynamics. Then, how can one generate natural surrogates? The answer to this question has been addressed in [50], where an algorithm based on recurrence has been proposed to generate natural surrogates without knowing the underlying equations of the system. These recurrence based surrogates are called *twin surrogates* and have been applied in [50] to tackle the problem of passive experiments in phase synchronisation.

We now show the performance of the MCR method applying the twin surrogates test to analyse the estimation of the asymmetry of the coupling in [Eqs.(3.8)

¹Note that considering $M_{CR}(x^s|y)$ and $M_{CR}(y|x^s)$ would yield the same result.

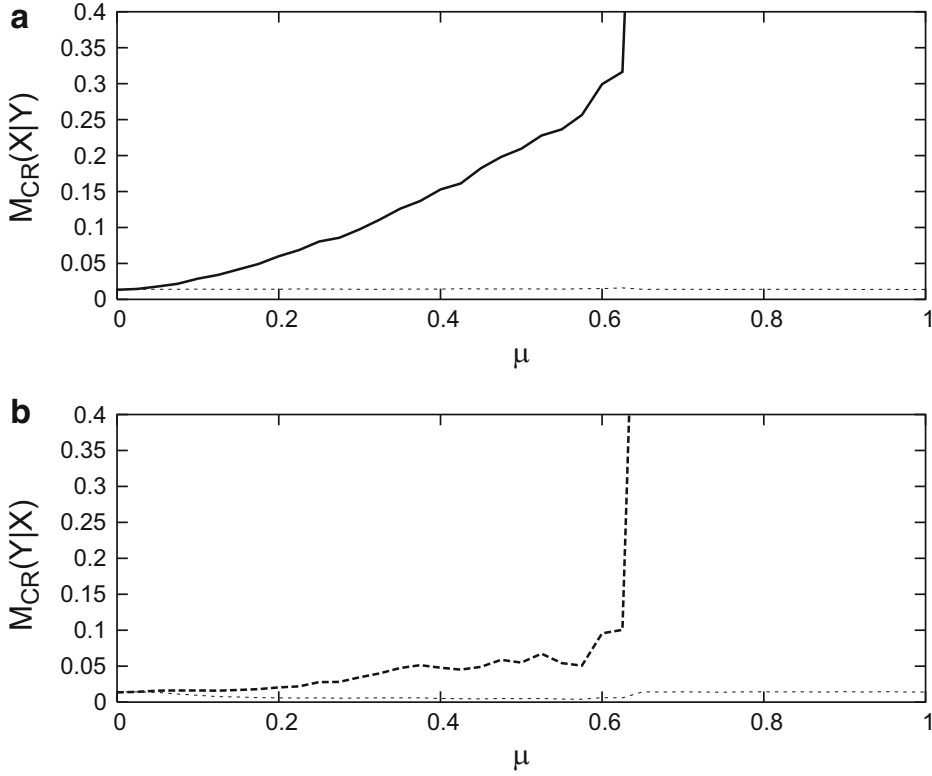


Fig. 3.8 (a) Mean conditional probability of recurrence $M_{CR}(X|Y)$ (solid) and 1 %-significance level (dotted) for two identical Hénon maps unidirectionally coupled. (b) $M_{CR}(Y|X)$ (dashed) and 1 %-significance level (dotted) for the same system. One hundred twin surrogates have been generated to estimate the significance level. The data have been normalised beforehand to have zero mean and standard deviation equal to one and $\varepsilon_X = \varepsilon_Y = 0.1$

and (3.9)]. We therefore generate 100 twin surrogates using the algorithm presented in [50]. We assume that we have scalar time series, and hence, use delay embedding to reconstruct the trajectory. We use embedding dimension $m = 3$ and delay $\tau = 1$, as in Sect. 3.1.3. The threshold for the generation of the surrogates is chosen to be $\delta = 0.09$ (see [50] for further details), according to the procedure given in [35]. Summarising, the following steps have to be undertaken for each value of the coupling strength μ :

- Choose the significance level α (i.e., $1 - \alpha$ quantile of the distribution).
- Compute $M_{CR}(x|y)$ and $M_{CR}(y|x)$.
- Generate L twin surrogate time series $\{x_i^{sj}\}_{i=1}^N$ and $\{y_i^{sj}\}_{i=1}^N$, with $j = 1, \dots, L$.
- Compute $M_{CR}(x|y^{sj})$ and $M_{CR}(y^{sj}|x)$ for $j = 1, \dots, L$.
- Compute the α -significance value based on the distribution obtained in the former step.
- If $M_{CR}(x|y)$ and $M_{CR}(y|x)$ are larger than the corresponding α -significance values, reject the null hypothesis.

The results for [Eqs. (3.8) and (3.9)] are shown in Fig. 3.8.

The values of $M_{CR}(X|Y)$ and $M_{CR}(Y|X)$ are above the significance level in both cases for all values of the coupling strength. Hence, the null hypothesis is correctly rejected. Therefore, in the case that we have a passive experiment, this procedure can be applied to assess the significance of the results about the asymmetry of the coupling.

Note that in the case of passive experiments, we cannot apply the criterion proposed in Sect. 3.1.2 to choose the thresholds ε_X and ε_Y , such that for coupling strength equal to zero we have $\langle p(\mathbf{x}_i) \rangle = \langle p(\mathbf{y}_i) \rangle$, because we do not know the value of the coupling strength. Therefore, one has to apply another criterion to choose the thresholds ε_X and ε_Y . In the example shown in Fig. 3.8 we have normalised the data beforehand to have zero mean and standard deviation equal to one, and then we have chosen $\varepsilon_X = \varepsilon_Y = 0.1$. If both interacting systems are structurally similar, then $M_{CR}(X|Y)$ will be approximately equal to $M_{CR}(Y|X)$ for coupling strength equal to zero. However, if the interacting systems are structurally different, this approach might not hold anymore.

3.1.7 Comparison with Other Methods

As mentioned in the introduction, several methods have been proposed in the literature to estimate the direction of the coupling. Most of these methods can be divided in the following categories:

1. In order to apply the method introduced in [38], one has to estimate first the phases of the interacting systems and then fit a functional relationship between them. From this function, the directionality variables are then derived. The main disadvantage of this method is that it is not always possible to assign a phase to a system based on a scalar time series, especially if the power spectrum of the signal does not present a predominant peak (i.e., one cannot speak of a main frequency of rotation of the system).
2. The state-space methods are based on the relationship between neighbours in the respective phase spaces of the interacting systems X and Y . At a first glance, these methods might seem to be very close to the recurrence based method introduced in this chapter. However, there are some important differences between them. For example, the computed indices in [4, 32] are based on the mean distances between a certain number q of nearest neighbours, i.e., they use the matrix of distances $|\mathbf{x}_i - \mathbf{x}_j|$ between all points of the trajectory. In contrast, the MCR indices do not use the distance matrix explicitly but rather the matrix of inequalities $|\mathbf{x}_i - \mathbf{x}_j| < \varepsilon$. Another way to express this difference is the following: in the state-space based methods, the threshold used to compute the neighbours is different for each point of the trajectory \mathbf{x}_i , i.e., $\varepsilon = \varepsilon(i)$, whereas to compute MCR the threshold is the same for all points of the trajectory. Another important difference is that in the MCR method, once the threshold is fixed, it remains the same for all different values of the coupling strength. In contrast, in the

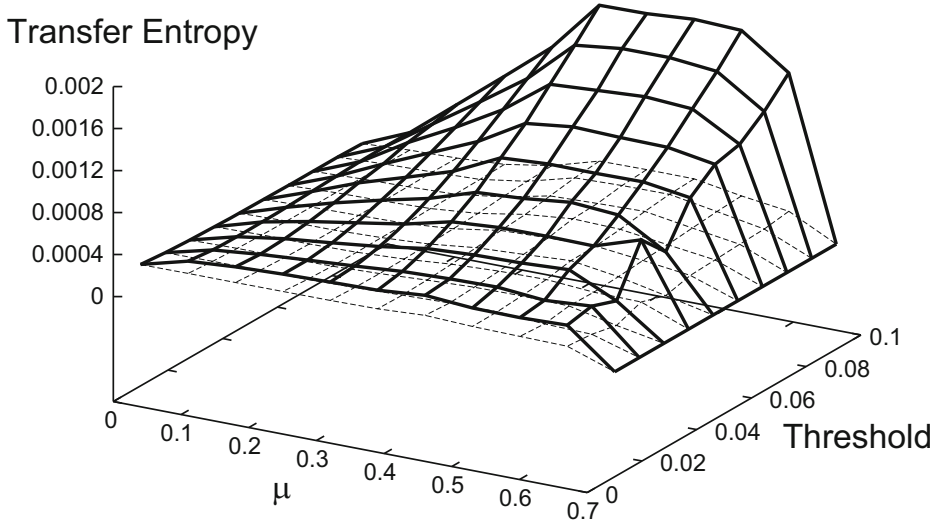


Fig. 3.9 Transfer entropy for two identical unidirectionally coupled Hénon maps in dependence on the coupling strength μ and the threshold. For each value of the coupling strength μ , the mean value over ten trajectories for uniformly distributed initial conditions has been computed. *Solid surface*: transfer entropy from X to Y , *dashed surface*: transfer entropy from Y to X

state-space based methods, the threshold does not only depend on the point of the trajectory, but also on the coupling strength.

3. Actually, the MCR method is closer to the information theory based methods, e.g., the transfer entropy [41]. In both cases, conditional probabilities of recurrence are estimated. But in the case of the transfer entropy, transition probabilities are considered, rather than static ones. This has the advantage of incorporating dynamical structure. The disadvantage compared to the MCR indices is that the number of data points needed for the estimation is considerable, and this might hamper the application of this method to experimental time series. For example, using the same number of data points (10,000) and the same range of values of the threshold as with the MCR method (Fig. 3.6) for the analysis of the direction of the coupling of [Eqs. (3.8) and (3.9)], we obtain the results for the transfer entropy as given in Fig. 3.9. The transfer entropy from X to Y is represented by the solid surface and the transfer entropy from Y to X by the dashed one. Note that even though the coupling is purely unidirectional, the transfer entropy from Y to X is larger than zero (it becomes only zero for coupling strength $\mu \geq 0.7$, when synchronisation sets in). That means that the transfer entropy does not detect that the coupling is purely unidirectional. This problem might be overcome using longer data sets. In the case of the MCR method, a purely unidirectional coupling can be easily detected by computing the recurrence rate of the driver in dependence on the coupling strength, which will then be constant.
4. A comparison between MCR and the recently introduced methods by [12, 13, 18] will be a subject for future work.

3.1.8 *Conclusions: Part I*

In this chapter, we have proposed recurrence based indices for the detection of the asymmetry of the coupling between interacting systems. The quantification of the asymmetry of the coupling can be very helpful in identifying driver-response relationships, which is a relevant problem in many fields, especially when dealing with experimental time series. The proposed indices are based on the mean conditional probabilities of recurrence (MCR). We have exemplified their applicability by several numerical examples which are representative of strong and weak coupled systems. Furthermore, we have shown that the MCR indices can also cope with the more challenging case of structurally different systems. We have studied the dependence of the MCR indices on the parameters needed for their estimation and we have found out that the choice of the parameters is not crucial for the correct detection of the asymmetry of the coupling. Moreover, we have addressed the very relevant problem of the quantification of the direction of the coupling in passive experiments and proposed an algorithm to assess the statistical significance of the results. Furthermore, we have studied the influence of observational noise on our method and compared it with other existing techniques for the detection of the asymmetry of the coupling. The numerical examples we considered in this chapter were mainly low dimensional. This technique is promising as shown by an application to experimental time series in the cardio-respiratory system [23].

3.2 Part II: Inferring Indirect Coupling

3.2.1 *Introduction: Part II*

The first part of this chapter introduced a new method to detect and quantify the asymmetry of the coupling between two interacting systems based on their recurrence properties [33]. Originally, it has focused on bivariate situations. It is crucial to extend it to multivariate time series analysis as this occurs quite often in many real applications. Inferring the coupling configuration at a local scale can be of substantial help to explain the global functioning of the network, e.g., the finding of motifs can be crucial for the understanding of the whole system [3, 44]. Therefore, now we focus on the inference of the coupling configuration of small networks consisting of three nodes.

Let us start considering the following small network as in Fig. 3.10, showing six different coupling settings for three unidirectionally interacting nodes. A pairwise analysis (bivariate) is often insufficient in addressing the possible indirect coupling (e.g., the coupling between X and Z in Fig. 3.10d). Hence, one main objective here is to identify the indirect coupling by means of recurrences. More specifically, we will identify the difference among these six coupling cases by studying the recurrence properties. The advantage of the extension from two to three coupled

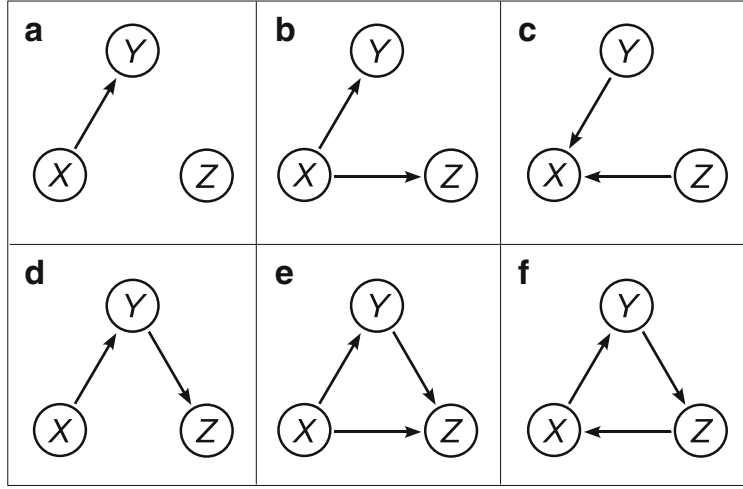


Fig. 3.10 Coupling configuration settings for three systems, that only uni-directional couplings are considered. (a) Z is independent of both X and Y . (b) X is the common driver for both Y and Z . (c) X is co-regulated by Y and Z . (d) X drives Y , while Y further drives Z . (e) Direct coupling with X being the common driver and Z being the common response. (f) Direct coupling in a ring way

systems is that it makes possible to analyze data measured from small networks, such as the EEG recordings on the scalp, so frequently used in neuroscience and cognitive psychology. In such experimental situations we have access to time series from typically of the order of ten nodes. Furthermore, it is crucial to identify the indirect coupling between X and Z as illustrated by Fig. 3.10d since it is one of big challenges for multivariate analysis. Therefore, we extend the study of [33] to three oscillators.

We show numerical studies for the application to three coupled Lorenz systems with six different coupling configurations (Fig. 3.10). Specifically, we consider the following system

$$X : \begin{cases} \dot{x}_1 = \sigma(x_2 - x_1), \\ \dot{x}_2 = rx_1 - x_2 - x_1x_3 + \mu_{21}y_2^2 + \mu_{31}z_3^2, \\ \dot{x}_3 = x_1x_2 - bx_3, \end{cases} \quad (3.15)$$

$$Y : \begin{cases} \dot{y}_1 = \sigma(y_2 - y_1), \\ \dot{y}_2 = ry_1 - y_2 - y_1y_3 + \mu_{12}x_2^2 + \mu_{32}z_3^2, \\ \dot{y}_3 = y_1y_2 - by_3, \end{cases} \quad (3.16)$$

$$Z : \begin{cases} \dot{z}_1 = \sigma(z_2 - z_1), \\ \dot{z}_2 = rz_1 - z_2 - z_1z_3 + \mu_{13}x_2^2 + \mu_{23}y_2^2, \\ \dot{z}_3 = z_1z_2 - bz_3. \end{cases} \quad (3.17)$$

Table 3.1 Variation of RR of each individual system

	(a)	(b)	(c)	(d)	(e)	(f)
RR_X	—	—	+	—	—	+
RR_Y	+	+	—	+	+	+
RR_Z	—	+	—	+	+	+

The symbols “+” correspond to the existence of variation, while “—” means there is no changes with increasing coupling strength

We integrate these equations numerically by a step 0.003 with sampling every 100 points leading to time step $\Delta t = 0.3$. We use standard parameters $\sigma = 10$, $r = 28$, $b = 8/3$ as in the uncoupled case $\mu_{i,j} = 0$ so that the oscillators are in a chaotic regime.

Our procedure to deal with the six coupling settings of Fig. 3.10 has three steps, which are explained in different sections: in Sect. 3.2.2, we apply a univariate analysis, namely analyzing each individual system separately; in Sect. 3.2.3, we perform a pairwise analysis, after which only the coupling configuration of Fig. 3.10d, e remains unclear. In Sect. 3.2.4, the partial mean conditional probability of recurrence is developed to cope with the last two remaining cases. Some conclusions are drawn in Sect. 3.2.7.

3.2.2 First Step: Univariate Analysis

Given the recurrence matrix Eq. (3.1), we are particularly interested in the mean probability of recurrence (recurrence rate), which is estimated by

$$\langle p(x_j) \rangle = RR = \frac{1}{N^2} \sum_{i,j=1}^N R_{i,j}. \quad (3.18)$$

The first step is to study the variations of the mean probability of recurrences with respect to an increase of the coupling strength (Fig. 3.10), separately. For instance, in the case of coupling as in Fig. 3.10a, both $\langle p(x) \rangle$ and $\langle p(z) \rangle$ remain unchanged while $\langle p(y) \rangle$ varies if the coupling strength μ is increased. This is because Z is independent of X and Y and the coupling from X to Y is unidirectional. A similar analysis can be achieved for each coupling setting of Fig. 3.10. We summarize the result in Table 3.1. We note that, throughout the chapter, the symbol “+” correspond to a change of the recurrence rate of the component, while “—” means there is no change with increasing the coupling strength. Furthermore “(a), . . . , (f)” denote the six different coupling configurations, as shown in Fig. 3.10.

The numerical simulation for the first step is shown in Fig. 3.11, which verifies the results presented in Table 3.1. By the first step of the univariate analysis, six different couplings are subdivided into three categories: $(+, -, -)$, $(+, +, -)$, and

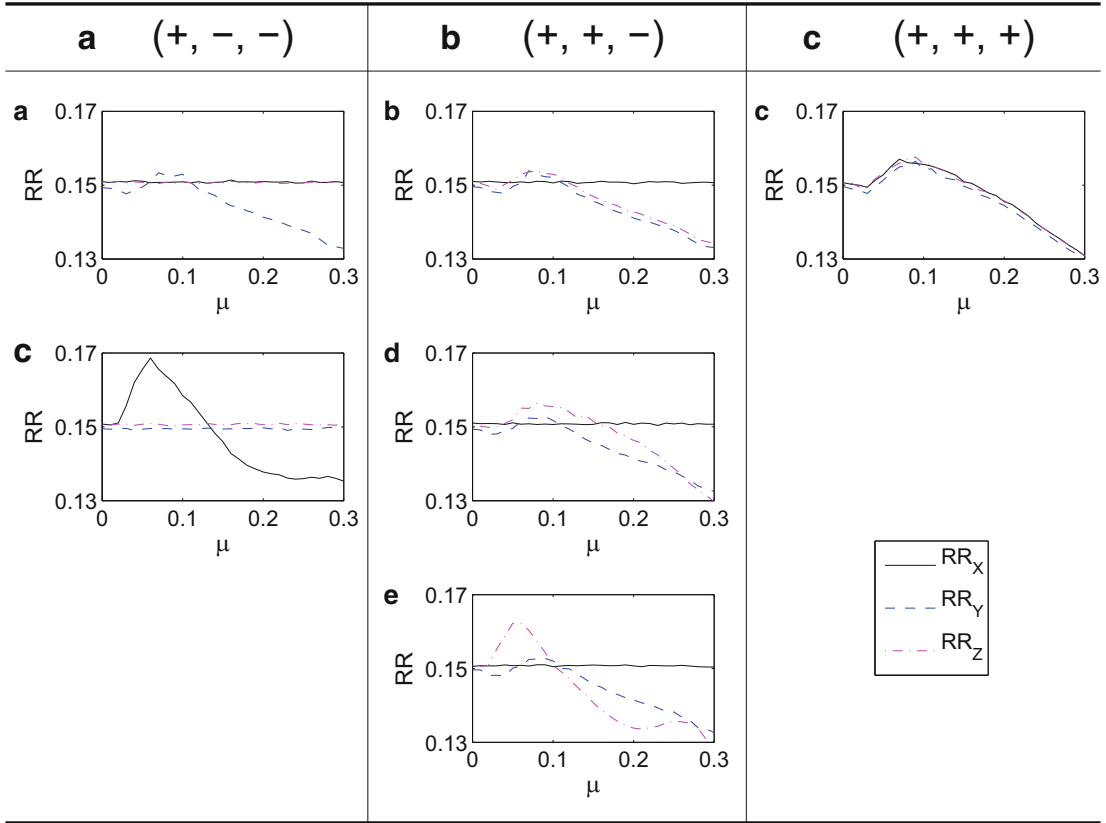


Fig. 3.11 Univariate analysis. Six coupling cases are classified into three groups A, B and C according to our univariate analysis

$(+, +, +)$. Note that $(+, -, -)$ denotes the configurations for which only the RR of one component changes with the coupling strength no matter which component. Therefore, the coupling configurations (a) and (c) are classified in the same group A.

Note that in principle one could also use other measures from Recurrence Quantification Analysis (RQA) [21] for this first step analysis (cf. Chap. 1). However it is not clear whether their use would allow identifying the coupling configuration. Moreover, the interpretation of the results might be difficult and rather empirical. Nevertheless, it is important to note that the probability of recurrence has a deep theoretical meaning as it is the basis for calculating many dynamical invariants, i.e., correlation dimension D_2 [17]. Furthermore, it has a clear relationship with information theoretic approaches (see [28]).

3.2.3 Second Step: Pairwise Analysis

In the second step, we perform a pairwise analysis, which is a generalisation of the RPs to joint recurrence plots (JRPs), namely calculating Eq. (3.2).

Based on the results obtained in the first step (Table 3.1 and Fig. 3.15), we do the pairwise analysis for cases A and B, following the notation of Fig. 3.11 since in

case C we have only one coupling configuration. Furthermore, we characterize the results of the pairwise analysis using the same notations as before. For instance, if $\Delta MCR(X|Y)$ changes with the coupling strength whereas $\Delta MCR(X|Z)$ and $\Delta MCR(Y|Z)$ stay constant, we write $(+, -, -)$. In particular, we have explicit expressions for Eq. (3.1.2) for each coupling:

A: couplings (a) and (c)

(a): $(+, -, -)$

$$\Delta MCR(Y|X) > 0, \quad (3.19)$$

$$\Delta MCR(Z|X) = RR_Z - RR_X \approx 0, \quad (3.20)$$

$$\Delta MCR(Z|Y) = RR_Y - RR_Z. \quad (3.21)$$

Note that RR_Y changes with the coupling strength, leading to $\Delta MCR(Z|Y) < 0$. However, $\Delta MCR(Z|Y)$ simply follows the same curve of univariate analysis since Y is independent of Z . In this respect, we denote it as “-”.

(c): $(+, +, -)$

$$\Delta MCR(Y|X) < 0, \quad (3.22)$$

$$\Delta MCR(Z|X) < 0, \quad (3.23)$$

$$\Delta MCR(Z|Y) = RR_Z - RR_Y \approx 0. \quad (3.24)$$

B: couplings (b), (d) and (e)

(b): $(+, +, -)$

$$\Delta MCR(Y|X) > 0, \quad (3.25)$$

$$\Delta MCR(Z|X) > 0, \quad (3.26)$$

$$\Delta MCR(Z|Y) \approx 0. \quad (3.27)$$

(d, e): $(+, +, +)$

$$\Delta MCR(Y|X) > 0, \quad (3.28)$$

$$\Delta MCR(Z|X) > 0, \quad (3.29)$$

$$\Delta MCR(Z|Y) > 0. \quad (3.30)$$

Therefore, after the pairwise analysis, only the cases (d) and (e) remain ambiguous.

The numerical simulations for the pairwise analysis of the second step are shown in Fig. 3.12 for (a, c), and Fig. 3.13 for (b, d, e), respectively.

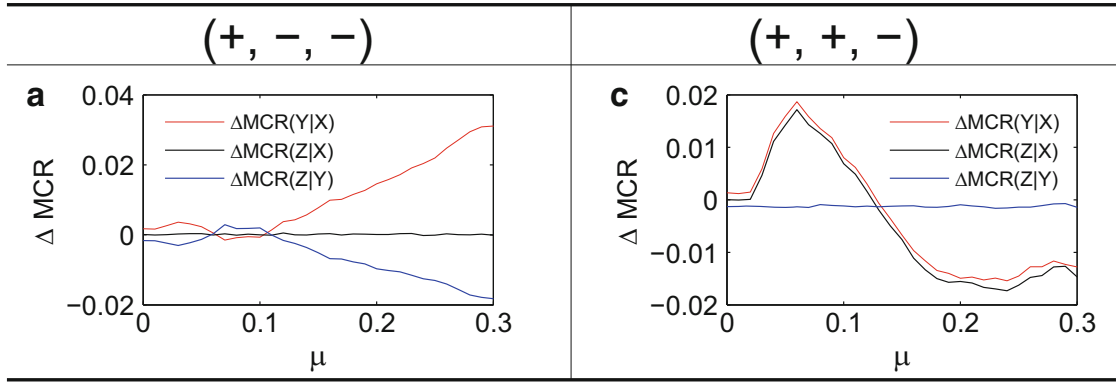


Fig. 3.12 Pairwise comparison for coupling (a, c). The *left panel* is for coupling configuration (a), the *right panel* is for coupling configuration (c). Note that in (a), $\Delta MCR(Z|Y)$ is completely overlapped with the difference between RR_Y and RR_Z , which can be obtained from the first step of univariate analysis (see Fig. 3.11)

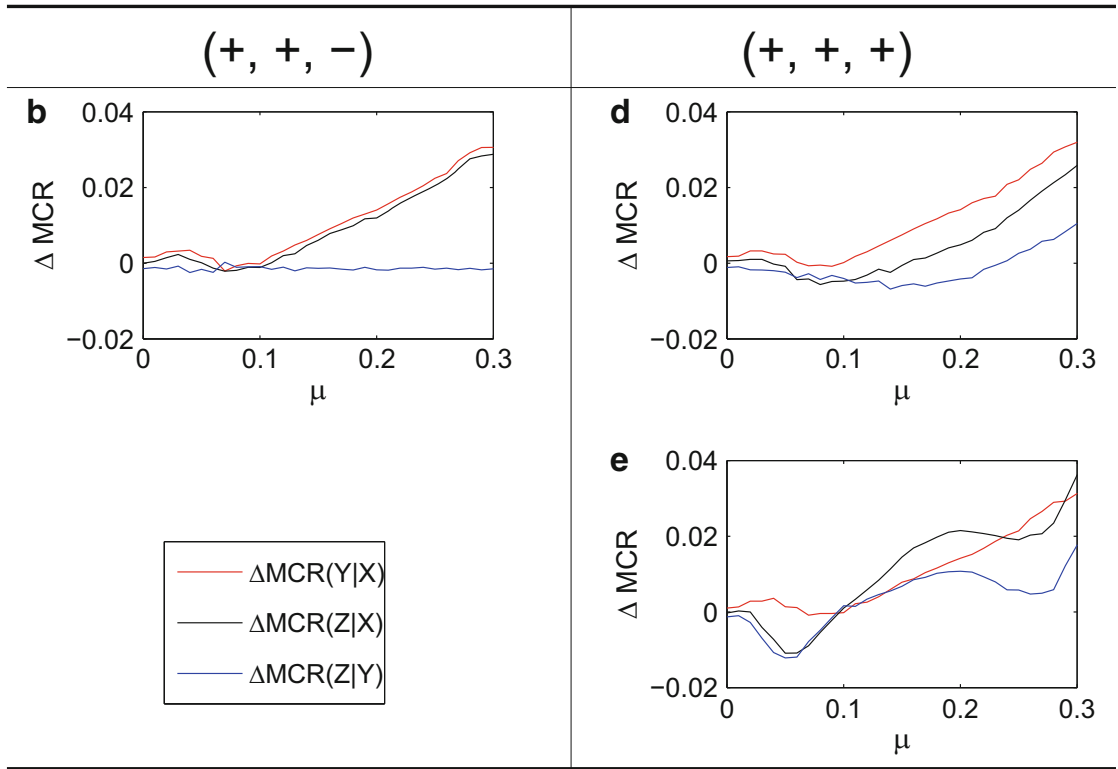


Fig. 3.13 Pairwise comparison for coupling configurations (b, d, e). The *left panel* is for (b), the *right panel* is for coupling configurations (d, e)

3.2.4 Third Step: Partial MCR

A recurrence-based bivariate analysis is generally not conclusive in addressing the existence of indirect coupling. Hence, we go to the crucial step in extending the method to deal with the remaining two coupling cases (d) and (e) after the previous two steps.

Let us first consider MCR of two variables used in our previous bivariate analysis, i.e., the pair of X and Y . We calculate $MCR(Y|X)$ and $MCR(X|Y)$, respectively. The second step is to quantify the influence of the third variable Z on $MCR(Y|X)$ by considering the difference between $MCR(Y|X)$ and the mean conditional probability of recurrence that Y recurs given that *both* X and Z recur,

$$\Delta MCR(Y|X)_Z = -(MCR(Y|X) - MCR(Y|X, Z)). \quad (3.31)$$

Similarly the contribution of Z to the $MCR(X|Y)$ is calculated by

$$\Delta MCR(X|Y)_Z = -(MCR(X|Y) - MCR(X|Y, Z)). \quad (3.32)$$

Before synchronisation sets in, we have $MCR(Y|X) < MCR(Y|X, Z)$ due to $\langle p(\mathbf{x}_j) \rangle > \langle p(\mathbf{x}_j, \mathbf{z}_j) \rangle$. Hence, for convenience the normalization factor (negative symbol) in front of the right hand side in Eq. (3.31) is introduced to keep the probability values to be positive. Note that $\Delta MCR(Y|X)_Z$ quantifies a subset of $MCR(Y|X)$, measuring the contribution of Z to the probabilities of recurrence of Y via X . In this regard, we call $\Delta MCR(Y|X)_Z$ *partial MCR*. Moreover, in general $\Delta MCR(Y|X)_Z$ is different from $\Delta MCR(X|Y)_Z$ because of the asymmetry between $MCR(Y|X)$ and $MCR(X|Y)$.

Analogously, the contribution of Y to $MCR(Z|X)$ and $MCR(X|Z)$ is calculated, respectively, by

$$\Delta MCR(Z|X)_Y = -(MCR(Z|X) - MCR(Z|X, Y)), \quad (3.33)$$

$$\Delta MCR(X|Z)_Y = -(MCR(X|Z) - MCR(X|Y, Z)). \quad (3.34)$$

The contribution of X to $MCR(Y|Z)$ and $MCR(Z|Y)$ is computed by

$$\Delta MCR(Y|Z)_X = -(MCR(Y|Z) - MCR(Y|X, Z)), \quad (3.35)$$

$$\Delta MCR(Z|Y)_X = -(MCR(Z|Y) - MCR(Z|X, Y)). \quad (3.36)$$

Depending on the particular coupling configuration (Fig. 3.10a–f), we obtain explicit expressions for measuring the contribution of one system to the other two. The details are presented in Sect. 3.2.6. In this section, however, we focus on the remaining two coupling configurations that have still to be distinguished, namely (d) and (e).

- (d) From the viewpoint of Z , the contribution of indirect coupling from X to Z is smaller than that of the direct coupling from Y to Z . Hence $MCR(Z|X) < MCR(Z|Y)$ and considering Eqs. (3.33) and (3.36) we have

$$\Delta MCR(Z|Y)_X < \Delta MCR(Z|X)_Y. \quad (3.37)$$

This is validated by showing a similar relationship from the X perspective, namely, $MCR(X|Z) < MCR(X|Y)$ implying

$$\Delta MCR(X|Y)_Z < \Delta MCR(X|Z)_Y. \quad (3.38)$$

However, from the mediator Y viewpoint, one has

$$\Delta MCR(Y|X)_Z \lesssim \Delta MCR(Y|Z)_X, \quad (3.39)$$

indicating Y to be a mediator.

- (e) In this case, from the perspective of system Z , the contribution of X to Z is larger than the contribution of Y to Z , which implies $MCR(Z|Y) < MCR(Z|X)$. This yields, considering again Eqs. (3.33) and (3.36),

$$\Delta MCR(Z|X)_Y < \Delta MCR(Z|Y)_X. \quad (3.40)$$

This relationship is validated from the X perspective, $MCR(X|Y) < MCR(X|Z)$, implying

$$\Delta MCR(X|Z)_Y < \Delta MCR(X|Y)_Z. \quad (3.41)$$

Moreover, from the viewpoint of Y , we have

$$\Delta MCR(Y|Z)_X < \Delta MCR(Y|X)_Z. \quad (3.42)$$

The numerical simulations for the partial MCR analysis for the coupling configurations (d, e) involved in the third step are shown in Fig. 3.14, which agree with our theoretical expectation in a reasonable range for coupling strengths.

3.2.5 Decision Tree

Following the discussions presented in previous sections, we conclude that the six different coupling configurations are clearly identified with our three-step procedure (plotted as a decision tree in Fig. 3.15):

1. Univariate analysis, namely the probability of recurrences for each individual systems, identifies the coupling configuration of (f). Note that couplings (a, c) are not able to be further classified because of the relabeling.
2. Pairwise analysis, allows distinguishing the coupling configuration of (a) from (c), and (b) from (d, e).
3. Partial MCR analysis identifies the difference between coupling (d) and (e).

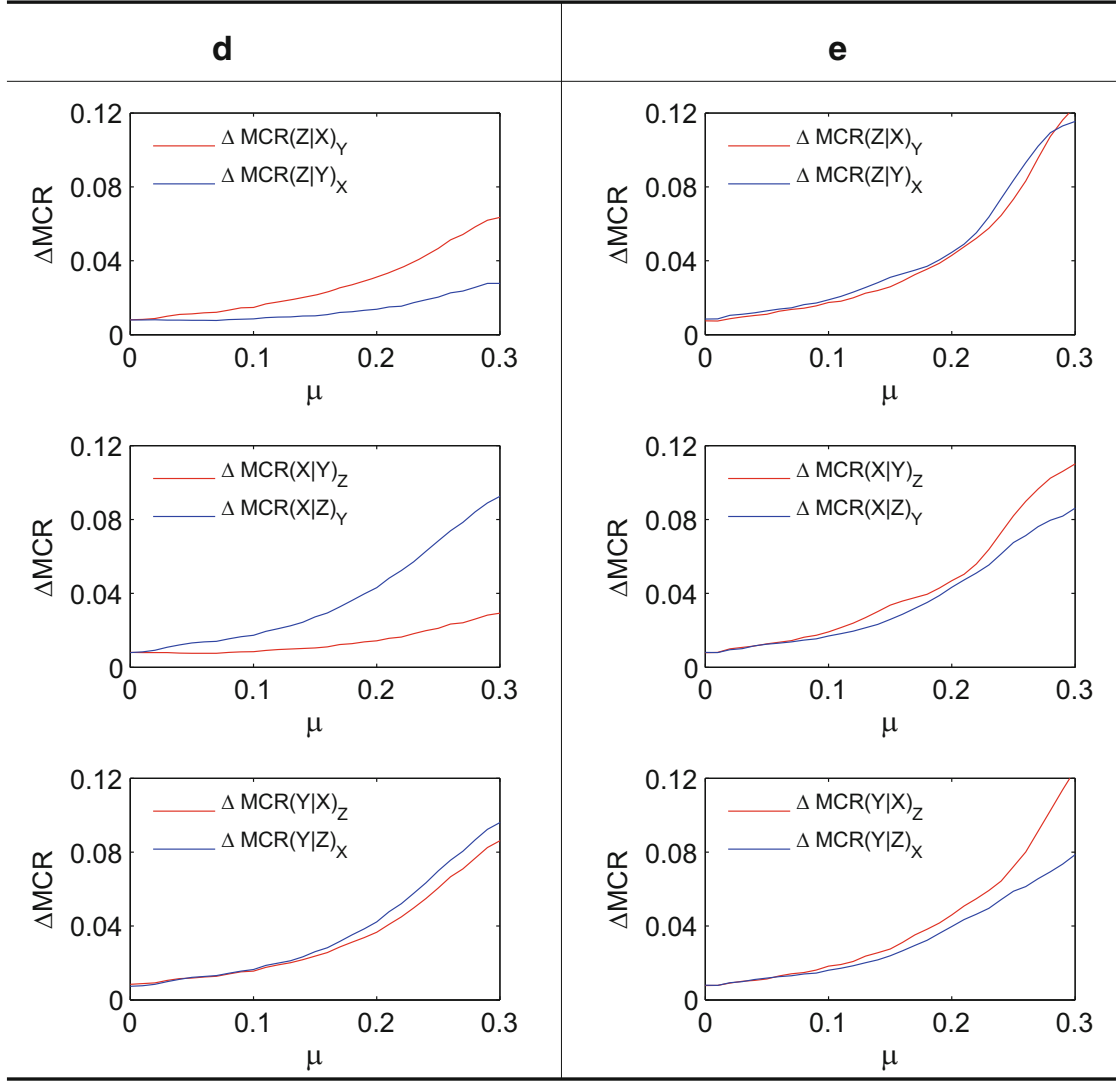


Fig. 3.14 Partial *MCR* shows distinct behavior for the coupling configurations (d) and (e)

3.2.6 Partial *MCR* for All Couplings of Fig. 3.10

The detailed analysis for the partial *MCR* for the coupling settings in Fig. 3.10 is performed in this section. We assume that the three systems are identical in the case of zero coupling and that the coupling strengths can be changed systematically, which is often the case for many active experiments. Note that the numerical results are based on Eqs. (3.15)–(3.17)).

(a) *Z* independent of both *X* and *Y* (Fig. 3.10a)

We obtain $MCR(Y|(X, Z)) = MCR(Y|X)$. This means that the contribution of *Z* to $MCR(Y|X)$ can be disregarded. Similarly, we have $MCR(X|(Y, Z)) = MCR(X|Y)$. Furthermore, $MCR(Z|X) = RR_Z$, $MCR(Z|X, Y) = RR_Z$, and $MCR(Y|Z) = RR_Y$ since *Z* is independent of both *X* and *Y*. Explicitly we have the following equations:

	Univariate: $\begin{pmatrix} RR_X \\ RR_Y \\ RR_Z \end{pmatrix}$		Bivariate: $\begin{pmatrix} \Delta MCR(X Y) \\ \Delta MCR(X Z) \\ \Delta MCR(Y Z) \end{pmatrix}$		Partial MCR	
$\left\{ \begin{matrix} (a) \\ (b) \end{matrix} \right\}$	$\begin{pmatrix} + \\ - \\ - \end{pmatrix}$	$\left\{ \begin{matrix} (a) \\ (c) \end{matrix} \right\}$	$\begin{pmatrix} + \\ - \\ - \end{pmatrix}$	(a)		(a)
			$\begin{pmatrix} + \\ + \\ - \end{pmatrix}$	(c)		(c)
$\left\{ \begin{matrix} (c) \\ (d) \\ (e) \end{matrix} \right\}$	$\begin{pmatrix} + \\ + \\ - \end{pmatrix}$	$\left\{ \begin{matrix} (b) \\ (d) \\ (e) \end{matrix} \right\}$	$\begin{pmatrix} + \\ + \\ - \end{pmatrix}$	(b)		(b)
			$\begin{pmatrix} + \\ + \\ + \end{pmatrix}$	$\left\{ \begin{matrix} (d) \\ (e) \end{matrix} \right\}$	$\frac{MCR(Z Y)}{MCR(Z X)} > 1$	(d)
					$\frac{MCR(Z Y)}{MCR(Z X)} < 1$	(e)
$\left\{ \begin{matrix} (f) \end{matrix} \right\}$	$\begin{pmatrix} + \\ + \\ + \end{pmatrix}$	(f)				(f)

Fig. 3.15 Decision tree derived from the theoretical analysis. Our procedure consists of three steps, namely univariate, pairwise and partial MCR analysis

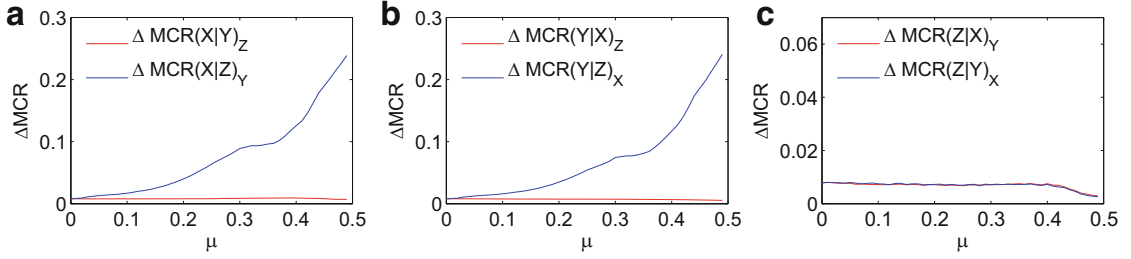


Fig. 3.16 Z independent of both X and Y (Fig. 3.10a, $\mu_{12} = \mu, \mu_{ij} = 0$ otherwise). (a) From the viewpoint of X , $\Delta MCR(X|Y)_Z = 0$, Eq. (3.43). (b) $\Delta MCR(Y|X)_Z = 0$, Eq. (3.45). However, due to the finite length of the time series used, these curves show some deviations from zero ($\mu = 0$). (c) $\Delta MCR(Z|Y)_X \approx 0, \Delta MCR(Z|X)_Y \approx 0$

$$\Delta MCR(X|Y)_Z = 0, \quad (3.43)$$

$$\Delta MCR(X|Z)_Y = -(RR_X - MCR(X|Y)), \quad (3.44)$$

$$\Delta MCR(Y|X)_Z = 0, \quad (3.45)$$

$$\Delta MCR(Y|Z)_X = -(RR_Y - MCR(Y|X)), \quad (3.46)$$

$$\Delta MCR(Z|X)_Y = 0, \quad (3.47)$$

$$\Delta MCR(Z|Y)_X = 0. \quad (3.48)$$

In general [Eqs. (3.44) and (3.46)] are not equal because of the asymmetry between X and Y . Figure 3.16 shows the numerical result.

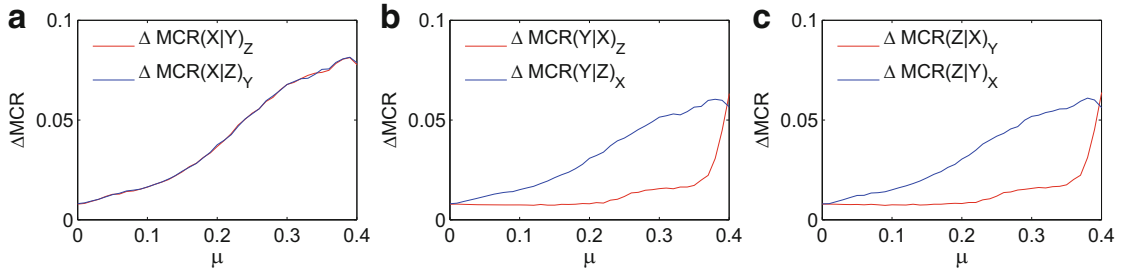


Fig. 3.17 (a) X is the common driver for both Y and Z (Fig. 3.10b, $\mu_{12} = \mu_{13} = \mu$, $\mu_{ij} = 0$ otherwise). (b) From X viewpoint, $\Delta MCR(Y|X)_Z < \Delta MCR(Y|Z)_X$, Eq. (3.52). (c) The same holds for Eq. (3.53), verified by $\Delta MCR(Z|X)_Y < \Delta MCR(Z|Y)_X$. Note that when $\mu > 0.39$, Y is synchronized with X suggested by the Lyapunov exponent spectrum (not shown here).

(b) X is a common driver (Fig. 3.10b)

From the viewpoint of X , $MCR(X|Y) = MCR(X|Z)$, since Y and Z are identical. Hence we have

$$\Delta MCR(X|Y)_Z = \Delta MCR(X|Z)_Y. \quad (3.49)$$

It is also obvious to see that Y is independent of Z under the condition of X , yielding $MCR(Y|Z) = RR_Y$, $MCR(Z|Y) = RR_Z$, thus,

$$\Delta MCR(Y|Z)_X = -(RR_Y - MCR(Y|X, Z)), \quad (3.50)$$

$$\Delta MCR(Z|Y)_X = -(RR_Z - MCR(Z|X, Y)). \quad (3.51)$$

In general before reaching synchronisation we have the mean recurrence probability relationship $p(\mathbf{x}_j) > p(\mathbf{y}_j) \approx p(\mathbf{z}_j)$, $p(\mathbf{x}_j, \mathbf{y}_j) > p(\mathbf{y}_j, \mathbf{z}_j) = p(\mathbf{y}_j) \cdot p(\mathbf{z}_j)$. Hence, $MCR(Y|X) > MCR(Y|Z) = RR_Y$, which leads to

$$\Delta MCR(Y|X)_Z < \Delta MCR(Y|Z)_X. \quad (3.52)$$

The same holds from the viewpoint of Z , namely,

$$\Delta MCR(Z|X)_Y < \Delta MCR(Z|Y)_X. \quad (3.53)$$

Figure 3.17 shows the numerical results.

(c) X is a common response (Fig. 3.10c)

When X is the common receiver, Y and Z are independent of each other. Hence, we derive the same theoretical results as the case that X is the common driver. Particularly, we have

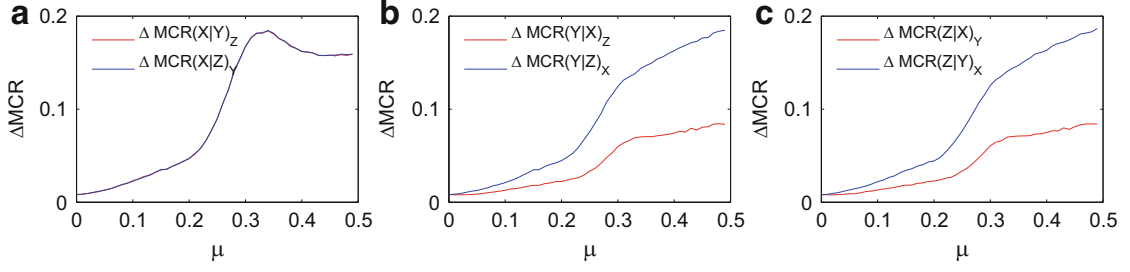


Fig. 3.18 X is the common receiver for both Y and Z (Fig. 3.10c, $\mu_{21} = \mu_{31} = \mu$, $\mu_{ij} = 0$ otherwise). Equations (3.54)–(3.56) are verified (see the caption of Fig. 3.17)

$$\Delta MCR(X|Y)_Z = \Delta MCR(X|Z)_Y, \quad (3.54)$$

$$\Delta MCR(Y|X)_Z < \Delta MCR(Y|Z)_X, \quad (3.55)$$

$$\Delta MCR(Z|X)_Y < \Delta MCR(Z|Y)_X. \quad (3.56)$$

Figure 3.18 shows the numerical results.

(d) Z indirectly coupled with X , but directly driven by Y (Fig. 3.10d)

Because of the asymmetry of $MCR(Y|X)$ and $MCR(X|Y)$, one expects the difference between $\Delta MCR(Y|X)_Z$ and $\Delta MCR(X|Y)_Z$. More specifically, the values of [Eqs. (3.31)–(3.36)] are not zero.

Note that within the coupling setting (Fig. 3.10d), we have $p(\mathbf{x}_j) > p(\mathbf{y}_j) > p(\mathbf{z}_j)$. Before synchronisation (neither two subsystems are synchronized), $p(\mathbf{x}_j) > p(\mathbf{y}_j, \mathbf{z}_j)$, $p(\mathbf{y}_j) > p(\mathbf{x}_j, \mathbf{z}_j)$, and $p(\mathbf{z}_j) > p(\mathbf{x}_j, \mathbf{y}_j)$. Furthermore, the complexity of (X, Z) is greater than the complexity of the joint space of (Y, Z) . This leads to the joint recurrence probability of $p(\mathbf{y}_j, \mathbf{z}_j)$ being greater than $p(\mathbf{x}_j, \mathbf{z}_j)$. Hence we have $MCR(Z|Y) > MCR(Z|X)$, which yields

$$\Delta MCR(Z|Y)_X < \Delta MCR(Z|X)_Y. \quad (3.57)$$

From the viewpoint of Z , the relationship of the inequality of equation (3.57) suggests that the contribution of indirect coupling from X to Z is smaller than that of the direct coupling from Y to Z . This is validated by showing a similar relationship from the X perspective, namely,

$$\Delta MCR(X|Y)_Z < \Delta MCR(X|Z)_Y. \quad (3.58)$$

However from the mediator Y viewpoint, it sends out the coupling information to Z when receiving the same amount of information from X . Hence, one has

$$\Delta MCR(Y|X)_Z \lesssim \Delta MCR(Y|Z)_X, \quad (3.59)$$

indicating Y to be the mediator. Figure 3.19 shows the numerical results.

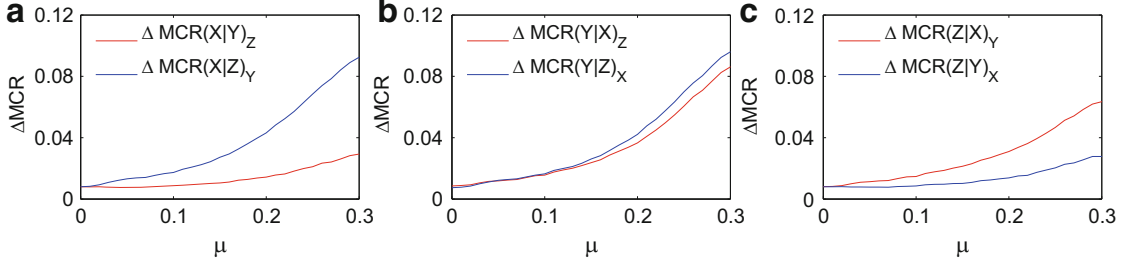


Fig. 3.19 X drives Y , and Y drives Z (Fig. 3.10d, $\mu_{12} = \mu_{23} = \mu$, $\mu_{ij} = 0$ otherwise). (a) $\Delta MCR(X|Y)_Z < \Delta MCR(X|Z)_Y$, (c) $\Delta MCR(Z|Y)_X < \Delta MCR(Z|X)_Y$. (b) However, Y is a mediator, having $\Delta MCR(Y|X)_Z \approx \Delta MCR(Y|Z)_X$

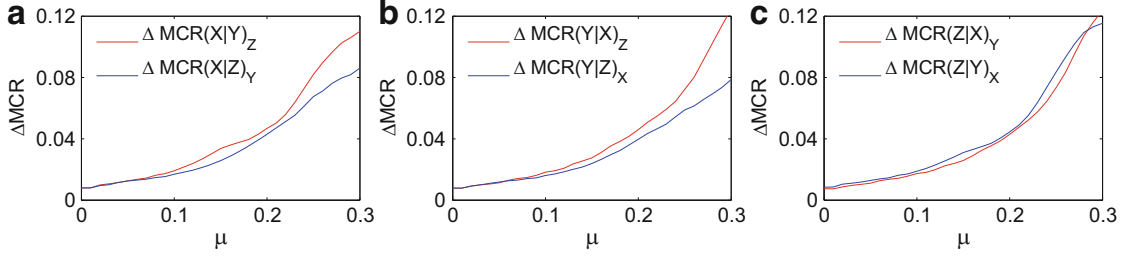


Fig. 3.20 X drives both Y and Z , and Y drives Z (Fig. 3.10e, $\mu_{12} = \mu_{13} = \mu_{23} = \mu$, $\mu_{ij} = 0$ otherwise). (a) $\Delta MCR(X|Z)_Y < \Delta MCR(X|Y)_Z$. (b) $\Delta MCR(Y|Z)_X < \Delta MCR(Y|X)_Z$ holds before synchronisation. (c) $\Delta(Z|X)_Y < \Delta(Z|Y)_X$

(e) Direct coupling setting: X is the common driver and Z is the common receiver (Fig. 3.10e)

From the perspective of system X , both Y and Z have the same contribution to X , namely, we have

$$\Delta MCR(X|Z)_Y < \Delta MCR(X|Y)_Z. \quad (3.60)$$

From the viewpoint of Z , it accepts two packages of information from the driver X . One is from the direct contribution sent by X to Z ; the other is received via the mediator Y . Z cannot distinguish where these two packages of information come from. Therefore, we have

$$\Delta MCR(Z|X)_Y < \Delta MCR(Z|Y)_X, \quad (3.61)$$

indicating that the transitivity property of Y is identified. Before onset of synchronisation, we again have $p(\mathbf{x}_j) > p(\mathbf{y}_j) > p(\mathbf{z}_j)$. Further, the relationship of $p(\mathbf{x}_j, \mathbf{y}_j) > p(\mathbf{y}_j, \mathbf{z}_j)$ holds. Hence, from the viewpoint of Y , we have

$$\Delta MCR(Y|Z)_X < \Delta MCR(Y|X)_Z. \quad (3.62)$$

Figure 3.20 shows the numerical results.

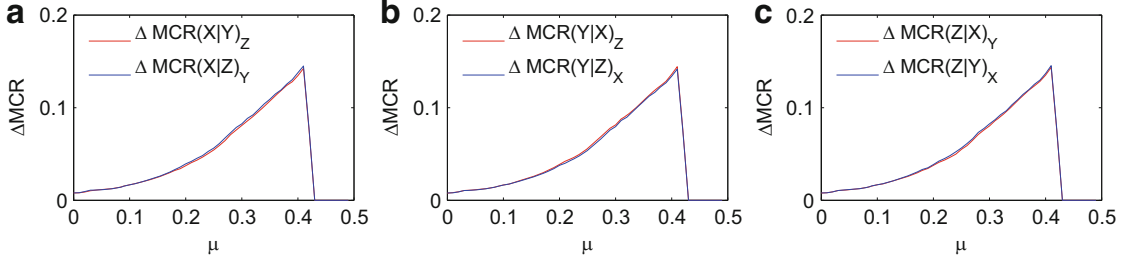


Fig. 3.21 X drives Y , Y drives Z , and Z drives X (Fig. 3.10f, $\mu_{12} = \mu_{23} = \mu_{31} = \mu$, $\mu_{ij} = 0$ otherwise) [Eqs. (3.63)–(3.65)]

(f) Direct coupling setting: in a ring way (Fig. 3.10f)

All measures are the same, since each system shows basically the same recurrence behavior, namely,

$$\Delta MCR(X|Y)_Z = \Delta MCR(X|Z)_Y, \quad (3.63)$$

$$\Delta MCR(Y|X)_Z = \Delta MCR(Y|Z)_X, \quad (3.64)$$

$$\Delta MCR(Z|X)_Y = \Delta MCR(Z|Y)_X. \quad (3.65)$$

Note that this only holds in the case when the three systems are identical, yielding the same transitivity ability. Figure 3.21 shows the numerical results.

3.2.7 Conclusions: Part II

We have shown a generalisation of MCR to three coupled systems, in particular including the extraction of indirect coupling. We have demonstrated our procedure using three Lorenz oscillators in chaotic regime in six different coupling configurations. Some issues might appear in the line of experimental studies, which have to be taken into consideration. Noise is ubiquitous in experimental time series and requires developing robust measures to identify the correct coupling configuration. It has been demonstrated that most recurrence structures are preserved if the free parameter ϵ (threshold in the recurrence matrix Eq. (3.1)) is chosen as $\epsilon \approx 5\sigma$, where σ corresponds to the standard deviation of the observational noise [48]. For the application to a large network of coupled units, the method has to be adapted to keep a good statistical power and to reduce the computation time. A further issue that will be considered in the future is the presence of asymmetric bidirectional coupling, as it is rather common in experimental complex networks.

Acknowledgements This work has been partially funded by the Federal Ministry for Education and Research (BMBF) via the Potsdam Research Cluster for Georisk Analysis, Environmental Change and Sustainability (PROGRESS), the National Natural Science Foundation of China (Grant No. 11305062, 11135001, 11075056), Specialized Research Fund for the Doctoral Program

(20130076120003), the Innovation Program of Shanghai Municipal Education Commission under grant No. 12ZZ043, SRF for ROCS, SEM, and the Hong Kong Polytechnic University Postdoctoral Fellowship. We acknowledge the discussion with N. Marwan, and M. Small.

References

1. V. Afraimovich, Pesin's dimension for Poincaré recurrences. *Chaos* **7**(1), 12–20 (1997)
2. R. Albert, A.L. Barabasi, Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**(1) (2002). doi:10.1103/RevModPhys.74.47
3. U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman & Hall/CRC Mathematical & Computational Biology (Chapman & Hall, Boca Raton, 2006)
4. J. Arnhold, K. Lehnertz, P. Grassberger, C.E. Elger, A robust method for detecting interdependencies: application to intracranially recorded EEG. *Physica D* **134**, 419 (1999)
5. A. Bahraminasab, F. Ghasemi, A. Stefanovska, P.V.E. McClintock, H. Kantz, Direction of coupling from phases of interacting oscillators: a permutation information approach. *Phys. Rev. Lett.* **100**(8), 084101 (2008)
6. S. Boccaletti, J. Kurths, G. Osipov, D.L. Valladares, C.S. Zhou, The synchronization of chaotic systems. *Phys. Rep.* **366**(1–2), 1–101 (2002)
7. M. Ding, Y. Chen, S.L., in *Bressler, Granger Causality, Basic Theory and Application to Neuroscience*, ed. by B. Schelter, M. Winterhalder, J. Timmer. Handbook of Time Series Analysis (Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 2007), pp. 437–460
8. J. Donges, Y. Zou, N. Marwan, J. Kurths, The backbone of the climate network. *Europhys. Lett.* **87**(4), 48,007 (2009)
9. J.F. Donges, Y. Zou, N. Marwan, J. Kurths, Complex networks in climate dynamics: comparing linear and nonlinear network construction methods. *Eur. Phys. J. ST* **174**, 157–179 (2009)
10. J.P. Eckmann, S.O. Kamphorst, D. Ruelle, Recurrence plots of dynamical systems. *Europhys. Lett.* **4**(9), 973–977 (1987)
11. P. Faure, H. Korn, A new method to estimate the kolmogorov entropy from recurrence plots: its application to neuronal signals. *Physica D* **122**(1–4), 265–279 (1998). doi:10.1016/S0167-2789(98)00177-8
12. J.H. Feldhoff, R.V. Donner, J.F. Donges, N. Marwan, J. Kurths, Geometric detection of coupling directions by means of inter-system recurrence networks. *Phys. Lett. A* **376**(46), 3504–3513 (2012). doi:http://dx.doi.org/10.1016/j.physleta.2012.10.008
13. J.H. Feldhoff, R.V. Donner, J.F. Donges, N. Marwan, J. Kurths, Geometric signature of complex synchronisation scenarios. *Europhys. Lett. (EPL)* **102**(3), 30,007 (2013)
14. S. Frenzel, B. Pompe, Partial mutual information for coupling analysis of multivariate time series. *Phys. Rev. Lett.* **99**(20), 204101 (2007)
15. R. Gilmore, M. Lefranc, in *The Topology of Chaos: Alice in Stretch and Squeezeland* (Wiley, New York, 2002)
16. C.W.J. Granger, Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**(3), 424–438 (1969)
17. P. Grassberger, I. Procaccia, Characterization of strange attractor. *Phys. Rev. Lett.* **50**(5), 346–349 (1983)
18. Y. Hirata, K. Aihara, Identifying hidden common causes from bivariate time series: a method using recurrence plots. *Phys. Rev. E* **81**, 016,203 (2010). doi:10.1103/PhysRevE.81.016203
19. H. Kantz, T. Schreiber, in *Nonlinear Time Series Analysis*, 2nd edn. (Cambridge University Press, Cambridge, 2004)
20. L. Kocarev, U. Parlitz, Generalized synchronization, predictability, and equivalence of unidirectionally coupled dynamical systems. *Phys. Rev. Lett.* **76**(11), 1816–1819 (1996)

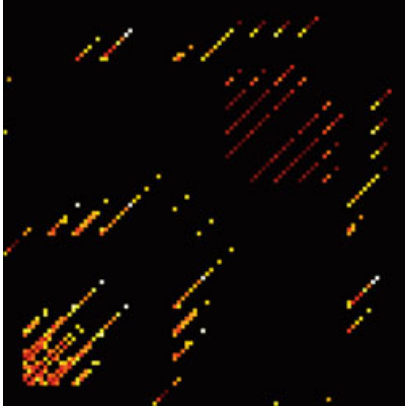
21. N. Marwan, M. Romano, M. Thiel, J. Kurths, Recurrence plots for the analysis of complex systems. *Phys. Rep.* **438**(5–6), 237–329 (2007)
22. N. Marwan, N. Wessel, U. Meyerfeldt, A. Schirdewan, J. Kurths, Recurrence plot based measures of complexity and its application to heart rate variability data. *Phys. Rev. E* **66**(2), 026,702 (2002)
23. N. Marwan, Y. Zou, N. Wessel, M. Riedl, J. Kurths, Estimating coupling directions in the cardiorespiratory system using recurrence properties. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **371**(1997) (2013). doi:10.1098/rsta.2011.0624
24. J. Nawrath, M.C. Romano, M. Thiel, I.Z. Kiss, M. Wickramasinghe, J. Timmer, J. Kurths, B. Schelter, Distinguishing direct from indirect interactions in oscillatory networks with multiple time scales. *Phys. Rev. Lett.* **104**(3), 038,701 (2010). doi:10.1103/PhysRevLett.104.038701
25. H. Osterhage, F. Mormann, T. Wagner, K. Lehnertz, Detecting directional coupling in the human epileptic brain: limitations and potential pitfalls. *Phys. Rev. E* **77**(1), 011914 (2008)
26. M. Palus, V. Komarek, Z. Hrnčir, K. Sterbova, Synchronization as adjustment of information rates: detection from bivariate time series. *Phys. Rev. E* **63**, 046,211 (2001)
27. M. Palus, A. Stefanovska, Direction of coupling from phases of interacting oscillators: an information-theoretic approach. *Phys. Rev. E* **67**, 055,201(R) (2003)
28. M. Paluš, M. Vejmelka, Directionality of coupling from bivariate time series: how to avoid false causalities and missed connections. *Phys. Rev. E* **75**(5), 056211 (2007)
29. A. Pikovsky, M. Rosenblum, J. Kurths, in *Synchronization: A Universal Concept in Nonlinear Sciences*. (Cambridge University Press, Cambridge, 2001)
30. H. Poincaré, Sur la probleme des trois corps et les équations de la dynamique. *Acta Math.* **13**, 1–271 (1890)
31. A. Porta, G. Baselli, N. Montano, T. Gneccchi-Ruscone, F. Lombardi, A. Malliani, S. Cerutti, Classification of coupling patterns among spontaneous rhythms and ventilation in the sympathetic discharge of decerebrate cats. *Biol. Cybern.* **75**(2), 163–172 (1996). doi:10.1007/s004220050284
32. R.Q. Quiroga, J. Arnhold, P. Grassberger, Learning driver-response relationships from synchronization patterns. *Phys. Rev. E* **61**(5), 5142–5148 (2000)
33. M.C. Romano, M. Thiel, J. Kurths, C. Grebogi, Estimation of the direction of the coupling by conditional probabilities of recurrence. *Phys. Rev. E* **76**(3), 036211 (2007)
34. M.C. Romano, M. Thiel, J. Kurths, I.Z. Kiss, J. Hudson, Detection of synchronization for non-phase-coherent and non-stationary data. *Europhys. Lett.* **71**(3), 466–472 (2005)
35. M.C. Romano, M. Thiel, J. Kurths, K. Mergenthaler, R. Engbert, Hypothesis test for synchronization: twin surrogates revisited. *Chaos* **19**(1), 015,108 (2009). <http://link.aip.org/link/?CHA/19/015108/1>
36. M.C. Romano, M. Thiel, J. Kurths, W. von Bloh, Multivariate recurrence plots. *Phys. Lett. A* **330**(3–4), 214–223 (2004)
37. M.G. Rosenblum, L. Cimponeriu, A. Bezerianos, A. Patzak, R. Mrowka, Identification of coupling direction: application to cardiorespiratory interaction. *Phys. Rev. E* **65**(4), 041909 (2002)
38. M.G. Rosenblum, A.S. Pikovsky, Detecting direction of coupling in interacting oscillators. *Phys. Rev. E* **64**(4), 045202 (2001)
39. N.F. Rulkov, M.M. Sushchik, L.S. Tsimring, H.D.I. Abarbanel, Generalized synchronization of chaos in directionally coupled chaotic systems. *Phys. Rev. E* **51**(2), 98–994 (1995)
40. B. Saussol, J. Wu, Recurrence spectrum in smooth dynamical systems. *Nonlinearity* **16**(6), 1991–2001 (2003)
41. T. Schreiber: Measuring information transfer. *Phys. Rev. Lett.* **85**(2), 461–464 (2000)
42. S.J. Schiff, P. So, T. Chang, R.E. Burke, T. Sauer, Detecting dynamical interdependence and generalized synchrony through mutual prediction in a neural ensemble. *Phys. Rev. E* **54**, 6708 (1996)
43. D.A. Smirnov, R.G. Andrzejak, Detection of weak directional coupling: Phase-dynamics approach versus state-space approach. *Phys. Rev. E* **71**(3), 036207 (2005)

44. O. Sporns, R. Kötter, Motifs in brain networks. *PLoS Biol.* **2**(11), e369 (2004). doi:10.1371/journal.pbio.0020369
45. C. Stam, B.F. Jones, G. Nolte, M. Breakspear, P. Scheltens, Small-world networks and functional connectivity in Alzheimer's disease. *Cerebral Cortex* **17**, 92 (2007)
46. C. Stam, B. van Dijk, Synchronization likelihood: an unbiased measure of generalized synchronization in multivariate data sets. *Physica D* **163**, 236 (2002)
47. M. Staniek, K. Lehnertz, Symbolic transfer entropy. *Phys. Rev. Lett.* **100**(15), 158101 (2008)
48. M. Thiel, M. Romano, J. Kurths, R. Meucci, E. Allaria, F. Arecchi, Influence of observational noise on the recurrence quantification analysis. *Physica D* **171**, 138–152 (2002)
49. M. Thiel, M.C. Romano, J. Kurths, How much information is contained in a recurrence plot? *Phys. Lett. A* **330**(5), 343–349 (2004)
50. M. Thiel, M.C. Romano, J. Kurths, M. Rolf, R. Kliegl, Twin surrogates to test for complex synchronisation. *Europhys. Lett.* **75**(4), 535 (2006)
51. M. Thiel, M.C. Romano, P.L. Read, J. Kurths, Estimation of dynamical invariants without embedding by recurrence plots. *Chaos* **14**(2), 234–243 (2004)
52. A.A. Tsonis, K.L. Swanson, Topology and predictability of El Niño and La Niña networks. *Phys. Rev. Lett.* **100**(22), 228502 (2008)
53. P. Van Leeuwen, D. Geue, M. Thiel, D. Cysarz, S. Lange, M.C. Romano, N. Wessel, J. Kurths, D.H. Grönemeyer, Influence of paced maternal breathing on fetal–maternal heart rate coordination. *Proc. Natl. Acad. Sci. USA* **106**(33), 13,661–13,666 (2009). doi:10.1073/pnas.0901049106
54. M. Vejmelka, M. Paluš, Inferring the directionality of coupling with conditional mutual information. *Phys. Rev. E* **77**(2), 026214 (2008)
55. K. Yamasaki, A. Gozolchiani, S. Havlin, Climate networks around the globe are significantly affected by El Niño. *Phys. Rev. Lett.* **100**(22), 228501 (2008)
56. Y. Zou, M.C. Romano, M. Thiel, N. Marwan, J. Kurths, Inferring indirect coupling by means of recurrences. *Int. J. Bifurcation Chaos Appl. Sci. Eng.* **21**(04), 1099–1111 (2011). doi:10.1142/S0218127411029033

Chapter 4

Complex Network Analysis of Recurrences

Reik V. Donner, Jonathan F. Donges, Yong Zou, and Jan H. Feldhoff



Abstract We present a complex network-based approach to characterizing the geometric properties of chaos by exploiting the pattern of recurrences in phase space. For this purpose, we utilize the basic definition of a recurrence as the mutual proximity of two state vectors in phase space (disregarding time information) and re-interpret the recurrence plot as a graphical representation of the adjacency matrix of a random geometric graph governed by the system's invariant density. The resulting recurrence networks contain exclusively geometric information about the

R.V. Donner (✉)

Potsdam Institute for Climate Impact Research, Telegrafenberg A31, 14473 Potsdam, Germany

Max Planck Institute for Biogeochemistry, Hans-Knöll-Straße 10, 07745 Jena, Germany

e-mail: reik.donner@pik-potsdam.de

J.F. Donges

Potsdam Institute for Climate Impact Research, Telegrafenberg A31, 14473 Potsdam, Germany

Stockholm Resilience Centre, Stockholm University, Kräftriket 2B, 11419 Stockholm, Sweden

e-mail: donges@pik-potsdam.de

Y. Zou

Department of Physics, East China Normal University, 200062 Shanghai, China

Potsdam Institute for Climate Impact Research, Telegrafenberg A31, 14473 Potsdam, Germany

e-mail: yzou@phy.ecnu.edu.cn

J.H. Feldhoff

Potsdam Institute for Climate Impact Research, Telegrafenberg A31, 14473 Potsdam, Germany

Department of Physics, Humboldt University, Newtonstraße 15, 12489 Berlin, Germany

e-mail: feldhoff@pik-potsdam.de

system under study, which can be exploited for inferring quantitative information on the geometric properties of the system's attractor without explicitly studying scaling characteristics as in the case of "classical" fractal dimension estimates.

Similar as the established recurrence quantification analysis, recurrence networks can be utilized for studying dynamical transitions in non-stationary systems, as well as for automatically discriminating between chaos and order without the necessity of extensive computations typically necessary when inferring this distinction based on the systems' maximum Lyapunov exponents. Moreover, we provide a thorough re-interpretation of two bi- and multivariate generalizations of recurrence plots in terms of complex networks, which allow tracing geometric signatures of asymmetric coupling and complex synchronization processes between two or more chaotic oscillators.

4.1 Introduction

With the recent increase in available computational capacities and rising data volumes in various fields of science, complex networks have become an interesting and versatile tool for describing structural interdependencies between mutually interacting units [1, 5, 9, 70]. Besides "classical" areas of research (such as sociology, transportation systems, computer sciences, or ecology), where these units are clearly (physically) identifiable, the success story of complex network theory has recently lead to a variety of "non-conventional" applications.

One important class of such non-traditional applications of complex network theory are *functional networks*, where the considered connectivity does not necessarily refer to "physical" vertices and edges, but reflects statistical interrelationships between the dynamics exhibited by different parts of the system under study. The term "functional" was originally coined in neuroscientific applications, where contemporaneous neuronal activity in different brain areas is often recorded using a set of standardized EEG channels. These data can be used for studying statistical interrelationships between different brain regions when performing certain tasks, having the idea in mind that the functional connectivity reflected by the strongest statistical dependencies can be taken as a proxy for the large-scale anatomic connectivity of different brain regions [107, 108]. Similar approaches have been later utilized for identifying dominant interaction patterns in other multivariate data sets, such as climate data [13, 14, 96].

Besides functional networks derived from multivariate time series, there have been numerous efforts for utilizing complex network approaches for quantifying structural properties of individual time series. In the last decade, several conceptually different approaches have been developed, see [27] for a recent review. One important class of approaches make use of ideas from symbolic dynamics and stochastic processes by discretizing the dynamics and then studying the transition probabilities between the obtained groups in a Markov chain-like approach [73]. A second class are visibility graphs and related concepts, which characterize

local convexity or record-breaking properties within univariate time series data [21, 51, 62]. The latter approach has important applications, such as providing new estimates of the Hurst exponent of fractal and multi-fractal stochastic processes [52, 72] or statistical tests for time series irreversibility [19, 53]. Finally, a third important class of time series networks make use of similarities or proximity relationships between different parts of a dynamical system's trajectory [27], including such diverse approaches as cycle networks [105], correlation networks [102], and phase space networks based on a certain definition of nearest neighbors [101]. One especially important example of proximity networks are recurrence networks (RNs) [24, 25, 66], which provide a reinterpretation of recurrence plots in network-theoretic terms and are already widely applied in a variety of fields.

In this chapter, we review the current state of knowledge on the theoretical foundations and potential applications of recurrence networks. We demonstrate that this type of time series networks naturally arises as random geometric graphs in the phase space of dynamical systems, which determines their structural characteristics and gives rise to a dimensionality interpretation of clustering coefficients and related concepts. In this spirit, the rich toolbox of complex network measures [5, 9, 70] provides various quantities that can be used for characterizing the system's dynamical complexity from an exclusively geometric viewpoint and allow discriminating between different types of dynamics. Beyond the single-system case, we also provide a corresponding in-depth discussion of cross- and joint recurrence plots from the complex network viewpoint. As a new aspect not previously reported in the literature, we provide a first-time theoretical treatment of a unification of single-system and cross-recurrence plots in a complex network context. Moreover, we discuss some new ideas related to the utilization of multivariate recurrence network-based approaches for studying geometric signatures of coupling and synchronization processes.

4.2 From Recurrence Plots to Recurrence Networks

In this section, we introduce and discuss RNs as an alternative framework for studying recurrences in phase space from a geometric point of view. We start with the basic setting suitable for single dynamical systems, followed by some detailed considerations on two different multivariate generalizations, taking advantage of corresponding recent extensions [82, 104] of the recurrence plot (RP) concept [30, 65]. Moreover, we provide a short overview on complex network characteristics and their meaning for RNs, highlighting the type of information that can be obtained using this approach—as opposed to other recurrence based techniques like recurrence quantification analysis (RQA) [95, 103], recurrence time statistics, or estimation of dynamical invariants from RPs.

In the remainder of this chapter, the properties of all approaches to be discussed will be illustrated for the specific example of a Rössler system [86]

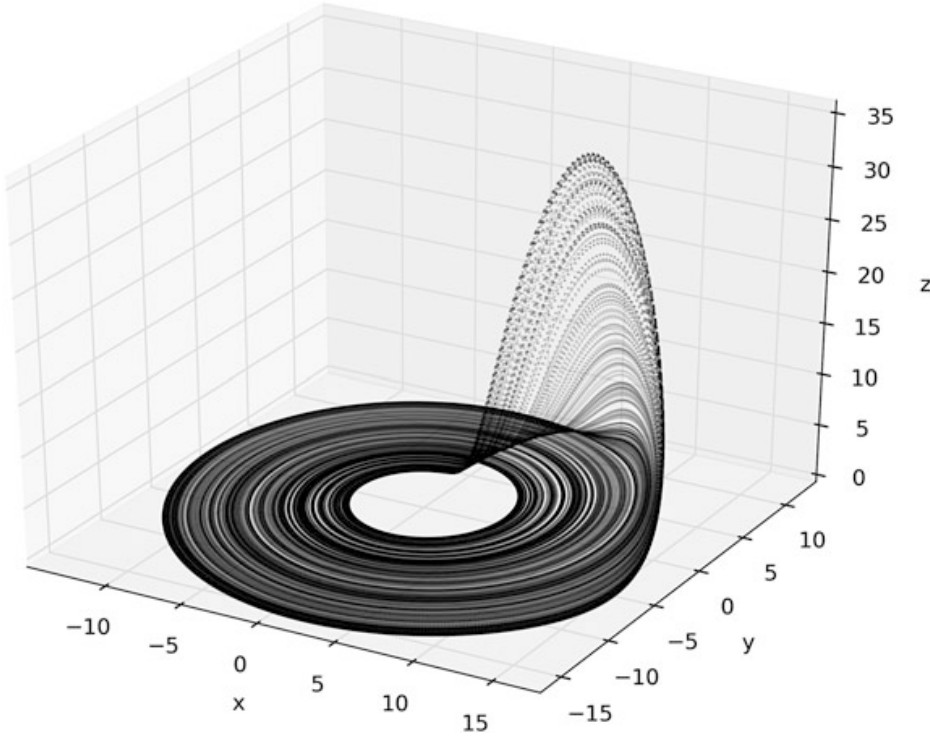


Fig. 4.1 Example trajectory on the chaotic attractor of the Rössler system Eq.(4.1) at the canonical parameter values

$$\begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{pmatrix} = \begin{pmatrix} -y - z \\ x + ay \\ b + z(x - c) \end{pmatrix}, \quad (4.1)$$

where a , b and c are control parameters of the system, which will be varied if necessary. In the “canonical” case, $a = 0.15$, $b = 0.2$ and $c = 10$. An example trajectory on the system’s chaotic attractor at these parameter values is shown in Fig. 4.1. A similar phase-coherent attractor¹ can be observed at $a = 0.16$, $b = 0.1$ and $c = 8.5$, whereas for $a = 0.2925$, $b = 0.1$ and $c = 8.5$, the system exhibits a structurally different chaotic attractor, the so-called funnel regime.

For the multivariate extensions, we consider the paradigmatic case of two slightly detuned Rössler systems (i.e., otherwise identical systems whose natural oscillation frequencies differ from each other) that are diffusively coupled via their second (y) component:

¹Here, phase-coherent describes chaotic oscillations around a well-defined center in a suitably chosen two-dimensional projection of the system [111, 112], whereas such a projection is not possible in the funnel regime.

$$\begin{aligned}
\dot{x}^{(1)} &= -(1 + \nu)y^{(1)} - z^{(1)} \\
\dot{y}^{(1)} &= (1 + \nu)x^{(1)} + ay^{(1)} + \mu_{21}(y^{(2)} - y^{(1)}) \\
\dot{z}^{(1)} &= b + z^{(1)}(x^{(1)} - c) \\
\dot{x}^{(2)} &= -(1 - \nu)y^{(2)} - z^{(2)} \\
\dot{y}^{(2)} &= (1 - \nu)x^{(2)} + ay^{(2)} + \mu_{12}(y^{(1)} - y^{(2)}) \\
\dot{z}^{(2)} &= b + z^{(2)}(x^{(2)} - c).
\end{aligned} \tag{4.2}$$

Here, ν measures the detuning, μ_{kl} is the coupling strength with which system k acts on system l ($k, l \in \{1, 2\}$), and the control parameters a, b, c are chosen as being identical for both systems. A natural extension is taking different parameter values for both coupled systems, e.g., studying the situation with one phase-coherent and one funnel system coupled to each other.

In what follows, trajectories are generated using numerical integration of Eq. (4.1) with a step size of $h = 0.001$. As far as univariate (single-system) analyses are concerned, we will utilize the canonical parameter values, whereas we will study the attractors of the phase-coherent and funnel regime at the above given parameters in the coupled systems setting.

4.2.1 Recurrence Networks from Single Dynamical Systems

4.2.1.1 Basic Idea

As in the classical RP analysis, we start with a (possibly multivariate) time series $\{x_i\}_{i=1}^N$ with $x_i = x(t_i)$, which we interpret as a finite representation of the trajectory of some (deterministic or stochastic) dynamical system. For a discrete system (map), the sampling of the time series is directly given by the map, whereas for a continuous-time system, the time series values correspond to a temporally discretized sampling of a finite part of one trajectory of the system determined by some initial conditions. In the case of observation functions not representing the full variety of dynamically relevant components, we additionally assume that attractor reconstruction has been performed (e.g., using time-delay embedding or some related technique) [37, 48, 49, 92] and that the x_i are state vectors in the corresponding (reconstructed) phase space of the dynamical system under study.

From the set of (original or reconstructed) state vectors representing a discrete sampling of the underlying system's trajectory (e.g., the chaotic attractor of a dissipative system), we compute the recurrence matrix \mathbf{R} in the standard way [30, 65] as

$$R_{ij}(\varepsilon) = \Theta(\varepsilon - \|x_i - x_j\|), \tag{4.3}$$

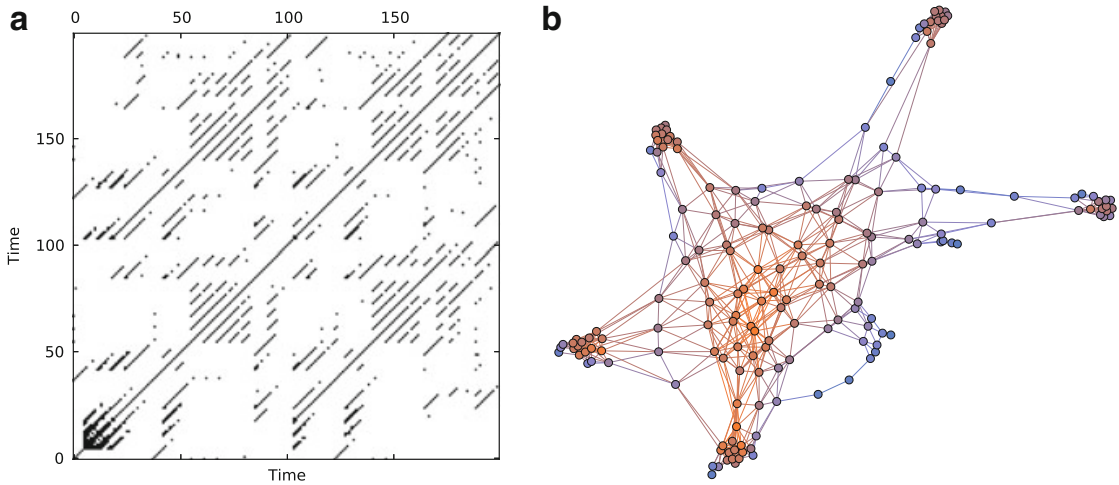


Fig. 4.2 (a) Recurrence plot and (b) recurrence network for some finite trajectory ($N = 200$, downsampled to a sampling interval of $\Delta t = 1.0$) of the Rössler system Eq. (4.1). Recurrences have been defined using maximum norm with a recurrence threshold ε resulting in a recurrence rate of $RR = 0.05$. The visual network representation has been obtained using a force-directed placement algorithm in the free software GEPHI. *Vertex colors* indicate the respective vertex degree in the network (*orange*: high degree, *blue*: low degree). Note that due to the peculiar three-dimensional structure of the Rössler attractor, the original attractor shape is not very well recovered by the network obtained from the short realization (this is distinctively different for other chaotic systems with a more two-dimensional structure, e.g., the Lorenz system [27])

where $\Theta(x)$ is the Heaviside function (we use here the convention $\Theta(x) = 1$ for $x \geq 0$, $\Theta(x) = 0$ otherwise), and $\|\cdot\|$ can be any norm in phase space (e.g., Manhattan, Euclidean, or maximum norm). For convenience, we will use the maximum norm in all following examples.

We can re-interpret the mathematical structure $\mathbf{R}(\varepsilon)$ as the adjacency matrix of some adjoint complex network embedded in phase space by setting

$$\mathbf{A}(\varepsilon) = \mathbf{R}(\varepsilon) - \mathbf{1}_N, \quad (4.4)$$

where $\mathbf{1}_N$ is the N -dimensional identity matrix. The complex network defined this way is called ε -recurrence network (RN), as opposed to other types of proximity-based networks in phase space making use of different definitions of geometric closeness, e.g., considering k -nearest neighbors [27]. Specifically, the sampled state vectors $\{x_i\}$ are interpreted as vertices of a complex network, which are connected by undirected edges if they are mutually close in phase space (i.e., describe recurrences). Notably, the binary matrix $\mathbf{A}(\varepsilon)$ retains the symmetry properties of $\mathbf{R}(\varepsilon)$, which implies that the RN is a *simple graph*, i.e., a complex network without multiple edges or self-loops (note that $A_{ii} = 0$ according to definition (4.4)). For an example of a RP and associated RN, see Fig. 4.2.

4.2.1.2 Complex Network Characteristics

Based on the re-interpretation of the recurrence matrix $\mathbf{R}(\varepsilon)$ as the adjacency matrix of an adjoint RN, we can utilize the large toolbox of complex network measures [1, 5, 9, 70] for characterizing the structural organization of a dynamical system in its phase space. Notably, this viewpoint is complementary to other concepts of nonlinear time series analysis making use of RPs. For example, RQA characterizes the statistical properties of line structures in the RP, which implies explicit consideration of dynamical aspects (i.e., sequences of state vectors). In turn, RNs do not make use of time information, since network properties are generally invariant under vertex relabelling (i.e., permutations of the order of observations) [25]. In this spirit, RN analysis provides geometric instead of dynamical characteristics. This idea of a geometric analysis is similar to some classical concepts of fractal dimensions (e.g., box-counting or correlation dimensions), where certain scaling laws in dependence on the spatial scale of resolution (corresponding here to ε) are exploited. In turn, RN analysis can be performed (as RQA) using only a single fixed scale (ε) instead of explicitly studying scaling properties over a range of threshold values.

The distinction between dynamical and geometric information implies that in case of RN analysis, the typical requirement of a reasonable (typically uniform) temporal sampling of the considered trajectory is replaced by the demand for a suitable spatial coverage of the system's attractor in phase space. Specifically, under certain conditions the latter could also be obtained by considering an ensemble of short trajectories instead of a single long one. If the trajectory under study is relatively densely sampled, trivial serial correlations can lead to a selection bias in the set of sampled state vectors; the latter could be avoided by reasonable downsampling. In the same context, the possibility of utilizing Theiler windows for removing edges representing short-term auto-correlations (e.g., recurrence points close to the main diagonal in the RP) should be mentioned as another strategy based on a somewhat different rationale [25]. However, from a conceptual perspective, downsampling can provide an unbiased sampling of the attractor as long as the fixed sampling time does not correspond to any integer multiple of some of the system's natural frequencies. As an alternative, bootstrapping from the set of available state vectors provides another feasible option, which should be preferred if a sufficiently long time series is available. In general, numerical experiments and different applications suggest that stable estimates of RN characteristics can often already be obtained using a sample size of $N \sim \mathcal{O}(10^2 \dots 10^3)$ data points [15, 16].

When considering quantitative characteristics of complex networks, different classifications of measures are possible. First, we may distinguish measures based on the concept of graph neighborhoods from those making use of shortest path-based characteristics. (This is not an exhaustive classification, since it potentially neglects other important network measures, e.g., such based on diffusion processes or random walks on the network.) Second, network measures can be classified into such making use of local, meso-scale and global information. This scheme is widely equivalent to the first one in that local information refers to

properties determined by the graph neighborhood of a given vertex, whereas global information takes contributions due to all vertices of the network into account, which is common for shortest path-based measures. Finally, we can differentiate between measures quantifying properties of single vertices, pairs of vertices, and the network as a whole. In this chapter, we will utilize the latter way of classification, since it appears most instructive from the applied point of view (i.e., we are commonly interested in either the local or the global geometry of an attractor).

In what follows, we will denote all properties computed from a RN consisting of a finite number N of state vectors as \hat{f} , pointing to the fact that they are estimated from a given sample of state vectors but shall characterize the entire trajectory of the system under study. We will discuss a corresponding continuous framework generalizing all network characteristics described below in Sect. 4.3. In order to focus the following discussion, we review only the possibly most relevant characteristics associated with RNs. More details including further measures can be found in [18, 25].

4.2.1.3 Shortest Paths in Recurrence Networks

In addition to the concepts of vertices (in the RN, the state vectors of a sampled dynamical system) and edges (recurrences between state vectors in phase space), the notion of shortest paths is a third important ingredient in complex network theory.

In general, a *path* between two specified vertices i and j is an ordered sequence of edges starting at i and ending at j , with its *path length* $\hat{l}_{ij}(\varepsilon)$ given by the number of edges in this sequence. In the context of RNs, we can thus understand a path as a sequence of mutually overlapping ε -balls² $B_\varepsilon(x_i), B_\varepsilon(x_{k_1}), \dots, B_\varepsilon(x_{k_{l_{i,j}-1}}), B_\varepsilon(x_j)$, where

$$B_\varepsilon(x) = \{y \mid \|x - y\| < \varepsilon\}$$

is an open set describing a volume with maximum distance ε (measured in a given norm) from x , and $B_\varepsilon(x_i) \cap B_\varepsilon(x_{k_1}) \neq \emptyset, \dots, B_\varepsilon(x_{k_{l_{i,j}-1}}) \cap B_\varepsilon(x_j) \neq \emptyset$.

Following these considerations, a *shortest path* is a minimum sequence of edges (mutually overlapping ε -balls) between two fixed vertices (state vectors) i and j . Note that a shortest path does not need to be unique. In turn, due to the discrete character of a network, it is rather typical that there are multiple shortest paths between some specific pair of vertices. In what follows, the shortest path length will be denoted as \hat{d}_{ij} , and the multiplicity of such shortest paths as \hat{o}_{ij} .

²Here, ε -balls refers to general (hyper-)volumes according to the specific norm chosen for measuring distances in phase space, e.g., hypercubes of edge length 2ε in case of the maximum norm, or hyperballs of radius ε for the Euclidean norm.

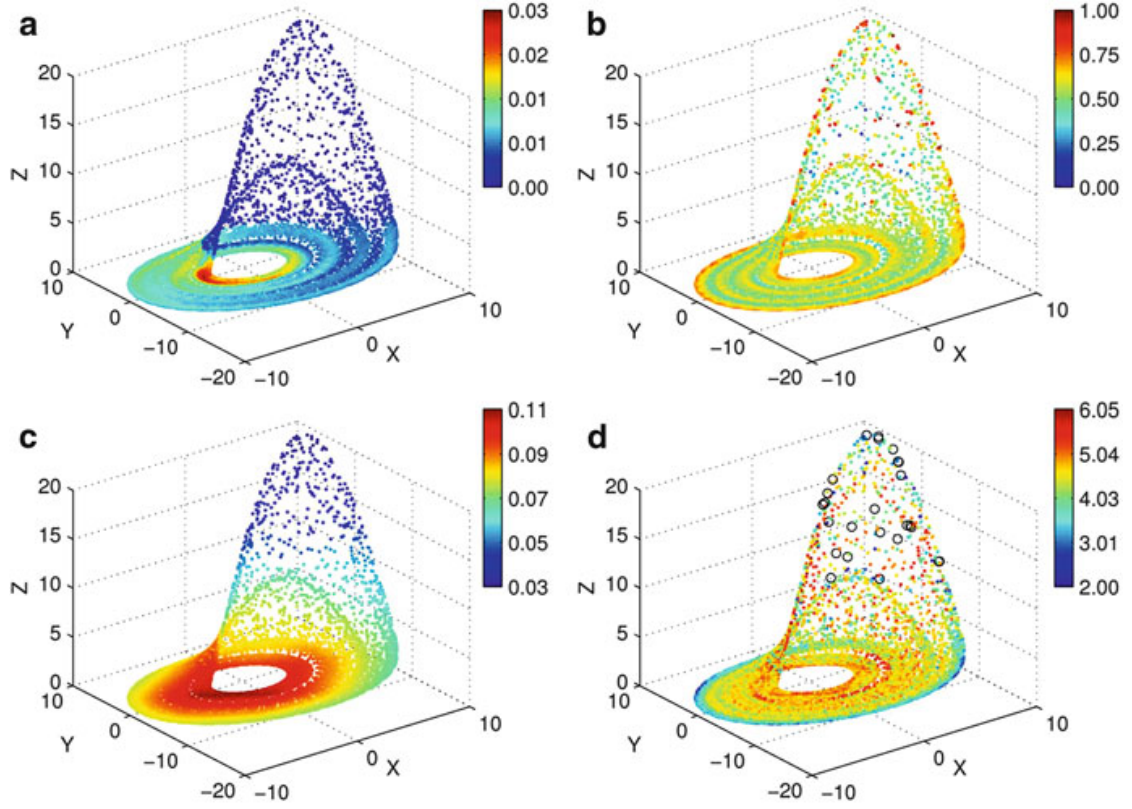


Fig. 4.3 Spatial distributions of vertex characteristics of the RN obtained for the Rössler system Eq.(4.1) at the canonical parameters (RPs computed using the maximum norm with $\hat{\rho} = 0.01$, $N = 10,000$ and sampling time $\Delta t = 0.2$): (a) degree centrality, (b) local clustering coefficient, (c) closeness centrality, (d) betweenness centrality. Open circles in (d) highlight vertices that are disconnected from the rest of the RN

4.2.1.4 Local (Vertex-Based) Measures

The conceptually simplest measure characterizing the connectivity properties of a single vertex in a complex network is the *degree* or *degree centrality*

$$\hat{k}_v(\varepsilon) = \sum_{i=1}^N A_{iv}(\varepsilon), \quad (4.5)$$

which simply counts the number of edges associated with a given vertex v . From the perspective of recurrences, it is reasonable to replace the degree by a normalized characteristic, the *degree density*

$$\hat{\rho}_v(\varepsilon) = \frac{\hat{k}_v(\varepsilon)}{N-1} = \frac{1}{N-1} \sum_{i=1}^N A_{iv}(\varepsilon), \quad (4.6)$$

which corresponds to the definition of the local recurrence rate of the state x_v . $\hat{\rho}_v(\varepsilon)$ quantifies the density of states in the ε -ball around x_v , i.e., the probability that a

randomly chosen member of the available sample of state vectors is ε -close to x_v . An illustration of this fact for the Rössler system is presented in Fig. 4.3a; here, phase space regions with a high density of points (i.e., a high residence probability of the sampled trajectory) are characterized by a high degree density.

In order to characterize the density of connections among the neighbors of a given vertex v , we can utilize the *local clustering coefficient*³

$$\hat{\mathcal{C}}_v(\varepsilon) = \frac{1}{\hat{k}_v(\varepsilon)(\hat{k}_v(\varepsilon) - 1)} \sum_{i,j=1}^N A_{vi}(\varepsilon) A_{ij}(\varepsilon) A_{jv}(\varepsilon), \quad (4.7)$$

which measures the fraction of pairs of vertices in the ε -ball around x_v that are mutually ε -close. For vertices with $\hat{k}_v(\varepsilon) < 2$, we define $\hat{\mathcal{C}}_v(\varepsilon) = 0$. It can be argued (see Sect. 4.4.2.2) that the local clustering coefficient in a RN is associated with the geometric alignment of state vectors. Specifically, close to dynamically invariant objects such as unstable periodic orbits (UPOs) of low period, the dynamics of the system is effectively lower-dimensional, which results in a locally enhanced fraction of closed paths of length 3 (“triangles”) and, thus, a higher local clustering coefficient. The latter behavior is exemplified in Fig. 4.3b for the Rössler system, where we recognize certain bands with higher values of $\hat{\mathcal{C}}_v$ corresponding to the positions of known low-periodic UPOs [25].

While degree and local clustering coefficient characterize network structures on the local and meso-scale, there are further vertex characteristics that make explicit use of the concept of shortest paths and, thus, provide measures relying on the connectivity of the entire network. Two specific properties of this kind are the *closeness* or *closeness centrality*

$$\hat{c}_v(\varepsilon) = \left(\frac{1}{N-1} \sum_{i=1}^N \hat{d}_{vi}(\varepsilon) \right)^{-1}, \quad (4.8)$$

which gives the inverse arithmetic mean of the shortest path lengths between vertex v and all other vertices, and the *local efficiency*

$$\hat{e}_v(\varepsilon) = \frac{1}{N-1} \sum_{i=1}^N \hat{d}_{vi}(\varepsilon)^{-1}, \quad (4.9)$$

which gives the inverse harmonic mean of these shortest path lengths. Notably, the latter quantity has the advantage of being well-behaved in the case of disconnected network components, where there are no paths between certain pairs of vertices (i.e., $\hat{d}_{ij} = \infty$). In order to circumvent divergences of the closeness due to the existence

³In Eq. (4.7) as well as several following definitions, we skip the condition $i \neq j \neq v$ for simplicity, since $A_{ii} = 0$ for all vertices i according to our RN definition Eq. (4.4).

of disconnected components, it is convenient to always set \hat{d}_{ij} to the highest possible value of $N - 1$ for pairs of vertices that cannot be mutually reached. Both $\hat{c}_v(\varepsilon)$ and $\hat{e}_v(\varepsilon)$ characterize the geometric centrality of vertex v in the network, i.e., closeness and local efficiency exhibit the highest values for such vertices which are situated in the center of the RN (see Fig. 4.3c for an illustration for the Rössler system).

Another frequently studied path-based vertex characteristic is the so-called *betweenness* or *betweenness centrality*, which measures the fraction of shortest paths in a network traversing a given vertex v . Let $\hat{\sigma}_{ij}$ denote the total number of shortest paths between two vertices i and j and $\hat{\sigma}_{ij}(v)$ the cardinality of the subset of these paths that include a given vertex v , betweenness centrality is defined as

$$\hat{b}_v(\varepsilon) = \sum_{i,j=1; i,j \neq v}^M \frac{\hat{\sigma}_{ij}(v)}{\hat{\sigma}_{ij}}. \quad (4.10)$$

Betweenness centrality is commonly used for characterizing the importance of vertices for information propagation in networks. In the RN context, it can be interpreted as indicating the local degree of fragmentation of the underlying attractor [25]. To see this, consider two densely populated regions of phase space that are separated by a poorly populated one. Vertices in the latter will “bundle” the shortest paths between vertices in the former ones, thus forming geometric bottlenecks in the RN. In this spirit, we may understand the spatial distribution of betweenness centrality for the Rössler system (Fig. 4.3d) which includes certain bands with higher and lower residence probability reflected in lower and higher betweenness values.

4.2.1.5 Pairwise Vertex and Edge Measures

In contrast to vertices, whose properties can be characterized by a multitude of graph characteristics, there are fewer measures that explicitly relate to the properties of edges or, more generally, pairs of vertices. One such measure is the *matching index*, which quantifies the overlap of the network neighborhoods of two vertices v and w :

$$\hat{m}_{vw}(\varepsilon) = \frac{\sum_{i=1}^N A_{vi}(\varepsilon) A_{wi}(\varepsilon)}{\hat{k}_v(\varepsilon) + \hat{k}_w(\varepsilon) - \sum_{i=1}^N A_{vi}(\varepsilon) A_{wi}(\varepsilon)}. \quad (4.11)$$

From the above definition, it follows that $\hat{m}_{vw}(\varepsilon) = 0$ if $\|x_v - x_w\| \geq 2\varepsilon$. In turn, there can be mutually unconnected vertices v and w ($A_{vw} = 0$) with $\varepsilon \leq \|x_v - x_w\| < 2\varepsilon$ that have some common neighbors and, thus, non-zero matching index. In the context of recurrences in phase space, $\hat{m}_{vw}(\varepsilon) = 1$ implies that the states x_v and x_w are twins, i.e., share the same neighborhood in phase space. In this spirit, we interpret $\hat{m}_{vw}(\varepsilon)$ as the degree of twinness of two state vectors. Note that twins have important applications in creating surrogates for testing for the presence of complex synchronization [85, 94].

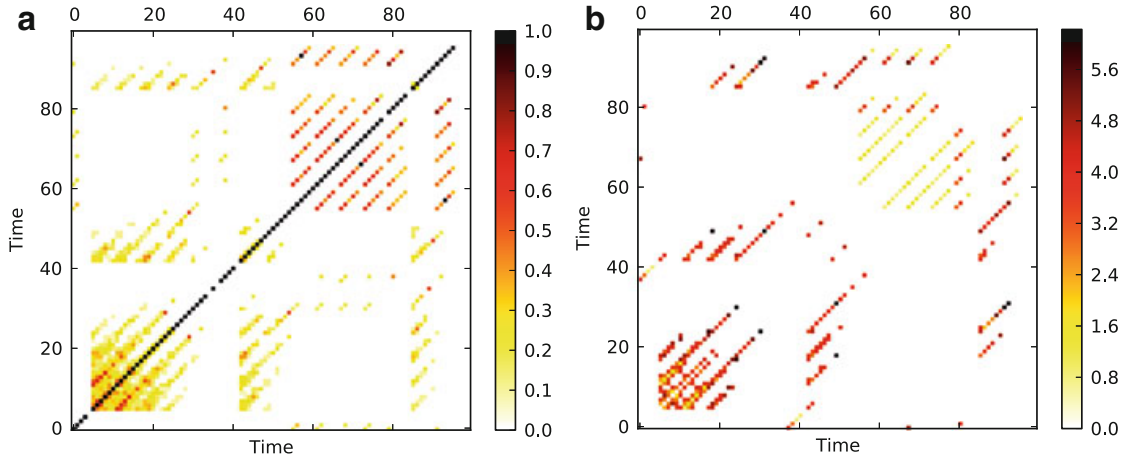


Fig. 4.4 Color-coded representations of (a) matching index \hat{m}_{ij} and (b) logarithmic edge betweenness $\log_{10} \hat{b}_{ij}$ for the RN of the Rössler system (using $N = 100$ vertices corresponding to a sampling with $\Delta t = 1.0$, RP obtained using the maximum norm with ε chosen such as to yield a recurrence rate of $RR = 0.05$)

While the concept of matching index does not require the presence of an edge between two vertices v and w , there are other characteristics that are explicitly edge-based. To this end, we only mention that the concept of betweenness centrality can also be transferred to edges, leading to the *edge betweenness* measuring the fraction of shortest paths on the graph traversing through a specific edge (v, w) :

$$\hat{b}_{vw}(\varepsilon) = \sum_{i,j=1; i,j \neq v,w}^M \frac{\hat{\sigma}_{ij}(v, w)}{\hat{\sigma}_{ij}}, \quad (4.12)$$

where $\hat{\sigma}_{ij}(v, w)$ gives the total number of shortest paths between two vertices i and j that include the edge (v, w) . If there is no edge between two vertices v and w , we set $\hat{b}_{vw} = 0$ for convenience. As the (vertex-based) betweenness centrality, in a RN edge betweenness characterizes the local fragmentation of the studied dynamical system in its phase space.

Figure 4.4 illustrates the two aforementioned concepts for one example trajectory of the Rössler system. As can be seen, there is no simple correspondence between matching index and edge betweenness, since both quantify distinctively different aspects of phase space geometry. Specifically, there are more pairs of vertices with non-zero matching index than edges, even though there are also pairs of vertices with $\hat{b}_{vw}(\varepsilon) > 0$ but $\hat{m}_{vw}(\varepsilon) = 0$ (i.e., there is an edge between v and w , but both have no common neighbors). However, for those pairs of vertices for which both characteristics are non-zero, we find a clear anti-correlation. One interpretation of this finding is that a large matching index typically corresponds to very close vertices in phase space; such pairs of vertices can in turn be easily exchanged as members of shortest paths on the network, which implies lower edge betweenness values. A similar argument may explain the coincidence of high edge betweenness and low non-zero matching index values.

4.2.1.6 Global Network Measures

Some, but not all useful global network characteristics can be derived by averaging certain local-scale (vertex) properties. Prominently, the *edge density*

$$\hat{\rho}(\varepsilon) = \frac{1}{N} \sum_{v=1}^N \hat{\rho}_v(\varepsilon) = \frac{1}{N(N-1)} \sum_{i,j=1}^N A_{ij}(\varepsilon) \quad (4.13)$$

is defined as the arithmetic mean of the degree densities of all vertices and characterizes the fraction of possible edges that are present in the network. Notably, for a RN the edge density equals the recurrence rate $RR(\varepsilon)$ of the underlying RP.⁴ It is trivial to show that $\hat{\rho}(\varepsilon)$ is a monotonically increasing function of the recurrence threshold ε : the larger the threshold, the more neighbors can be found, and the higher the edge density.

In a similar way, we can consider the arithmetic mean of the local clustering coefficients $\hat{\mathcal{C}}_v(\varepsilon)$ of all vertices, resulting in the (Watts–Strogatz) *global clustering coefficient* [97]

$$\hat{\mathcal{C}}(\varepsilon) = \frac{1}{N} \sum_{v=1}^N \hat{\mathcal{C}}_v(\varepsilon) = \frac{1}{N} \sum_{v=1}^N \frac{\sum_{i,j=1}^N A_{vi}(\varepsilon) A_{ij}(\varepsilon) A_{jv}(\varepsilon)}{\hat{k}_v(\varepsilon)(\hat{k}_v(\varepsilon) - 1)}. \quad (4.14)$$

The global clustering coefficient measures the mean fraction of triangles that include the different vertices of the network. Given our interpretation of the local clustering coefficient in a RN, $\hat{\mathcal{C}}(\varepsilon)$ can be interpreted as a proxy for the average local dimensionality of the dynamical system in phase space.

Notably, in the case of a very heterogeneous degree distribution, the global clustering coefficient will be dominated by contributions from the most abundant type of vertices. For example, for a scale-free network with $p(k) \sim k^{-\gamma}$, vertices with small degree will contribute predominantly, which can lead to an underestimation of the actual fraction of triangles in the network, since $\hat{\mathcal{C}}_v(\varepsilon) = 0$ if $\hat{k}_v(\varepsilon) < 2$ by definition. In order to correct for such effects, Barrat and Weigt [3] proposed an alternative definition of the clustering coefficient, which is nowadays frequently referred to as *network transitivity* [5] and is defined as

$$\hat{\mathcal{T}}(\varepsilon) = \frac{\sum_{v,i,j=1}^N A_{vi}(\varepsilon) A_{ij}(\varepsilon) A_{jv}(\varepsilon)}{\sum_{v,i,j=1}^N A_{vi}(\varepsilon) A_{jv}(\varepsilon)}. \quad (4.15)$$

When interpreting $\hat{\mathcal{C}}(\varepsilon)$ as a proxy for the average local dimensionality, $\hat{\mathcal{T}}(\varepsilon)$ characterizes the effective global dimensionality of the system.

⁴Strictly speaking, this is only true if the recurrence rate is defined such that the main diagonal in the RP is excluded in the same way as potential self-loops from the RN's adjacency matrix.

Finally, turning to shortest path-based characteristics, we define the *average path length*

$$\hat{\mathcal{L}}(\varepsilon) = \frac{1}{N(N-1)} \sum_{i,j=1}^N \hat{d}_{ij}(\varepsilon) = \frac{1}{N} \sum_{v=1}^N \hat{c}_v(\varepsilon)^{-1} \quad (4.16)$$

as the arithmetic mean of the shortest path lengths between all pairs of vertices, and the *global efficiency*

$$\hat{\mathcal{E}}(\varepsilon) = \left(\frac{1}{N(N-1)} \sum_{i,j=1}^N \hat{d}_{ij}(\varepsilon)^{-1} \right)^{-1} = \left(\frac{1}{N} \sum_{v=1}^N \hat{e}_v(\varepsilon) \right)^{-1} \quad (4.17)$$

as the associated harmonic mean. Notably, the average path length can be rewritten as the arithmetic mean of the inverse closeness, and the global efficiency as the inverse arithmetic mean of the local efficiency. We can easily convince ourselves that the average path length must exhibit an inverse relationship with the recurrence threshold, since it approximates (constant) distances in phase space in units of ε [25].

4.2.2 Inter-System Recurrence Networks

In the last decade, two different widely applicable bi- and multivariate extensions of RPs have been proposed [65]: cross-recurrence plots [63, 104] and joint recurrence plots [82]. In the following, we discuss some possibilities for utilizing these approaches in a complex network framework, following previous considerations in [33–35]. For this purpose, let us consider K (possibly multivariate) time series $\{x_i^k\}_{i=1}^{N_k}$ with $x_i^k = x^k(t_i^k)$ sampled at times $\{t_i^k\}$ from dynamical systems $\{X_k\}$ with $k = 1, \dots, K$.

4.2.2.1 Cross-Recurrences and Cross-Recurrence Networks

One way of extending recurrence analysis to the study of multiple dynamical systems is looking at *cross-recurrences*,⁵ i.e., encounters of the trajectories of two

⁵It is important to realize that cross-recurrences are not to be understood in the classical sense of Poincaré's considerations, since they do not indicate the return of an isolated dynamical system to some previously assumed state. In contrast, they imply an arbitrarily delayed close encounter of the trajectories of two *distinct* systems and, therefore, should be better named *cross encounters* instead. Following the same reasoning, terms such as *cross-recurrence plot* or *cross-recurrence rate* are suggestive, but potentially misleading. However, to comply with the existing literature on cross-recurrence plots, we will adopt the established terms even despite their conceptual ambiguities.

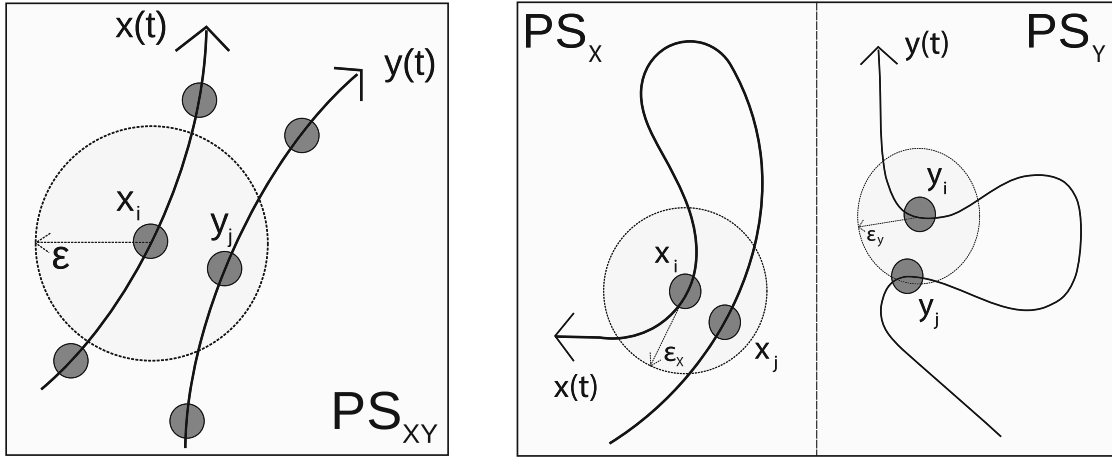


Fig. 4.5 Schematic representation of the notions of cross-recurrence (*left*) and joint recurrence (*right*) between the trajectories $x(t)$ and $y(t)$ of two dynamical systems X and Y . PS_X and PS_Y denote the individual phase spaces of systems X and Y within the joint recurrence framework, whereas PS_{XY} indicates the joint phase space of X and Y necessary for considering cross-recurrences

systems X_k and X_l sharing the same phase space, where $x_i^k \approx x_j^l$ [63, 104] (see Fig. 4.5 for some illustration). Unlike the traditional recurrence matrix \mathbf{R} of a single system, the elements of the cross-recurrence matrix \mathbf{CR}^{kl} are defined as

$$CR_{ij}^{kl}(\varepsilon_{kl}) = \Theta(\varepsilon_{kl} - \|x_i^k - x_j^l\|), \quad (4.18)$$

where $i = 1, \dots, N_k$, $j = 1, \dots, N_l$, and ε_{kl} is a prescribed threshold distance in the joint phase space of both systems. As in the single-system case, ε_{kl} determines the number of mutual neighbors in phase space, quantified by the *cross-recurrence rate*

$$RR_{kl}(\varepsilon_{kl}) = \frac{1}{N_k N_l} \sum_{i=1}^{N_k} \sum_{j=1}^{N_l} CR_{ij}^{kl}(\varepsilon_{kl}), \quad (4.19)$$

which is a monotonically increasing function of ε_{kl} (i.e., the larger the distance threshold in phase space, the more neighbors are found). Note that unlike \mathbf{R}^k and \mathbf{R}^l , the cross-recurrence matrix \mathbf{CR}^{kl} is asymmetric, since we typically have $\|x_i^k - x_j^l\| \neq \|x_i^l - x_j^k\|$. Even more, it can be non-square if time series of different lengths ($N_k \neq N_l$) are considered.

Due to the aforementioned characteristics, \mathbf{CR}^{kl} cannot be directly interpreted as the adjacency matrix of a network with similar properties as single-system RNs. This is because the indices i and j label two distinct sets of state vectors belonging to systems X_k (i) and X_l (j), respectively. In turn, we can interpret the state vectors $\{x_i^k\}$ and $\{x_j^l\}$ as two distinct groups of vertices, and \mathbf{CR}^{kl} as being an adjacency matrix of a *cross-recurrence network* (CRN) providing a binary encoding of the presence of edges between vertices belonging to different groups. This is the defining property of bipartite graphs [70].

Bipartite networks are found in a wide range of fields [46, 50] and can be understood as a generic way for describing arbitrary complex networks [44, 45]. The large variety of applications of bipartite graphs has triggered great interest in models describing their properties in an appropriate way. Particular attention has been spent on the problem of community detection [36], involving new definitions for the modularity function [2, 46, 69, 91] and the development of proper algorithms for community detection [2, 28, 56, 87], partially relating to the spectral properties of the networks. However, their specific structure renders some traditional definitions of network-theoretic measures non-applicable, calling for generalizations or even re-definitions of quantities such as the clustering coefficient [60, 106]. This is why we do not further consider explicit quantification of the properties of the bipartite CRN, but follow a different approach detailed below.

4.2.2.2 Combining Single-System and Cross-Recurrence Networks

As mentioned in Sect. 4.2.2.1, there is a lack of appropriate measures for characterizing explicit bipartite network structures as compared with the rich toolbox of general-purpose complex network characteristics [5, 9]. Therefore, instead of explicitly investigating the bipartite structure of the CRN, it is more useful to combine the information contained in the single-system recurrence matrices $\mathbf{R}^k(\varepsilon_k)$ and the cross-recurrence matrices $\mathbf{CR}^{kl}(\varepsilon_{kl})$ to construct an inter-system recurrence matrix [34]

$$\mathbf{IR}(\boldsymbol{\varepsilon}) = \begin{pmatrix} \mathbf{R}^1(\varepsilon_{11}) & \mathbf{CR}^{12}(\varepsilon_{12}) & \dots & \mathbf{CR}^{1K}(\varepsilon_{1K}) \\ \mathbf{CR}^{21}(\varepsilon_{21}) & \mathbf{R}^2(\varepsilon_{22}) & \dots & \mathbf{CR}^{2K}(\varepsilon_{2K}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{CR}^{K1}(\varepsilon_{K1}) & \mathbf{CR}^{K2}(\varepsilon_{K2}) & \dots & \mathbf{R}^K(\varepsilon_{KK}) \end{pmatrix}. \quad (4.20)$$

Here, $\boldsymbol{\varepsilon} = (\varepsilon_{kl})_{kl}$ is a $K \times K$ matrix containing the single-system recurrence thresholds $\varepsilon_{kk} = \varepsilon_k$ and (cross-recurrence) distance thresholds ε_{kl} . The corresponding *inter-system recurrence network (IRN)* [34] (see Fig. 4.6 for an example) is fully described by its adjacency matrix

$$\mathbf{A}(\boldsymbol{\varepsilon}) = \mathbf{IR}(\boldsymbol{\varepsilon}) - \mathbf{1}_N, \quad (4.21)$$

where $N = \sum_{k=1}^K N_k$ is the number of vertices and $\mathbf{1}_N$ the N -dimensional identity matrix. As in the case of single-system RNs, the IRN is an undirected and unweighted simple graph, which additionally obeys a natural partition of its vertex and edge sets (see Sect. 4.2.2.3). Vertices represent state vectors in the phase space common to all systems X_k and edges indicate pairs of state vectors from *either* the same *or* two different systems that are mutually close, whereby the definition of closeness can vary between different pairs of systems. To this end, we briefly mention two specific choices that may be convenient:

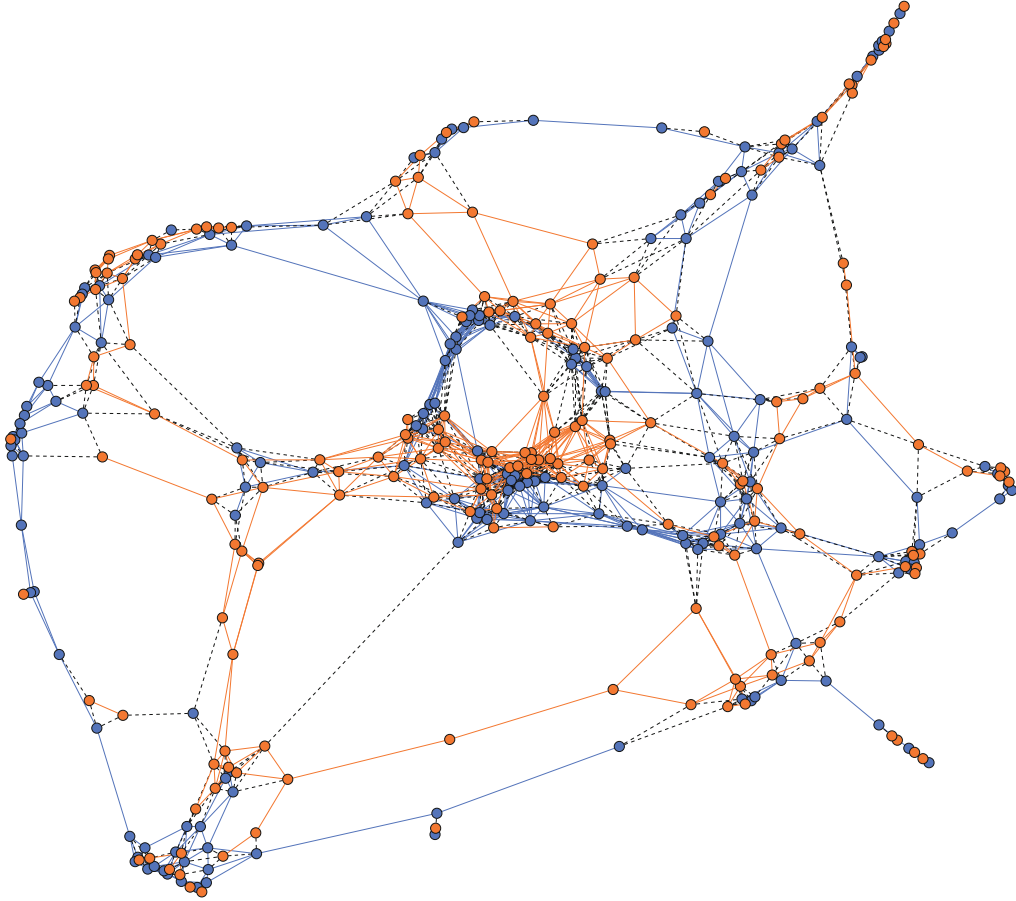


Fig. 4.6 Example of an inter-system recurrence network of two coupled Rössler systems ($N_{1,2} = 100$, $RR = 0.03$, $CRR = 0.02$, $\mu_{1,2} = 0.1$). The vertex sets as well as internal edge sets corresponding to both systems are indicated by *red* and *blue* colors, respectively, while cross-edges are displayed as *dashed gray lines*

- Since we assume the considered systems to share the same phase space, it can be reasonable to measure distances in a way disregarding the specific membership of vertices to the different systems under study. This would imply choosing $\varepsilon_{kl} = \varepsilon$ as equal values for all $k, l = 1, \dots, K$. In such a case, we can reinterpret the IRN as the RN constructed from the concatenated time series

$$\{y_i\}_{i=1}^N = (x_1^1, \dots, x_{N_1}^1, x_1^2, \dots, x_{N_2}^2, \dots, x_1^K, \dots, x_{N_K}^K).$$

In this situation, we can reconsider the general framework of single-system RN analysis as discussed above for studying the geometric properties of the combined system as reflected in a RN. Note, however, that in this case it is hardly possible to explicitly exploit the given natural partitioning of the concatenated data.⁶ In contrast, all state vectors are treated in exactly the same way.

⁶One corresponding strategy could be utilizing methods for community detection in networks [36], such as consideration of modularity [71]. Notably, such idea has not yet been explored in the

Table 4.1 Comparison of two multivariate generalizations of RN analysis regarding the principal requirements on the time series to be analyzed

	IRN	JRN
Length	Arbitrary	Identical
Sampling	Arbitrary	Identical
Physical units	Identical	Arbitrary
Phase space dimension	Identical	Arbitrary

Identical means that a specific property must be the same for all involved time series, while *arbitrary* implies that this does not need to be the case

- An alternative choice of recurrence and distance thresholds is based on considering the individual single-system RNs as quantitatively comparable. Since some of the network measures discussed in Sect. 4.2.1.2 explicitly depend on the number of existing edges in the network, this requirement calls for networks with the same edge density $\rho_k = \rho$ for all $k = 1, \dots, K$. In other words, the recurrence thresholds ϵ_{kk} ($k = 1, \dots, K$) should be chosen such that the (single-system) recurrence rates are equal ($RR_1 = \dots = RR_K = RR$). Given the natural partitioning of the IRN vertex set, such network can be viewed and statistically analyzed as a network of networks (see Sect. 4.2.2.3). In this case, in order to highlight the interconnectivity structure of the individual RNs, it is beneficial to choose the distance thresholds ϵ_{kl} for $k \neq l$ such that the resulting cross-recurrence rates RR_{kl} yield $RR_{kl} < RR_k = RR_l = RR$ and possibly also take the same values $RR_{kl} = CRR < RR$ for all $k \neq l$.

As already stated above, the meaningful construction and analysis of IRNs requires time series $\{x_i^k\}_{i=1}^{N_k}$ that share the same phase space and, hence, describe the same observables with identical physical units (Table 4.1). However, the time series under study can in principle be sampled at arbitrary times $\{t_i^k\}_{i=1}^{N_k}$ and have different lengths N_k , because the method discards all information on time and focuses exclusively on neighborhood relationships in phase space. This type of geometric information is what can be exploited for studying coupling structures between different dynamical systems as reflected by the spatial arrangement of state vectors in the joint phase space (see Sect. 4.4.4).

4.2.2.3 Interacting Network Characteristics

In order to define characteristics that are specifically tailored for analyzing the interdependence structure between two or more complex networks (also called

context of RN analysis, and it is unclear to what extent the inferred possible community structure of an IRN could exhibit relevant information for studying any geometric signatures associated with the mutual interdependences between different dynamical systems. To this end, we leave this problem for future research.

interacting, interdependent, or networks of networks) [7], we utilize a recently introduced general framework [17, 99]. For this purpose, let us consider an arbitrary undirected and unweighted simple graph $G = (V, E)$ with adjacency matrix $\mathbf{A} = \{A_{ij}\}_{i,j=1}^N$. Furthermore, let us assume that there is a given partition of G with the following properties:

1. The vertex set V is decomposed into K disjunct subsets $V_k \subseteq V$ such that $\bigcup_{k=1}^K V_k = V$ and $V_k \cap V_l = \emptyset$ for all $k \neq l$. The cardinality of V_k will be denoted as N_k .
2. The edge set E consists of mutually disjoint sets $E_{kl} \subseteq E$ with $\bigcup_{k,l=1}^K E_{kl} = E$ and $E_{kl} \cap E_{mn} = \emptyset$ for all $(k, l) \neq (m, n)$.
3. Let $E_{kl} \subseteq V_k \times V_l$. Specifically, for all $k = 1, \dots, K$, $G_k = (V_k, E_{kk})$ is the induced subgraph of the vertex set V_k with respect to the full graph G .

Under these conditions, E_{kk} comprises the (internal) edges within G_k , whereas E_{kl} contains all (cross-) edges connecting G_k and G_l . Specifically, for the “natural” partition of an IRN, the G_k correspond to the single-system RNs constructed from the systems X_k , whereas the cross-recurrence structure is encoded in E_{kl} for $k \neq l$.

We are now in a position to study the interconnectivity structure between two subnetworks G_k, G_l on several topological scales drawing on the lineup of local and global graph-theoretical measures generalizing those used for single network characterization (Sect. 4.2.1.2). In this context, local measures \hat{f}_v^{kl} characterize a property of vertex $v \in V_k$ with respect to subnetwork G_l , while global measures \hat{f}_{kl} assign a single real number to a pair of subnetworks G_k, G_l to quantify a certain aspect of their mutual interconnectivity structure. Most interconnectivity characteristics discussed below have been originally introduced in [17] which see for more detailed discussions.

4.2.2.4 Local Measures

The *cross-degree* (or *cross-degree centrality*)

$$\hat{k}_v^{kl} = \sum_{i \in V_l} A_{vi} \quad (4.22)$$

counts the number of neighbors of v within G_l , i.e., direct connections between G_k and G_l (Fig. 4.7a). Thus, this measure provides information on the relevance of v for the network “coupling” between G_k and G_l .⁷ For the purpose of the present

⁷In the specific case of an IRN, we interpret this as geometric signatures of the coupling between the underlying dynamical systems X_k and X_l [33, 34].

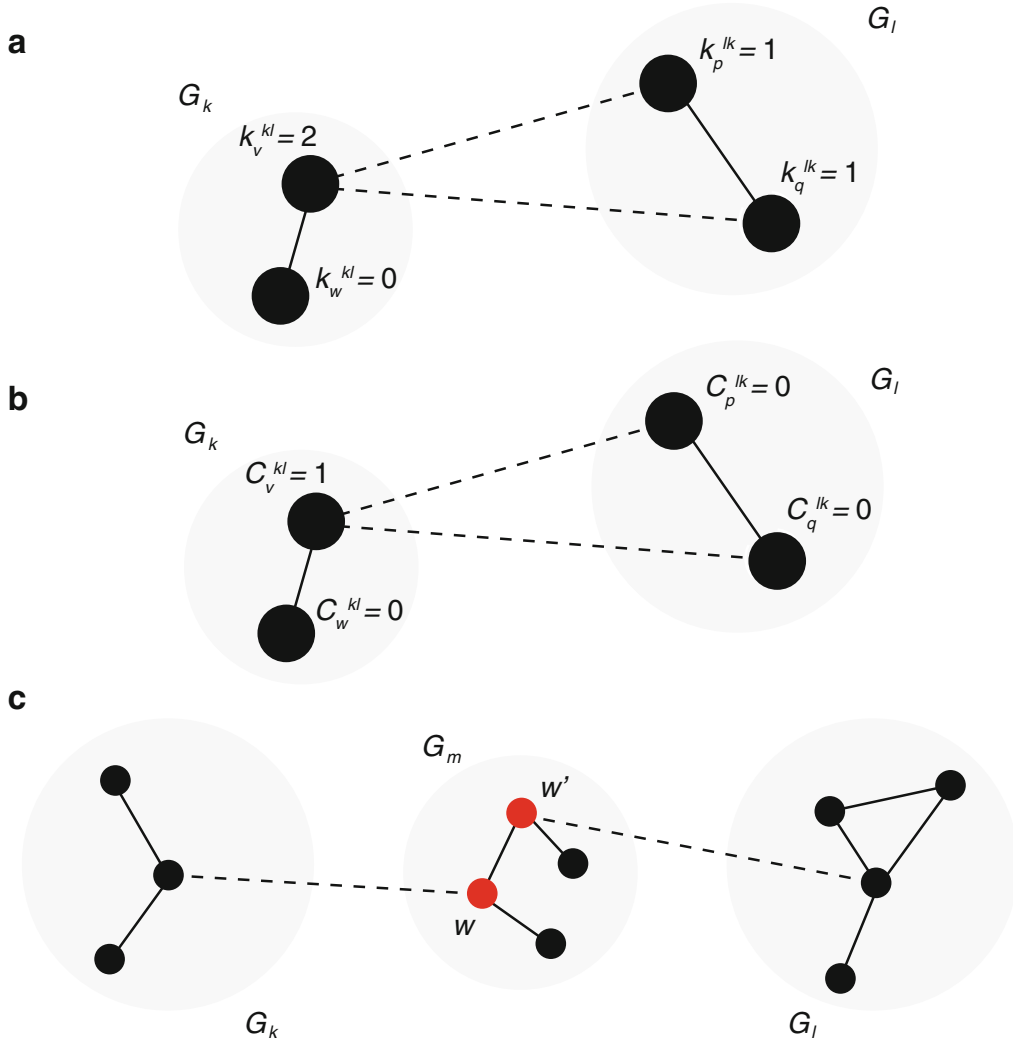


Fig. 4.7 Schematic illustration of some characteristics of interdependent networks: **(a)** the cross-degree k_v^{kl} counts the number of neighbors of vertex $v \in V_k$ within the subnetwork G_l . In the example, the associated cross-edge density is $\rho_{kl} = 0.5$. **(b)** The local cross-clustering coefficient C_v^{kl} is the probability that two randomly drawn neighbors of vertex $v \in V_k$ from subnetwork G_l are mutually connected. In the example, the associated global cross-clustering coefficients are given by $C_{kl} = 0.5$ and $C_{lk} = 0$, whereas the cross-transitivities are $\mathcal{T}_{kl} = 1$ and $\mathcal{T}_{lk} = 0$. **(c)** The cross-betweenness centrality b_w^{kl} measures the fraction of shortest paths between vertices from subnetworks G_k and G_l that traverse vertex $w \in V_m$ (note that G_m can coincide here with G_k or G_l). In the example, $w, w' \in V_m$ (red) have a large cross-betweenness, whereas the remaining vertices $p \in V_m \setminus \{w, w'\}$ from subnetwork G_m do not participate in shortest paths between G_k and G_l and therefore have vanishing cross-betweenness $b_p^{kl} = 0$.

work, it is useful studying a normalized version of this measure, the *cross-degree density*

$$\hat{\rho}_v^{kl} = \frac{1}{N_l} \sum_{i \in V_l} A_{vi} = \frac{1}{N_l} \hat{k}_v^{kl}. \quad (4.23)$$

As for the single network case, important information is governed by the presence of triangles in the network. Given two subnetworks, the *local cross-clustering coefficient*

$$\hat{\mathcal{C}}_v^{kl} = \frac{1}{\hat{k}_v^{kl}(\hat{k}_v^{kl} - 1)} \sum_{i,j \in V_l} A_{vi} A_{ij} A_{jv}, \quad (4.24)$$

measures the relative frequency of two randomly drawn neighbors $i, j \in V_l$ of $v \in V_k$ that are mutually connected (Fig. 4.7b). For $\hat{k}_v^{kl} < 2$, we define $\hat{\mathcal{C}}_v^{kl} = 0$. In general, $\hat{\mathcal{C}}_v^{kl}$ characterizes the tendency of vertices in G_k to connect to clusters of vertices in G_l .

The *cross-closeness centrality*

$$\hat{c}_v^{kl} = \left(\frac{\sum_{i \in V_l} d_{vi}}{N_l} \right)^{-1} \quad (4.25)$$

(where d_{vi} is the graph-theoretical shortest-path length between v and i) characterizes the topological closeness of $v \in G_k$ to G_l , i.e., the inverse arithmetic mean of the shortest path lengths between v and all vertices $i \in V_l$. If there exist no such paths, d_{vi} is commonly set to the maximum possible value $N - 1$ given the size of G . As in the single network case, replacing the arithmetic by the harmonic mean yields the *local cross-efficiency*

$$\hat{e}_v^{kl} = \frac{\sum_{i \in V_l} d_{vi}^{-1}}{N_l}, \quad (4.26)$$

which can be interpreted in close analogy to \hat{c}_v^{kl} . Note that in the case of IRNs, topological closeness directly implies geometric closeness.

As a final vertex characteristic, we may generalize the betweenness concept to the case of coupled subnetworks, which results in the *cross-betweenness centrality*

$$\hat{b}_v^{kl} = \sum_{i \in V_k, j \in V_l; i, j \neq v} \frac{\hat{\sigma}_{ij}(v)}{\hat{\sigma}_{ij}}. \quad (4.27)$$

Here, $\hat{\sigma}_{ij}(v)$ and $\hat{\sigma}_{ij}$ are defined as in the case of a single network. Note that unlike the other vertex characteristics discussed above, in the case of cross-betweenness centrality, we do not require v belonging to G_k or G_l (Fig. 4.7c). The reason for this is that vertices belonging to any subnetwork may have a non-zero betweenness regarding two given subgraphs G_k and G_l , in the sense that shortest paths between $i \in V_k$ and $j \in V_l$ can also include vertices in other subnetworks.

4.2.2.5 Global Measures

The density of connections between two subnetworks can be quantified by taking the arithmetic mean of the local cross-degree density Eq. (4.23), yielding the *cross-edge density*

$$\hat{\rho}^{kl} = \frac{1}{N_k N_l} \sum_{i \in V_k, j \in V_l} A_{ij} = \hat{\rho}^{lk}. \quad (4.28)$$

Notably, $\hat{\rho}^{kl}$ corresponds to the definition of the *cross-recurrence rate* RR_{kl} Eq. (4.19). Since we consider here only undirected networks (i.e., bidirectional edges), the cross-edge density is invariant under mutual exchange of the two considered subnetworks.

The *global cross-clustering coefficient*

$$\hat{\mathcal{C}}^{kl} = \left\langle \hat{\mathcal{C}}_v^{kl} \right\rangle_{v \in V_k} = \frac{1}{N_k} \sum_{v \in V_k, \hat{k}_v^{kl} > 1} \frac{\sum_{i, j \in V_l} A_{vi} A_{ij} A_{jv}}{\sum_{i \neq j \in V_l} A_{vi} A_{vj}} \quad (4.29)$$

estimates the probability of vertices in G_k to have mutually connected neighbors in G_l . Unlike the cross-edge density, the corresponding “cross-transitivity” structure is typically asymmetric, i.e., $\hat{\mathcal{C}}^{kl} \neq \hat{\mathcal{C}}^{lk}$. As in the single network case, we need to distinguish $\hat{\mathcal{C}}^{kl}$ from the *cross-transitivity*

$$\hat{\mathcal{T}}^{kl} = \frac{\sum_{v \in V_k; i, j \in V_l} A_{vi} A_{ij} A_{jv}}{\sum_{v \in V_k; i \neq j \in V_l} A_{vi} A_{vj}}, \quad (4.30)$$

for which we generally have $\hat{\mathcal{T}}^{kl}(\varepsilon) \neq \hat{\mathcal{T}}^{lk}(\varepsilon)$ as well. Again, we have to underline that cross-transitivity and global cross-clustering coefficient are based on a similar concept, but capture distinctively different network properties.

Regarding the quantification of shortest path-based characteristics, we define the *cross-average path length*

$$\hat{\mathcal{L}}^{kl} = \frac{1}{N_k N_l} \sum_{i \in V_k, j \in V_l} d_{ij} \quad (4.31)$$

and the *global cross-efficiency*

$$\hat{\mathcal{E}}^{kl} = \left(\frac{1}{N_k N_l} \sum_{i \in V_k, j \in V_l} d_{ij}^{-1} \right)^{-1} \quad (4.32)$$

Unlike $\hat{\mathcal{C}}^{kl}$ and $\hat{\mathcal{T}}^{kl}$, $\hat{\mathcal{L}}^{kl}$ and $\hat{\mathcal{E}}^{kl}$ are (as shortest path-based measures) symmetric by definition, i.e., $\hat{\mathcal{L}}^{kl}(\varepsilon) = \hat{\mathcal{L}}^{lk}(\varepsilon)$ and $\hat{\mathcal{E}}^{kl}(\varepsilon) = \hat{\mathcal{E}}^{lk}(\varepsilon)$. In the case of

disconnected network components, the shortest path length d_{ij} is defined as discussed for the corresponding local measures.

In the same spirit as shown above, other single network characteristics [5, 9] can be adopted as well for defining further interdependent network measures. This includes measures characterizing edges or, more generally, pairs of vertices like edge betweenness or matching index, further global network characteristics (assortativity, network diameter and radius), mesoscopic structures (motifs), or even characteristics associated with diffusion processes on the network instead of shortest paths (e.g., eigenvector centrality or random walk betweenness). The selection of measures introduced above reflects those characteristics which have the most direct interpretation in the context of IRNs and have also been utilized in studying the interdependence structure between complex networks in other contexts [17, 99].

4.2.3 Joint Recurrence Networks

4.2.3.1 Basic Idea

Besides cross-recurrences, another possible multivariate generalization of RPs is studying joint recurrences of different systems in their individual (possibly different) phase spaces. Here, the basic idea is that the simultaneous occurrence of recurrences in two or more systems $\{X_k\}$ (see Fig. 4.5) contains information on possible interrelationships between their respective dynamics, for example, the emergence of generalized synchronization [82, 83]. Consequently, based on time series $\{x_i^k\}$, the joint recurrence matrix **JR** with elements

$$JR_{ij}(\varepsilon_1, \dots, \varepsilon_K) = \prod_{k=1}^K R_{ij}^k(\varepsilon_k) \quad (4.33)$$

is defined as the element-wise product of the single-system recurrence matrices \mathbf{R}^k with elements

$$R_{ij}^k(\varepsilon_k) = \Theta(\varepsilon_k - \|x_i^k - x_j^k\|), \quad (4.34)$$

where $(\varepsilon_1, \dots, \varepsilon_K)$ is the vector of recurrence thresholds that can be selected for each time series individually, typically such as to yield the same global recurrence rates $RR_k = RR$ for all $k = 1, \dots, K$.

Analogously to single-system recurrence network analysis, we can take a graph-theoretical perspective by defining a *joint recurrence network (JRN)* by its adjacency matrix

$$\mathbf{A}(\varepsilon_1, \dots, \varepsilon_K) = \mathbf{JR}(\varepsilon_1, \dots, \varepsilon_K) - \mathbf{1}_N, \quad (4.35)$$

where $\mathbf{1}_N$ again denotes the N -dimensional identity matrix. Hence, edges (i, j) of a JRN indicate joint recurrences occurring simultaneously in *all* K time series under study. Alternatively, $\mathbf{A}(\varepsilon_1, \dots, \varepsilon_K)$ may be viewed as the element-wise product of the single-system recurrence networks' adjacency matrices $\mathbf{A}^k(\varepsilon_k)$.

As single-system RN and IRN, the JRN describes an undirected and unweighted simple graph. However, due to the temporal simultaneity condition of the joint recurrence concept, vertices i are explicitly associated with points in time $t_i^k = t_i^l$ common to the K considered time series (cf. Table 4.1). This is conceptually different from RNs and IRNs where time information is not taken into account so that network characteristics are invariant under permutations of the state vectors (i.e., the—possibly embedded—observations). More specifically, it is not possible to independently relabel the observations in the underlying time series prior to the computation of the JRN, whereas the JRN vertices can be shuffled again without altering the resulting network properties.

By construction, the time series $\{x_i^k\}$ used for constructing a JRN need to be sampled at identical times $\{t_i^k\}$ and have to have the same length, i.e., $N_1 = N_2 = \dots = N_K = N$. However, since recurrences are compared instead of state vectors, the $\{x_i^k\}$ neither have to represent the same physical quantity measured in identical units, nor need they reside in the same phase space (Table 4.1).

From a conceptual perspective, a JRN can be regarded as a simple RN for the combined system $(X_1 \otimes \dots \otimes X_K)$ in its higher-dimensional phase space spanned by all state variables. However, recurrences are defined here in some non-standard way taking distances in the subspaces associated with the individual systems separately into account. This implies that the properties of JRN can be studied in essentially the same way as those of single-system RNs (but with more subtle geometric interpretations of the respective network characteristics). In turn, comparing the same properties for JRN and single-system RNs provides important information about the similarity of neighborhood relationships in the combined phase space and projections on the individual systems' subspaces. Specifically, we can gain insights about the effective degrees of freedom of the combined system, which may be reduced in comparison with the sum of the degrees of freedom of the uncoupled systems due to dynamical interdependences between its components. We will further detail this idea in Sect. 4.4.5.

4.2.3.2 α -Joint Recurrence Networks

Equivalently to their interpretation outlined in Sect. 4.2.3.1, we can also consider JRN as the reduction of a generalized graph, where the vertices correspond to time points t_i , which can be connected by at most K different types of (labelled) edges representing the mutual closeness of states in the K different systems. In this viewpoint, the reduction towards the JRN follows from the requirement that for a given pair of vertices, in the generalized graph *all* K possible labelled edges must be present. With other words, in terms of Boolean logics the entries of the binary

recurrence matrices \mathbf{R}^k are connected by a logical AND for defining the elements of \mathbf{JR} .

Notably, the presence of a joint recurrence becomes increasingly unlikely as the number of possibly interacting systems K increases. Even in the case of very strong interdependences, there may be stochastic fluctuations in the individual systems (e.g., observational noise) that mask recurrences in individual systems and, thus, subsequently reduce the *joint recurrence rate*

$$JRR(\varepsilon_1, \dots, \varepsilon_K) = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^N JR_{ij}(\varepsilon_1, \dots, \varepsilon_K) \quad (4.36)$$

aka JRN edge density ρ_J .

One possibility to circumvent the problem sketched above is relaxing the requirement of having simultaneous recurrences in all systems (i.e., the logical AND operation connecting the recurrence matrices of the individual systems in a component-wise way), but considering the case where at least a fraction $\alpha \in (0, 1]$ of all systems exhibit recurrences (the standard JRN follows for $\alpha = 1$). This point of view allows defining a hierarchy of networks, which we call *α -joint recurrence networks* (α -JRN). Starting from the union of the single-system RNs providing a network with K different edge types corresponding to recurrences of the individual systems, we require that there exist at least $\lceil \alpha K \rceil$ edges between two specified vertices (i.e., time points). In the specific case of $K = 2$ systems and $\alpha \in (0, 0.5]$ (or, more generally, for $\alpha \in (0, 1/K]$), we can rewrite this requirement with a simple logical (Boolean) operation connecting the single-system recurrence matrices in a component-wise way as $JR_{ij}^\alpha(\varepsilon_1, \varepsilon_2) = R_{ij}^{X_1}(\varepsilon_1) \text{ OR } R_{ij}^{X_2}(\varepsilon_2)$.

For the more general case, in order to mathematically formulate the requirement of $\lceil \alpha K \rceil$ simultaneous recurrences, it is convenient to start from a practically equivalent re-definition of the joint recurrence matrix,

$$JR_{ij}^{(1)}(\varepsilon_1, \dots, \varepsilon_K) = \Theta \left(\sum_{k=1}^K R_{ij}^k(\varepsilon_k) - K \right), \quad (4.37)$$

$$JR_{ij}^{(\alpha)}(\varepsilon_1, \dots, \varepsilon_K) = \Theta \left(\sum_{k=1}^K R_{ij}^k(\varepsilon_k) - \alpha K \right), \quad (4.38)$$

to be the *α -joint recurrence matrix*. We can use the latter definition to define α -joint recurrence plots as well as α -JRN in full analogy to the classical case $\alpha = 1$.

Trivially, the number of edges in an α -JRN decreases monotonically for increasing α if all single-system recurrence thresholds ε_k are kept fixed. We note that a similar relaxation of the strict requirement of a conjunction (AND relation) between the (Boolean) entries of different recurrence matrices has been recently discussed in the framework of symbolic recurrence plots [22]. Moreover, it might be interesting (but has not yet been explored) to use concepts from fuzzy logic as the basis for

somewhat weaker requirements than in the rather restrictive definition of the original JRN.

The conceptual idea of α -JRN has not yet been further developed and studied elsewhere. One possible field of application could be finding proper values of α (for example, in dependence on the magnitude of some observational noise) for which results commonly obtained using “normal” JRN become stable in the case of real-world time series. To this end, we only emphasize the possibility of defining α -JRN and studying the properties of these entities (e.g., the scaling of network characteristics as a function of α), but leave a corresponding investigation as a subject for future research.

4.3 Analytical Description of Recurrence Networks

As we will demonstrate in the following, the properties of RNs can be described analytically supporting their better understanding and, hence, applicability. For this purpose, we can exploit the formal equivalence of RNs and random geometric graphs (RGGs), a well-studied concept in graph theory and computational geometry. In this section, we motivate this equivalence and demonstrate how the variety of RN characteristics can be reformulated in the continuum limit $N \rightarrow \infty$ for any finite ε . This framework allows gaining deep insights into the geometric organization of chaotic attractors by exploring the multitude of characteristics provided by complex network theory. Moreover, we present a first-time extension [12] of the previous analytical considerations [18] to IRNs.

4.3.1 Random Geometric Graphs

Random geometric graphs [77] are based on a (finite) set of vertices randomly positioned in a d -dimensional ($d \in \mathbb{N}^+$) metric space according to some probability density function $p(x)$ with $x \in \mathbb{R}^d$. In general, the connectivity among this set of vertices is taken to be distance-dependent, i.e., for two vertices i and j , the probability of being connected in the RGG has the form $P(A_{ij} = 1) = f(\|x_i - x_j\|)$ with some predefined function f , which is monotonically decreasing. As a consequence, spatially close vertices are more likely to be connected than distant ones. A particularly well studied special case is $f(\delta) = \Theta(\varepsilon - \delta)$ (δ denoting here the distance between any two points in the considered metric space), often referred to as RGG (in the strict sense). Notably, the latter definition has fundamental real-world importance (e.g., in terms of ad-hoc communication networks or, more general, contact networks). Moreover, it matches that of the adjacency matrix of a RN in Eq. (4.4) if we identify $p(x)$ with the invariant density of the dynamically invariant object under study (e.g., some attractor in case of a dissipative system), and take the space in which the RGG is embedded as that spanned by the respective

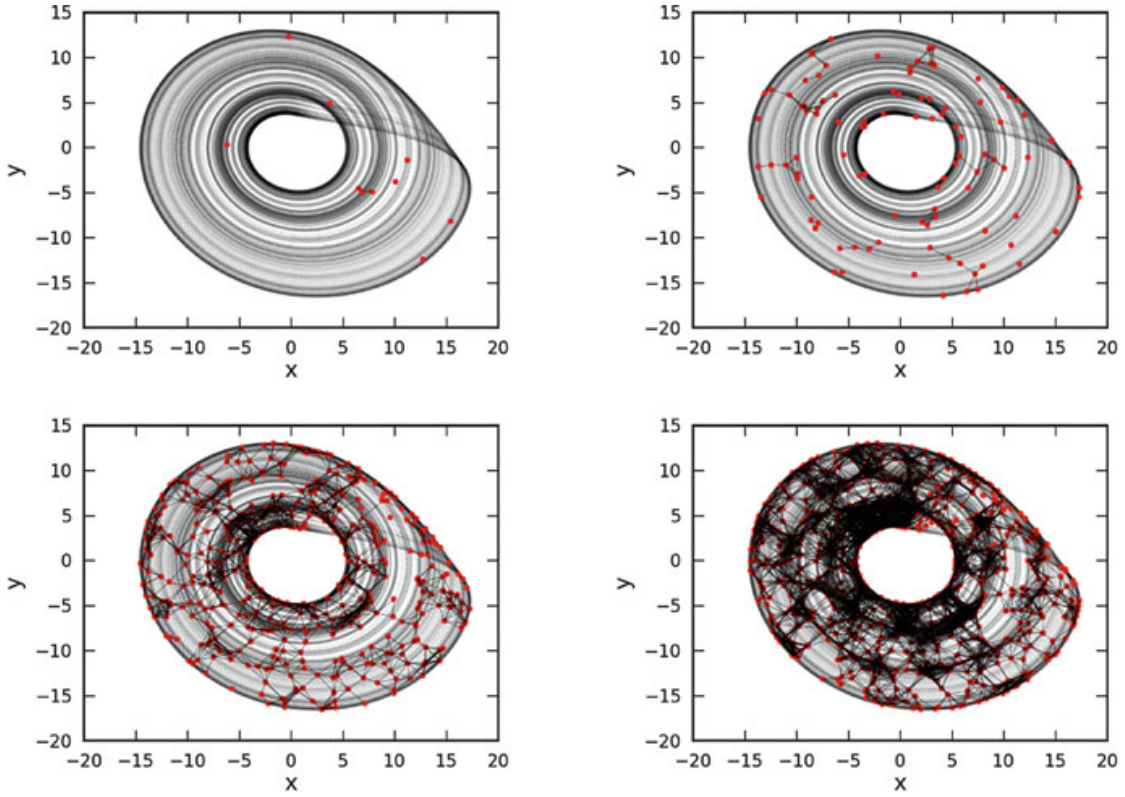


Fig. 4.8 Example of RGGs constructed from the invariant density of the Rössler system Eq. (4.1) using sample sizes of $N = 10, 100, 500$ and $1,000$ and a fixed $\varepsilon = 2.0$. In order to obtain a better visualization, only projections of the attractor and resulting graph to the (x, y) -plane are shown. Note that although the attractor's invariant density is not yet well sampled at $N = 1,000$, most network characteristics obtained from the corresponding RN (not shown) already provide reasonable approximations of the expected values for the underlying RGG.

dynamical variables of the system. In this respect, for all following considerations, it is sufficient restricting attention to the support of $p(x)$ (respectively its closure), which is described by some manifold $S = \text{supp}(p)$ embedded in the considered metric space (e.g., the attractor manifold).

Figure 4.8 shows some example of RGGs obtained using the invariant density of the Rössler system Eq. (4.1) as $p(x)$. Specifically, the RGG has been obtained from the numerical realization of a single trajectory of the system, which has been downsampled to a certain fixed sampling rate.

From a practical perspective, the spatial coverage of $p(x)$ by the RGG's vertices can be strongly affected by the sampling, leading to a spatial clustering of vertices if the sampling frequency is close to an integer multiple of the chaotic attractor's characteristic frequency. In such a situation, it is advantageous to follow alternative sampling strategies for $p(x)$.⁸ As already mentioned above, generating the RGG/RN

⁸For ergodic systems, sampling from one long trajectory, ensembles of short independent realizations of the same system, or directly from the invariant density should lead to networks with the same properties at sufficiently large N .

representation based on bootstrapping from the ensemble of available state vectors can eventually provide better results than a regular sampling of a given trajectory.

As outlined above, the importance of RGGs for the considerations on RNs is that some of their properties (like the degree distribution [47] or transitivity [10]) have been intensively studied in the past for the generic case of a hard distance threshold in f and arbitrary probability density functions $p(x)$ for metric spaces of various integer dimensions. For example, Hermann et al. [47] give a closed-form expression of the degree distribution for arbitrary $p(x)$, whereas Dall and Christensen [10] provide a deep discussion of the transitivity properties of RGGs. Notably, the latter aspect has become particularly relevant in the interpretation of RN properties as well as their multivariate generalizations, as will be further discussed in Sect. 4.4.

4.3.2 *Single-System Recurrence Network Characteristics*

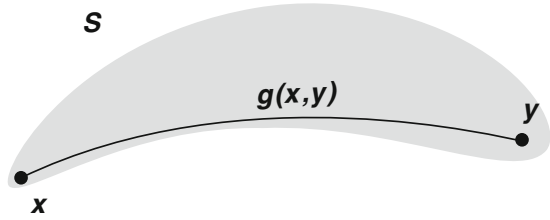
By making use of the fact that RNs are a specific type of RGGs, all relevant graph-theoretical measures for recurrence networks can be seen as discrete approximations of more general and continuous geometrical properties of a dynamical system's underlying attractor characterized by a set S together with an associated invariant density $p(x)$, $x \in S$. This point of view allows obtaining deeper insights into the geometrical meaning of the network quantifiers introduced in Sect. 4.2.1.2 and enables us to establish surprising connections to other fields, e.g., the close relationship of transitivity measures like the local clustering coefficient and global network transitivity to the local and global fractal dimension of the dynamical system's attractor, respectively [26], see Sect. 4.4.2.2. In the following, we review a corresponding analytical framework for general spatially embedded networks which is specifically tailored for defining continuous variants of the common discrete complex network characteristics [18].

4.3.2.1 *General Setting*

Let S be a compact and smooth manifold with a non-vanishing continuous probability density function $p : S \rightarrow (0, \infty)$ with $\int_S dx p(x) = 1$. For the purpose of the present work, we identify S with the set of points defining the attractor of a (dissipative) dynamical system. In case of chaotic attractors in time-continuous systems, we obtain a closure of the open attractive set by considering its union with the set of (infinitely many) unstable periodic orbits embedded in the attractor.

Continuous analogs of the discrete complex network characteristics introduced in Sect. 4.2.1.2 should be approximated by taking the limit $N \rightarrow \infty$ and $\varepsilon \rightarrow 0$ (note that the latter limit may not be assessible in the case of fractal sets S , which we will not further consider in the following). Here, “continuous” refers to a network with uncountably many vertices and edges, which is determined by the *adjacency function*

Fig. 4.9 Schematic illustration of a set S (gray), where $g(x, y)$ denotes the geodesic distance between $x, y \in S$ (after [18])



$$A(x, y) = \Theta(\varepsilon - \|x - y\|) - \delta(x - y) \quad (4.39)$$

for all $x, y \in S$, which is a continuous analog of the adjacency matrix. In the latter expression, $\delta(x - y) = 1$ if $x = y$, and 0 otherwise.

4.3.2.2 Shortest Paths and Geodesics

A large variety of complex network characteristics introduced in Sect. 4.2.1.2 relies on the concept of shortest paths. Examples include closeness and betweenness centrality, local and global efficiency, and average path length. In the continuum limit, we consider a path in S as a closed curve described by a properly parametrized function $f : [0, 1] \rightarrow S$, and define the associated path length

$$l(f) = \sup_{n > 0; \{t_i\}_{i=1}^n} \left\{ \sum_{i=1}^n d(f(t_{i-1}), f(t_i)) \mid 0 = t_0 \leq t_1 \leq \dots \leq t_n = 1 \right\} \in [0, \infty] \quad (4.40)$$

where $d(\cdot)$ denotes some metric used for defining distances on S . The *geodesic distance* between two points $x, y \in S$, which serves as the analog of the shortest path length on a network, is then defined as (cf. Fig. 4.9).

$$g(x, y) = \inf_f \{l(f) \mid f : [0, 1] \rightarrow S, f(0) = x, f(1) = y\}. \quad (4.41)$$

A path of length $g(x, y)$ is called a *global geodesic* on S . Depending on the specific geometry of the considered set S , the multiplicity of global geodesics connecting x and y may differ, including no, one, or even infinitely many distinct global geodesics.

Regarding a continuum limit for RNs, we note that shortest paths in such networks approximate global geodesics on the underlying invariant set S in the limit of $\varepsilon \rightarrow 0$ and $N \rightarrow \infty$. Specifically, in the latter limit the shortest path length $l_{ij}(\varepsilon)$ between two points $x(t_i), x(t_j) \in S$ behaves as

$$\varepsilon l_{ij}(\varepsilon) \rightarrow g(x(t_i), x(t_j)) \quad (4.42)$$

independently of the chosen metric [18].

For defining betweenness centrality, we do not only require information on the lengths of global geodesics, but also their total multiplicity $\sigma(y, z; \varepsilon)$ as well as their multiplicity conditional on a third point $x \in S$ being part of the curve, denoted as $\sigma(y, z|x; \varepsilon)$ in the following. The definition of the latter quantity is, however, not unique for a given finite ε . Two possible, yet generally not equivalent expressions read [18]

$$\sigma_1(y, z|x; \varepsilon) = \sum_{k=1}^{\sigma(y, z; \varepsilon)} \int_0^1 dt \delta(f_k(t) - x) \quad (4.43)$$

$$\sigma_2(y, z|x; \varepsilon) = \sum_{k=1}^{\sigma(y, z; \varepsilon)} \int_0^1 dt \Theta(\varepsilon - \|f_k(t) - x\|), \quad (4.44)$$

where $f_k(t)$ denotes the family of global geodesics between y and z . Note that this family can have uncountably many members (to see this, consider, for example, the set of geodesics between the two poles on a sphere). In this case, the sum in Eqs. (4.43) and (4.44) should be replaced by an integral. Furthermore, we emphasize that the ε -dependence in the multiplicities of shortest paths is implicit rather than explicit, since the chosen discretization level ε can affect the effective “shape” of S and, hence, the positions of possible edges in the considered space.

4.3.2.3 Local (Vertex-Based) Measures

The *continuous ε -degree density*

$$\rho(x; \varepsilon) = \int_{B_\varepsilon(x)} d\mu(y) \quad (4.45)$$

gives the probability that a point $y \in S$ randomly drawn according to p falls into an ε -neighborhood $B_\varepsilon(x) = \{y \in S \mid \|x - y\| < \varepsilon\}$ around x . Its discrete estimator is given by the classical degree density $\hat{\rho}_v(\varepsilon)$ Eq. (4.6).

In order to quantify the density of closed paths of length 3 in the network, we can consider the *continuous local ε -clustering coefficient*

$$\mathcal{C}(x; \varepsilon) = \frac{\int \int_{B_\varepsilon(x)} d\mu(y) d\mu(z) \Theta(\varepsilon - \|y - z\|)}{\rho(x; \varepsilon)^2}. \quad (4.46)$$

This measure characterizes the probability that two points y and z randomly drawn according to p from the ε -neighborhood of $x \in S$ are mutually closer than ε . Its discrete approximation is provided by the classical local clustering coefficient $\hat{\mathcal{C}}_v(\varepsilon)$ Eq. (4.7).

Let $y \in S$ be drawn randomly according to p . For a fixed $x \in S$, the *continuous ε -closeness centrality*

$$c(x; \varepsilon) = \left(\int_S d\mu(y) \frac{g(x, y)}{\varepsilon} \right)^{-1} \quad (4.47)$$

and the *continuous local ε -efficiency*

$$e(x; \varepsilon) = \int_S d\mu(y) \left(\frac{g(x, y)}{\varepsilon} \right)^{-1} \quad (4.48)$$

give the inverse expected geodesic distance and the expected inverse geodesic distance of y to x , respectively. Hence, both measures quantify the geometric closeness of x to any other point in S according to the probability density function p . By making use of RNs, they can be approximated by the classical closeness centrality $\hat{c}_v(\varepsilon)$ Eq. (4.8) and local efficiency $\hat{e}_v(\varepsilon)$ Eq. (4.9).

Finally, the probability that a point x lies on a randomly chosen global geodesic connecting two points $y, z \in S$ according to p is measured by the *continuous ε -betweenness centrality*

$$b(x; \varepsilon) = \int \int_S d\mu(y) d\mu(z) \frac{\sigma(y, z|x; \varepsilon)}{\sigma(y, z; \varepsilon)}. \quad (4.49)$$

Its discrete estimator is given by the standard RN betweenness centrality $\hat{b}_v(\varepsilon)$ Eq. (4.10) with the different possible expressions for $\sigma(y, z|x; \varepsilon)$ [Eqs. (4.43) and (4.44)] [18].

4.3.2.4 Pairwise Vertex and Edge Measures

The *continuous ε -matching index*

$$m(x, y; \varepsilon) = \frac{\int_{B_\varepsilon(x) \cap B_\varepsilon(y)} d\mu(z)}{\int_{B_\varepsilon(x) \cup B_\varepsilon(y)} d\mu(z)} \quad (4.50)$$

quantifies the mutual overlap between the neighborhoods of two vertices $x, y \in S$. In other words, $m(x, y; \varepsilon)$ is the probability that a point $z \in S$ randomly chosen from $B_\varepsilon(x)$ according to p is also contained in $B_\varepsilon(y)$ and vice versa. For $x \rightarrow y$, we have $B_\varepsilon(x) \rightarrow B_\varepsilon(y)$ and, consequently, $m(x, y; \varepsilon) \rightarrow 1$, whereas $m(x, y; \varepsilon) = 0$ if $\|x - y\| > 2\varepsilon$. As in the case of the other measures described above, $m(x, y; \varepsilon)$ can be approximated by the discrete RN matching index $\hat{m}_{vw}(\varepsilon)$ Eq. (4.11).

$m(x, y; \varepsilon)$ does not require mutual closeness between x and y (i.e., $\|x - y\| \in (\varepsilon, 2\varepsilon)$ is possible). In contrast, the *continuous ε -edge betweenness*

$$b(x, y; \varepsilon) = \int \int_S d\mu(z) d\mu(z') \frac{\sigma(z, z'|x, y; \varepsilon)}{\sigma(z, z'; \varepsilon)} \quad (4.51)$$

(with $\sigma(z, z'|x, y; \varepsilon)$ denoting the number of global geodesics between z and z' containing both x and y under the condition $\|x - y\| \leq \varepsilon$, and $\sigma(z, z'; \varepsilon)$ the total number of global geodesics between z and z') is a measure whose discrete estimator $\hat{b}_{vw}(\varepsilon)$ Eq. (4.12) is related to the presence of an edge between $x(t_v)$ and $x(t_w)$, i.e., $\|x(t_v) - x(t_w)\| < \varepsilon$. However, although this property has been originally introduced as an explicit edge property, it can be understood in a more general way as a two-vertex property such that $b(x, y; \varepsilon)$ measures the probability that two specific (not necessarily ε -close) points x and y both lie on a p -randomly drawn global geodesic connecting two points $z, z' \in S$ and are mutually closer than ε . Further generalizations towards n -point relationships are possible, but not instructive within the scope of this work.

4.3.2.5 Global Network Measures

The *continuous ε -edge density*

$$\rho(\varepsilon) = \int_S d\mu(x) \rho(x; \varepsilon) \quad (4.52)$$

is the p -expectation value of the continuous ε -degree density and approximated by the discrete edge density $\hat{\rho}(\varepsilon)$ of a RN Eq. (4.13).

In the same spirit, the *continuous global ε -clustering coefficient*

$$\mathcal{C}(\varepsilon) = \int_S d\mu(x) \mathcal{C}(x; \varepsilon) \quad (4.53)$$

is the p -expectation value of the continuous local ε -clustering coefficient. Its associated discrete estimator is the classical global (Watts–Strogatz) clustering coefficient $\hat{\mathcal{C}}(\varepsilon)$ Eq. (4.14). As an alternative measure characterizing geometric transitivity, we can define the *continuous ε -transitivity*

$$\mathcal{T}(\varepsilon) = \frac{\int \int \int_S d\mu(x) d\mu(y) d\mu(z) \Theta(\varepsilon - \|x - y\|) \Theta(\varepsilon - \|y - z\|) \Theta(\varepsilon - \|z - x\|)}{\int \int \int_S d\mu(x) d\mu(y) d\mu(z) \Theta(\varepsilon - \|x - y\|) \Theta(\varepsilon - \|z - x\|)}, \quad (4.54)$$

which gives the probability that among three points $x, y, z \in S$ randomly drawn according to p , y and z are mutually closer than ε given they are both closer than ε to x . The corresponding discrete estimator is the RN transitivity $\hat{\mathcal{T}}(\varepsilon)$ Eq. (4.15).

As examples of shortest path-based characteristics, we can define the *continuous ε -average path length*

$$\mathcal{L}(\varepsilon) = \int \int_S d\mu(x) d\mu(y) \frac{g(x, y)}{\varepsilon} \quad (4.55)$$

and the *continuous global ε -efficiency*

$$\mathcal{E}(\varepsilon) = \left(\int \int_S d\mu(x) d\mu(y) \left(\frac{g(x, y)}{\varepsilon} \right)^{-1} \right)^{-1}, \quad (4.56)$$

which quantify the expected geodesic distance and the inverse of the expected inverse geodesic distance, respectively, both measured in units of ε between two points $x, y \in S$ drawn randomly according to p . Their discrete estimators are given by the classical RN average path length $\hat{\mathcal{L}}(\varepsilon)$ Eq.(4.16) and global efficiency $\hat{\mathcal{E}}(\varepsilon)$ Eq.(4.17), respectively. Notably, we can reformulate $\mathcal{L}(\varepsilon)$ as the p -expectation value of the inverse continuous ε -closeness centrality,

$$\mathcal{L}(\varepsilon) = \int_S d\mu(x) c(x; \varepsilon)^{-1}, \quad (4.57)$$

and $\mathcal{E}(\varepsilon)$ the inverse p -expectation value of the continuous local ε -efficiency

$$\mathcal{E}(\varepsilon) = \left(\int_S d\mu(x) e(x; \varepsilon) \right)^{-1}. \quad (4.58)$$

4.3.2.6 Further Characteristics

The selection of measures discussed above is far from being complete. Continuous versions of further complex network characteristics, such as assortativity, network diameter and radius, and network motifs are discussed in [18], where also some outlook on corresponding generalizations of other measures like eigenvector centrality or random walk betweenness has been given. To this end, we restrict ourselves to the measures discussed above, since they have been most commonly used in recent applications of the RN framework.

4.3.3 Inter-System Recurrence Network Characteristics

In the same spirit as for the single-system RNs (Sect.4.3.2), we can consider the graph-theoretical measures for studying the interconnections between subnetworks within IRNs (Sect.4.2.2.3) as discrete approximations of continuous geometric properties [12]. Let $S_k \subset Y$ be a subset of an m -dimensional compact smooth manifold Y and $p_k(x)$ represent its invariant density for all $k = 1, \dots, K$, where $x \in S_k$. In the following, the S_k and p_k are assumed to fulfill the same requirements that are stated for S and p in Sect.4.3.2. Notably, the S_k are assumed to have considerable non-empty pairwise intersections. We will use the abbreviation $\int d\mu_k(x) = \int_{S_k} d^m x p_k(x)$, where μ_k is a probability measure on S_k . For

simplicity, only a single recurrence threshold $\varepsilon = \varepsilon_{kl}$ for all k, l will be used in the following. The generalization to different values of ε_{kl} is straightforward.

4.3.3.1 Local Measures

The *continuous ε -cross-degree density*

$$\rho^{kl}(x; \varepsilon) = \int_{B_\varepsilon(x) \cap S_l} d\mu_l(y) = \int d\mu_l(y) \Theta(\varepsilon - \|x - y\|) \quad (4.59)$$

measures the probability that a randomly chosen point in S_l is found in the neighborhood $B_\varepsilon(x)$ of $x \in S_k$. Its discrete version is the cross-degree density $\hat{\rho}_v^{kl}(\varepsilon)$ Eq. (4.23).

The *continuous local ε -cross-clustering coefficient*

$$\mathcal{C}^{kl}(x; \varepsilon) = \frac{\iint_{B_\varepsilon(x) \cap S_l} d\mu_l(y) d\mu_l(z) \Theta(\varepsilon - \|y - z\|)}{\rho^{kl}(x; \varepsilon)^2} \quad (4.60)$$

gives the probability that two randomly chosen points $y, z \in S_l$ are ε -close to each other ($\|y - z\| < \varepsilon$) if they both lie in the neighborhood of $x \in S_k$. $\mathcal{C}^{kl}(x; \varepsilon)$ is approximated by the discrete local cross-clustering coefficient $\hat{\mathcal{C}}_v^{kl}(\varepsilon)$ Eq. (4.24).

Considering the mutual global geometry of the sets S_k, S_l , we furthermore introduce *continuous ε -cross-closeness centrality*

$$c^{kl}(x; \varepsilon) = \left(\int d\mu_l(y) \frac{g(x, y)}{\varepsilon} \right)^{-1} \quad (4.61)$$

quantifying the closeness of $x \in S_k$ to all points of the set S_l along geodesics together with the related harmonic *continuous local ε -cross-efficiency*

$$e^{kl}(x; \varepsilon) = \int d\mu_l(y) \left(\frac{g(x, y)}{\varepsilon} \right)^{-1}. \quad (4.62)$$

Here, geodesics are defined with respect to the union of all involved systems' attractors $S = \bigcup_{k=1}^K S_k$ and $g(x, y)$ is a suitable distance metric on such geodesics (Sect. 4.3.2). The proposed local path-based measures for interdependent networks are approximated by the discrete cross-closeness centrality $\hat{c}_v^{kl}(\varepsilon)$ Eq. (4.25) and local cross-efficiency $\hat{e}_v^{kl}(\varepsilon)$ Eq. (4.26).

Finally, we define the *continuous ε -cross-betweenness centrality*

$$b^{kl}(x; \varepsilon) = \int \int d\mu_k(y) d\mu_l(z) \frac{\sigma(y, z|x; \varepsilon)}{\sigma(y, z; \varepsilon)}. \quad (4.63)$$

As in the single network case, $\sigma(y, z|x; \varepsilon)$ denotes the number of times $x \in S$ (i.e., from any arbitrary subnetwork) lies on a geodesic between $y \in S_k$ and $z \in S_l$, and $\sigma(y, z; \varepsilon)$ denotes the total number of such geodesics. Regarding the appropriate parametrization of $\sigma(y, z|x; \varepsilon)$, we refer to our discussion for the single network case in Sect. 4.3.2. The discrete estimator $\hat{b}_v^{kl}(\varepsilon)$ of $b^{kl}(x; \varepsilon)$ is given in Eq. (4.27).

4.3.3.2 Global Measures

The simplest continuous global property describing the geometric overlap between the sets S_k and S_l is the *continuous ε -cross-edge density*

$$\rho^{kl}(\varepsilon) = \iint d\mu_k(x) d\mu_l(y) \Theta(\varepsilon - \|x - y\|) = \rho^{lk}(\varepsilon) \quad (4.64)$$

that is empirically estimated by the discrete cross-edge density $\hat{\rho}^{kl}(\varepsilon)$ Eq. (4.28).

The expectation value of the continuous local ε -cross-clustering coefficient $\mathcal{C}^{kl}(x; \varepsilon)$ is referred to as the *continuous global ε -cross-clustering coefficient*

$$\mathcal{C}^{kl}(\varepsilon) = \int d\mu_k(x) \mathcal{C}^{kl}(x; \varepsilon), \quad (4.65)$$

which is approximated by the discrete global cross-clustering coefficient $\hat{\mathcal{C}}^{kl}(\varepsilon)$ Eq. (4.29). Moreover, designed for quantifying transitivity in the cross-recurrence structure, the *continuous ε -cross-transitivity*

$$\mathcal{T}^{kl}(\varepsilon) = \frac{\iiint d\mu_k(x) d\mu_l(y) d\mu_l(z) \Theta(\varepsilon - \|x - y\|) \Theta(\varepsilon - \|y - z\|) \Theta(\varepsilon - \|z - x\|)}{\iint d\mu_k(x) d\mu_l(y) d\mu_l(z) \Theta(\varepsilon - \|x - y\|) \Theta(\varepsilon - \|x - z\|)} \quad (4.66)$$

gives the probability that two randomly chosen points $y, z \in S_l$ which are ε -close to a randomly chosen point $x \in S_k$ are also ε -close with respect to each other. $\mathcal{T}^{kl}(\varepsilon)$ is approximated by the discrete cross-transitivity $\hat{\mathcal{T}}^{kl}(\varepsilon)$ Eq. (4.30). As in the case of the discrete estimators, the two latter quantities are in general not symmetric, i.e., $\mathcal{C}^{kl}(\varepsilon) \neq \mathcal{C}^{lk}(\varepsilon)$ and $\mathcal{T}^{kl}(\varepsilon) \neq \mathcal{T}^{lk}(\varepsilon)$.

While the two former measures depend only on the local overlap structure between S_k and S_l together with the invariant densities $p_k(x)$ and $p_l(x)$, path-based measures contain information on the global geometry of both sets. The *continuous ε -cross-average path length*

$$\mathcal{L}^{kl}(\varepsilon) = \iint d\mu_k(x) d\mu_l(y) \frac{g(x, y)}{\varepsilon} = \mathcal{L}^{lk}(\varepsilon) \quad (4.67)$$

gives the average length of geodesic paths starting in S_k and ending in S_l or vice versa. Similarly, we define the *continuous global ε -cross-efficiency*

$$\mathcal{E}^{kl}(\varepsilon) = \left(\iint d\mu_k(x) d\mu_l(y) \left(\frac{g(x, y)}{\varepsilon} \right)^{-1} \right)^{-1} = \mathcal{E}^{lk}(\varepsilon) \quad (4.68)$$

which is the harmonic mean geodesic distance between S_k and S_l . Discrete approximations of these global path-based quantifiers are provided by the cross-average path length $\hat{\mathcal{L}}^{kl}(\varepsilon)$ Eq. (4.31) and global cross-efficiency $\hat{\mathcal{E}}^{kl}(\varepsilon)$ Eq. (4.32), respectively. As for their discrete estimators, the path-based characteristics $\mathcal{L}^{kl}(\varepsilon)$ and $\mathcal{E}^{kl}(\varepsilon)$ are invariant under an exchange of S_k and S_l .

4.4 Recurrence Networks: General Properties and Applications

With the general RN framework (Sect. 4.2) and the associated analytical treatment of RNs (Sect. 4.3) in mind, it is possible to study the properties of RNs as well as their multivariate generalizations from a solid theoretical basis. In the following, we will first discuss some general aspects of complex networks often found in real-world systems, such as small-world effects, the emergence of scale-free degree distributions, or assortative mixing (i.e., the tendency of vertices to connect with other vertices that exhibit a similar degree), regarding their presence or absence in RNs. Subsequently, we will turn to the transitivity characteristics of RNs, motivating their particular usefulness for detecting geometric signatures of qualitative changes in the dynamics of a single system, as well as the dynamical interrelationships between two or more mutually interacting systems.

4.4.1 Generic Network Characteristics

4.4.1.1 Absence of Small-World Effects

A first generic property shared by many real-world networks is the so-called small-world effect, first described as the outcome of studies on social interrelationships, predominantly Milgram's famous chain-letter experiment in the 1960s [68]. In the spirit of the latter studies, the term "small-world effect" originally denoted the fact that average shortest path lengths in social networks, but also other real-world networks, are much shorter than we would expect from random connectivity configurations. Given the importance of redundancy in such networks, Watts and Strogatz [97] suggested including the presence of a high clustering coefficient (i.e., higher than in random graphs) as a second criterion for identifying the small-world effect in real-world networks.

From the latter considerations, it is clear that RNs cannot obey a small-world effect: although they may exhibit a high degree of transitivity (typically depending

on the specific system under study; see our corresponding considerations in Sect. 4.4.2), for any *fixed* value of ε the average path lengths can only take specific values, which become independent of the network size N in case of sufficiently large samples. On the one hand, for any chosen pair of vertices i and j at positions x_i and x_j , the shortest path length is bounded from below as $\hat{l}_{ij} \geq \lceil \|x_i - x_j\|/\varepsilon \rceil$ (respectively, the geodesic distance on the attractor S divided by the recurrence threshold ε). Specifically, each shortest path length will converge to a finite value for $N \rightarrow \infty$. On the other hand, due to the finite diameter of chaotic attractors, the average path length $\hat{\mathcal{L}}(\varepsilon)$ cannot exceed a certain maximum value. Hence, the average path length is bounded from above, which is distinct from the common behavior of small-world networks ($\hat{\mathcal{L}} \sim \log N$) for $N \rightarrow \infty$ [97]. Moreover, as another immediate consequence of the latter considerations, we observe that $\hat{\mathcal{L}} \sim \varepsilon^{-1}$ [25] as long as $\hat{\mathcal{L}} \gg 1$ and $\varepsilon > \varepsilon_c$ (the percolation threshold of the RN [18]). This implies that by tuning ε , it is possible to achieve any desired average shortest path length $\hat{\mathcal{L}} \gg 1$; this fact notably reduces the explanatory power of this global network characteristic.

4.4.1.2 Emergence of Scale-Free Distributions

A general analytical expression for the degree distribution $P(k)$ of a RGG and, hence, a RN has been given by Herrmann et al. [47]. For this purpose, let us make the following assumptions: (1) The system under study is ergodic. (2) The sampled trajectory is sufficiently close to its attractor, i.e., we exclude the presence of transient behavior. (3) The sampling interval is co-prime to any possible periods of the system. If these three conditions are met, the vertices of the RN can be considered as being randomly sampled from the probability density function $p(x)$ associated with the invariant measure μ of the attractor given that N is sufficiently large [29].

For a RGG with arbitrary $p(x)$, the degree distribution $P(k)$ can be derived from $p(x)$ in the limit of large sample size N as

$$P(k) = \int dx p(x) e^{-\alpha p(x)} (\alpha p(x))^k / k! \quad (4.69)$$

(representing an n -dimensional integral in case of an n -dimensional system) with $\alpha = \langle k \rangle / \int dx p(x)^2$ [47]. In order to understand this relationship, note that for each x , the probability that a sampled point falls into the ε -ball centered at x is approximately proportional to $p(x)$. Hence, the degree of a node at x has a binomial distribution. For sufficiently large N , the latter can be approximated by a Poissonian distribution with parameter $\alpha p(x)$, leading to Eq. (4.69).

For the specific case of one-dimensional maps, Eq. (4.69) can be explicitly evaluated, leading to a general characterization of the conditions under which scale-free distributions can emerge in RNs. When projecting higher-dimensional time-continuous systems to such one-dimensional maps by making use of appropri-

ate (Poincaré) return maps, the corresponding considerations can be generalized to such systems, given the specific Poincaré surface is “representative” for the system’s geometric structure. A detailed discussion has been presented in [113]. To this end, we only recall the main result that when the system’s invariant density $p(x)$ exhibits a singularity with power-law shape, Eq. (4.69) implies that the resulting RN’s degree distribution must also display a power-law in the limit $N \rightarrow \infty$ for sufficiently small ε . In turn, if ε is chosen too large, the scale-free behavior cannot be detected anymore, since it is masked by too large neighborhoods of the points close to the singularity. Figure 4.10 demonstrates the latter effect for the specific case of the Rössler system Eq. (4.1).

Notably, it is not trivial to provide an exhaustive characterization of the conditions under which scale-free distributions can emerge for higher-dimensional systems. As a consequence, generally applicable necessary and sufficient conditions for the presence of power-laws in the degree distributions of RNs have not been established so far.

We note that in general complex systems, the emergence of power-laws is often associated with a hierarchical organization related to certain fractal properties. In contrast, for RNs it has been shown that the presence of power-laws is not directly related to some (global) fractal structure of the system, but rather the local shape of its invariant density. Consequently, although there are examples of dynamical systems where the scaling exponent of the degree distribution coincides well with the associated fractal dimension, there is no such relationship in general. It will be subject of future studies under which conditions regarding the structural organization of the attractor, fractal structure and power-law singularities are sufficiently closely related so that the RN’s degree distribution allows quantifying the system’s fractal properties.

4.4.1.3 Assortative vs. Disassortative Mixing

Unlike regarding small-world effects and scale-free degree distributions, there are hardly any available results regarding the mixing properties of RNs. In general, RNs obey a tendency towards showing assortative mixing (i.e., vertices tend to link to other vertices with similar degree), which is reasonable in situations where the invariant density $p(x)$ is continuous or even differentiable as supported by recent numerical results [15, 25].

4.4.2 Characterization of Dynamical Complexity

The main field of application of RQA as well as other quantitative approaches to characterizing the distribution of recurrences in phase space (e.g., recurrence time statistics) is identifying and quantifying different degrees of dynamical complexity among realizations of the same system under different conditions (e.g., different

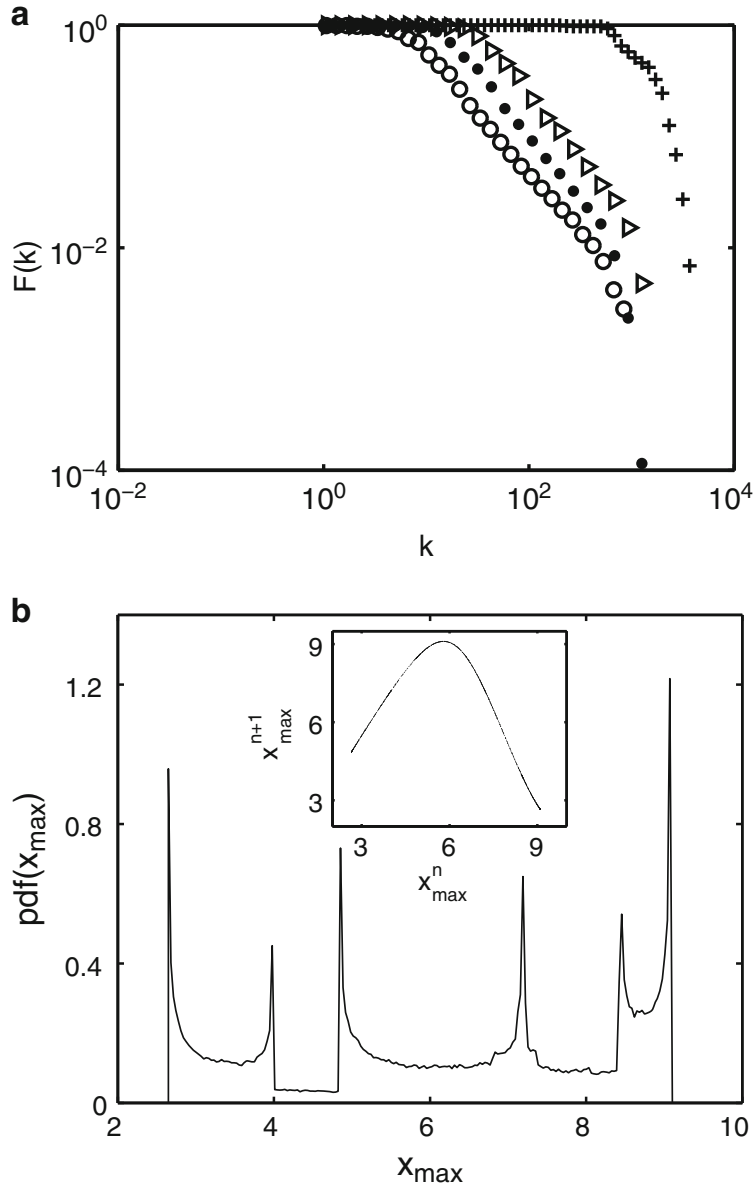


Fig. 4.10 (a) Complementary cumulative distribution function $F(k) = \sum_{k'=k}^{\infty} p(k')$ for RNs obtained from the x -component of the first return map of the Rössler system (with $a = b = 0.2$, $c = 5.7$) through the $y = 0$ plane, using edge densities $\hat{\rho}_1 = 0.02\%$ (open circle), $\hat{\rho}_2 = 0.03\%$ (filled circle), $\hat{\rho}_3 = 0.05\%$ (triangle right), and $\hat{\rho}_4 = 3\%$ (+). All curves have been obtained as mean values taken from five independent realizations of the system with length $N = 2 \times 10^5$ and using the Euclidean norm. For $\hat{\rho}_1$ to $\hat{\rho}_3$, we find power-law behavior with a characteristic exponent of $\gamma = 2.16 \pm 0.03$, whereas no clear scaling region is found in the denser RN with edge density $\hat{\rho}_4$. (b) PDF of the x values of the corresponding return map (insert). Power-law shaped singularities of the PDF (observable spikes) correspond to supertrack functions [64, 74, 75] of the return map. Redrawn after [113]

values of the control parameter(s)), or even within a single time series given the system is non-stationary. While the line-based characteristics of RQA are founded on heuristic considerations (e.g., the higher the predictability of the observed dynamics, the longer the diagonal line structures off the main diagonal should

be), we have argued in Sect. 4.3 that RNs have an analytical foundation in RGGs. Notably, the corresponding characteristics are based on the same binary structure (the recurrence matrix) as the RQA measures. Hence, both concepts should allow deriving a similar kind of information, with the important difference being that RQA quantifies dynamical properties whereas RNs encode topological/geometric characteristics. However, since both aspects are ultimately linked in the case of chaotic attractors, this general observation suggests that RN analysis is in principle suitable for characterizing dynamical complexity in the same way as other established concepts. Therefore, one natural question arises: How do RN measures perform in this task, and which of the multiple possible network measures are particularly suited for this purpose?

The latter questions have been the main motivation behind much of the early work on RNs focussing on numerical studies of various paradigmatic model systems for low-dimensional chaos [23–25, 27, 66, 109, 111]. The latter studies suggest that for characterizing dynamical complexity, global network characteristics are conceptually easier to use and could provide potentially more stable and distinctive results than certain statistics over local network properties such as the distributions of vertex degrees [113] or local clustering coefficients [111]. Among the set of possible global RN measures, two properties have been found particularly useful: network transitivity $\hat{\mathcal{T}}$ and average path length $\hat{\mathcal{L}}$.

4.4.2.1 Average Path Length

Regarding for the average path length, the discriminatory skills regarding different degrees of dynamical complexity can be understood by the fact that for time-continuous systems, chaotic systems can display different degrees of spatial filling of the “populated” area in phase space. In this spirit, a high (fractal) dimension of a chaotic attractor close to the (integer) dimension of the corresponding phase space gives rise to a more homogeneous filling than lower ones, which has a natural geometric consequence for the possible path lengths between pairs of sampled state vectors on the attractor. However, it needs to be noted that quantifying dynamical complexity by means of $\hat{\mathcal{L}}$ suffers from two important drawbacks.

On the one hand, the measure is not normalized and depends crucially on the choice of ε . Hence, working in different methodological settings (e.g., using fixed recurrence thresholds ε vs. fixed recurrence rates $RR = \hat{\rho}$) can provide potentially ambiguous results, since numerical values of $\hat{\mathcal{L}}$ cannot necessarily be directly compared with each other.

On the other hand, even the qualitative behavior of $\hat{\mathcal{L}}$ in dependence on the system’s dynamical complexity depends on whether the system is a discrete map or time-continuous. In the latter case, a periodic orbit would result in a higher average path length than a chaotic one, since a chaotic attractor is a “spatially extended” object in phase space on which there are “shortcuts” between any two state vectors connecting points corresponding to different parts of the trajectory [25]. In turn, for discrete maps, a periodic orbit contains only a finite set of p mutually different state

vectors, so that for sufficiently low ε and large N , the RN is decomposed into p disjoint, fully connected components. In such a situation with not just a few isolated vertices, but a completely decomposed network, a reasonable redefinition of $\hat{\mathcal{L}}$ would be summing up only over pairs of mutually reachable vertices in Eq. (4.16). Consequently, we approach the minimum possible value of $\hat{\mathcal{L}} = 1$ [66], whereas chaotic orbits typically lead to larger $\hat{\mathcal{L}}$.

According to the above observations, there is no fully developed theoretical understanding and description of the influence of attractor dimensionality on the resulting average path length beyond the general considerations presented in Sect. 4.3. Corresponding further investigations might be an interesting subject for future studies.

4.4.2.2 Network Transitivity

As mentioned in Sect. 4.4.1.2, the scaling exponent of a possible power-law degree distribution has no direct relationship to the fractal dimension of the system. In turn, such a relationship naturally exists when studying the corresponding integrated measure (i.e., the edge density $\hat{\rho}(\varepsilon)$) in terms of its scaling properties as the recurrence threshold is systematically varied. The latter approach has been extensively discussed in the literature in connection with the estimation of dynamical invariants from RPs [31, 93] and gives rise to estimates of the correlation dimension D_2 . Notably, one of the classical approaches to estimating D_2 from time series data, the Grassberger-Procaccia algorithm [42, 43], makes use of the correlation sum, which can be easily formulated in terms of the recurrence rate or RN edge density.

The relatively high computational complexity of approaches to estimating the correlation dimension from a RP stems from the fact that a sequence of RPs for different values of ε needs to be studied for obtaining a proper scaling relationship. In turn, as shown by us in previous studies [26], network transitivity provides an alternative approach to defining and estimating a different notion of fractal dimension. For this purpose, note that for a classical RGG embedded in some integer-dimensional metric space, the expected network transitivity (which is numerically estimated as the ensemble mean over sufficiently many realizations of the stochastic generation of the RGG) is an analytical function of the dimension m , which decays (exactly when using the maximum norm, otherwise approximately) exponentially with m [10]. This analytical relationship can be generalized to attractor manifolds with non-integer fractal dimensions, which can in turn be estimated from the RN transitivity by inverting this function.

For the general case, the latter idea leads to a pair of quantities referred to as upper and lower transitivity dimensions [26],

$$D_{\mathcal{T}}^u = \limsup_{\varepsilon} \frac{\log(\mathcal{T}(\varepsilon))}{\log(3/4)}, \quad (4.70)$$

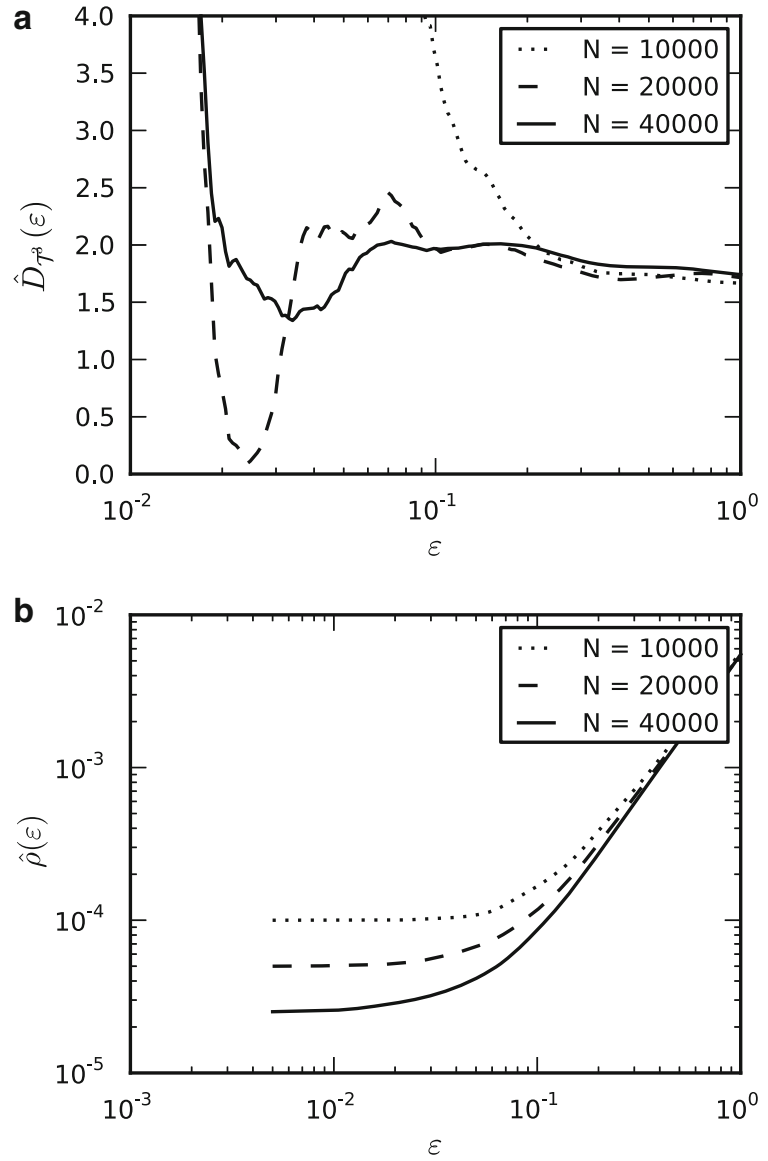
$$D_{\mathcal{T}}^l = \liminf_{\varepsilon} \frac{\log(\mathcal{T}(\varepsilon))}{\log(3/4)}, \quad (4.71)$$

where the two definitions originate from the fact that certain systems (in particular, chaotic maps whose attractors form Cantor sets in at least one direction in phase space [26]) can exhibit an oscillatory behavior between some upper and lower accumulation point of $\mathcal{T}(\varepsilon)$ as the recurrence threshold ε is varied. For systems without such fragmented structure, upper and lower transitivity dimension seem to coincide, which allows estimating them from the sample RN transitivity with reasonable accuracy using only a single network instance with one suitably chosen value of ε . A detailed analytical investigation of the qualitatively different behavior of the RN transitivity for chaotic attractors with continuous and fragmented invariant densities in dependence on ε will be subject of future work. Note that in the above definition, we do not explicitly consider a scaling behavior for $\varepsilon \rightarrow 0$, since the definition does not explicitly contain ε (as it is the case for other classical notions of fractal dimensions), but makes use of normalized characteristics with a probabilistic interpretation (cf. Sect. 4.3). In this spirit, the fraction on the right-hand side of the former equations is a well-defined object for each value of ε (i.e., the specific scale under which the system is viewed) individually.

Figure 4.11a shows the behavior of the scale-dependent transitivity dimension estimate $\hat{D}_{\mathcal{T}}(\varepsilon) = \log(\hat{\mathcal{T}}(\varepsilon))/\log(3/4)$ for the Rössler system Eq. (4.1) for three different RN sizes. We clearly recognize that $\hat{D}_{\mathcal{T}}(\varepsilon)$ assumes approximately stable (i.e., N - and ε -independent) values if the recurrence threshold is chosen sufficiently large. In general, there exist two limits that need to be taken into account: For too large recurrence rates, the RN characteristics lose their discriminatory skills, since too many edges are present masking subtle small-scale properties of the attractor [18, 24]. In turn, if ε is too low (e.g., if $\hat{\rho}$ is below the RN's percolation threshold) [18], the network decomposes into mutually disjoint components, and the resulting network characteristics can become ambiguous. In the considered example of the Rössler system, this decomposition is mainly caused by the rare excursions of some cycles towards larger z values (cf. Fig. 4.1), which give rise to a poorly populated region (low $p(x)$) of the attractor. In order to properly cover this part of the attractor for a given ε , many samples (i.e., a large network size N) are necessary. Otherwise, the edge density $\hat{\rho}$ starts saturating as ε gets smaller (at least in the regime where most vertices close to the $z = 0$ plane are still connected, cf. Fig. 4.11b), and the transitivity dimension estimates strongly deviate from their expected values.

Notably, the analytical relationship [Eqs. (4.70) and (4.71)] between the effective (geometric) dimension of chaotic attractors and RN transitivity provides the theoretical justification and foundation for applying $\hat{\mathcal{T}}$ as a characteristic discriminating between high and low dynamical complexity of chaotic attractors. Unlike for $\hat{\mathcal{L}}$, the transitivity shows qualitatively the same behavior for discrete and time-continuous systems and is normalized, so that its values can be directly used as a quantitative measure of dynamical complexity associated with the effective geometric dimensionality and, hence, structural complexity of the attractor in phase space.

Fig. 4.11 (a) Dependence of the transitivity dimension $\hat{D}_{\mathcal{T}}(\varepsilon)$ on the recurrence threshold ε for realizations of the Rössler system Eq. (4.1) with different lengths N (sampling time $\Delta t = 0.05$, first part of the trajectory removed to avoid possible transient dynamics). (b) Corresponding behavior of the edge density $\hat{\rho}$. Note the different scale on the x axis



4.4.2.3 Other Network Characteristics

In addition to the aforementioned characteristics, there are also other RN measures on both local and global scale that are able to trace changes in the dynamical complexity of a given system [15, 111, 112]. In some cases, there are strong conceptual interrelationships with the previously discussed properties [15], whereas other measures (for example, the assortativity) require more complex considerations for providing potential interpretations of the observed variability. In general, we emphasize that as of today, transitivity and path length based characteristics provide the computationally simplest and theoretically best understood tracers of dynamical complexity based on RNs.

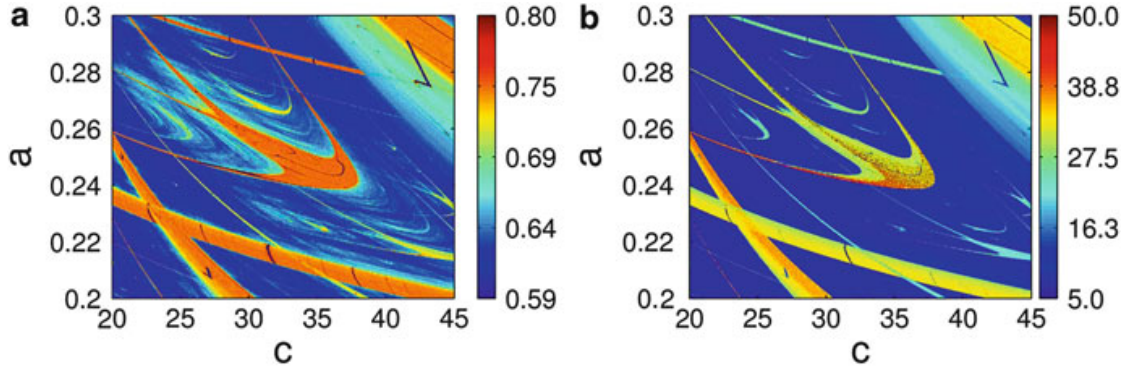


Fig. 4.12 RN transitivity $\hat{\mathcal{T}}$ (a) and average path length $\hat{\mathcal{L}}$ (b) for a two-dimensional intersection ($a = b$) of the three-dimensional parameter space of the Rössler system Eq. (4.1), displaying “shrimp” structures (i.e., self-similar periodic windows with complex shape). For details, see [109]

4.4.2.4 Example: Tracing Bifurcations in the Rössler System

In order to illustrate the performance of RN transitivity $\hat{\mathcal{T}}$ and average path length $\hat{\mathcal{L}}$ as tracers for qualitative changes in the dynamics of complex systems, we briefly recall results originally obtained by the authors [109]. In the latter work, the RN properties have been successfully used to discriminate between periodic and chaotic solutions in a two-dimensional subspace ($a = b$) of the original three-dimensional parameter space of the Rössler system.

As Fig. 4.12 reveals, there are sequences of transitions between periodic and chaotic solutions. Specifically, we clearly see from the figure that the periodic windows are characterized by higher values of $\hat{\mathcal{T}}$ and $\hat{\mathcal{L}}$ than the chaotic solutions, which is in agreement with the general considerations discussed above. Specifically, for the periodic windows, we find $\hat{\mathcal{T}}$ close to 0.75, the theoretical value for periodic dynamics (i.e., a system with effective dimension of 1).

In a similar way, we may use the RN framework for capturing the signatures of qualitative changes in the attractor’s shape and invariant density as a single control parameter is varied systematically. In a previous study using the Rössler system, we have investigated the RN properties across the transition from the classical phase-coherent Rössler attractor to the non-coherent funnel regime [111]. Our results indicate that phase coherence—in a similar spirit as fractal dimension—should be characterized from a geometric rather than a dynamics viewpoint. However, as of today there is no single RN-based index for phase coherence that has been explicitly derived from theoretical considerations.

While the aforementioned results have been obtained for stationary systems, i.e., independent realizations of the system at fixed parameter values, tracing temporal changes in dynamical complexity of non-stationary systems is another interesting field of application with numerous examples in the real-world. Using model systems with drifting parameters such as the Lorenz [15] or Rössler systems (see Fig. 4.13), it is possible to systematically evaluate the performance of RN characteristics in a sliding windows framework, underlining their capabilities for discriminating between qualitatively different types of dynamics and different

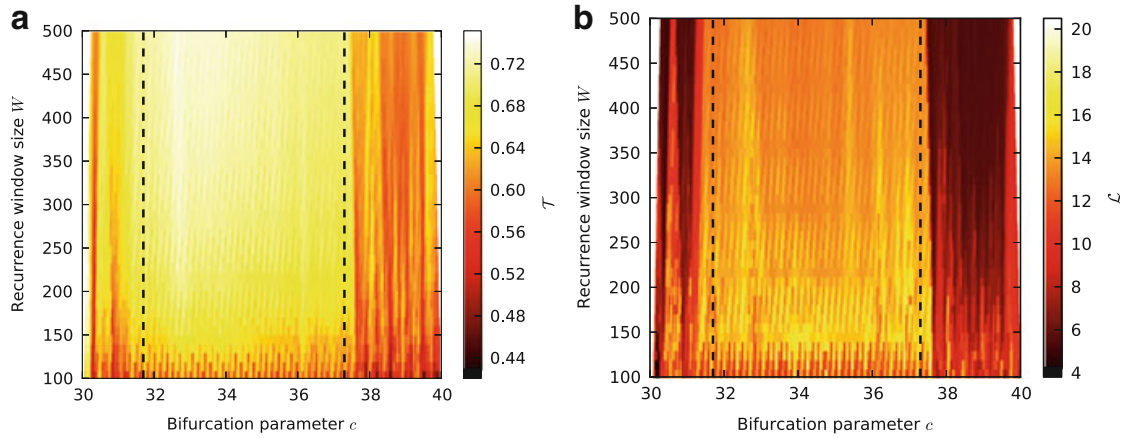


Fig. 4.13 (a) RN transitivity \mathcal{T} and (b) average path length \mathcal{L} for varying recurrence window size W for the Rössler system Eq. (4.1) at $a = b = 0.25$ with linearly drifting bifurcation parameter c over 10,000 time steps (sampling interval $\Delta t = 0.2$). For constructing the RN, a single long trajectory with the three original coordinates (i.e., no embedding) and initial values $(x_0, y_0, z_0) = (0, -10, 0)$ was used. W was varied linearly in the interval (100, 500), the recurrence window step size was fixed to $\Delta W = 10$. The threshold ε has been set such as to yield $\hat{\rho} = 0.05$ in all windows. Vertical dashed lines indicate the critical values of c marking transitions between periodic and chaotic windows of the stationary system at about $c = 31.7$ and 37.3 , respectively (cf. Fig. 4.12)

degrees of complexity in non-stationary (transient) runs as well. For the example of a linearly drifting control parameter c of the Rössler system (Fig. 4.13), we find that the values at which bifurcations between periodic and chaotic behavior occur in the non-stationary system do well coincide with the numerically estimated bifurcation points of the autonomous system inferred from Fig. 4.12, indicating that in the considered example, transient dynamics close to the bifurcation points does not play a major role as long as the considered RNs are still sufficiently large to obtain a reliable statistics.

4.4.3 Characterization of Local Dimensionality

With the same rationale as for the global network transitivity, we can make use of the local clustering properties of RNs for defining local measures of attractor dimensionality, referred to as upper and lower clustering dimensions [26]:

$$D_{\mathcal{C}}^u(x) = \limsup_{\varepsilon} \frac{\log(\mathcal{C}(x; \varepsilon))}{\log(3/4)}, \quad (4.72)$$

$$D_{\mathcal{C}}^l(x) = \liminf_{\varepsilon} \frac{\log(\mathcal{C}(x; \varepsilon))}{\log(3/4)}. \quad (4.73)$$

Following the same argument as for the (global) transitivity dimensions, we do not need to consider the limit $\varepsilon \rightarrow 0$ here.

With similar considerations regarding the possible existence of two distinct accumulation points of $\mathcal{C}(x)$ as ε varies, we may utilize this framework for characterizing the point-wise dimension of chaotic attractors in a unique way without making explicit use of scaling characteristics as in the common point-wise dimensions [26]. However, we need to keep in mind that the considered concept of (geometric) dimensionality is largely affected by the profile of the invariant density, e.g., the existence of sharp attractor boundaries or supertrack functions [23, 25, 26]. For example, if the attractor has distinct tips (e.g., in the case of the Hénon system [25, 26]), the geometric dimension at these points is effectively reduced to zero, which is reflected by $\hat{\mathcal{C}}_v = 1$ for vertices v sufficiently close to the tips. A similar behavior can be observed for the logistic map at the attractor boundaries and the supertrack functions [23, 25, 26].

The latter observations point to a prospective application of the local clustering properties of RNs. In case of chaotic attractors of time-continuous dynamical systems, it is known that an infinite number of unstable periodic orbits (UPOs) provide the skeleton of the chaotic dynamics and are densely embedded in the attractor. The localization of such UPOs is, however, known to be a challenging task. Since UPOs are relatively weakly repulsive (from a practical perspective, those UPOs with low periods are typically least unstable), a trajectory getting close to the vicinity of an UPO will stay close to this orbit for some finite amount of time [55]. As a result, the dynamics close to UPOs is quasi one-dimensional, and state vectors sampled from the trajectories approximate some lower-dimensional (in the limiting case one-dimensional) subset of the attractor manifold. In such case, the above theoretical considerations suggest that the local clustering coefficient $\hat{\mathcal{C}}_v$ of vertices v close to low-periodic UPOs should be higher than the values typical for other parts of the chaotic attractor. This conceptual idea is supported by numerical results from our previous work [24, 25] (cf. also the band structures with increased $\hat{\mathcal{C}}_v$ in Fig. 4.3b), but has not yet been systematically applied to the problem of UPO localization. Notably, the detection limit of UPOs should be ultimately determined by the recurrence threshold ε in conjunction with the RN size N . Specifically, for every finite $\varepsilon > 0$, there are infinitely many UPOs intersecting with the ε -neighborhood of some point x_v in phase space, whereas we will (for a finite sample of state vectors) only resolve the signatures of the least unstable orbits.

4.4.4 *Cross-Transitivity Properties and Coupling Asymmetry*

The new class of statistical network measures designed for investigating the topology of networks of networks discussed in Sect. 4.2.2 is readily applicable for analyzing the interdependency structure of multiple complex dynamical systems. For the special case of two coupled systems X and Y , we have demonstrated numerically that in an IRN, the asymmetry intrinsic to the global measures cross-transitivity $\hat{\mathcal{T}}^{XY}$ and global cross-clustering coefficient $\hat{\mathcal{C}}^{XY}$ can be exploited to reliably detect the direction of coupling between chaotic oscillators over a wide

range of coupling strengths, requiring only a relatively small number of samples $N_{X,Y} \sim \mathcal{O}(10^2 \dots 10^3)$ [34]. For this purpose, we make again use of the fact that transitivity-based characteristics quantify subtle geometric properties that can be easily evaluated both analytically and numerically.

In order to see how cross-transitivities and global cross-clustering coefficients capture dynamical signatures of asymmetric vs. symmetric coupling configurations, let us assume a diffusive coupling with positive sign (i.e., an attractive interaction) as in Eq. (4.2). In the uncoupled case, cross-triangles arise randomly according to the sampling from the systems' respective invariant densities. In this case, eventual asymmetries between $\hat{\mathcal{T}}^{XY}$ and $\hat{\mathcal{T}}^{YX}$ (or, equivalently, $\hat{\mathcal{C}}^{XY}$ and $\hat{\mathcal{C}}^{YX}$) originate from the geometry of the respective sets and the associated $p(x)$, which should already be reflected in the single-system RN transitivities and global clustering coefficients. In turn, if both systems are represented by the same set of state variables (a prerequisite for the application of IRNs) and obey similar values of $\hat{\mathcal{T}}^X$ and $\hat{\mathcal{T}}^Y$ ($\hat{\mathcal{C}}^X$ and $\hat{\mathcal{C}}^Y$), it is likely that also $\hat{\mathcal{T}}^{XY}$ and $\hat{\mathcal{T}}^{YX}$ ($\hat{\mathcal{C}}^{XY}$ and $\hat{\mathcal{C}}^{YX}$) take similar values. Note that minor asymmetries in the interdependent network characteristics can already occur if both systems are only weakly non-identical, e.g., when considering uncoupled identical Rössler systems with just a small detuning of their natural frequencies [34].

Let us suppose now that there is a unidirectional coupling $X \rightarrow Y$. In this case, the trajectory of the driven system Y is attracted by that of the driver X due to the considered form of coupling. As a result, it is likely to find more states in Y that are close to mutually connected pairs of states in X than in the uncoupled case. This implies that $\hat{\mathcal{T}}^{YX}$ ($\hat{\mathcal{C}}^{YX}$) increases since X is “pulling” the trajectory of Y and, hence, the number of triangles having their baseline in system X increases relatively to those having their baseline in Y . Consequently, we expect to have $\hat{\mathcal{T}}^{YX} > \hat{\mathcal{T}}^{XY}$ and $\hat{\mathcal{C}}^{YX} > \hat{\mathcal{C}}^{XY}$, which is confirmed by numerical studies [34]. An alternative way for understanding the observed asymmetry of the interdependent network characteristics is illustrated in Fig. 4.14: moderate unidirectional coupling (below the onset of synchronization) increases the driven system's dimension [84, 110] (we will numerically demonstrate this behavior in Sect. 4.4.5), so that some former neighbors of pairs of recurrent states in X are not mutually close in Y anymore. In this case, the number of “cross-triangles” with baseline in Y decreases in comparison with those having their baseline in X . In fact, a corresponding decrease in $\hat{\mathcal{T}}^{XY}$ ($\hat{\mathcal{C}}^{XY}$) and an increase in $\hat{\mathcal{T}}^{YX}$ ($\hat{\mathcal{C}}^{YX}$) can often be observed in parallel (see Fig. 4.15).

Figure 4.15 shows the illustrative example of global cross-clustering coefficients for two unidirectionally coupled Rössler systems in the funnel regime with the same parameters a , b and c , but a weak detuning of $\nu = 0.02$, following the setting of [34]. The obtained results are consistent with our above heuristic explanation for the emergence of asymmetries between the interdependent network characteristics in the presence of unidirectional coupling. Specifically, for a wide range of moderate coupling strengths, the difference between the two global cross-clustering coefficients allows to correctly identify the direction of the imposed coupling. At large coupling strengths (i.e., close to and beyond the onset of generalized synchronization, which is indicated by the second largest Lyapunov exponent of

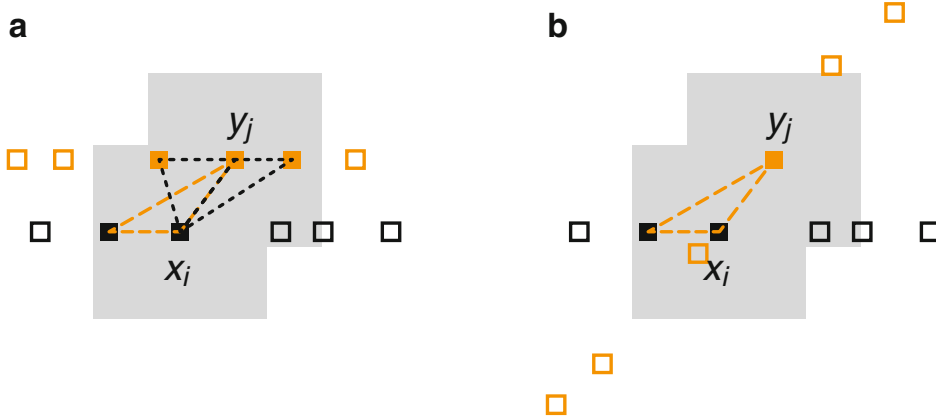


Fig. 4.14 Schematic illustration of the dimensionality and thus resulting mutual neighborhoods of state vectors x_i and y_j in the case of (a) uncoupled and (b) unidirectionally coupled systems. Shaded areas represent the neighborhoods of the respective state vectors, filled squares indicate recurrent states. In case (b), the coupling increases the driven system's dimension, so that formerly recurrent states are now outside of the neighborhood of y_j . Thus, the number of “cross-triangles” with baseline in Y (i.e., “from X to Y ”) decreases

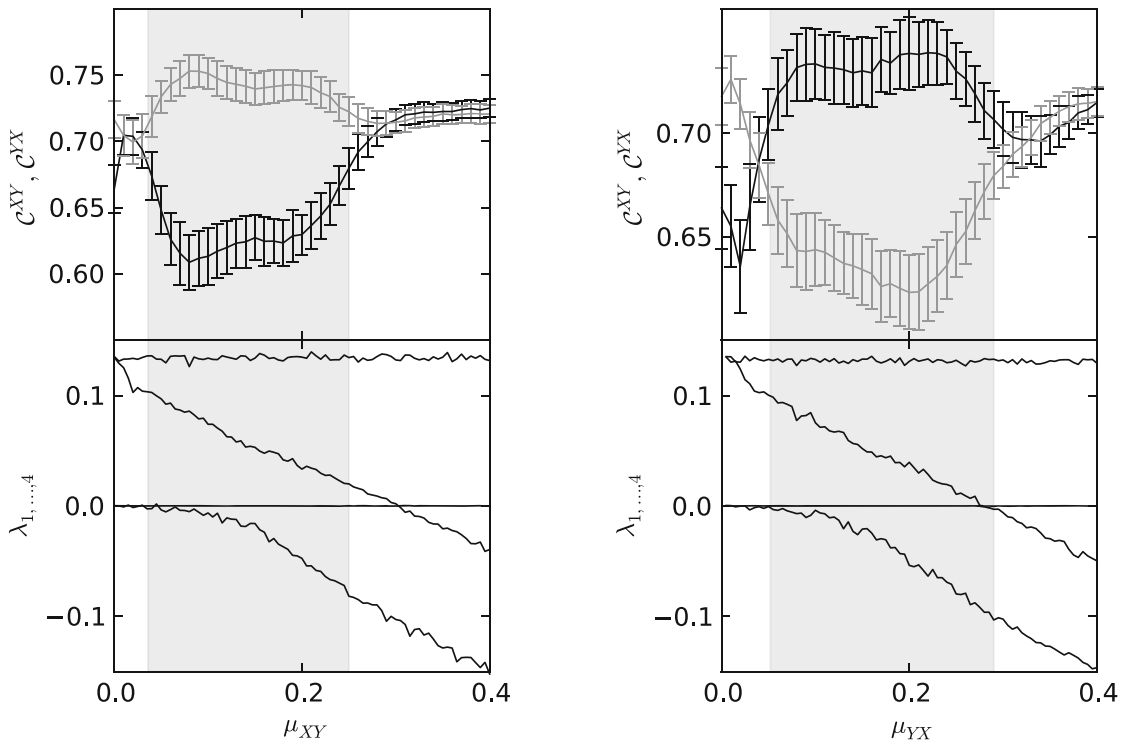


Fig. 4.15 Global cross-clustering coefficients \mathcal{C}^{XY} (black), \mathcal{C}^{YX} (gray) and the four largest Lyapunov exponents $\lambda_{1,\dots,4}$ estimated using the Wolf algorithm [100] for two identically but slightly detuned ($\nu = 0.02$) Rössler oscillators in the funnel regime subject to unidirectional coupling $X \rightarrow Y$ (left) and $Y \rightarrow X$ (right). The shaded regions mark the values of the coupling strength for which a correct identification of the coupling direction is achieved. Error bars represent mean values and standard deviations taken from an ensemble of 200 independent network realizations (with $N = 1,500$ data points per system)

the system approaching zero as shown in Fig. 4.15), both global cross-clustering coefficients become statistically indistinguishable, which is consistent with the fact that the behavior of the driven system is completely locked to the dynamics of the driver (cf. Sect. 4.4.5). In turn, the indistinguishability of both coupling directions at very low coupling strengths is most likely due to the fact that the geometric deformations of the driven system's attractor are too small to be detected for the given finite values of ε_X , ε_Y and ε_{XY} and the chosen network size. We expect that for larger IRNs and smaller distance thresholds, the lower boundary of the interval of coupling strengths for which the two global cross-clustering coefficients differ statistically significantly from each other will shift towards zero.

We emphasize that the same results can be obtained using the cross-transitivity replacing the global cross-clustering coefficient. Moreover, it is notable that the reported distinction can already be obtained at comparably small network sizes of some hundred vertices [34].

4.4.5 Joint Transitivity Properties and Synchronization

The concept of joint recurrence plots (JRPs) has been found very useful for studying the otherwise hard to detect emergence of generalized synchronization (GS) between two coupled chaotic systems X and Y [83]. GS describes the presence of a general functional relationship between the trajectories of both systems, $y(t) = f(x(t))$, which can arise at sufficiently large coupling strengths in both uni- and bidirectional coupling configurations. Most available methods for identifying GS from time series data have been developed for driver-response relationships, and only few approaches are also suitable for studying GS in the presence of symmetric couplings [35]. Among the latter, JRPs have recently attracted specific interest.

Romano et al. [83] argued that in case of GS, recurrences in the two coupled systems need to occur simultaneously (or with a given fixed time lag in the special case of lag synchronization, $y(t) = f(x(t - \tau))$). Hence, comparing the joint recurrence rate JRR with the recurrence rates of the individual single-system RPs (taken to be the same for both systems) should show convergence of both values. The latter fact is quantified in terms of the *joint probability of recurrence (JPR) index*

$$JPR = \max_{\tau} \frac{S(\tau) - RR}{1 - RR} \quad (4.74)$$

with the lagged joint recurrence rate ratio

$$S(\tau) = \frac{1}{N^2 RR} \sum_{i,j=1}^N \Theta(\varepsilon_X - \|x_i - x_j\|) \Theta(\varepsilon_Y - \|y_{i+\tau} - y_{j+\tau}\|) \quad (4.75)$$

and RR being the recurrence rate taken equal for both considered systems. Since for GS, we can expect that $S(\tau) \rightarrow 1$ for some τ , $JPR \rightarrow 1$. However, the latter measure has some disadvantages. On the one hand, testing for significance of a specific value of JPR usually requires complex surrogate data approaches for properly approximating the distribution of the underlying null hypothesis (no synchronization) adapted to the specific time series under study [94]. On the other hand, comparing the single-system and joint recurrence rates may be insufficient since due to the complexity of fluctuations or the presence of stochastic components (observational noise), we hardly ever capture all single-system recurrence in the JRP. Consequently, a solely RR -based characterization does not necessarily lead to the expected “optimum” value of the synchronization index ($JPR = 1$) in case of fully developed GS.

As an alternative, we have suggested that looking at higher-order characteristics (specifically, three-point instead of two-point relationships) may improve the results [35]. One convenient way is utilizing again the concept of transitivities from RN and JRN. The exploitation of alternative higher-order characteristics might be possible, but has not yet been explored. Notably, the specific requirements on the time series data render JRN a promising approach for detecting intricate interconnections between qualitatively distinct observables in observational or experimental real-world data.

As a heuristic indicator for the presence of GS, we have proposed using the *transitivity ratio* [35]

$$\hat{Q}_{\mathcal{T}} = \frac{\hat{\mathcal{T}}^J}{(\hat{\mathcal{T}}^X + \hat{\mathcal{T}}^Y)/2}, \quad (4.76)$$

i.e., the ratio between the JRN transitivity and the arithmetic mean of the single-system RN transitivities. The rationale behind this definition is that for systems exhibiting GS, all degrees of freedom are completely locked, implying that both approach the same effective (fractal) dimension and should thus have the same RN transitivities, which approximately equal the JRN transitivity. Alternatively, we could also use other means of $\hat{\mathcal{T}}^{X,Y}$, such as the geometric or harmonic means, for obtaining a meaningful ratio. Numerical experiments show that using the arithmetic mean provides values of $\hat{Q}_{\mathcal{T}}$ that are mostly confined to the interval $[0, 1]$ with only minor exceedances in the fully developed GS regime [35]. Since the arithmetic mean is always larger than the geometric one, normalizing with respect to the geometric mean $\sqrt{\hat{\mathcal{T}}^X \hat{\mathcal{T}}^Y}$ would lead to higher values of $\hat{Q}_{\mathcal{T}}$ and, hence, an even stronger violation of the desired normalization of the transitivity ratio. However, even when considering the normalization by the arithmetic mean of single-system RN transitivities, the thus defined transitivity ratio has two major drawbacks:

On the one hand, if the single-system RN transitivities are essentially different (a case that has not been studied in [35]), the contribution of the lower-dimensional system (higher transitivity) dominates the arithmetic mean in the denominator of Eq.(4.76) and, hence, the transitivity ratio itself irrespective of a possible well-defined driver-response relationship.

On the other hand, there is no rigorous theoretical justification for $\hat{Q}_{\mathcal{T}}$ being a good indicator of GS. Notably, the definition of the transitivity ratio is based on the idea that the transivities are related with the effective dimensions of the individual systems. In the uncoupled case, the degrees of freedom of both systems are independent; hence, the effective dimension of the composed system $X \otimes Y$ is expected to read $D^{X \otimes Y} = D^X + D^Y$ (notably, due to the logarithmic transform between RN transitivity and transitivity dimension, this additivity does *not* apply to the RN transivities). In turn, in case of GS, the degrees of freedom of both systems become mutually locked, probably leading to $D^{X \otimes Y} = D^X = D^Y$ (i.e., one system can be viewed as a—possibly nonlinear—projection of the other), with D^X and D^Y eventually differing from their values in the uncoupled case depending on the specific coupling configuration (e.g., uni- versus bidirectional coupling). Taking the estimated transitivity dimensions $\hat{D}_{\mathcal{T}^{X,Y}}$ as proxies for $D^{X,Y}$ and the *pseudo-dimension* $\hat{D}_{\mathcal{T}^J} = \log(\hat{\mathcal{T}}^J)/\log(3/4)$ as an approximation of the true dimension $D^{X \otimes Y}$ of the composed system $X \otimes Y$,⁹ the latter case would translate into $\hat{Q}_{\mathcal{T}} = 1$, which is approximately attained in numerical studies for coupled Rössler systems in different dynamical regimes [35].

In order to circumvent both problems, we suggest here utilizing an alternative indicator, which is directly based on the concept of effective dimensions (degrees of freedom) of the individual systems. In analogy with the mutual information (sometimes also called redundancy [76,78]) frequently used in nonlinear time series analysis, we define the *transitivity dimension redundancies*

$$\hat{D}_{\mathcal{T}^R} = \hat{D}_{\mathcal{T}^X} + \hat{D}_{\mathcal{T}^Y} - \hat{D}_{\mathcal{T}^J}, \quad (4.77)$$

$$\hat{D}_{\mathcal{T}^R} = \hat{D}_{\mathcal{T}^X} + \hat{D}_{\mathcal{T}^Y} - \hat{D}_{\mathcal{T}^{X \otimes Y}}, \quad (4.78)$$

which should assume zero values in the uncoupled case and exhibit $\hat{D}_{\mathcal{T}^X} = \hat{D}_{\mathcal{T}^Y} = \hat{D}_{\mathcal{T}^{X \otimes Y}} = \hat{D}_{\mathcal{T}^J}$ in case of GS. In order to obtain a normalized measure for the presence of GS, we define the *dimensional locking index (DLI)*

$$\widehat{DLI} = \frac{\hat{D}_{\mathcal{T}^R}}{\hat{D}_{\mathcal{T}^J}}, \quad (4.79)$$

$$\widehat{DLI} = \frac{\hat{D}_{\mathcal{T}^R}}{\hat{D}_{\mathcal{T}^{X \otimes Y}}}. \quad (4.80)$$

⁹In fact, we should take here the transitivity dimension of the RN obtained for $X \otimes Y$, i.e., $\hat{D}_{\mathcal{T}^{X \otimes Y}} = \log(\hat{\mathcal{T}}^{X \otimes Y})/\log(3/4)$, which is in general not identical to the pseudo-dimension $\hat{D}_{\mathcal{T}^J}$ due to the different metrics used for the definition of recurrences of $X \otimes Y$ and joint recurrences of X and Y .

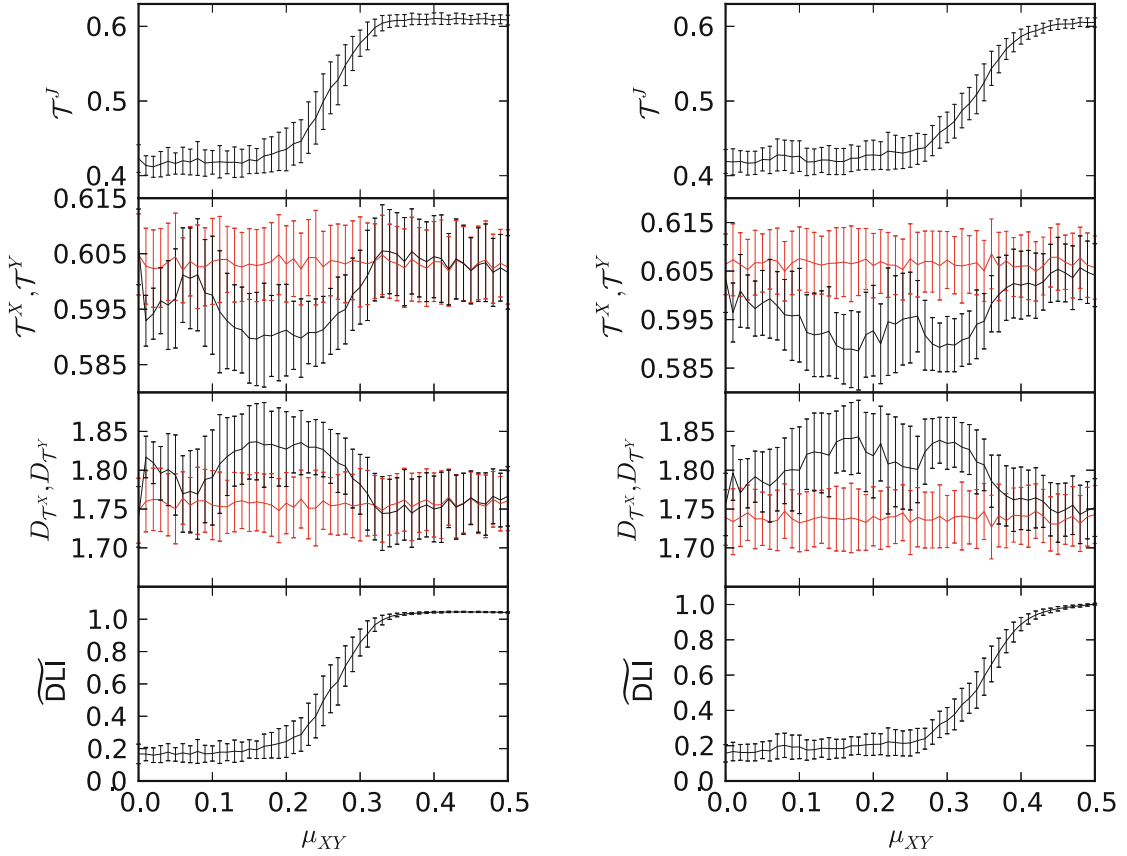


Fig. 4.16 Joint transitivity $\hat{\mathcal{T}}^J$, single-system RN transivities $\hat{\mathcal{T}}^{X,Y}$, corresponding transitivity dimensions $\hat{D}_{\mathcal{T}^X}$, $\hat{D}_{\mathcal{T}^Y}$ and derived dimensional locking index \widehat{DLI} (from *top to bottom*) for two unidirectionally coupled Rössler systems ($X \rightarrow Y$) in the funnel regime with $\nu = 0.02$ (driver oscillates faster than driven system, *left panels*) and $\nu = -0.02$ (*right panels*). The error bars indicate mean values and standard deviations estimated from 100 independent network realizations for each value of the coupling strength μ_{XY} . For transivities and transitivity dimensions, *red (black) lines* correspond to the values for system X (Y)

Notably, this index is tailored to the dimensionality interpretation of RN transitivity. In a strict sense, this argument only applies if using the single-system RN transitivity (dimension) of the composed system $X \rightarrow Y$ instead of the JRN transitivity (dimension) $\hat{D}_{\mathcal{T}^J}$. However, at this point, we suggest using the latter as an approximation. A detailed comparison between the two definitions will be subject to future research.

In order to further illustrate the behavior of the (J)RN-based characteristics for detecting the emergence of GS, we reconsider the example of two unidirectionally coupled identical but slightly detuned Rössler systems from Sect. 4.4.4. In contrast to [35], who studied different settings for uni- and bidirectional configurations with single realizations of the same system, we present here results obtained from ensembles of realizations. The results shown in Fig. 4.16 demonstrate that the estimated values of \mathcal{T}^J and \widehat{DLI} exhibit a marked increase at the onset of GS. Specifically, the new *DLI* index approaches one (with little overshooting) in the

synchronized regime as expected, but takes values of only about 0.2 or lower in the non-synchronous case (in comparison with values of about 0.7 exhibited by $Q_{\mathcal{T}}$, cf. Fig. 2b in [35]).

As a second important observation, we find a systematic and significant decrease in the RN transitivity of the driven system at moderate coupling strengths before the onset of GS, which corresponds to an increase of the associated transitivity dimension. This behavior is precisely what was claimed in the context of coupling analysis in Sect. 4.4.4 for providing an explanation of the numerically observed asymmetry between the transitivity-based interdependent network characteristics. These results underline that some integrated utilization of single-system, inter-system and joint recurrence networks can eventually provide deep insights into the coupling regime and strength from bivariate observations.

4.4.6 Real-World Applications

Although much recent work on RNs and multivariate generalizations thereof has been focused on the development of the theoretical framework and its numerical exploration using simple low-dimensional model systems, there have already been some first successful applications to characterizing system's properties from experimental or observational time series.

4.4.6.1 Applications in Climatology

One important field of recent applications is paleoclimatology, which has already been taken as an illustrative example in the seminal paper by Marwan et al. [66]. The corresponding study was later extended to some systematic investigation of the temporal variability profile of RN-based complexity measures for three marine sediment records of terrigenous dust flux off Africa during the last 5 million years. Donges et al. [15] argued that RNs can be used for characterizing dynamics from non-uniformly sampled or age-uncertain data, since this methodological approach does not make explicit use of time information. In turn, due to the necessity of using time-delay embedding, there is implicit time information entering the analysis, which has been recognized but widely neglected in previous works. Notably, disregarding age uncertainty and sampling heterogeneity appears a reasonable approximation only in cases where the distribution of instantaneous sampling rates remains acceptably narrow.

The results of Donges et al. [16] pointed to the existence of spatially coherent changes in the long-term variability of environmental conditions over Africa, which have probably influenced the evolution of human ancestor species. Specifically, RN transitivity and average path length have been interpreted as indicators for “climate regularity” (i.e., the complexity of fluctuations as captured by the transitivity dimensions) and “abrupt dynamical changes”, respectively. By identifying three

time intervals with consistent changes of the RN properties obtained from spatially widely separated records, it has been possible to attribute the corresponding long-term changes in the dynamics to periods characterized by known or speculated mechanisms for large-scale climate shifts such as changes in the Indian ocean circulation patterns, the intensification of the atmospheric Walker circulation, or changes in the dominant periodicity of Northern hemispheric glacial cycles. Moreover, Donges et al. [15] demonstrated a good robustness of the results of RN analysis obtained in a sliding windows framework when varying the corresponding parameters (e.g., window size or embedding delay) over a reasonable range.

As another methodological step towards better understanding climatic mechanisms, the authors have used two speleothem records for studying interdependencies between the two main branches of the Asian summer monsoon (the Indian and East Asian summer monsoon) by means of IRNs [34]. For this purpose, they selected two data sets of oxygen isotope anomalies from speleothems obtained from two caves in China and the Oman, respectively, which can be considered as proxies for the annual precipitation and, hence, the overall strength of the two monsoon branches over the last about 10,000 years. The asymmetries of the IRN cross-transitivities and global cross-clustering coefficients provided clear evidence for a marked influence of the Indian summer monsoon on the East Asian branch rather than vice versa, which is in good agreement with existing climatological theories. As a subsequent extension of this work, we emphasize the possibility of repeating the same kind of analysis in a sliding windows framework, thereby gaining information on possible temporal changes of the associated climatic patterns during certain time periods as recently revealed using correlation-based complex network analysis applied to a larger set of paleoclimate records from the Asian monsoon domain [81].

In order to characterize dynamical complexity associated with more recent environmental variability, Lange and Böse [6,54] used RQA as well as RN analysis for studying global photosynthetic activity from remote sensing data in conjunction with global precipitation patterns. Specifically, they studied 14-years long time series (1998–2011) of the fraction of absorbed photosynthetically active radiation (faPAR) with a spatial resolution of 0.5° around the Earth and a temporal sampling of about 10 days, providing time series of $N = 504$ data points. Their results revealed very interesting spatial complexity patterns, which have been largely, but not exclusively determined by the amplitude of the annual cycle of vegetation growth in different ecosystems.

4.4.6.2 Applications in Fluid Dynamics

In a series of papers, Gao et al. investigated the emerging complexity of dynamical patterns in two-phase gas–liquid or oil–water flows in different configurations using RN techniques. In general, multiple sensors measuring fluctuations of electrical conductance have been used for obtaining signals that are characteristic for the different flow patterns. For gas–liquid two-phase upward flows in vertical pipes, different types of complex networks generated from observational data have been

proposed, among which the degree correlations (assortativity) of RNs was proven to be particularly useful for distinguishing between qualitatively different flow types [38]. For oil–water two-phase upward flows in a similar configuration, the global clustering coefficient of RNs reveals a marked increase in dynamical complexity (detectable in terms of a decreasing $\hat{\mathcal{C}}$) as the flow pattern changes from slug flow over coarse to very finely dispersed bubble flow [40]. In case of oil–water two-phase flows in inclined pipes [39], the motif distributions of RNs (specifically, the frequency distributions of small subgraphs containing exactly four vertices) revealed an increasing degree of heterogeneity, where the motif ranking was conserved in all experimental conditions, whereas the absolute motif frequency dramatically changed. The corresponding results were independently confirmed using some classical measures of complexity, which indicated increasing complexity in conjunction with increasing heterogeneity of the RN motif distributions. Finally, for characterizing horizontal oil–water flows [41], RN and IRN analysis were combined for studying conductance signals from multiple sensors. Specifically, cross-transitivity was found a useful measure for tracing the transitions between stable stratified and unstable states associated with the formation of droplets.

4.4.6.3 Applications in Electrochemistry

Zou et al. [112] studied the complexity of experimental electrochemical oscillations as one control parameter of the experiments (temperature) was systematically varied. By utilizing a multitude of complementary RN characteristics, they could demonstrate a systematic rise in dynamical complexity as temperature increased, but an absence of a previously speculated phase transition [98] separating phase-coherent from noncoherent chaotic oscillations. The latter results were independently confirmed using other classical indicators for phase coherence, as well as studies of a corresponding mathematical model of the specific electrochemical processes.

4.4.6.4 Applications in Medicine

Finally, there have been a couple of successful applications in a medical context. Marwan et al. [67] demonstrated that the global clustering coefficients of RNs obtained from heartbeat intervals, diastolic and systolic blood pressure allowed a reliable identification of patients with pre-eclampsia, a cardiovascular disease during pregnancy with a high risk of fetal and maternal morbidity. Their results were further improved by Ramírez et al. [79,80] who considered combinations of various RN-based network characteristics. In a similar spirit as for cardiovascular diseases, recent results point to the capability of RN characteristics for discriminating between the EEG signals of healthy and epileptic patients [90].

4.5 Related Approaches

Besides the recurrence definition based on a fixed distance threshold ε in phase space (i.e., equal neighborhood volumes around all available state vectors), there are alternative ways for defining recurrences and, hence, RPs and RNs. For example, the original definition of a RP by Eckmann et al. [30] makes use of k -nearest neighbors (i.e., a fixed probability mass of the considered neighborhoods). Re-interpreting the resulting recurrence matrix as the adjacency matrix of a complex network leads to a different type of RN [88], typically referred to as *k-nearest neighbor network* [27]. Since in this definition, the neighborhood relation is not symmetric (i.e., x_j being among the k nearest neighbors of x_i does not imply x_i also being among the k nearest neighbors of x_j), the resulting networks are in general directed graphs, and the local density of unidirectional edges (as opposed to bidirectional ones) is related to the gradient of the invariant density.

In order to circumvent the directedness of k -nearest neighbor networks, Xu et al. [89, 101] proposed an algorithm for balancing the neighborhood relationships in such a way that they become symmetric again. The resulting networks embedded in phase space, sometimes also referred to as *adaptive nearest neighbor networks* [27], are conceptually more similar to classical (ε -)RNs, but still exhibit somewhat different topological characteristics. In particular, the motif distribution of adaptive nearest neighbor networks has been shown to allow a discrimination between different types of dynamics in terms of a different motif ranking [61, 101]. Consequently, this approach has been mainly used for such discriminatory tasks, including applications to turbulence phenomena, stock markets [61] or instrumental music [27].

For a detailed discussion of the differences between ε -RNs, k -nearest neighbor and adaptive nearest neighbor networks, we refer to [27]. While these three classes of time series networks exhibit very strong conceptual similarities (the same applies to correlation networks [102] if interpreting the correlation coefficient between two sufficiently high-dimensional state vectors as a generalized distance), the approach proposed by Li et al. [57, 58] can be understood as being derived from the RN idea. Here, for a set of m -dimensional embedding vectors, all mutual Euclidean distances are computed. Based on the maximum point-wise Euclidean distance $d_{\max}(m)$, the threshold distance of a RN is taken as $\varepsilon(m) = d_{\max}(m)/(N - 1)$. This procedure is repeated for different m , and the critical value of the embedding dimension for which the resulting network gets completely disconnected is interpreted as a complexity index [8]. However, it has not yet been demonstrated that this algorithmic approach has any conceptual benefits in comparison with the classical RN transitivity obtained for a fixed embedding dimension.

Another conceptual approach loosely related to RNs provides the foundation of the frequency-degree mapping algorithm introduced by Li et al. [59]. Here, the resulting time series networks contain two types of edges: (1) temporal edges connecting subsequent points in time, and (2) proximity edges containing observations of similar values, where similarity is defined by an initial grouping of the data

into a discrete set of classes, and observed values being connected if and only if they belong to the same class. Notably, the latter approach combines the classical recurrence idea and basic concepts of symbolic dynamics [11]. In this spirit, the resulting network's adjacency matrix is given as the recurrence matrix associated with a symbolic recurrence plot [4, 22, 32] plus a “stripe” around the matrix' main diagonal. The frequency-degree mapping algorithm has been successfully applied to characterizing signatures of various types of ventricular arrhythmias in human heart beat time series [59].

4.6 Summary

This work has provided an overview on the current state of complex network approaches to quantifying nonlinear dynamics based on recurrence plots. This emerging field has already proven its great potentials by addressing methodological questions common to various fields of research, such as the characterization of complexity, the identification of dynamical transitions in non-stationary time series, and the detection of asymmetric coupling and generalized synchronization from bivariate records. In all these areas, recurrence networks and multivariate generalizations thereof have been successfully applied to model systems and at least partially also to real-world observational or experimental data. The results available so far suggest reasonable numerical properties and, hence, a wide applicability of this approach. Moreover, the quantitative analysis of recurrence networks is supported by a rigorous analytical framework derived from the theory of random geometric graphs. In turn, unlike other recurrence plot-based analysis concepts, recurrence networks characterize geometric properties of the system under study rather than dynamical ones. However, since geometry and dynamics are known to be closely interrelated, the network-based approach is still useful for obtaining information on dynamical complexity and related aspects as reflected in the system's structural features.

Beyond exclusively reviewing recent results, we have also discussed some new aspects not yet found in the previous peer-reviewed literature: (1) an analytical framework for inter-system recurrence networks providing a possible foundation for a future analytical understanding of geometric signatures of coupling between dynamical systems, (2) a straightforward extension of joint recurrence plots and joint recurrence networks to a less restrictive definition, which could allow studying synchronization processes between noisy systems based on recurrence plots, and (3) the introduction of a new index for generalized synchronization, which is expected to yield a better performance than other recent recurrence-based indices. The latter three aspects open new perspectives for both an improved theoretical understanding of geometric effects of coupling and synchronization and applications in various contexts. We outline corresponding further investigations as an important research avenue for our future work.

Acknowledgements The reported development of the recurrence network framework has been a community effort. Among other colleagues, we particularly acknowledge important contributions by Jobst Heitzig and Norbert Marwan, as well as multiple inspiring discussions with Jürgen Kurths. Financial support of this work has been granted by the German Federal Environmental Agency, the European Union Seventh Framework Program, the Max Planck Society, the Stordalen Foundation, the German Research Association via the IRTG 1740 “Dynamical phenomena in complex networks”, the National Natural Science Foundation of China (Grant No. 11305062, 11135001), and the German Federal Ministry for Science and Education (project CoSy-CC², grant no. 01LN1306A). Numerical codes used for estimating recurrence network properties can be found in the software package pyunicorn [20], which is available at <http://tocsy.pik-potsdam.de/pyunicorn.php>.

References

1. R. Albert, A.L. Barabasi, Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**(1), 47–97 (2002). doi:[10.1103/RevModPhys.74.47](https://doi.org/10.1103/RevModPhys.74.47)
2. M.J. Barber, Modularity and community detection in bipartite networks. *Phys. Rev. E* **76**(6), 066102 (2007). doi:[10.1103/PhysRevE.76.066102](https://doi.org/10.1103/PhysRevE.76.066102)
3. A. Barrat, M. Weigt, On the properties of small-world network models. *Eur. Phys. J. B* **13**, 547–560 (2000). doi:[10.1007/s100510050067](https://doi.org/10.1007/s100510050067)
4. P. beim Graben, A. Hutt, Detecting recurrence domains of dynamical systems by symbolic dynamics. *Phys. Rev. Lett.* **110**, 154101 (2013). doi:[10.1103/PhysRevLett.110.154101](https://doi.org/10.1103/PhysRevLett.110.154101)
5. S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.U. Hwang, Complex networks: structure and dynamics. *Phys. Rep.* **424**(4–5), 175–308 (2006). doi:[10.1016/j.physrep.2005.10.009](https://doi.org/10.1016/j.physrep.2005.10.009)
6. S. Böse, Recurrence network analysis of remote sensing data. Master’s thesis, University of Bayreuth (2012)
7. S.V. Buldyrev, R. Parshani, G. Paul, H.E. Stanley, S. Havlin, Catastrophic cascade of failures in interdependent networks. *Nature* **464**(7291), 1025–1028 (2010). doi:[10.1038/nature08932](https://doi.org/10.1038/nature08932)
8. H. Cao, Y. Li, Unraveling chaotic attractors by complex networks to measure the complexity of stock markets. *Chaos*. **24**(1), 013134 (2014). doi:[10.1063/1.4868258](https://doi.org/10.1063/1.4868258)
9. L. da F. Costa, F.A. Rodrigues, G. Traverso, P.R. Villas Boas: Characterization of complex networks: a survey of measurements. *Adv. Phys.* **56**(1), 167–242 (2007). doi:[10.1080/00018730601170527](https://doi.org/10.1080/00018730601170527)
10. J. Dall, M. Christensen, Random geometric graphs. *Phys. Rev. E* **66**(1), 016121 (2002). doi:[10.1103/PhysRevE.66.016121](https://doi.org/10.1103/PhysRevE.66.016121)
11. C.S. Daw, C.E.A. Finney, E.R. Tracy, A review of symbolic analysis of experimental data. *Rev. Sci. Instrum.* **74**(2), 915–930 (2003). doi:[10.1063/1.1531823](https://doi.org/10.1063/1.1531823)
12. J. Donges, Functional network macroscopes for probing past and present earth system dynamics: complex hierarchical interactions, tipping points, and beyond. Ph.D. thesis, Humboldt University, Berlin, 2012
13. J.F. Donges, Y. Zou, N. Marwan, J. Kurths, The backbone of the climate network. *Europhys. Lett.* **87**(4), 48007 (2009). doi:[10.1209/0295-5075/87/48007](https://doi.org/10.1209/0295-5075/87/48007)
14. J.F. Donges, Y. Zou, N. Marwan, J. Kurths, Complex networks in climate dynamics: comparing linear and nonlinear network construction methods. *Eur. Phys. J. Spec. Top.* **174**, 157–179 (2009). doi:[10.1140/epjst/e2009-01098-2](https://doi.org/10.1140/epjst/e2009-01098-2)
15. J.F. Donges, R.V. Donner, K. Rehfeld, N. Marwan, M.H. Trauth, J. Kurths, Identification of dynamical transitions in marine palaeoclimate records by recurrence network analysis. *Nonlinear Process. Geophys.* **18**(5), 545–562 (2011). doi:[10.5194/np-18-545-2011](https://doi.org/10.5194/np-18-545-2011)
16. J.F. Donges, R.V. Donner, M.H. Trauth, N. Marwan, H.J. Schellnhuber, J. Kurths, Nonlinear detection of paleoclimate-variability transitions possibly related to human evolution. *Proc. Natl. Acad. Sci. USA* **108**, 20422–20427 (2011). doi:[10.1073/pnas.1117052108](https://doi.org/10.1073/pnas.1117052108)

17. J.F. Donges, H.C.H. Schultz, N. Marwan, Y. Zou, J. Kurths, Investigating the topology of interacting networks. *Eur. Phys. J. B* **84**(4), 635–652 (2011). doi:[10.1140/epjb/e2011-10795-8](https://doi.org/10.1140/epjb/e2011-10795-8)
18. J.F. Donges, J. Heitzig, R.V. Donner, J. Kurths, Analytical framework for recurrence network analysis of time series. *Phys. Rev. E* **85**, 046105 (2012). doi [10.1103/PhysRevE.85.046105](https://doi.org/10.1103/PhysRevE.85.046105)
19. J.F. Donges, R.V. Donner, J. Kurths, Testing time series irreversibility using complex network methods. *Europhys. Lett.* **102**(1), 10004 (2013). doi:[10.1209/0295-5075/102/10004](https://doi.org/10.1209/0295-5075/102/10004)
20. J.F. Donges, J. Heitzig, J. Runge, H.C.H. Schultz, M. Wiedermann, A. Zech, J. Feldhoff, A. Rheinwalt, H. Kutza, A. Radebach, et al.: Advanced functional network analysis in the geosciences: the pyunicorn package. *Geophys. Res. Abstr.* **15**, 3558 (2013)
21. R.V. Donner, J.F. Donges, Visibility graph analysis of geophysical time series: potentials and possible pitfalls. *Acta Geophysica* **60**(3), 589–623 (2012). doi:[10.2478/s11600-012-0032-x](https://doi.org/10.2478/s11600-012-0032-x)
22. R. Donner, U. Hinrichs, B. Scholz-Reiter, Symbolic recurrence plots: a new quantitative framework for performance analysis of manufacturing networks. *Eur. Phys. J. Spec. Top.* **164**, 85–104 (2008). doi:[10.1140/epjst/e2008-00836-2](https://doi.org/10.1140/epjst/e2008-00836-2)
23. R.V. Donner, J.F. Donges, Y. Zou, N. Marwan, J. Kurths, Recurrence-based evolving networks for time series analysis of complex systems, in *Proceedings of the International Symposium on Nonlinear Theory and its Applications (NOLTA 2010)* (2010), pp. 87–90
24. R.V. Donner, Y. Zou, J.F. Donges, N. Marwan, J. Kurths, Ambiguities in recurrence-based complex network representations of time series. *Phys. Rev. E* **81**(1), 015101(R) (2010). doi:[10.1103/PhysRevE.81.015101](https://doi.org/10.1103/PhysRevE.81.015101)
25. R.V. Donner, Y. Zou, J.F. Donges, N. Marwan, J. Kurths, Recurrence networks: a novel paradigm for nonlinear time series analysis. *New J. Phys.* **12**(3), 033025 (2010). doi:[10.1088/1367-2630/12/3/033025](https://doi.org/10.1088/1367-2630/12/3/033025)
26. R.V. Donner, J. Heitzig, J.F. Donges, Y. Zou, N. Marwan, J. Kurths, The geometry of chaotic dynamics: a complex network perspective. *Eur. Phys. J. B* **84**(4), 653–672 (2011). doi:[10.1140/epjb/e2011-10899-1](https://doi.org/10.1140/epjb/e2011-10899-1)
27. R.V. Donner, M. Small, J.F. Donges, N. Marwan, Y. Zou, R. Xiang, J. Kurths, Recurrence-based time series analysis by means of complex network methods. *Int. J. Bifurcat. Chaos* **21**(4), 1019–1046 (2011). doi:[10.1142/S0218127411029021](https://doi.org/10.1142/S0218127411029021)
28. N. Du, B. Wang, B. Wu, Y. Wang, Overlapping community detection in bipartite networks, in *Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, vol. 01 (IEEE Computer Society, Washington, DC, 2008), pp. 176–179. doi:[10.1109/WIIAT.2008.98](https://doi.org/10.1109/WIIAT.2008.98)
29. J.P. Eckmann, D. Ruelle, Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys.* **57**, 617–656 (1985). doi:[10.1103/RevModPhys.57.617](https://doi.org/10.1103/RevModPhys.57.617)
30. J.P. Eckmann, S.O. Kamphorst, D. Ruelle, Recurrence plots of dynamical systems. *Europhys. Lett.* **4**(9), 973–977 (1987). doi:[10.1209/0295-5075/4/9/004](https://doi.org/10.1209/0295-5075/4/9/004)
31. P. Faure, H. Korn, A new method to estimate the Kolmogorov entropy from recurrence plots: its application to neuronal signals. *Physica D* **122**(1–4), 265–279 (1998). doi:[10.1016/S0167-2789\(98\)00177-8](https://doi.org/10.1016/S0167-2789(98)00177-8)
32. P. Faure, A. Lesne, Recurrence plots for symbolic sequences. *Int. J. Bifurcat. Chaos* **20**(6), 1731–1749 (2010). doi:[10.1142/S0218127410026794](https://doi.org/10.1142/S0218127410026794)
33. J. Feldhoff, Multivariate extensions of recurrence network analysis. Master’s thesis, Humboldt University, Berlin (2011)
34. J.H. Feldhoff, R.V. Donner, J.F. Donges, N. Marwan, J. Kurths, Geometric detection of coupling directions by means of inter-system recurrence networks. *Phys. Lett. A* **376**(46), 3504–3513 (2012). doi:[10.1016/j.physleta.2012.10.008](https://doi.org/10.1016/j.physleta.2012.10.008)
35. J.H. Feldhoff, R.V. Donner, J.F. Donges, N. Marwan, J. Kurths, Geometric signature of complex synchronisation scenarios. *Europhys. Lett.* **102**(3), 30007 (2013). doi:[10.1209/0295-5075/102/30007](https://doi.org/10.1209/0295-5075/102/30007)
36. S. Fortunato, Community detection in graphs. *Phys. Rep.* **486**(3–5), 75–174 (2010). doi:[10.1016/j.physrep.2009.11.002](https://doi.org/10.1016/j.physrep.2009.11.002)

37. A.M. Fraser, H.L. Swinney, Independent coordinates for strange attractors from mutual information. *Phys. Rev. A* **33**(2), 1134–1140 (1986). doi:[10.1103/PhysRevA.33.1134](https://doi.org/10.1103/PhysRevA.33.1134)
38. Z. Gao, N. Jin, Flow-pattern identification and nonlinear dynamics of gas-liquid two-phase flow in complex networks. *Phys. Rev. E* **79**(6), 066303 (2009). doi:[10.1103/PhysRevE.79.066303](https://doi.org/10.1103/PhysRevE.79.066303)
39. Z.K. Gao, N.D. Jin, W.X. Wang, Y.C. Lai, Motif distributions in phase-space networks for characterizing experimental two-phase flow patterns with chaotic features. *Phys. Rev. E* **82**, 016210 (2010). doi:[10.1103/PhysRevE.82.016210](https://doi.org/10.1103/PhysRevE.82.016210)
40. Z.K. Gao, X.W. Zhang, M. Du, D.D. Jin, Recurrence network analysis of experimental signals from bubbly oil-in-water flows. *Phys. Lett. A* **377**, 457–462 (2013). doi:[10.1016/j.physleta.2012.12.017](https://doi.org/10.1016/j.physleta.2012.12.017)
41. Z.K. Gao, X.W. Zhang, D.D. Jin, R. Donner, N. Marwan, J. Kurths, Recurrence networks from multivariate signals for uncovering dynamic transitions of horizontal oil-water stratified flows. *Europhys. Lett.* **103**(5), 50004 (2013). doi:[10.1209/0295-5075/103/50004](https://doi.org/10.1209/0295-5075/103/50004)
42. P. Grassberger, Generalized dimensions of strange attractors. *Phys. Lett. A* **97**(6), 227–230 (1983). doi:[10.1016/0375-9601\(83\)90753-3](https://doi.org/10.1016/0375-9601(83)90753-3)
43. P. Grassberger, I. Procaccia, Characterization of strange attractors. *Phys. Rev. Lett.* **50**(5), 346–349 (1983). doi:[10.1103/PhysRevLett.50.346](https://doi.org/10.1103/PhysRevLett.50.346)
44. J.L. Guillaume, M. Latapy, Bipartite structure of all complex networks. *Inf. Process. Lett.* **90**(5), 215–221 (2004). doi:[10.1016/j.ipl.2004.03.007](https://doi.org/10.1016/j.ipl.2004.03.007)
45. J.L. Guillaume, M. Latapy, Bipartite graphs as models of complex networks. *Physica A* **371**(2), 795–813 (2006). doi:[10.1016/j.physa.2006.04.047](https://doi.org/10.1016/j.physa.2006.04.047)
46. R. Guimerà, M. Sales-Pardo, L.A.N. Amaral, Module identification in bipartite and directed networks. *Phys. Rev. E* **76**(3), 036102 (2007). doi:[10.1103/PhysRevE.76.036102](https://doi.org/10.1103/PhysRevE.76.036102)
47. C. Herrmann, M. Barthélemy, P. Provero, Connectivity distribution of spatial networks. *Phys. Rev. E* **68**(2), 026128 (2003). doi:[10.1103/PhysRevE.68.026128](https://doi.org/10.1103/PhysRevE.68.026128)
48. H. Kantz, T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, 1997)
49. M.B. Kennel, R. Brown, H.D.I. Abarbanel, Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys. Rev. A* **45**(6), 3403–3411 (1992). doi:[10.1103/PhysRevA.45.3403](https://doi.org/10.1103/PhysRevA.45.3403)
50. M. Kitsak, D. Krioukov, Hidden variables in bipartite networks. *Phys. Rev. E* **84**, 026114 (2011). doi:[10.1103/PhysRevE.84.026114](https://doi.org/10.1103/PhysRevE.84.026114)
51. L. Lacasa, B. Luque, F. Ballesteros, J. Luque, J.C. Nuno, From time series to complex networks: the visibility graph. *Proc. Natl. Acad. Sci. USA* **105**(13), 4972–4975 (2008). doi:[10.1073/pnas.0709247105](https://doi.org/10.1073/pnas.0709247105)
52. L. Lacasa, B. Luque, J. Luque, J.C. Nuno, The visibility graph: a new method for estimating the Hurst exponent of fractional Brownian motion. *Europhys. Lett.* **86**(3), 30001 (2009). doi:[10.1209/0295-5075/86/30001](https://doi.org/10.1209/0295-5075/86/30001)
53. L. Lacasa, A. Nuñez, E. Roldán, J.M.R. Parrondo, B. Luque, Time series irreversibility: a visibility graph approach. *Eur. Phys. J. B* **85**, 217 (2012). doi:[10.1040/epjb/e2012-20809-8](https://doi.org/10.1040/epjb/e2012-20809-8)
54. H. Lange, S. Böse, Recurrence quantification and recurrence network analysis of global photosynthetic activity, in *Recurrence Quantification Analysis: Theory and Best Practices*, ed. by C.L. Webber, N. Marwan (Springer, Berlin, 2014, Chap. 12 of this volume)
55. D.P. Lathrop, E.J. Kostelich, Characterization of an experimental strange attractor by periodic-orbits. *Phys. Rev. A* **40**(7), 4028–4031 (1989). doi:[10.1103/PhysRevA.40.4028](https://doi.org/10.1103/PhysRevA.40.4028)
56. S. Lehmann, M. Schwartz, L.K. Hansen, Biclique communities. *Phys. Rev. E* **78**(1), 016108 (2008). doi:[10.1103/PhysRevE.78.016108](https://doi.org/10.1103/PhysRevE.78.016108)
57. Y. Li, H. Cao, Y. Tan, A comparison of two methods for modeling large-scale data from time series as complex networks. *AIP Adv.* **1**, 012103 (2011). doi:[10.1063/1.3556121](https://doi.org/10.1063/1.3556121)
58. Y. Li, H. Cao, Y. Tan, Novel method of identifying time series based on network graphs. *Complexity* **17**, 13–34 (2011). doi:[10.1002/cplx.20384](https://doi.org/10.1002/cplx.20384)

59. X. Li, D. Yang, X. Liu, X.M. Wu, Bridging time series dynamics and complex network theory with application to electrocardiogram analysis. *IEEE Circuits Syst. Mag.* **12**(4), 33–46 (2012). doi:[10.1109/MCAS.2012.2221521](https://doi.org/10.1109/MCAS.2012.2221521)
60. P.G. Lind, M.C. González, H.J. Herrmann, Cycles and clustering in bipartite networks. *Phys. Rev. E* **72**(5), 056127 (2005). doi:[10.1103/PhysRevE.72.056127](https://doi.org/10.1103/PhysRevE.72.056127)
61. C. Liu, W.X. Zhou, Superfamily classification of nonstationary time series based on DFA scaling exponents. *J. Phys. A* **43**, 495005 (2009). doi:[10.1088/1751-8113/43/49/495005](https://doi.org/10.1088/1751-8113/43/49/495005)
62. B. Luque, L. Lacasa, F. Ballesteros, J. Luque, Horizontal visibility graphs: exact results for random time series. *Phys. Rev. E* **80**(4), 046103 (2009). doi:[10.1103/PhysRevE.80.046103](https://doi.org/10.1103/PhysRevE.80.046103)
63. N. Marwan, J. Kurths, Nonlinear analysis of bivariate data with cross recurrence plots. *Phys. Lett. A* **302**(56), 299–307 (2002). doi:[10.1016/S0375-9601\(02\)01170-2](https://doi.org/10.1016/S0375-9601(02)01170-2)
64. N. Marwan, N. Wessel, U. Meyerfeldt, A. Schirdewan, J. Kurths, Recurrence plot based measures of complexity and its application to heart rate variability data. *Phys. Rev. E* **66**(2), 026702 (2002). doi:[10.1103/PhysRevE.66.026702](https://doi.org/10.1103/PhysRevE.66.026702)
65. N. Marwan, M.C. Romano, M. Thiel, J. Kurths, Recurrence plots for the analysis of complex systems. *Phys. Rep.* **438**(5–6), 237–329 (2007). doi:[10.1016/j.physrep.2006.11.001](https://doi.org/10.1016/j.physrep.2006.11.001)
66. N. Marwan, J.F. Donges, Y. Zou, R.V. Donner, J. Kurths, Complex network approach for recurrence analysis of time series. *Phys. Lett. A* **373**(46), 4246–4254 (2009). doi:[10.1016/j.physleta.2009.09.042](https://doi.org/10.1016/j.physleta.2009.09.042)
67. N. Marwan, N. Wessel, J. Kurths, Recurrence based complex network analysis of cardiovascular variability data to predict pre-eclampsia, in *Proceedings of Biosignals 2010*, 2010, 022
68. S. Milgram, Small-world problem. *Psychol. Today* **1**(1), 61–67 (1967)
69. T. Murata, Detecting communities from bipartite networks based on bipartite modularities, in *Proceedings of the 2009 International Conference on Computational Science and Engineering*, vol. 04 (IEEE Computer Society, Washington, DC, 2009), pp. 50–57. doi:[10.1109/CSE.2009.81](https://doi.org/10.1109/CSE.2009.81)
70. M.E.J. Newman, The structure and function of complex networks. *SIAM Rev.* **45**(2), 167–256 (2003). doi:[10.1137/S003614450342480](https://doi.org/10.1137/S003614450342480)
71. M.E.J. Newman, Detecting community structure in networks. *Eur. Phys. J. B* **38**(2), 321–330 (2004). doi:[10.1140/epjb/e2004-00124-y](https://doi.org/10.1140/epjb/e2004-00124-y)
72. X.H. Ni, Z.Q. Jiang, W.X. Zhou, Degree distributions of the visibility graphs mapped from fractional Brownian motions and multifractal random walks. *Phys. Lett. A* **373**(42), 3822–3826 (2009). doi:[10.1016/j.physleta.2009.08.041](https://doi.org/10.1016/j.physleta.2009.08.041)
73. G. Nicolis, A. García Cantú, C. Nicolis, Dynamical aspects of interaction networks. *Int. J. Bifurcat. Chaos* **15**(11), 3467–3480 (2005). doi:[10.1142/S0218127405014167](https://doi.org/10.1142/S0218127405014167)
74. E.M. Obloy, Supertracks, supertrack functions and chaos in the quadratic map. *Phys. Lett. A* **128**(8), 406–412 (1988). doi:[10.1016/0375-9601\(88\)90119-3](https://doi.org/10.1016/0375-9601(88)90119-3)
75. E. Ott, *Chaos in Dynamical Systems*, 2nd edn. (Cambridge University Press, Cambridge, 2002)
76. M. Paluš, Testing for nonlinearity using redundancies: quantitative and qualitative aspects. *Physica D* **80**, 186–205 (1995). doi:[10.1016/0167-2789\(95\)90079-9](https://doi.org/10.1016/0167-2789(95)90079-9)
77. M. Penrose, *Random Geometric Graphs*, (Oxford University Press, Oxford, 2003)
78. D. Prichard, J. Theiler, Generalized redundancies for time series analysis. *Physica D* **84**, 476–493 (1995). doi:[10.1016/0167-2789\(95\)00041-2](https://doi.org/10.1016/0167-2789(95)00041-2)
79. G. Ramírez Ávila, A. Gapelyuk, N. Marwan, T. Walther, H. Stepan, J. Kurths, N. Wessel, Classification of cardiovascular time series based on different coupling structures using recurrence networks analysis. *Phil. Trans. P. Soc. A* **371**, 20110623 (2013). doi:[10.1098/rsta.2011.0623](https://doi.org/10.1098/rsta.2011.0623)
80. G. Ramírez Ávila, A. Gapelyuk, N. Marwan, H. Stepan, J. Kurths, T. Walther, N. Wessel, Classifying healthy women and preeclamptic patients from cardiovascular data using recurrence and complex network methods. *Autonom. Neurosci.* **178**(1–2), 103–110 (2013). doi:[10.1016/j.autneu.2013.05.003](https://doi.org/10.1016/j.autneu.2013.05.003)
81. K. Rehfeld, N. Marwan, S.F.M. Breitenbach, J. Kurths, Late Holocene Asian summer monsoon dynamics from small but complex networks of paleoclimate data. *Clim. Dyn.* **41**(1), 3–19 (2013). doi:[10.1007/s00382-012-1448-3](https://doi.org/10.1007/s00382-012-1448-3)

82. M.C. Romano, M. Thiel, J. Kurths, W. von Bloh, Multivariate recurrence plots. *Phys. Lett. A* **330**(3–4), 214–223 (2004). doi:[10.1016/j.physleta.2004.07.066](https://doi.org/10.1016/j.physleta.2004.07.066)
83. M.C. Romano, M. Thiel, J. Kurths, I.Z. Kiss, J.L. Hudson, Detection of synchronization for non-phase-coherent and non-stationary data. *Europhys. Lett.* **71**(3), 466–472 (2005). doi:[10.1209/epl/i2005-10095-1](https://doi.org/10.1209/epl/i2005-10095-1)
84. M.C. Romano, M. Thiel, J. Kurths, C. Grebogi, Estimation of the direction of the coupling by conditional probabilities of recurrence. *Phys. Rev. E* **76**(3), 036211 (2007). doi:[10.1103/PhysRevE.76.036211](https://doi.org/10.1103/PhysRevE.76.036211)
85. M.C. Romano, M. Thiel, J. Kurths, K. Mergenthaler, R. Engbert, Hypothesis test for synchronization: twin surrogates revisited. *Chaos* **19**(1), 015108 (2009). doi:[10.1063/1.3072784](https://doi.org/10.1063/1.3072784)
86. O.E. Rössler, An equation for continuous chaos. *Phys. Lett. A* **57**(5), 397–398 (1976). doi:[10.1016/0375-9601\(76\)90101-8](https://doi.org/10.1016/0375-9601(76)90101-8)
87. E.N. Sawardecker, C.A. Amundsen, M. Sales-Pardo, L.A.N. Amaral, Comparison of methods for the detection of node group membership in bipartite networks. *Eur. Phys. J. B* **72**, 671–677 (2009). doi:[10.1140/epjb/e2009-00397-6](https://doi.org/10.1140/epjb/e2009-00397-6)
88. Y. Shimada, T. Kimura, T. Ikeguchi, Analysis of chaotic dynamics using measures of the complex network theory, in *Artificial Neural Networks - ICANN 2008, Pt. I*, ed. by V. Kurkova, R. Neruda, J. Koutnik. Lecture Notes in Computer Science, vol. 5163, (Springer, New York, 2008), pp. 61–70
89. M. Small, J. Zhang, X. Xu, Transforming time series into complex networks, in *Complex Sciences. First International Conference, Complex 2009, Shanghai, China, February 2009*, ed. by J. Zhou. Revised Papers, Part 2. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 5, (Springer, Berlin, 2009), pp. 2078–2089. doi:[10.1007/978-3-642-02469-6_84](https://doi.org/10.1007/978-3-642-02469-6_84)
90. N.P. Subramaniam, J. Hyttinen, Analysis of nonlinear dynamics of healthy and epileptic eeg signals using recurrence based complex network approach, in *Proceedings of the 6th International IEEE EMBS Conference on Neural Engineering*, 2013, pp. 605–608. doi:[10.1109/NER.2013.6696007](https://doi.org/10.1109/NER.2013.6696007)
91. K. Suzuki, K. Wakita, Extracting multi-facet community structure from bipartite networks, in *Proceedings of the 2009 International Conference on Computational Science and Engineering*, vol. 04 (IEEE Computer Society, Washington, DC, 2009), pp. 312–319. doi:[10.1109/CSE.2009.451](https://doi.org/10.1109/CSE.2009.451)
92. F. Takens, Detecting strange attractors in turbulence, in *Dynamical Systems and Turbulence, Warwick 1980*, ed. by D.A. Rand, L.S. Young. Lecture Notes in Mathematics, vol. 898, (Springer, New York, 1981), pp. 366–381. doi:[10.1007/BFb0091924](https://doi.org/10.1007/BFb0091924)
93. M. Thiel, M.C. Romano, P.L. Read, J. Kurths, Estimation of dynamical invariants without embedding by recurrence plots. *Chaos* **14**(2), 234–243 (2004). doi:[10.1063/1.1667633](https://doi.org/10.1063/1.1667633)
94. M. Thiel, M.C. Romano, J. Kurths, M. Rolf, R. Kliegl, Twin surrogates to test for complex synchronisation. *Europhys. Lett.* **75**(4), 535–541 (2006). doi:[10.1209/epl/i2006-10147-0](https://doi.org/10.1209/epl/i2006-10147-0)
95. L.L. Trulla, A. Giuliani, J.P. Zbilut, C.L. Webber Jr., Recurrence quantification analysis of the logistic equation with transients. *Phys. Lett. A* **223**(4), 255–260 (1996). doi:[10.1016/S0375-9601\(96\)00741-4](https://doi.org/10.1016/S0375-9601(96)00741-4)
96. A.A. Tsonis, P.J. Roebber, The architecture of the climate network. *Physica A* **333**, 497–504 (2004). doi:[10.1016/j.physa.2003.10.045](https://doi.org/10.1016/j.physa.2003.10.045)
97. D.J. Watts, S.H. Strogatz, Collective dynamics of “small-world” networks. *Nature* **393**(6684), 440–442 (1998). doi:[10.1038/30918](https://doi.org/10.1038/30918)
98. M. Wickramasinghe, I.Z. Kiss, Effect of temperature on precision of chaotic oscillations in nickel electrodisolution. *Chaos* **20**(2), 023125 (2010). doi:[10.1063/1.3439209](https://doi.org/10.1063/1.3439209)
99. M. Wiedermann, J.F. Donges, J. Heitzig, J. Kurths, Node-weighted interacting network measures improve the representation of real-world complex systems. *Europhys. Lett.* **102**(2), 28007 (2013). doi:[10.1209/0295-5075/102/28007](https://doi.org/10.1209/0295-5075/102/28007)
100. A. Wolf, J.B. Swift, H.L. Swinney, J.A. Vastano, Determining Lyapunov exponents from a time series. *Physica D* **16**(3), 285–317 (1985). doi:[10.1016/0167-2789\(85\)90011-9](https://doi.org/10.1016/0167-2789(85)90011-9)

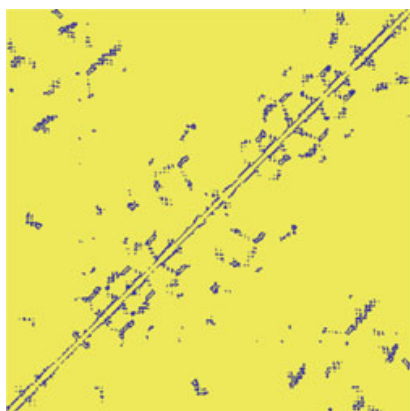
101. X. Xu, J. Zhang, M. Small, Superfamily phenomena and motifs of networks induced from time series. *Proc. Natl. Acad. Sci. USA* **105**(50), 19601–19605 (2008). doi:[10.1073/pnas.0806082105](https://doi.org/10.1073/pnas.0806082105)
102. Y. Yang, H. Yang, Complex network-based time series analysis. *Physica A* **387**(5–6), 1381–1386 (2008). doi:[10.1016/j.physa.2007.10.055](https://doi.org/10.1016/j.physa.2007.10.055)
103. J.P. Zbilut, C.L. Webber Jr., Embeddings and delays as derived from quantification of recurrence plots. *Phys. Lett. A* **171**(3–4), 199–203 (1992). doi:[10.1016/0375-9601\(92\)90426-M](https://doi.org/10.1016/0375-9601(92)90426-M)
104. J.P. Zbilut, A. Giuliani, C.L. Webber Jr., Detecting deterministic signals in exceptionally noisy environments using cross-recurrence quantification. *Phys. Lett. A* **246**(12), 122–128 (1998). doi:[10.1016/S0375-9601\(98\)00457-5](https://doi.org/10.1016/S0375-9601(98)00457-5)
105. J. Zhang, M. Small, Complex network from pseudoperiodic time series: topology versus dynamics. *Phys. Rev. Lett.* **96**(23), 238701 (2006). doi:[10.1103/PhysRevLett.96.238701](https://doi.org/10.1103/PhysRevLett.96.238701)
106. P. Zhang, J. Wang, X. Li, M. Li, Z. Di, Y. Fan, Clustering coefficient and community structure of bipartite networks. *Physica A* **387**(27), 6869–6875 (2008). doi:[10.1016/j.physa.2008.09.006](https://doi.org/10.1016/j.physa.2008.09.006)
107. C. Zhou, L. Zemanova, G. Zamora, C.C. Hilgetag, J. Kurths, Hierarchical organization unveiled by functional connectivity in complex brain networks. *Phys. Rev. Lett.* **97**(23), 238103 (2006). doi:[10.1103/PhysRevLett.97.238103](https://doi.org/10.1103/PhysRevLett.97.238103)
108. C. Zhou, L. Zemanova, G. Zamora-Lopez, C.C. Hilgetag, J. Kurths, Structure-function relationship in complex brain networks expressed by hierarchical synchronization. *New J. Phys.* **9**(6), 178 (2007). doi:[10.1088/1367-2630/9/6/178](https://doi.org/10.1088/1367-2630/9/6/178)
109. Y. Zou, R.V. Donner, J.F. Donges, N. Marwan, J. Kurths, Identifying shrimps in continuous dynamical systems using recurrence-based methods. *Chaos* **20**(4), 043130 (2010). doi:[10.1063/1.3523304](https://doi.org/10.1063/1.3523304)
110. Y. Zou, M.C. Romano, M. Thiel, N. Marwan, J. Kurths, Inferring indirect coupling by means of recurrences. *Int. J. Bifurcat. Chaos* **21**(4), 1099–1111 (2011). doi:[10.1142/S0218127411029033](https://doi.org/10.1142/S0218127411029033)
111. Y. Zou, R.V. Donner, J. Kurths, Geometric and dynamic perspectives on phase-coherent and noncoherent chaos. *Chaos* **22**(1), 013115 (2012). doi:[10.1063/1.3677367](https://doi.org/10.1063/1.3677367)
112. Y. Zou, R.V. Donner, M. Wickramasinghe, I.Z. Kiss, M. Small, J. Kurths, Phase coherence and attractor geometry of chaotic electrochemical oscillators. *Chaos* **22**(3), 033130 (2012). doi:[10.1063/1.4747707](https://doi.org/10.1063/1.4747707)
113. Y. Zou, J. Heitzig, R.V. Donner, J.F. Donges, J.D. Farmer, R. Meucci, S. Euzzor, N. Marwan, J. Kurths, Power-laws in recurrence networks from dynamical systems. *Europhys. Lett.* **98**(4), 48001 (2012). doi:[10.1209/0295-5075/98/48001](https://doi.org/10.1209/0295-5075/98/48001)

Part II
RQA Best Practices

Chapter 5

From Time to Space Recurrences in Biopolymers

Alfredo Colosimo and Alessandro Giuliani



Abstract The application of Recurrence-Based Techniques to biopolymers is here-with introduced with an emphasis on the differences holding between the analysis of strings endowed with a mainly logical (DNA) or chemico-physical (Proteins) information content. The unique features of RQA when applied to systems in which spatial order (sequence) takes the place of time are described, highlighting the emergence of ‘time distortions’. This is a metaphorical term stressing the fact that a monodimensional array of aminoacid residues (sequence) beside being formally identical to a discrete time series is a physical object that folds in the usual three dimensional space. This behavior allows to fully appreciate the fact that RQA as an analytical tool is flexible enough to deal with complex networks in either the spatial or the temporal dimension. The comparison of DNA sequences with text strings helps to shed light on the particular nature of biological information coding as well as on the role of RQA technique in bioinformatics and computational biology fields.

A. Colosimo (✉)

Department SAIMLAL at Sapienza University, Via A. Borelli 50, 00161 Roma, Italy

e-mail: alfredo.colosimo@uniroma1.it

A. Giuliani

Department EPP at Ist. Superiore Sanità, Viale Reg. Elena 299, 00161 Roma, Italy

e-mail: alessandro.giuliani@iss.it

5.1 Introduction

5.1.1 Generalities

Recurrence Quantification Analysis (RQA) was birthed out in the realm of time series analysis, following the inspiration of the seminal work by Eckmann, Kamphorst and Ruelle [1]. In this paper the authors introduced Recurrence Plots (RPs) as a purely graphical techniques for detecting recurrences in time series. Recurrences are reminiscent of quasi-attractors in phase spaces of physical systems, and the authors [1] suggested RPs could be used as a qualitative portrait of systems dynamics. Chuck Webber and Joe Zbilut [2] transformed this purely graphical technique into a quantitative methodology enlarging the reach of this RQA generating approach well beyond theoretical physics.

The quantitative version of RPs took the name of Recurrence Quantification Analysis and some numerical descriptors (Percent Recurrence, Determinism, Entropy, Maxline, etc.) were soon recognized to have a strong link with well known dynamical measures. In particular, the link between Maxline and Lyapunov exponent represented a versatile alternative to more demanding techniques (e.g. Fourier Analysis) to describe the periodicities present in a time series [3,4].

Webber and Zbilut [2] identified physiological time series (heartbeat series or tacograms, EEG and EMG traces, etc.) as an investigation field particularly suited for RQA. Physiological time series can be hardly considered as the output generated by a hidden deterministic, equation-driven, system. They reflect the continuous adjustments induced by the need to adapt to the contingent vagaries of both the internal and the external milieu. In other words, the fluctuations we observe in subsequent RR intervals of the heartbeat or in the voltage changes of an electromyography do not correspond to a wired (albeit unknown) law of functioning but to the homeostatic response of a perturbed system.

This change in perspective (consequent to the shift toward physiology) was in some sense at the basis of the success of RQA: the ability of the technique to cope with non-stationary and very short signals, its sensitivity to singularities, the independence of RQA descriptors from signal length demonstrated to be very well suited for the application of RQA to several different situations where more traditional (and in many case more mathematically intensive) methods failed.

A crucial feature of RQA is its discrete character. An RP (the starting point of RQA) is nothing more nothing less than a distance matrix made binary by the action of a threshold that assigns a unitary (a darkened pixel) value to the distance $d_{i,j}$ between two i and j vectors (epochs): if this is lower than a pre-defined radius: the darkened pixels correspond to recurrences. In order to compute distances the original signal needs to be sampled in discrete equi-spaced values whose number constitutes an epoch and corresponds to the Embedding Dimension (ED).

In such a context the heartbeat series or tacogram appears as a very favorable one, since the sampling frequency is directly imposed by Nature: the subsequent heartbeats are given by the heart itself and the series of the lengths of the subsequent beats is offered by the system as the natural signal, without the need of external discretization strategies such as the first zero of autocorrelation function.

The naturally discrete character of the corresponding series is the essential feature of biopolymers (actually of any sequence of alphanumeric symbols) making the application of RQA a straightforward choice in the analysis of both protein and nucleic acid molecules. This application is less demanding, given the monodimensional character of protein and DNA sequences, with respect to the bidimensional applications like those devised by Vasconcelos and collaborators [5].

The ordering parameter in numerical series describing biopolymers, is the defined sequence of monomers (nucleotides or amino-acids) along the primary structure [6]. Thus, for both DNA and protein molecules the natural order of monomers corresponds to time, and the lack of stationarity dependence and the ability to deal with very short series of RQA can be fully exploited.

The kind of information stored in DNA and protein molecules is very different: while DNA stores information in the form of long series made by the juxtapositions of four letters (the symbols of the four kinds of nucleotides), the 20 aminoacid species constituting proteins can be easily coded (in addition to the purely symbolic 20-letter code) by their chemico-physical properties. The reason is that while DNA is a passive store of information that, in order to become biologically active, must be translated into a protein by the agency of a complex enzymatic process, proteins are physiologically active molecules as such. Hence proteins are by themselves operational devices that, after a folding process converts the linear sequence into a complex three-dimensional object, exert their specific biological function strictly dependent on the acquired 3D configuration.

Since the 3D configuration can be intended as the microenvironment interpretation of the primary structure recipe [6], we can speculate that different chemico-physical coding of the same sequence can retain some relevant semantic information about protein structures. Although we know that nucleic acids (e.g. RNA molecules) undergo a folding process (that in any case is driven by base deterministic pairing a-la-Watson and Crick) and that the recognition of specific regulatory traits of DNA molecule is depending on its 3D configuration (that in turn has some relation with the underlying sequence) we are in presence of two basically different information storing devices:

- DNA: very long (billions of bases) symbolic sequences
- Proteins: relatively short (from 30 to 10,000) sequences endowed with chemico-physical semantics.

We will show in the following sections that this basic difference has a great impact in the evaluation of what a recurrence means in these two cases.

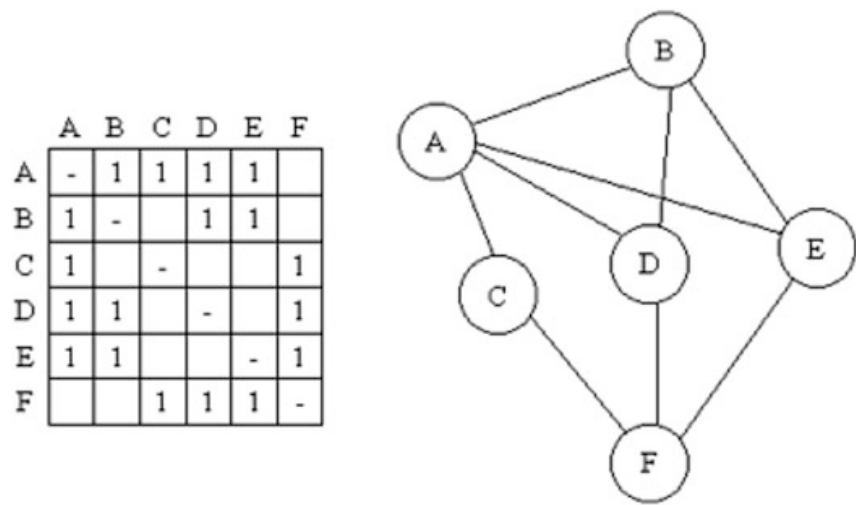


Fig. 5.1 Thresholded distance matrix (RP, equivalent to an adjacency matrix) and the corresponding graph (see the text for explanation)

5.1.2 Recurrence Plots as Distance Matrices

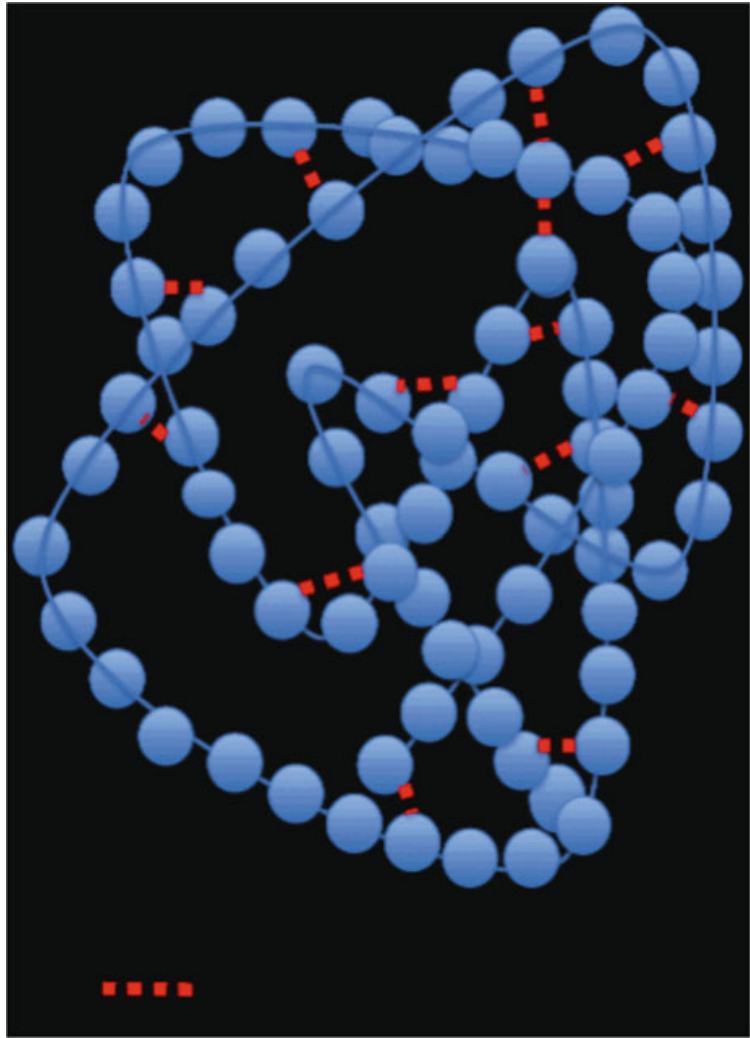
A distance between two vectors i,j can be expressed as:

$$d_{(i,j)} = \sqrt{(X_i - X_j)^2 + (Y_i - Y_j)^2 + (Z_i - Z_j)^2} \tag{5.1}$$

Formula (5.1) can be generalized to any number of dimensions and to metric paradigms different from Euclidean (e.g. Hamming, Manhattan, Minkowski, Mahalanobis distances, etc.): the crucial point is that we can collapse spaces of any dimensionality into a bi-dimensional symmetric matrix having as rows and columns the discrete elements (statistical units) whose mutual distances appear at the row-column crosses. After the thresholding, the distance matrix becomes the proper RP corresponding to the adjacency matrix of an undirected network as evident from Fig. 5.1 reporting on the left the adjacency matrix (RP) and on the right the corresponding graph.

This implies RQA is homologous to methods in graph theory whose quantitative descriptors correspond to topological invariants of the network. At odds with other networks that do not have a mono-dimensional privileged order of nodes, sequence (and clearly time-series) based networks entail an implicit metrics (in addition to the topology of contacts) corresponding to the sequence order. This feature has important consequences for both DNA and proteins RQA, as we will discuss in what follows.

Fig. 5.2 The figure reports a protein structure depicted as an intermingled thread of pearls corresponding to amino-acid residues. The *dashed lines* indicate contacts between residues near in space but far-away in sequence. These contacts are endowed with high ordK values and are the most critical for both folding and protein stability



5.2 Results

5.2.1 Folding Metrics

Figure 5.2 displays a cartoon representation of protein 3D structure equated to a bundled pearl thread kept in the bundled (folded) state by the pinching of some links between pearls (amino-acid residues) that are spatially close but far away along the sequence (broken red lines).

It is worth noting in Fig. 5.2 that the residues more relevant for the protein stability are those putting in contact (in the 3D space) aminoacid residues relatively distant along the sequence. This feature is described by the valued of the following descriptor (ordK) that weights each contact (recurrence) in 3D configuration by the relative distance of the nodes (residues) along the sequence [7]:

$$ordK = \sum_{s,t \in V} \frac{d_{s,t}}{n(n-1)} \quad (5.2)$$

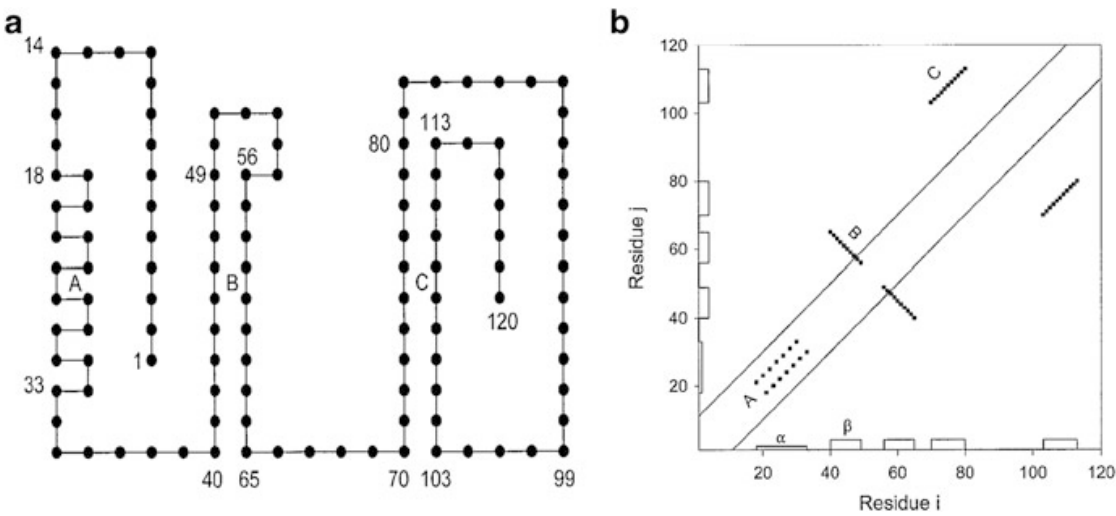


Fig. 5.3 A model protein on a 2-dimensional lattice together with its recurrence plot/contact map. The protein has 120 residues forming three secondary structures: alpha helix (A); antiparallel beta sheet (B); parallel beta sheet (C) that can be easily detected in its contact map

Being $d_{s,t}$ the distance between s and t residues along the chain, $d_{s,t} = 1$ if s and t are in contact (in the case of protein 3D structures the distance threshold for scoring a contact is usually set to a distance between 4 and 8 Å corresponding to two Van der Waals radii [8]), the number of contacts aminoacid i has with other residues. The folding rate of actual proteins was demonstrated (at equal size $N =$ number of residues) to be negatively correlated with average ordK (or Long Range contact Order (LRO) [7]. On the other hand, secondary structure patterns are exactly predicted by deterministic (corresponding to alfa helix) and anti-deterministic (corresponding to beta sheets) lines in the recurrence plot (see Fig. 5.3) [9].

Figures 5.3 and 5.4 allow us to appreciate how alpha-helix, when considered as time series features, correspond to periodic motions: the recurrence motifs (deterministic lines) go in the same direction of the time (sequence) axis, while beta sheets provoke a time inversion being oriented in the opposite direction with respect to sequence order [9].

High ordK recurrences are the contacts between residues far away in sequence, they can be made to correspond to singular time distortions (high distance in sequence corresponding to short distance in 3D space) from which the global tertiary specific structure of the molecule arises. In Fig. 5.5 high ordk recurrences are marked with an ellipse in the RP/contact map panel on the right side of ribbon diagram of recovering molecule [10].

Protein contact maps analysis allow to shed light on many emergent properties of protein molecules such as allostery, folding, lethal mutations (for a comprehensive review see [8]).

It is important to stress that protein contact maps reporting the between residues contacts in 3D space along the sequence metrics (the axes of the map) are totally equivalent to RPs. The only difference from ordinary RPs is that the metrics is not computed by a single value attached to each time point but by a triplet of values

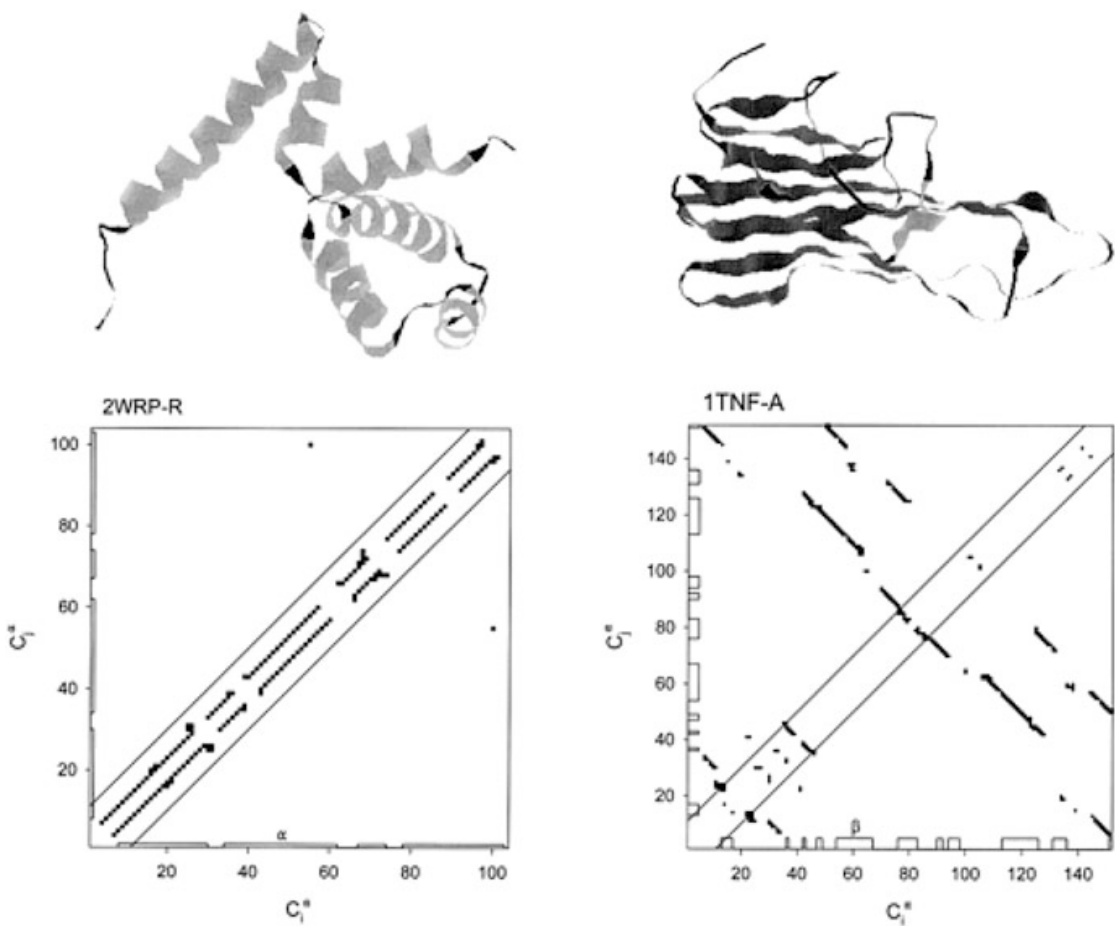


Fig. 5.4 Alternative representation of secondary structures in protein. The contact map (*bottom*) and the standard ribbon diagram (*top*) representation of Trp repressor protein (2WRP-R) and of Tumor Necrosis factor (1TNF-A) [9] are reported on the left and right side, respectively. Notice the quite different shapes associated to alpha helix motifs in 2WRP-R (deterministic lines parallel to main diagonal) and to beta sheets in 1TNF-A (anti-deterministic line orthogonal to main diagonal)

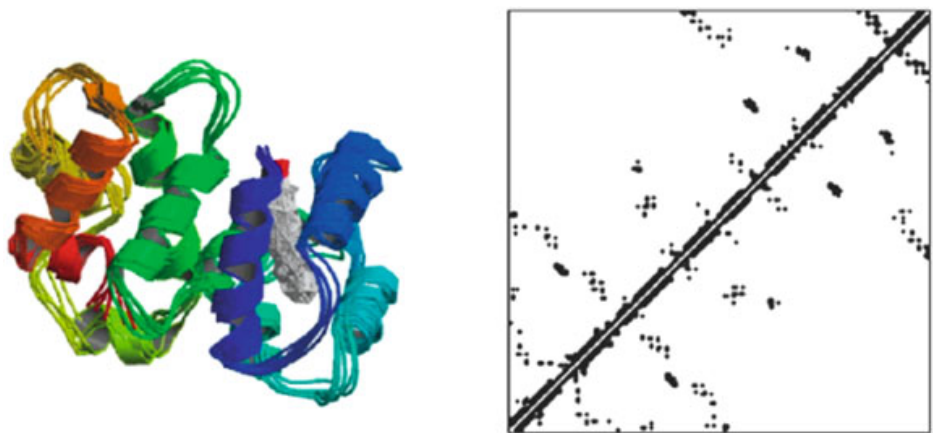


Fig. 5.5 Recovering, a calcium sensor. This calcium binding protein (PDB code: 1IKU) is finely regulated by small changes in calcium concentrations: the *left panel* reports the 3D structure in the usual ribbon diagram; the *right panel* reports the derived between residues adjacency matrix

corresponding to the X, Y, Z coordinates of the element (aminoacid) in the protein 3D structure. We can say the RP is computed by the protein folding process to its native state, in this sense is a spontaneously occurring natural RP [8, 9].

3D structure is by no means the only relevant feature of protein molecules, and the consideration of amino-acid chemico-physical properties is in many cases of utmost importance.

5.2.2 *Chemical Metrics*

3D structure may be considered as a time distortion, i.e. an alternative metrics superimposed to the primary structure of proteins that in turn corresponds to the internal time of the molecule. This interpretation has an intuitive counterpart in terms of the bundling of a thread (primary structure) generating a nonlinear geometry changing the mutual distances between amino-acid residues by means of short-cuts connecting far apart residues along the sequence [7, 8, 10, 11]. When focusing on the chemico-physical nature of residues instead of their positions in the 3D structure we face a different kind of transformation: what is in act is no more a distortion of a linear order induced by the bending and folding of a linear chain but the imposition of a semantics on the monomers allowing for the detection of similarities between different patches along the sequence not apparent in the symbolic code [6].

Among the myriads of different chemico-physical features, a particularly relevant property of amino-acids is their differential solubility in apolar (lipids) and polar (water) phases. This property can be coded in terms of partition coefficients with high values corresponding to a marked hydrophobic (and consequently lipophylic) character, while a negative value points to a hydrophylic behaviour [12]. This allows for a direct transformation of the text string corresponding to the primary structure of a protein into a numerical series of hydrophobicity values producing a characteristic profile as reported in Fig. 5.6 [6].

From a purely mathematical point of view the bottom panel is identical to a time series, thus it is amenable to an RQA: the recurrences are in this case patches (with a number of components equal to the chosen embedding dimension) with similar hydrophobic character variously distributed along the sequence; deterministic lines correspond, in turn, to longer patches of similar hydrophobic/hydrophilic distribution. The fact that patches with similar hydrophobic character tend to interact in solution assigns a relevant meaning to this kind of coding [13, 14].

The substitution of a continuous property (in the specific case the hydrophobic/hydrophylic character of each amino-acid residue) to a categorical code (letters) allows to shift from a pure identity metrics (the same patch of residues is present in two different regions of the sequence) to a proper continuous similarity metrics: two very similar hydrophobicity patches are detected along the chain. The enlargement of recurrence detection from strict identity to similarities allows for a deeper insight into biopolymers features. In what follows we discuss a couple of cases where this

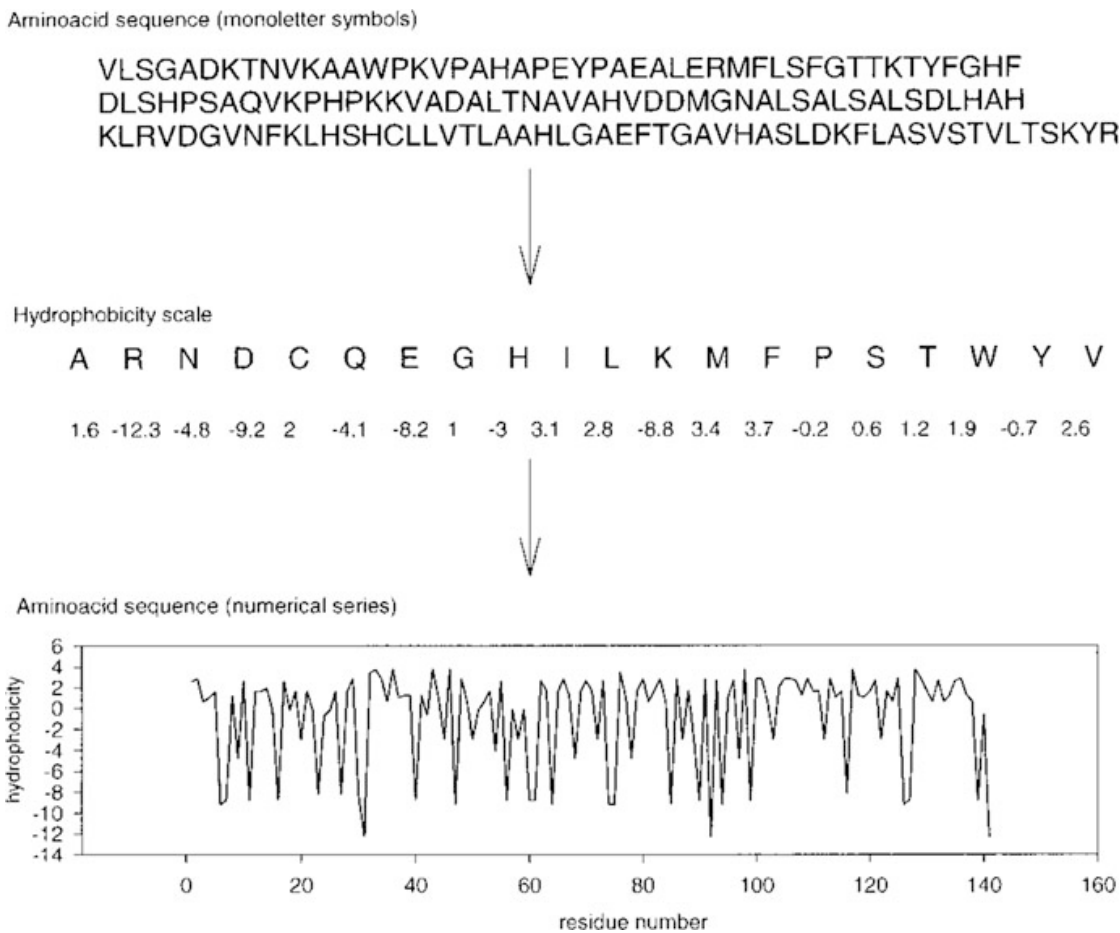


Fig. 5.6 Producing a hydrophobicity profile out of a protein primary structure. The primary structure of a protein (*upper panel*) is translated into the corresponding hydrophobicity profile (*bottom panel*) assigning to each aminoacid the hydrophobicity value defined by the MJ scale [13] (*middle panel*)

approach showed extremely useful : the identification of specific locations along the P53 protein involved in the interaction with other protein systems [15], and the subtle structural peculiarities associated to extensive functional differences among the elements of an apparently homogeneous set of proteins (bacterial rubredoxins) [16].

5.2.2.1 Functional Hot-Spots in the P53 Protein

Figure 5.7 reports the hydrophobicity profile corresponding to the P53 protein (a molecule endowed with a crucial role in cancer development) sequence and its RP.

It is worth noting that the deterministic and laminar lines indicated by arrows were *a posteriori* shown [6, 15] to correspond to specific locations in the molecule involved in the interaction with other protein systems. This is of crucial biological importance because P53 molecule (playing a very important role in cancer development [17] acts as a sort of platform for generating multi-proteins aggregates: the

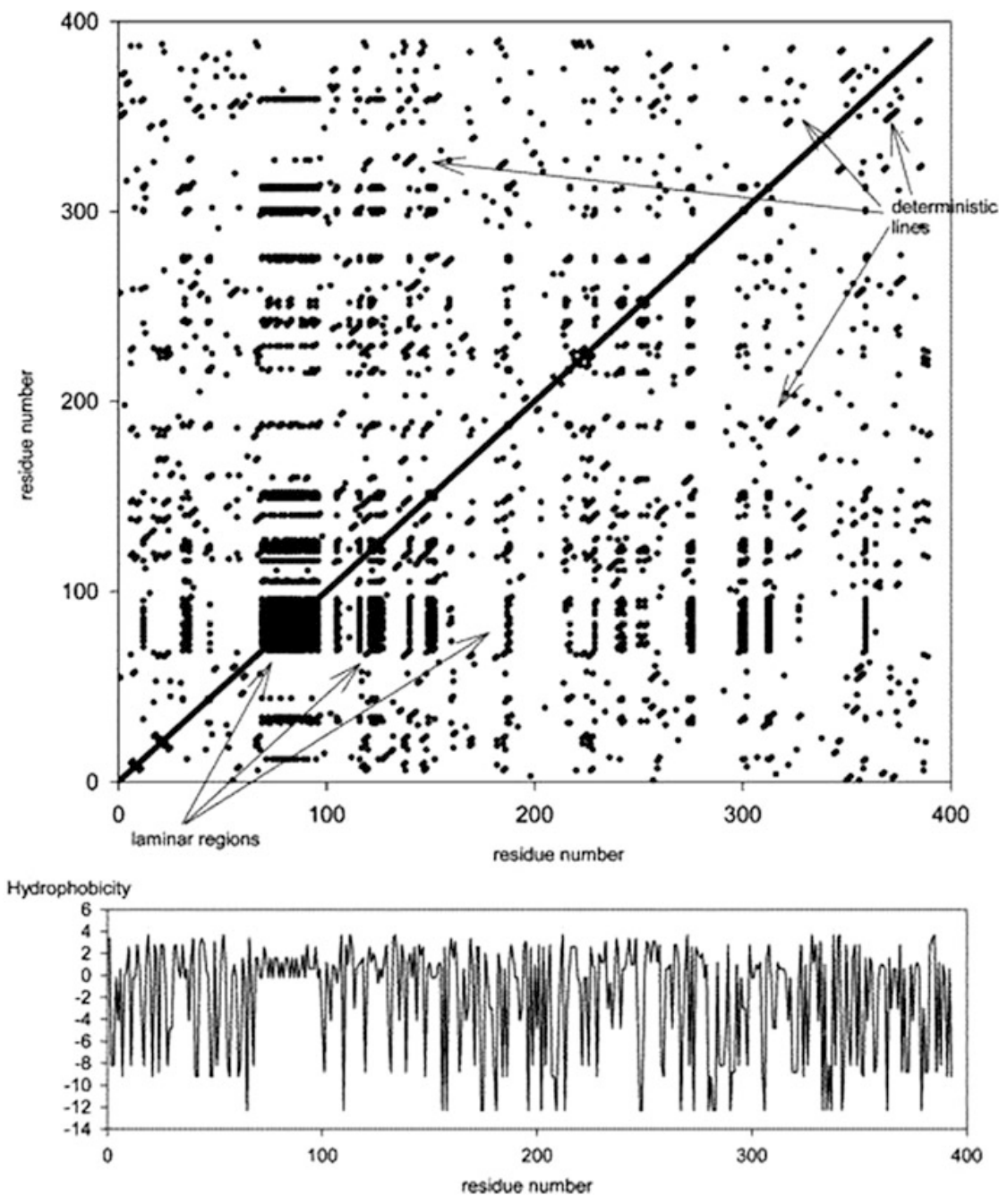


Fig. 5.7 RP and hydrophobicity profile of the P53 Protein. The *lines* indicated with *arrows* point to the locations of the interaction of P53 with other protein species

recognition of interaction points along the sequence is thus of utmost importance for rationalize the role of p53 mutations discovered in cancer patients [18].

The recognition of such interaction hot-spots is totally impossible in terms of pure symbolic coding : we need the set of a specific chemico-physical semantics (and consequently a hypothesis on the interaction mechanism) for the elucidation of physiological role of different tracts in the sequence.

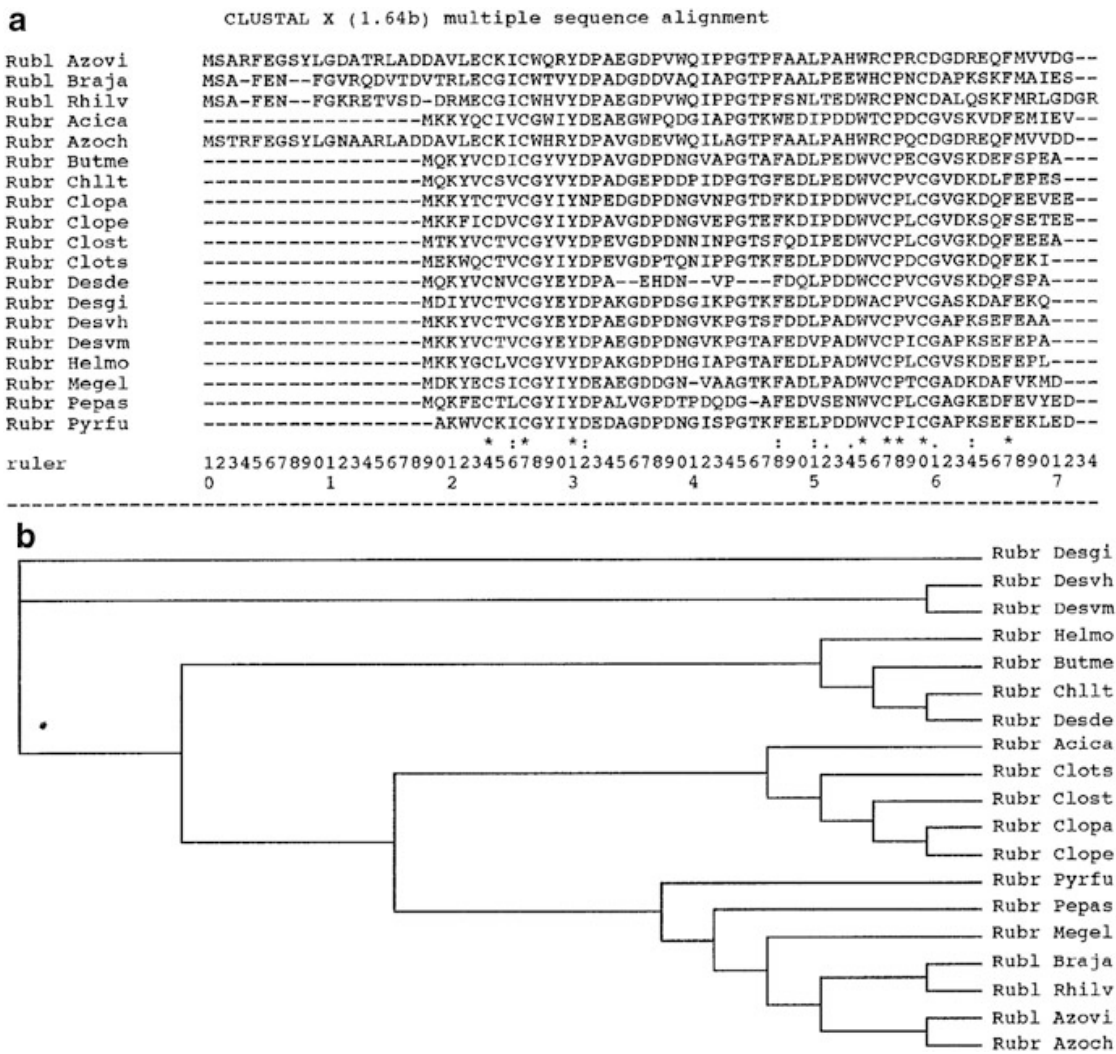


Fig. 5.8 Bacterial rubredoxin from thermophylic and mesophylic species. In panel **a** sequences of proteins pertaining to the family of rubredoxins are aligned in order to maximize their superposition (symbolic identity) and allow for the computation of the mutual distance between primary structures. Each sequence corresponds to a rubredoxin coming from a different bacterial species whose names are reported in the figure, the amino-acids are indicated by means of a single letter symbolic code, *dashed lines* corresponds to gaps, i.e. to the lack of amino-acid in the corresponding position along the sequence. Panel **b** reports the hierarchical clustering of the same protein set, on the basis of the similarities in the primary structures [16]

5.2.2.2 Thermophylic and Mesophylic Rubredoxins

Among the rubredoxin (a small protein species ubiquitary in prokariotes) sequences corresponding to mesophylic bacteria in Fig. 5.8, i.e. to organisms living at standard temperature condition, only one sequence refers to an organism living at extremely high temperatures in volcanic calderas, namely *Pyrococcus Furiosus* (Pyrfu). This organism lives at nearly 100 °C, a temperature at which the proteins of other bacteria denature (with the consequent death of the organism) in 2 or 3 min. Notwithstanding its peculiar thermophylic character, Pyrfu does not occupy a

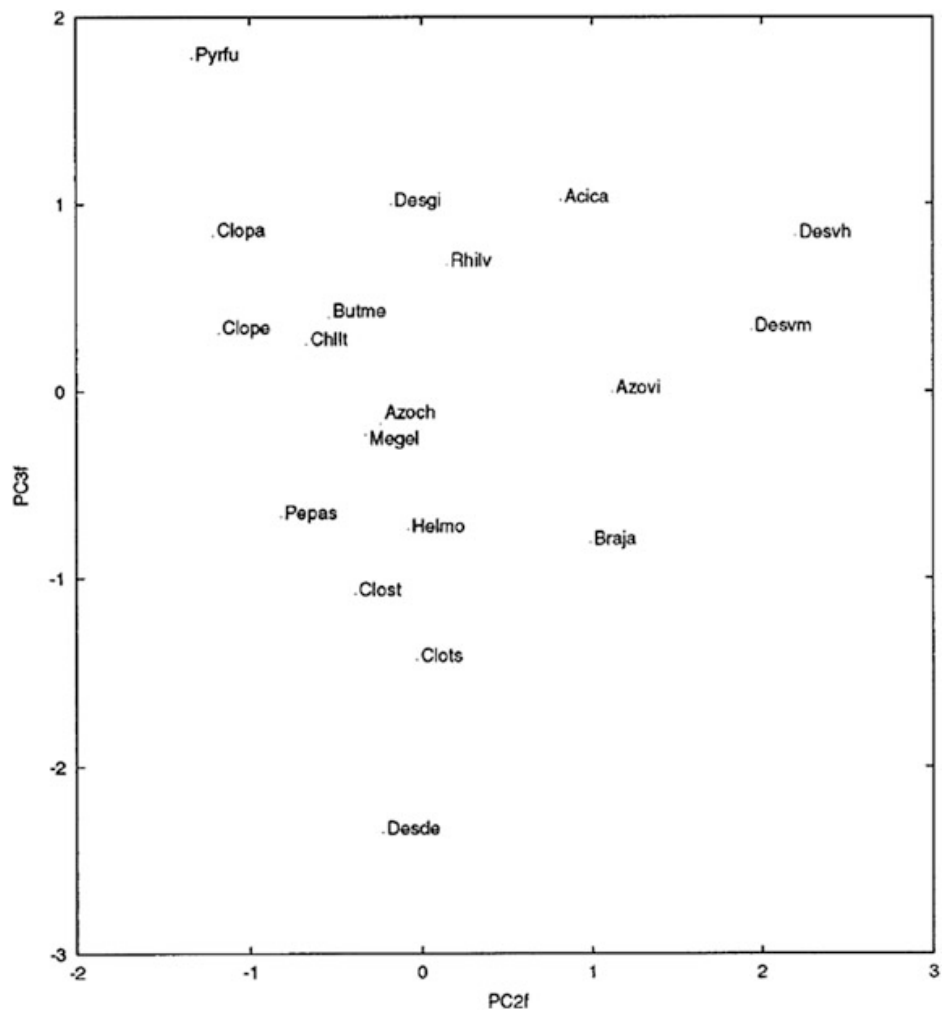


Fig. 5.9 Bacterial rubredoxins classified after PCA filtering of RQA descriptors (from [16]. it is worth noting the extreme position of the thermophylic structure (Pyrfu) at the upper left corner of the graph, while the same molecule was in a central position in the space of symbolic sequence homology

particularly peripheral position in the symbolic sequence space. as evident from the hierarchical clustering of sequences reported in panel B. This is in full agreement with the idea that amino-acid residues must be primarily considered as chemical entities endowed with peculiar chemico-physical properties, if one needs to infer functional properties from primary structure.

Thus, while in such an exemplary situation the simple sequence similarity fails (Fig. 5.6), the hydrophobicity coding of sequences followed by RQA allows to discover the extremal character of thermophylic sequence: Fig. 5.9 shows that Pyrfu is now clearly isolated in the left high corner of a bi-dimensional principal component space derived from the usual RQA descriptors (% Recurrence, % Determinism, MaxLine, Entropy).

In other words, when recurrences mark similar hydrophobicity patches , using a chemical (instead of a symbolic) metrics, RQA is able to discriminate thermophylic

system in terms of internal rules of construction of single molecules [16]. In such a context the following points are worth stressing:

- At odds with the sequence alignment, that needs the contemporary superposition of the entire group of molecules and consequently can be applied only at very similar sets of proteins (homologs), a set of RQA descriptors is built INDEPENDENTLY for each protein molecule: this allows to elucidate the chemico-physical mechanism at the basis of the observed discrimination and, on a more general perspective, to compare proteins sharing no sequence homology.
- The rubredoxin molecule corresponding to *Clostridium Pasteurianum* (CloPA) has a 3D structure practically identical to *Pyrococcus Furiosus* (PyrFu) but on the contrary has the ordinary mesophylic properties. Thus, the thermophylic character does not arise deterministically by a specific 3D structure but by a mesoscopic level constituted by hydrophobicity patterning located in between pure sequence and 3D structural information. The evidence of such an intermediate layer located between sequence and structure is of utmost importance for the study of protein physiology.

5.3 Mixed Metrics for Complex Behavior

The fact CloPA and PyrFu share the same 3D structure makes it possible to generate chimeric sequences by exchanging corresponding pieces between the two primary structures, remaining confident that the chimeric product will be in turn foldable and thus soluble in water. Figure 5.10 reports the sequences of six chimeric products, each one having the two parent (CLOPA and PyrFu) systems contributing for the 50 % to the resulting chimeric primary structure:

In the last column of Fig. 5.10 the lifetime (in minutes) of each protein at 92°C is reported. It is immediate to note how Chim1–Chim4 are mesophylic systems (they have very short lifetimes), while Chim5 and Chim6 are thermophylic. All the chimeras are, by construction, located halfway between the mesophylic and thermophylic poles in terms of pure symbolic sequence space, nevertheless they have two opposite behaviors and, what is still more intriguing, with nothing in between. There is no intermediate stability behavior: the mesophylic/thermophylic property is a yes/no emerging property of the system, strictly dependent from the peculiar internal chemical metrics and totally independent from structural features (all the proteins share the same 3D structure). The recurrence plots of hydrophobicity coded thermophylic and mesophylic rubredoxins are shown in the upper and bottom panels, respectively, of Fig. 5.11.

The two patterns allow for a clear separation of the thermal stability of the molecules and can be interpreted in mechanistic terms: the distributed pattern allows for richer dynamics enabling the structure to cope with stronger thermal fluctuations. The RPs shown in Fig. 5.11 reflect the bipartite character of thermal stability: there are two clearly distinct patterns of hydrophobicity distribution along the chain.

		Lifetime
Chim1	MKKYTCTVCGYIYNPDAGDPDNGISPGTKFEELPDDWVCPICGAPKSEFEKLED	3.6
Chim2	-AKYTCTVCGYIYNEDGDPDNGVNP GTDFKDIPDDWVCPICG V GK D Q FEEVEE	5.5
Chim3	MKKYTCTVCGYIYNPDAGDPDNGISPGTKFEELPDDWVCPICGAPKSQFEEVEE	8.3
Chim4	-AKWVCKICGYIYDEEDGDPDNGVNP GTDFKDIPDDWVCPICG V GK D E F E K L E D	8.9
Chim5	MAKVVCKICGYIYDEEDGDPDNGISPGTKFEELPDDWVCPICGAPKSEFEKLED	240
Chim6	-AKWVCKICGYIYDEEDGDPDNGISPGTKFEELPDDWVCPICGAPKSQFEEVEE	250

Fig. 5.10 Chimeric rubredoxins Sequences of six chimeric products originated from CloPa (mesophylic) and PyrFu (thermophylic) rubredoxins. Each parent system contribute for the 50 % to the chimeric structures

Mesophylic systems are characterized by hydrophobicity similarities concentrated in a specific part of the sequence (between 5–10 and 37–42 sequence patches), while thermophylic proteins have similarities distributed all over the sequence. It is worth noting the total amount of recurrence is almost identical while the distribution is different. As in the case of ordK index that was halfway between sequence and 3D structure metrics, even in this case the functional discrimination was achieved by a sort of mixed metrics index corresponding to the combination of two different metrics supporting the same system. The two metrics are in this case the distribution pattern of a chemical coding (hydrophobicity, chemical metrics) along a structural one (location of the similarity patterns in the molecule, topological metrics). From a statistical point of view this mixed metrics was made quantitative in terms of hydrophobicity recurrences kurtosis along the sequence that allowed to exactly discriminate mesophylic and thermophylic molecules [16]. The presence of a closed individual system (a protein molecule with a finite number of amino-acids and with a specific 3D structure in solution), that can be jointly defined on different but consistent metrics (topology, chemistry, sequence spaces) greatly enhances the ability of RQA inspired techniques to detect emerging properties and complex non-linear behavior of the studied systems. Protein science (with thousands of well known 3D structures and sequences) is an almost infinite playground for RQA and in general, complex network based, models.

5.4 Shifting from Protein to DNA Analysis: A Cryptography Tale

What happens when we leave the world of well defined, closed macromolecular systems and enter the huge mono-dimensional text of genomic DNA?

To make a long story short we can imagine the shift from protein to DNA analysis as entering a gigantic library collecting heterogeneous texts whose interpretation needs someone actively reading the code in the aim to put in light the possible existence of a basic informational architecture at a deeper level than that embedded into the genetic code. In other words we shift from the task of inferring the function of a given unknown artifact from the analysis of its shape and internal mechanism (proteins) to the task of understanding a book about the history of an

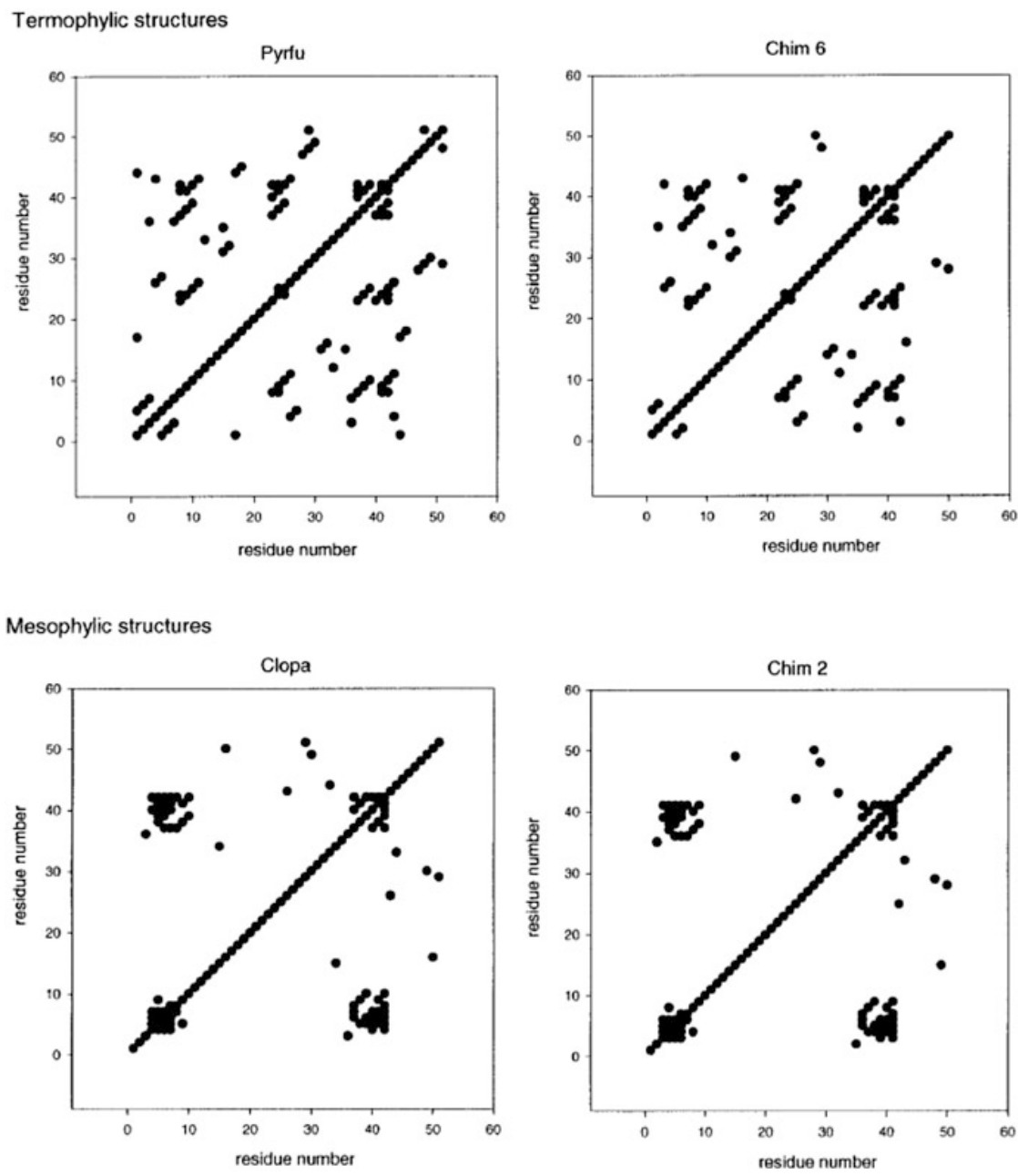


Fig. 5.11 RPs of thermophylic and mesophylic rubredoxins. It is worth noting the presence of two neatly defined hydrophobicity distribution patterns: a diffuse one typical of thermophylic structures (*upper panel*) in which the patches with similar hydrophobicity are distributed all over the sequence and the mesophylic one (*bottom panel*) in which the hydrophobic similarities are concentrated in a specific part of the molecule. The total number of recurrences (*darkened points*) is almost identical for the two patterns, while their relative positions are drastically different

ancient civilization written in an unknown language (DNA). The metaphor of the history book is not a case: the vagaries of billions of years of natural history are in some (still unknown) way written in the DNA text, which allows us to state that the application of RQA to DNA sequences has to do with cryptography (see Appendix 1). When not in presence of semantically rich metrics, the main avenue

to judge about the relevance of a recurrence pattern is its distance from a randomly shuffled counterpart. In the following we will look how this method, albeit necessary in many instances, can by no means be considered as definitive. As a typical example, the genomic DNA of the HIV virus, carrying basic biological information, is practically identical in terms of RQA descriptors to its shuffled counterpart. In this realm, the search for recurrences in long texts was pursued by many scientists especially in the form of long range correlations. A search light in the analysis of long range correlations in genomic DNA was the intron/exon and, more in general, the coding/non-coding discrimination [19, 20]. Nevertheless the coding/noncoding differences, even if by no means can be interpreted as functional/non functional but only in terms of having/not-having a corresponding protein sequence is very useful in trying and rationalize correlation structure in DNA molecules. Noncoding sequences can be located inside structural genes (in this case we properly speak of introns) or interspersed between genes (inter-genic sequences), in any case a basic question is if the long range correlations observed in DNA (that has a global scaling very similar to the Zipfs law we described for the natural languages, see Appendix 1) could be ascribed to these non-coding sequences [21].

5.4.1 Coding and Noncoding Regions in DNA

In eukaryotic genomes the by far major part of DNA molecule (80–95 % according to different species) does not code for corresponding protein sequences, but plays different (and for the large part still unknown) roles. For a minority of these non-coding sequences we are now able to assign very specialized biological (mainly regulatory) functions, for some other we suppose a structural role, while for most of them we have no relevant hypothesis. Recently, the completion of the so called ENCODE project [22] demonstrated that genomic DNA is almost completely transcribed into RNA. This means that the old concept of ‘junk’ DNA [23] as being noncoding, is discredited, although its function remains largely unknown.

DNA molecules were considered in terms of very rough descriptors (e.g. frequency of different length words coming from the juxtaposition of nucleotides) with no specific interest in the appearing timing (in this case time must be considered the order along the DNA linear molecule). This strategy is completely fit to an ergodic hypothesis or in any case to the functioning of a stationary dynamical system at equilibrium but, as we said before, DNA sequence is not the output of a dynamical system at equilibrium. Genomic DNA of a given organism can be equated to a palinsesto, one of those old books in which different writings and re-writings, errors and repetitions go together one upon the other, and not as the output of a dynamical system with an a priori established generating function. The RQA approach adds to this purely distributional approaches the idea of local self correlation i.e. of the repetition of particular motifs in different parts of the DNA molecule

In this respect a very interesting approach is reported in Frontali and Pizzi work [24] that were among the pioneers of RQA as applied to DNA sequence

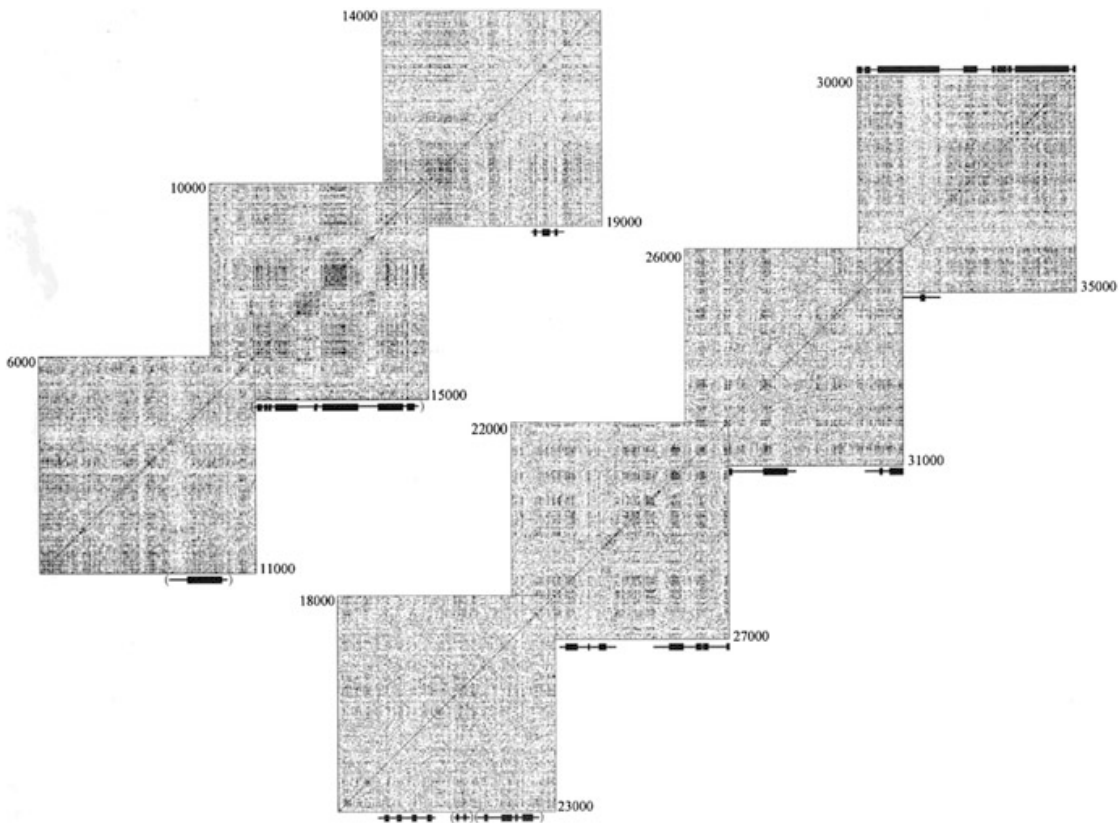


Fig. 5.12 Shifted recurrence plots for the sequence between nucleotides 6,000 and 35,000 of cosmid T05G5 from *C. elegans* chromosome III. the boxes below each graph correspond to the location of already known inter-genic (non coding) sequences: these tracts are endowed with laminar (more darkened) columns spanning the RP, so pointing to the repetition of the same sequence pattern along the entire genome scale

analysis. Figure 5.12 (coming from the above quoted paper) reports the subsequent RPs relative to the analysis of the genome of *Caenorabditis Elegans* where the repetition of non-coding regions was demonstrated to generate correlation structures extending over several megabases.

The numbers report the order of the bases along the chromosome, due to the text character of DNA alphabet, recurrences correspond to exact matches between seven bases long patches (but similar results can be obtained with an epoch of length variable from three to ten bases). The small squares at the basis of each RP indicate already established non-coding regions (in parentheses only putative regions), while lines refer to coding sequences. It is immediate to note the greater recurrence (more shaded regions) of non-coding with respect to coding regions.

Very similar results were obtained with other eukaryotic genomes (*Arabidopsis thaliana*, *Plasmodium falciparum*, *Saccharomyces cerevisiae*, etc.). What appears as a general long range correlation is due to the presence of high concentrations of repetitive patches in specific locations along the genome.

Introns’ pentamer vocabularies

<i>Caenorhabditis elegans</i>		<i>Drosophila melanogaster</i>	
AATTT*	AAATT*	AAAAT	ATTTT
AAAAT*	ATTTT*	AAAAG	CTTTT
GAAAA*	TTTTC*	GTTTT	AAAAC
TTTTT*	AAAAA*	AAATG	CATTT
TTTTG	CAAAA*	AAACA	TGTTT
TGAAA	TTTCA	AATGC	GCATT
GGAAA	TTTCC*	AACAA	TTGTT
CGAAA	TTTCG	TTTTG	CAAAA
GATTT*	AAATC*	GAAAA	TTTTC
TTGAA*	TTCAA	TTTGG	CCAAA
TTTGA	TCAA	TTTGC	GCAAA
GTTTT*	AAAAC*	TTGGC	GCCAA
		TTGCA	TGCAA
GAAAT* , AGAAA*		AGAAA , AGCAA , TTTTT TTTGT , TGAAA , TGGCA GAAAT , GGCAA , GCAGC GCAAC , CAAAT , CAATT CAACA , CAGCA	

Fig. 5.13 Pentamers significantly more represented in noncoding with respect to coding DNA patches. *Asterisks* mark pentamers significantly different (on the whole DNA sequence) from tetramer-based statistical expectation. Pentamers appearing in both species are in *bold* (from [25])

5.4.2 *Specific Repeated Words in DNA*

The recognition of specific repeated words made possible by RQA allowed the same team [25] to single out a vocabulary of repeated patches. The best choice for such a vocabulary was a word length equal to five, the analysis of the complete genomes of two very different organisms (*Caenorabditis E* and *Drosophila M.*) allowed the recognition of many shared repeated words as evident in Fig. 5.13 reporting the most frequent pentamers in the two organism (the shared ones are bolded).

The non-random appearance of such words along the genome calls for a biological explanation of such motifs that could play an epigenetic regulatory role in gene expression. As evident from the above results, the use that is made of RQA in the case of DNA sequences follows totally different methodological lines with respect to protein molecules analysis : while RQA of proteins was dominated by the network paradigm of mutual recognition between different parts of the sequence

so to give birth to an integrated whole, RQA approach to DNA molecules focused on the use of the technique as a general similarity detector so to let syntactic (as opposed to the strong semantic flavor of protein analysis) principles hidden in extremely long (and largely mysterious) texts.

5.5 Analyzing the Information Content of Biopolymers

The fundamental differences between proteins and DNA molecules we discussed above makes it possible to try and sketch a quantitative comparison between the two fields only in terms of very general syntactical rules, i.e. in terms of information coding. Assuming that both proteins and nucleic acids primary structures can be viewed as information carrying strings of symbols, leads to a number of questions concerning the difference between the two classes of biopolymers from the informational point of view. In this frame, it is impossible to overestimate the utility of the RQA descriptors, which have been successfully used to pick up structural (not semantic) differences in written texts or words in various languages (English, German, French, Italian, etc.) [26, 27] exclusively due to syntactic and grammatical subtleties.

In the following we will present a comparison of different symbolic strings analyzed with different coding strategies so to check the effect of coding on classical RQA descriptors. The biopolymers (DNA molecules) were contrasted with text strings expressed in natural languages (English and Italian) so to highlight purely syntactical similarities between series of heterogeneous origin that could point to relevant properties of the systems along the above mentioned approach inspired by cryptography.

Figure 5.14 reports the determinism distribution of different symbolic strings coming from literary texts at increasing values of the line parameter. The upper panel refers to the original texts, while the bottom panel reports the determinism distribution for randomly shuffled versions of the same texts. The choice of the texts was dictated by the need to compare semantic information rich texts (Dante (Ita) = verses from the first canto of Dante's poem and Dante(Eng) = the English translation of the same chant) with a semantic poor but prosodic rich (a lot of rhymes and repetitions) like the Dr. Suess poem Green Eggs and Ham (see Appendix 2). The Dr. Suess poem was subjected to two alternative codes : a letter based one (recurrences of letters, Suess(Lett)) and a word based one (recurrences of words, Suess(Word)), the use of two different symbolic coding for the same sequence allows us to check the relative effect of the change in periodicity (original vs. shuffled) and in the alphabet features (letter vs. word).

Looking at the top Panel of Fig. 5.14 it is evident how the Dr. Suess poem, thanks to its extremely repetitive prosody is much more deterministic than the semantically dense Dantes poem. The shift from letter to word coding provoked a marked decrease in determinism, looking at the poem in the Appendix 2 we can understand why: the poem is fed into the RQA algorithm as a continuous string of letters (letter coding) or as a continuous string of words (each word being correspondent to a given

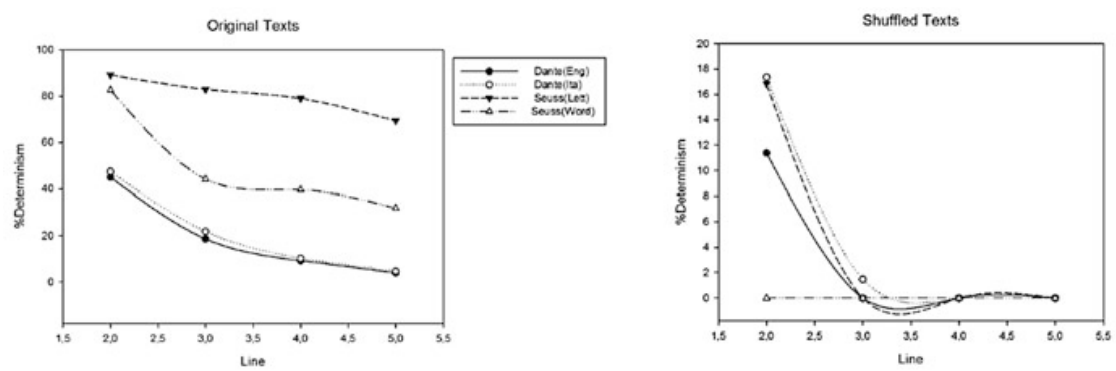


Fig. 5.14 Symbolic strings from spoken languages. The distribution of the % Determinism always decreases at increasing values of line before (*left panel*) and after (*right panel*) shuffling. See the text for further explanation. The analyzed strings are reported in Appendix 2

symbol). The possibility to score deterministic (it is important to stress determinism is the ratio between recurrence occurring in subsequent times with the total number of recurrences) patch for the letter coding is not restricted to the single word repetition but to all the different frameshifts of the underlying semantic recurrence. Thus, for example the recurrence in two parts of the poem of the statement I do not like (idonotlike in the compact version) results in the following consecutive (and thus deterministic) recurrences at $ED = 4$: idon; dono; notl; otli; tlik; like.

Notwithstanding this segmentation effect (especially relevant at high line values) both the Letter and Word coding of Dr. Suess poem are very different from their shuffled counterparts (see bottom panel) so demonstrating the presence of a strong periodic ordering. The amount of prosodic order of Dante’s poem was identical in the English and Italian versions (as a matter of fact English version was a poetic translation by the famous American poet Henry W. Longfellow who purportedly maintained the original prosody of the Italian text), is much lower than the Dr. Suess Green Eggs and Ham notwithstanding that (especially for higher line values) is markedly different from the shuffled versions. The above results indicate the decoupling of semantic and syntactic structures that strongly limits our possibilities to interpret long textual strings written in an unknown language. Notwithstanding that, RQA is sensible even to subtle nuances of the self-correlation structure of the analyzed series (e.g. the Dante prosody) even when presented with different symbols (English version identical to the Italian one), this high sensitivity can have a crucial importance when trying and use syntactic structure as a first step into the code cracking.

This point can be appreciated in the analyses reported in Fig. 5.15 where the determinism distributions are reported for three different DNA molecules.

The analyzed sequences are the following:

- the virus HIV genome (HIVDNA), that is an information dense string corresponding to the genes of nine viral proteins with no introns or noncoding regions. Thus, such DNA sequence fully corresponds to protein sequences and can be assimilated to the Dante texts in the linguistic examples.

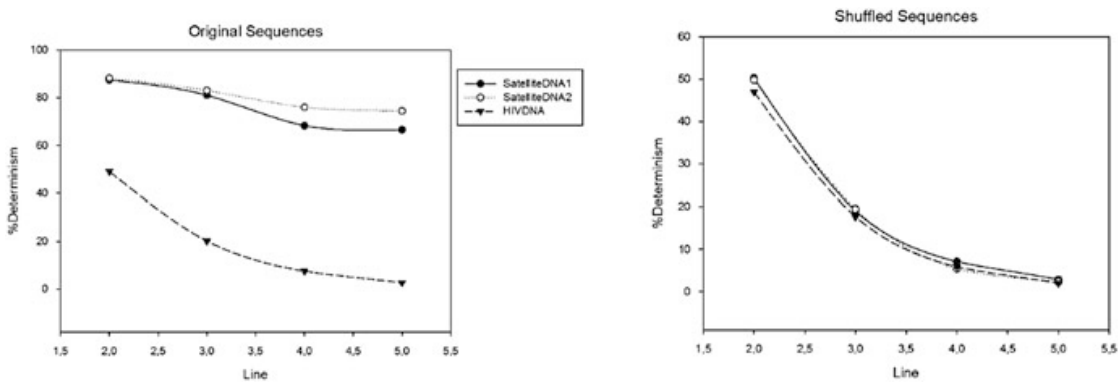


Fig. 5.15 % Determinism distribution of polynucleotidic strings. % Determinism distribution of alphanumeric strings from DNA molecules at increasing values of line before (*left panel*) and after (*right panel*) shuffling. See the text for further explanation. The strings of satelliteDNA1, 2 are reported in Appendix 3; as for the HIVDNA, the first 4,900 of the 9,700 points provided in the RQA software available at the C. Webber Jr. page (<http://homepages.luc.edu/~cwebber/>)

- two noncoding, very repetitive sequences of Satellite DNA (SatelliteDNA1, SatelliteDNA2).¹

Satellite DNA, consisting of very large arrays of tandemly repeating, noncoding DNA, can be taken as the analogue of Dr. Suess poem. It is worth noting that, while not involved in protein, nevertheless Satellite DNA is endowed with a biological role (even if probably linked to the chromatin packing) and thus its syntactical structure, at odds with coding DNA, is deeply intertwined with its semantics (just as in the case of Dr. Suess where the obtaining of a rhyme was the main goal of the text).

As a matter of fact, Figs. 5.14 and 5.15 are very similar: linguistic texts are not basically different from DNA molecules, the effect of rhymes is the same of the effect of repeats along the sequence, the semantic/syntactic link is masked (HIVDNA, Dante) when it is filtered by an autonomous receiver: chemico-physical relations with the environment, 3D structure of coded proteins for HIVDNA, human reader attaching a specific meaning to the symbols sequence in the case of Dante poem. On the contrary when the sense of the message is strictly intertwined with the syntactic structure of the string: tandem repeats for satellite DNA corresponding to the structural role of the molecule in chromatine packing, rhymes of Dr. Suess poem, RQA allows for an immediate quantification of relevant information.

As expected, the repetitive character of Satellite DNA was paralleled by a very strong deterministic structure (Fig. 5.15, left panel), practically absent in HIV DNA

¹Satellite DNA is the main component of functional centromeres, and form the main structural constituent of heterochromatin, i.e. the densely packed, non expressed part of DNA molecule. The name “satellite DNA” refers to how repetitions tend to produce a different frequency of the nucleotides adenine, cytosine, guanine and thymine, and thus have a different density from bulk DNA, such that they form a second or ‘satellite’ band when genomic DNA is separated on a density gradient [28].

where the semantics (protein coding) did not appear at the syntactic level. After shuffling (right panel) the three DNA sequences showed identical determinism, pointing to the coding commonality (four letters strings) of the DNA molecules.

5.6 Conclusion

To our opinion, any statistical modeling of structure-function relationships of biopolymers should, at the present stage, take advantage of the following two characteristic features of Recurrence Quantification Analysis: (a) the relatively easy dealing with different coding scheme of symbolic and chemico-physical information embedded into the primary structures, and (b) the sensitivity to local and subtle correlation structures (interaction sites as the case of in P53, short noncoding regions interspersed into genomic DNA, etc.).

In the presence of a closed and well defined system (like in the case of protein studies) the RQA potentiality can be fully exploited, especially since the same protein molecule hosts physically independent metrics (sequence, chemico-physical and spatial metrics) taken to interaction by the specific molecule of interest. RQA features allow for an immediate, model independent appreciation of such between-metrics relations, so locating RQA at the exact core of the biological problem as evidenced by the efficacy of mixed metrics descriptors as *ordK* or the kurtosis of chemico-physical recurrences along the sequence.

The analogy of the protein contact network with a signal transmission line maximizing the efficiency of information transit across the molecule, so that it can work as a coherent whole, is paralleled by the equivalence of the recurrence plots with graph adjacency matrixes, thus allowing for an immediate translation of complex network formalisms into RQA terms .

Appendix 1: Cryptography

Cryptography was a strategically crucial discipline during the Second World War: the decipherment of hidden information in encrypted messages (like in the case of the cracking by the allies of the German code generated by the system Enigma) was based upon the notion that any human language, despite its apparent randomness and arbitrariness, is endowed with regularities of various kinds (e.g. the relative abundance of words of given length, the juxtaposition of pairs of symbols, etc.) and that no masking code can obscure the code-independent features typical of the original language. These code-independent features are supposed to derive from some general invariants common to all languages, like the so called Zipfs law [29] stating the frequency of occurrence of the words in any kind of (sufficiently long) text written in any language is negatively correlated with the number of letters according to a power law (Fig. 5.16).

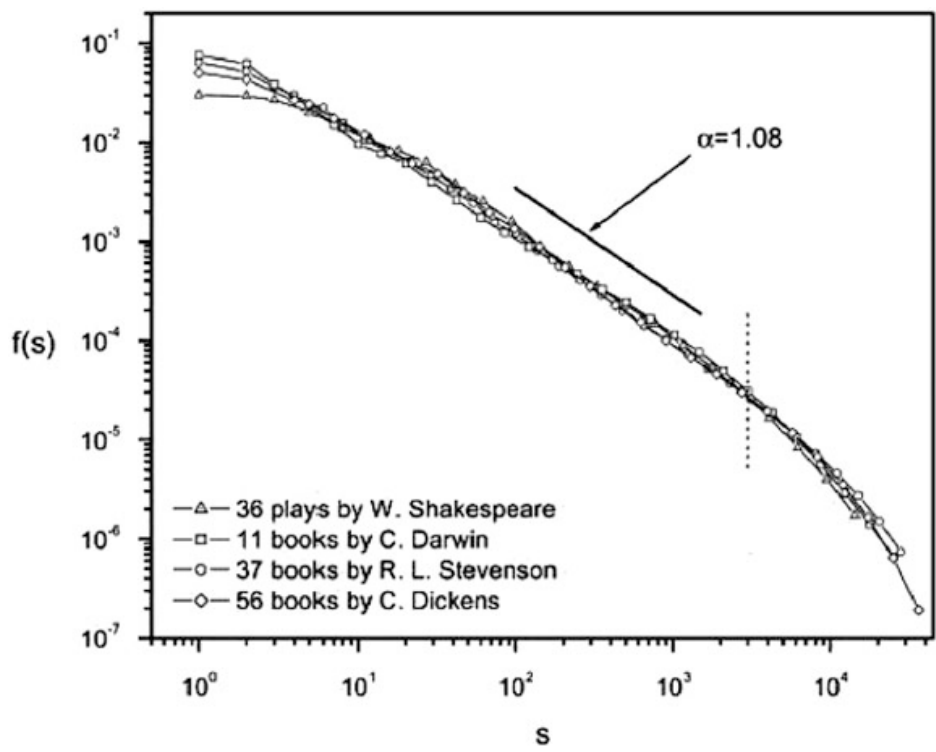


Fig. 5.16 Scaling of the frequency of occurrence of words on word length as computed from different book collection

Figure 5.16 reports (on a double-logarithm scale) the strictly invariant relation between frequency of occurrence and word length in different book collections. Such regularities are clearly independent from the rich semantic information present in the analyzed books: the observed scaling comes from global constraints linked to the general features of human languages. The fact these general features are largely content independent was considered very important for many investigators involved in the analysis of the DNA sequences: in this way they could skip the specific (and extremely heterogeneous) function of different patches so to concentrate on the global statistical features of the billions letters DNA text.

Appendix 2: Strings from Human Languages

(A) Dante Alighieri - Inferno - I Canto (tercets 1–3) - FILTERED

Nelmezzodelcammindinostravitamiritrovaiperunaselvaoscuracheladirittaviaerasmarrita
Ahiquantoadirqualeraecosadurastaselvaselvaggiaeaspraefortechenelpensierrinoyalapaura
Tanteamarachepocoepiumortemapertrattardelbenchivitrovaiddirodellaltrecosechivhoscorte

**(B) Dante Alighieri - Inferno - I Canto (tercets 1–3) -
NONFILTERED (English Translation by Henry Wadsworth
Longfellow))**

Midway upon the journey of our life
I found myself within a forest dark
For the straightforward pathway had been lost.

Ah me! how hard a thing it is to say
What was this forest savage, rough, and stern
Which in the very thought renews the fear.

So bitter is it, death is little more;
But of the good to treat, which there I found,
Speak will I of the other things I saw there.

(C) Dr. Suess Poem - NONFILTERED

I do not like eggs in the file.
I do not like them in any style.
I will not take them fried or boiled.
I will not take them poached or broiled.
I will not take them soft or scrambled,
Despite an argument well-rambled.
No fan I am of the egg at hand.
Destroy that egg! Today! Today!
Today I say!
Without delay!

(A), (B) and (C) refer to the strings from spoken languages whose % Det is shown in Fig. 5.14. Notice that in all cases RQA was applied after filtering the original texts as indicated in [30]: only in (A), however, the filtered text is shown. (A) and (B) show the first three of the 45 analyzed tercets in Fig. 5.14.

Appendix 3: Nucleotidic Strings

Satellite DNA 1 - GenBank: BI067039.1
Homo Sapiens Genomic Region Containing Hypervariable
Minisatellites, mRNA Sequence

GTCCTCCGCCCCACACTTATGGGGCAGAACCCACACTTCCGGTCCTCCGCTCCACACTTATGGGGCAG
CCCACACTTCTGGTCTCTGCCCCACACTTATGGGGCAGAGTTGGGTGTTCTGCCCCACACTTATGGGGC
ACAGACAGCAGTTCCGGACCTCCACCCACACTTATGGGGCAGAACCCACACTTCCGGTCCTCCGCCCCA
CACTTATGGGGCAGAACCCACAGTTTTGGTCCTCCGCTCCACACTTATGGGGCACAACAACCCACAGTTA
TGGGGCTTATGAGGTTCTGCCCCACACTTATGGGGCAGACAGCAGTTCTGGTCCTCCGCCCCACACTT
ATGGGGCAGAACCCACACTTCCGGTCCTCCGCCCCACACTTATGGGGCAGAACCCACACTTCCGGTCCTC
CGCTCCACACTTATGGGGCAGAGCCACACTTCTGGTCCTCTGCCCCACACTTATGGGGCAGAGCTGGGG
GTCCTACCCACACTTATGGGGCAGAACCCACAGTTCCGGTCCTCCACCCACACTTATGGGGCAGAGCT
GGGGATTCTGTGCCACACTTATGGGGCAGAACCCACAGTTCCGGCCCTCCGCCCCACACTTATGGGGCAG
NNCNNGCNGNCGGG

Satellite DNA 2 - GenBank: BM439581.1
Homo Sapiens Genomic Region Containing Hypervariable
Minisatellites, mRNA Sequence

GGCACAGCTGGGGATTCTGCCCCACACTTATGCGGCACAACCCACAGTTCTGGTCCTCTCCCCACACTT
ATGGGGCACAACAACCCACAGTTATGGGGCTTATGAGGTTCTGCCCCACACTTACGGGGCAGACAGCA
GTTCCAGTCTCCGCCCCACACTTATGGGGCAGAACCCACAATTCCGGACCTCTGCCCCACACTTACGGG
GCACAGCTGGGGATTCTGCCCCACACTTATGGGGCACAACCCACAGTTCTGGTCCTCTCCCCACACTTA
TGGGGCAGAACCCACACTTCCGGTCCTCCGCCCCACACTTAGGGAGCAGAACCCACACTTCCGGTCCTCC
GCCCCACACTTATGGGGCACAACAACCCACAGTTATGGGGCCTATGAGGTTCTGCCCCACACTTATGGGG
CACAGACAGCAGTTCCGGACCTCTGCCCCACACTTATGGGGCAGAGTTGGGGGTCTACCCACACTTAT
GGGGCAGAACCCACAGTTCCGGACCTCCGCCCCACACTTATGGGGCAGAACCCACACTTCCGNACCTCTG
CCCCACACTTATGGGGCACA

References

1. J.P. Eckmann, S.O. Kamps, D. Ruelle, Recurrence plots of dynamical systems. Eur. Phys. Lett. **4**, 973–977 (1987)
2. C.L. Webber Jr., J.P. Zbilut, Dynamical assessment of physiological systems and states using recurrence plot strategies. J. Appl. Physiol. **76**, 965–973 (1994)
3. N. Marwan, N. Wessel, U. Meyerfeldt, A. Schirdewan, J. Kurths, Recurrence plot based measures of complexity and its application to heart rate variability data. Phys. Rev. E **66**, 026702–1026702–7 (2002)

4. N. Marwan, M.C. Romano, M. Thiel, J. Kurths, Recurrence plots for the analysis of complex systems. *Phys. Rep.* **438**, 237–329 (2007)
5. D.B. Vasconcelos, S.R. Lopes, R.L. Viana, J. Kurths, Spatial recurrence plots. *Phys. Rev. E* **73**, 056207 (2006)
6. A. Giuliani, R. Benigni, J.P. Zbilut, C.L. Webber Jr., P. Sirabella, A. Colosimo, Nonlinear signal analysis methods in the elucidation of protein sequence structure relationships. *Chem. Rev.* **102**, 1471–1491 (2002)
7. G. Oliva, L. Di Paola, A. Giuliani, F. Pascucci, R. Setola. Assessing protein resilience via a complex network approach. In *Network Science Workshop (NSW), 2013 IEEE 2nd*, (IEEE 2013), pp. 131–137
8. L. Di Paola, M. De Ruvo, P. Paci, D. Santoni, A. Giuliani, Protein contact networks: an emerging paradigm in chemistry. *Chem. Rev.* **113**, 1598–1613 (2013)
9. C.L. Webber Jr., A. Giuliani, J.P. Zbilut, A. Colosimo, Elucidating protein secondary structures using alpha carbon recurrence quantifications. *Proteins Struct. Funct. Genet.* **44**, 292–303 (2001)
10. M. De Ruvo, A. Giuliani, P. Paci, D. Santoni, L. Di Paola, Shedding light on protein-ligand binding by graph theory: the topological nature of allostery. *Biophys. Chem.* **165–166**, 21–29 (2012)
11. S. Vishveshwara, K. Brinda, N. Kannan, Protein structure: insights from graph theory. *J. Theor. Comput. Chem.* **1**, 187–212 (2002)
12. C. Hansch, D. Hoekman, H. Gao, Comparative qsar: toward a deeper understanding of chemico-biological interactions. *Chem. Rev.* **96**, 1045–1075 (1996)
13. S. Miyazawa, R.L. Jernigan, Estimation of effective inter-residue contact energies from protein crystal structure: quasi-chemical approximation. *Macromolecules* **18**, 534–552 (1985)
14. J. Kyte, R.F. Doolittle, A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982)
15. A. Porrello, S. Soddu, J.P. Zbilut, M. Crescenzi, A. Giuliani, Discrimination of single aminoacid mutations of the p53 protein by means of recurrence quantification analysis. *Proteins Struct. Funct. Bioinf.* **55**, 743–755 (2004)
16. A. Giuliani, R. Benigni, P. Sirabella, J.P. Zbilut, A. Colosimo, Nonlinear methods in the analysis of protein sequences: a case study in rubredoxins. *Biophys. J.* **78**, 136–149 (2000)
17. S. Soddu, G. Blandino, R. Scardigli, R. Martinelli, M.G. Rizzo, M. Crescenzi, A. Sacchi, Wild-type p53 induces diverse effects in 32d cells expressing different oncogenes. *Mol. Cell. Biol.* **16**, 487–495 (1996)
18. T. Soussi, Y. Legros, R. Lubin, K. Ory, B. Schlichtholz, Multifactorial analysis of p53 alteration in human cancer: a review. *Int. J. Cancer* **57**, 1–9 (1994)
19. H.J. Jeffrey, Chaos game representation of gene structure. *Nucleic Acid Res.* **18**, 2163–2170 (1990)
20. O.C. Kulkarni, R. Vigneshwar, V.K. Jayaraman, B.D. Kulkarni, Identification of coding and noncoding sequences using local holder exponent formalism. *Bioinformatics* **21**, 3818–3822 (2005)
21. R.N. Mantegna, S.V. Buldyrev, A.L. Goldberger, S. Havlin, C.K. Peng, M. Simons, H.E. Stanley, Linguistic features of noncoding dna sequences. *Phys. Rev. Lett.* **73**, 3169–3175 (1994)
22. E.A. Feingold, P.J. Good, M.S. Guyer, S. Kamholz, L. Liefer, K. Wetterstrand, F.S. Collins et al., The encode (encyclopedia of dna elements) project. *Science* **306**, 636–640 (2004)
23. J.O. Andersson, S.G. Andersson, Pseudogenes, junk dna, and the dynamics of rickettsia genomes. *Mol. Biol. Evol.* **18**(5), 829–839 (2001)
24. C. Frontali, E. Pizzi, Similarity in oligonucleotide usage in introns and intergenic regions contributes to long-range correlation in the *Caenorhabditis elegans* genome. *Gene* **232**, 87–95 (1999)
25. E. Bultrini, E. Pizzi, P. Del Giudice, C. Frontali, Pentamer vocabularies characterizing introns and intron-like intergenic tracts from *Caenorhabditis elegans* and *Drosophila melanogaster*. *Gene* **304**, 183–192 (2003)

26. F. Orsucci, A. Giuliani, C.L. Webber, J.P. Zbilut, P. Fonagy, M. Mazza, Combinatorics and synchronization in natural semiotics. *Phys. A* **361**, 665–676 (2006)
27. G. Leonardi, The study of language and conversation with recurrence analysis methods. *Psychol. Lang. Commun.* **16**, 165–183 (2012)
28. B. John, G.L. Miklos, Functional aspects of satellite dna and heterochromatin. *Int. Rev. Cytol.* **58**, 1–114 (1979)
29. M.A. Montemurro, Beyond the Zipf-Mandelbrot law in quantitative linguistics. *Physica A* **300**, 567–578 (2001)
30. C.L. Webber Jr., J.P. Zbilut, Recurrence quantification analysis of nonlinear dynamical systems, in *Tutorials in Contemporary Nonlinear Methods for the Behavioral Sciences*, Chap. 2 (National Science Foundation, Washington, DC, 2005) pp. 26–94

Chapter 6

Dynamic Coupling Between Respiratory and Cardiovascular System

Federica Censi, Giovanni Calcagnini, and Sergio Cerutti



Abstract The analysis of non-linear dynamics of the coupling among interacting quantities can be very useful for understanding the cardiorespiratory and cardiovascular control mechanisms. In this chapter RP is used to detect and quantify the degree of non-linear coupling between respiration and spontaneous rhythms of both heart rate and blood pressure variability signals. RQA turned out to be suitable for a quantitative evaluation of the observed coupling patterns among rhythms, both in simulated and real data, providing different degrees of coupling. The results from the simulated data showed that the increased degree of coupling between the signals was marked by the increase of PR and PD, and by the decrease of ER. When the RQA was applied to experimental data, PD and ER turned out to be the most significant variables, compared to PR. A remarkable finding is the detection of transient 1:2 PL episodes between respiration and cardiovascular variability signals. This phenomenon can be associated to a sub-harmonic synchronization between the two main rhythms of HR and BP variability series.

F. Censi (✉) • G. Calcagnini

Italian National Institute of Health, Viale Regina Elena 299, 00161 Roma, Italy

e-mail: federica.censi@iss.it; giovanni.calcagnini@iss.it

S. Cerutti

Polytechnic University of Milan, Piazza Leonardo da Vinci 32, 20133 Milano, Italy

e-mail: sergio.cerutti@polimi.it

6.1 Introduction

Physiological rhythms are central to life: complex or simple rhythms usher the activities of living systems, involving several phenomena such as locomotion, metabolism, and hormonal regulation. Some rhythms are maintained throughout life, and even a brief interruption leads to death. Other rhythms may be expression of pathological conditions and thus indicate an unpaired equilibrium. Oscillating phenomena in living systems range within a wide frequency band, from firing rates of nervous fibers to circadian oscillations up to monthly hormonal changes. The study of biological oscillations is important for a reliable comprehension of the living system organization and functioning. The rhythms interact with one another and with the external environment. The analysis of the genesis of a biological rhythm and the interpretation of the conditions which change its temporal and morphological features allow to advance hypotheses on the mechanisms underlying several living processes. In addition, variations of rhythms and/or of their interaction outside normal limits, or appearance of new rhythms and coupling conditions where none existed previously, are associated with malfunctioning, disease, transient or permanent pathological states. An understanding of the mechanisms of physiological rhythms requires an approach that integrates mathematics and physiology.

Cardiovascular variables are characterized by oscillations even in stationary physiological conditions. In the range 0.03–0.4 Hz, such oscillations reflect complex control mechanisms mediated by the autonomic nervous system (ANS). Over the last 20 years a great deal of research regarding short-term heart rate and blood pressure variability has been carried out using linear and non-linear approaches.

Although both approaches for the analysis of the heart rate variability provide important pieces of information about the ANS, other physiological variables as well as external stimuli are involved in the autonomic nervous system control mechanisms. The interactions between respiration, heart rate and blood pressure fluctuations, which reflect the cardiovascular and cardiorespiratory couplings, are considered to be of paramount importance for the study of the functional organization of the ANS. The cardiovascular control is performed by a number of mechanisms and pathways which simultaneously act in either synergic or antagonistic manner. Feedback control loops mediated by baroreceptors, chemoreceptors or mechanoreceptors activity regulate blood pressure, heart rate and respiratory dynamics; oscillators located in the brain stem and in the heart modify respiration and vasomotion; local peripheral oscillators control microvascular resistances and flows. When responding to the need of an organism the interactions among these oscillator can provoke prevalence of one or more oscillating structures on the others, and/or synchronization phenomena.

The analysis of non-linear dynamics of the coupling among interacting quantities turned out to be more adequate for understanding the cardiorespiratory and cardiovascular couplings than traditional linear methods [1–5]. Several non-linear techniques have been applied to cardiorespiratory data in order to analyze their coupling mechanisms [6–8].

The recurrence plot analysis was introduced by Eckmann et al. [9] in physics as a graphical method designed to locate recurrent and intermittent patterns in experimental time series. This technique was then enhanced by the introduction of several variables to quantify the recurrence plot features [10], which turn out to be useful in assessing transient behavior of experimental data sets in different scientific explorations [11–18].

The recurrence plots were first introduced to analyze the regularity of single time series. According to this traditional approach a recurrence plot (RP) is a graphical tool which can be used to observe whether recurrences characterize the dynamics of a given system. After the introduction of the recurrence map to analyze single time series, Porta et al. [19] suggested the use of this tool to analyze the interaction among signals. Nowadays, the approaches to study the interaction among different time series by using the recurrence map, are the recurrence, the cross-recurrence and the joint recurrence plot analysis [20, 21]. The first merely compares the periodicity of two (or more) time series, without employing the time embedding procedure [19, 22, 23], and it could be considered a sort of multivariate RP; the other approaches are somewhat more complex and quantify the recurrences in a distance matrix between the two embedded series.

In this chapter RP will be used to detect and quantify the degree of non-linear coupling between respiration and spontaneous rhythms of both heart rate and blood pressure variability signals. The analysis could of course be extended by evaluating the recurrence of the three signals together, but this approach would render the clinical interpretation of the results much more difficult.

6.2 Recurrence Plot Analysis for the Interaction Between Two Series

The interactions among rhythms causing frequency and/or phase synchronization represent a particular class of non-linear phenomena. Two interacting oscillators are N:M phase locked (PL) if marked events of one oscillator occur at fixed phases of the other oscillator. For this particular interaction, the spectral coherence (linear method) is often found not to be significant (but for a 1:1 coupling), since no phase synchronization or weak correlation between amplitudes are present. The notion of phase synchronization, which is often referred to as entrainment or phase locking, has been extensively used for analyzing and modeling interaction between different physiological sub-systems. Entrainment phenomena, arising from periodic stimulation of non-linear oscillators, have been studied in a number of biomedical applications [24–26]. Particularly, a number of studies focused on the synchronization between heartbeat or heart rate fluctuations and respiratory rhythm [27–30].

In 1994, Porta et al. [19] first suggested the use of the RP to measure synchronization among signals and to analyze the interactions between two or more signals.

Following this approach, in a bidimensional case (time series $x(t)$ and $y(t)$) RP is a representation of the normalized distance between the points $[x(i), y(i)]$ and $[x(j), y(j)]$, plotted in the time-to-time domain (i, j) . The time series “belong” to different systems: time series $x(t)$ is a measurement from system X and y is the measurement of the system Y .

The normalized distance $D(i, j)$ is computed as follows:

$$D(i, j) = \sqrt{\frac{[x(i) - x(j)]^2}{\text{var}(x)} + \frac{[y(i) - y(j)]^2}{\text{var}(y)}} \quad (6.1)$$

where $\text{var}(\cdot)$ indicates the time series variance. If the two points are sufficiently close to each other, i.e. the distance $D(i, j)$ is lower than a fixed cut-off value, a dot is plotted in (i, j) . A recurrent point in (i, j) means that the interaction between the signals in the instant i is almost the same as in the instant j , i.e. the interaction is recurring. If $D(i, j)$ is higher than the threshold, (i, j) is a not recurrent point.

The distance threshold is usually set to as a percentage of the maximum distance value, computed as the 95th percentile [10]. Since the distance between a point and itself is zero ($D(i, i) = 0$), the main diagonal line in the plot is always visible. Moreover the plot is symmetric, since $D(i, j) = D(j, i)$ according to the definition of distance given in (6.1).

A recurrent point indicates an isolated recurrence of an amplitude relationship between the signals. Line segments parallel to the main diagonal come from points close to each other successively forward in time $((i, j), (i + 1, j + 1), \dots, (i + T, j + T))$. Thus, a recurrent diagonal line indicates a stable recurrence of the relationship for a time interval corresponding to the length of the diagonal (T). The time interval separating adjacent diagonals is the recurrence period. Recurrent points organized in rows and columns do not convey any information about the timing of the periodicity, since they are not successively forward in time.

6.2.1 Application to Test Data Set

Figure 6.1 makes it clear how a RP can detect recurrent structures on a test data set. In this figure the RP of the x - and the z -components of the Lorentz system is shown (lower panel), together with the time courses of the two variables (upper panel). In the RP, the isolated recurrent points are plotted in gray while those organized in diagonal lines are in black.

Four sections have been selected, in the plot, each showing a different kind of coupling. In section I the RP features diagonal segments recurring without temporal periodicity, indicating that there exist non-periodical intervals of synchronization (i.e. the signals show a stable phase relationship). Such a periodicity is instead present in section III, where diagonal lines are separated by the same time intervals (recurrent period). In section II and IV both signal amplitudes are close to zero and

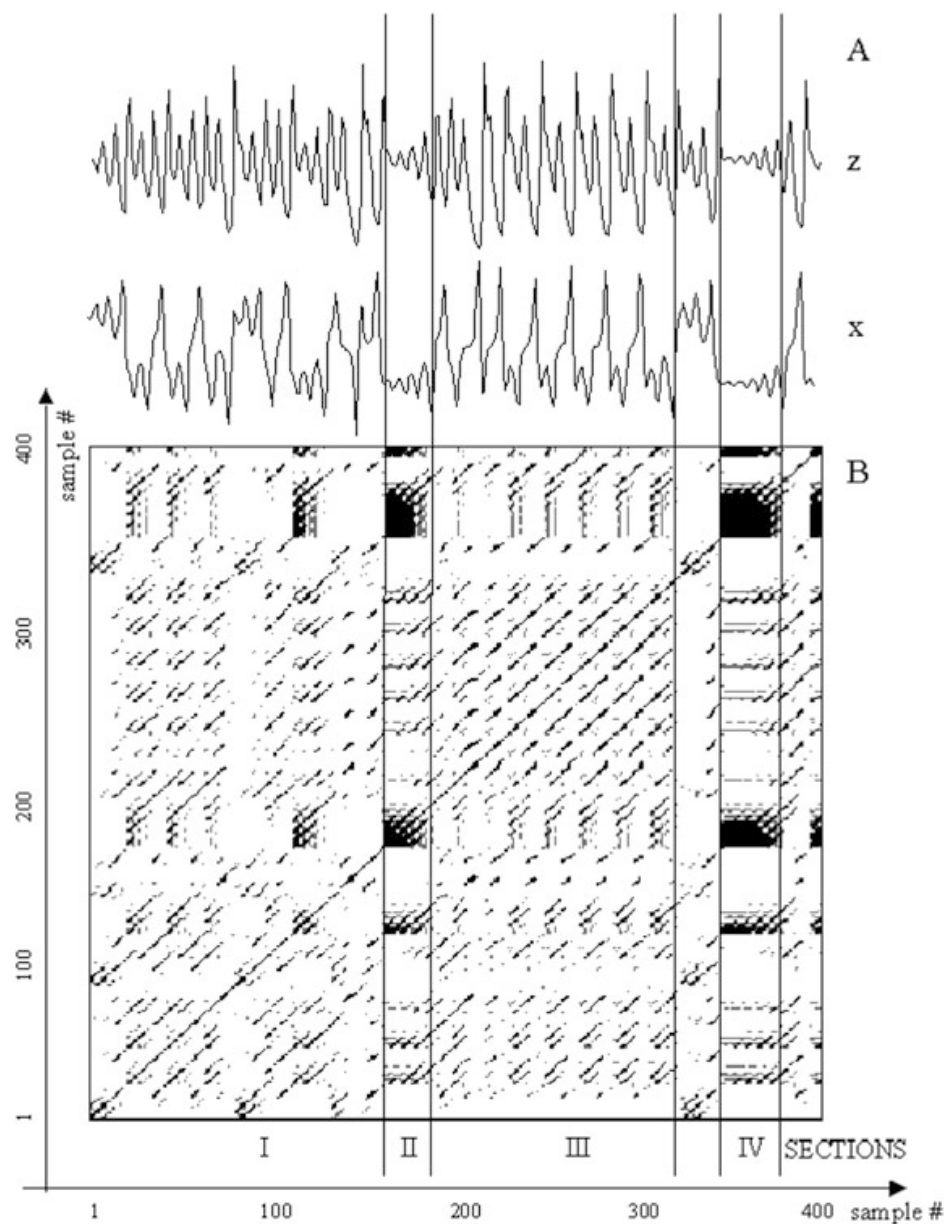


Fig. 6.1 Trackings of the x- and z-component of the Lorentz system (panel A) and the recurrence plot associated (panel B), with the isolated recurrent points in *gray* and those forming diagonal lines in *black*. The *vertical lines* indicate four sections, each showing a different kind of coupling between the two series. The values of the RPQ variables for the overall recurrence plot are: PD = 65.2 %, PR = 15.1 %, ER = 2.9

recurrent points in the map are organized not only in diagonals but also in rows and columns. In fact all points within these two sections are recurrent with each other and not only with those successively forward in time.

Figure 6.2 shows the deterministic structure of a 1:1 Phase locking (PL) phenomena (panels A and B), with diagonals recurring every period. RP of a 1:2 phase locking (panel D) shows diagonal lines recurring every period of the slower oscillator (series 2, panel B), and isolated points recurring each period of the faster one (series 1, panel A).

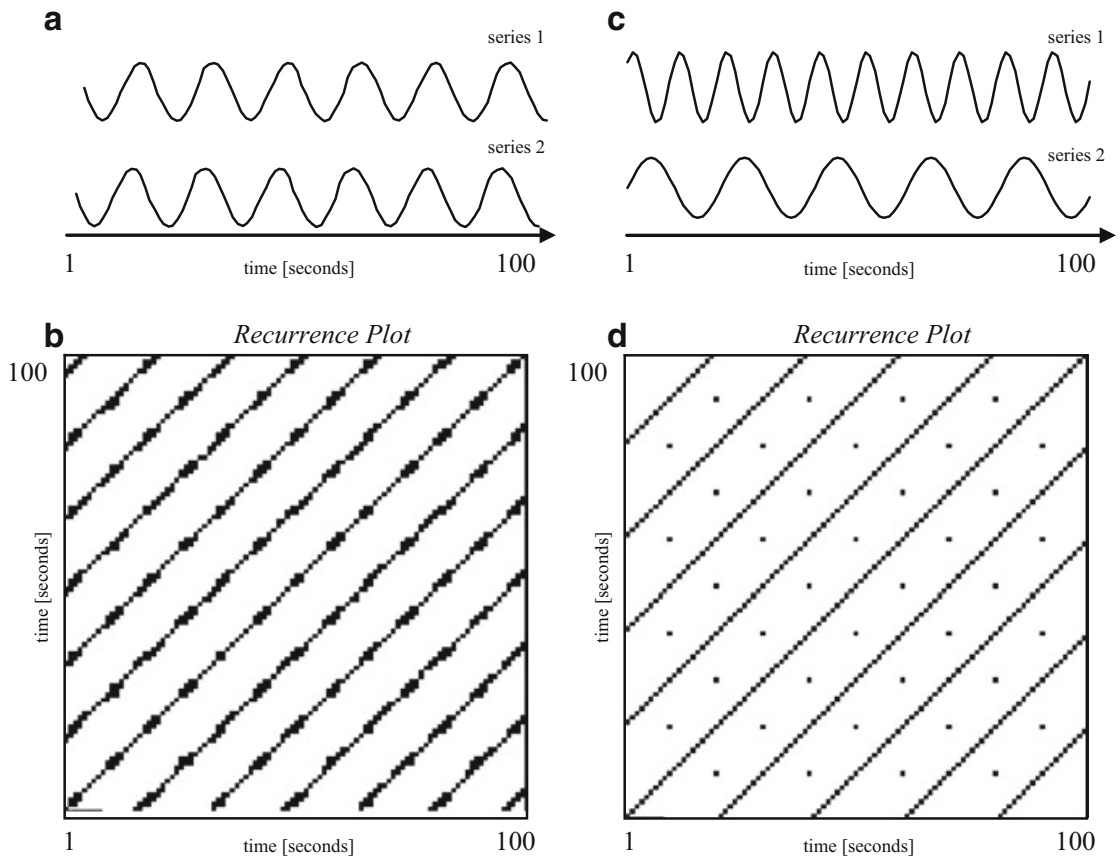
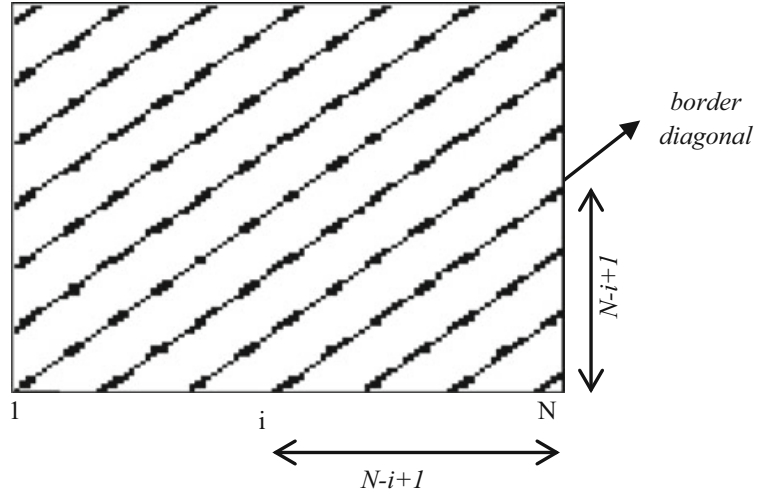


Fig. 6.2 Example of RP for a PL 1:1 and PL 1:2 (panels **B** and **D** respectively) between the two series shown in panels **A** and **C**, respectively

6.2.2 Quantification

Although these plots can provide a qualitative assessment of the system, as recurrent points and diagonals increase, the visual inspection is not easy. Thus, there is the need to quantify the organization of recurrence points into specific patterns.

Three descriptors of the RP plot features are commonly used [10]. One is the percentage of plot occupied by recurrent points, namely Percent Recurrence (PR), defined as the percent ratio between the number of recurrent points and the total number of all possible points. It quantifies the number of time instants characterized by a recurrence in the signals interaction: the more frequent the signal dynamics, the higher the PR value. Another descriptor is the percentage of recurrent points forming diagonal lines called Percent Determinism (PD), defined as the percent ratio between the number of recurrent points forming diagonals longer than a fixed threshold (l) and the total number of recurrent points. This variable discriminates between the isolated recurrent points and those forming diagonals. Since a diagonal represents points close to each other successively forward in time, PD also contains the information about the duration of a stable interaction: the longer the interactions, the higher the PD value. A third variable is the Entropy of Recurrence (ER), which is the Shannon entropy of diagonal length distribution, defined as follows:

Fig. 6.3 Border diagonals in the RP

$$ER = - \sum_{L=2}^N p_L \log_2 p_L \quad (6.2)$$

p_L is the probability that a diagonal is L -point long, with L ranging from the chosen threshold l to the main diagonal length (which correspond to the total number of points of the analyzed time series, i.e. N). p_L is defined as the ratio between the number of L -point long diagonals, and the total number of diagonals. ER is measured in bits of information, because of the base-2 logarithm in Eq. (6.2). Thus, whereas PD accounts for the number of the diagonals, ER quantifies the distribution of the diagonal line lengths. The more different the lengths of the diagonals, the more complex the deterministic structure of the RP.

A more complex dynamics will require a larger number of bits (ER) to be represented than a deterministic dynamics such as a 1:1 phase locking phenomena reported in Fig. 6.2, panels A and B. However the estimation of the ER on the basis of Eq. (6.2) may lead to its over-estimation since in 1:1 phase locking for example, the diagonal lengths are different because of a border effect and not because of the recurrence end. A diagonal starting from position $(i, 1)$ and ending at position $(N, N-i+1)$ is called a border diagonal (Fig. 6.3). A border diagonal is long $N-i+1$ points. Of course, the distribution of the border diagonals leads to a misleading (over-) estimation of the ER, since for the deterministic 1:1 phase locking pattern shown in Fig. 6.2 (panels A and B), six border diagonals with different lengths are detectable.

In order to avoid the border effect, a corrected diagonal length (CDL) has been introduced in such a way that a border diagonal is considered to be long as much as the main diagonal (N samples). In particular the correction has been applied to all L -long border diagonals (with $L > l$) according to the following formula:

$$CDL = L \frac{N}{N-i+1} \quad (6.3)$$

Unfortunately, also the number of diagonals may affect the estimation of ER. When the diagonal are a few, ER has a relatively low value, regardless the lengths of the diagonal. In order to take into account the number of diagonals, ER has been divided by the ratio between the number of recurrent points forming diagonals and the total number of recurrent points. This ratio (which is the not-percent PD) varies between 0 and 1, in such a way that if the number of diagonals increases ER decreases and vice versa.

The choice of the threshold value l for the length of the diagonals is crucial for the correct estimation of both PD and ER. The threshold l should be chosen according to the frequency content of the analyzed time series: high values for l lead to the under-estimation of PD and ER, while low values to their over-estimation. The choice of l value implies the definition of the minimum duration of a stable recurrence. A recurrence is considered stable if it lasts at least half period of the faster oscillator. This choice allows the classification of various coupling phenomena, either simulated or experimentally observed, as showed in the following paragraphs.

6.3 Quantification of Recurrences of the Van Der Pol Model

In order to test how the RPs descriptors work, we used the Van Der Pol model. It is described by the following non-linear differential second order equation:

$$\ddot{x} + \varepsilon (x^2 - 1) \dot{x} + \omega_0^2 x = f(t) \quad (6.4)$$

$f(t)$ is the forcing term, and ε is a model parameter (it must be a positive constant). In the absence of forcing term the Van der Pol oscillator exhibits periodic oscillations at ω_0 , sustained by the non-linear term. When $f(t)$ is a periodic term, different kinds of interference occur between the forcing term $f(t)$ and the Van Der Pol oscillator $x(t)$, by varying the amplitude and frequency of the external stimulus. Although this model is not strictly interpretable from a physiological point of view, it well reproduces interaction phenomena observed in experimental data, such as sub-harmonic synchronization and full entrainment.

By varying the amplitude and the frequency of the sinusoidal forcing term $f(t)$, we analyzed three coupling patterns between the forcing term and the Van der Pol oscillator: no interaction, sub-harmonic synchronization, and full entrainment (Fig. 6.4, upper part, panels A–C, respectively). Sub-harmonic synchronization can be obtained by setting the frequency of the sinusoidal forcing term close to twice ω_0 (self-oscillations). Full-entrainment was obtained by applying a forcing term with a relatively high amplitude.

Figure 6.4 shows the RPs and the RQA results for each analyzed coupling pattern (middle part). The degrees of coupling of the three analyzed phenomena can be distinguished by values of RQA (Fig. 6.4, lower part). Particularly, the increase of synchronization between signals, from no interaction to full entrainment, is marked by an increased number of recurrence points and of diagonals in the RPs, which lead

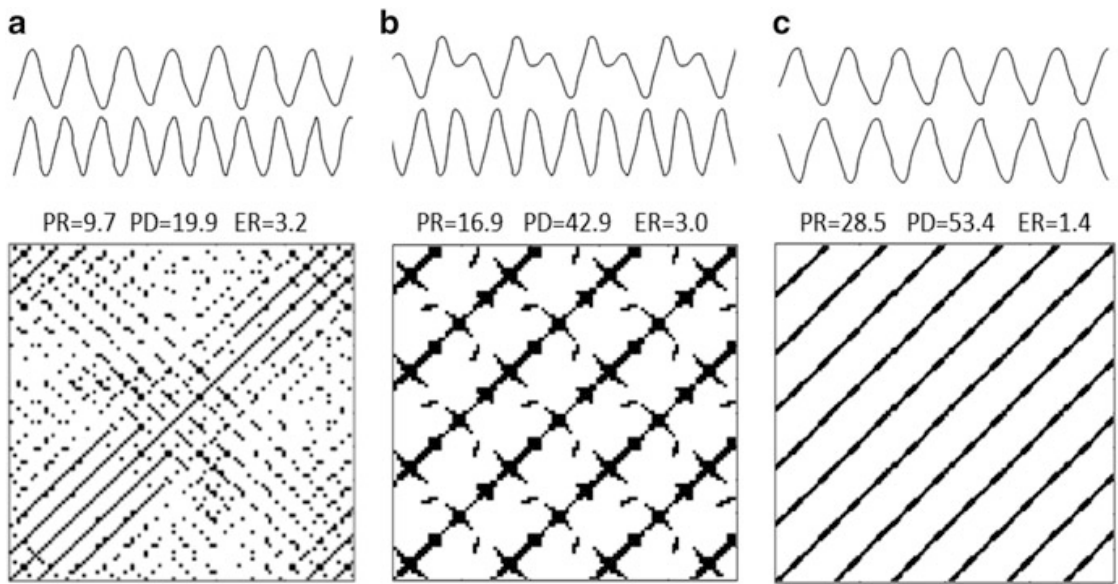


Fig. 6.4 Simulation results obtained with the Van der Pol oscillator. From top to the bottom: the tracks of the oscillator and the forcing term, the RQA results, and the RPs. **A:** no interaction phenomena; **B:** sub-harmonic synchronization; **C:** full entrainment

to increased values of PR and PD. Conversely, ER decreases from no-interaction to full-entrainment. The higher the synchronization between the time series, the more deterministic their interaction, the smaller the ER value. ER is higher when the diagonals have different lengths, regardless the number of the diagonals themselves.

6.4 Cardiorespiratory Synchronization Experiment

The results showed in this chapter refer to an experimental protocol aimed at analyzing the interference between heart rate variability (HRV), blood pressure variability (BPV) and respiration. This protocol was based on the simultaneous acquisition of ECG, Arterial Blood Pressure (ABP) and Respiratory Activity (RA). Surface ECG (II lead) was recorded by an analog electrocardiograph (MCR I, Esaote, Italy). APB was continuously and non-invasively recorded by plethysmographic technique (Finapres, Omheda, USA), and RA was monitored using a pletismographic thoracic belt. The signals were sampled in real-time (sampling frequency: 500 Hz, resolution: 12 bit, DT2801A, Data Translation, USA) and stored for off-line analysis.

We extracted the HRV signal from the low-pass filtered event series (ES) as it is a reliable time domain representation with high temporal resolution. In ES, each beat is replaced by a δ function, and to obtain the HRV series, the ES is low-pass filtered at 0.5 Hz and re-sampled at 2 Hz.

Blood pressure variability (BPV) and the RA series were obtained by a similar low-pass filtering procedure (FIR, 10 order, cut-off frequency 0.5 Hz) and

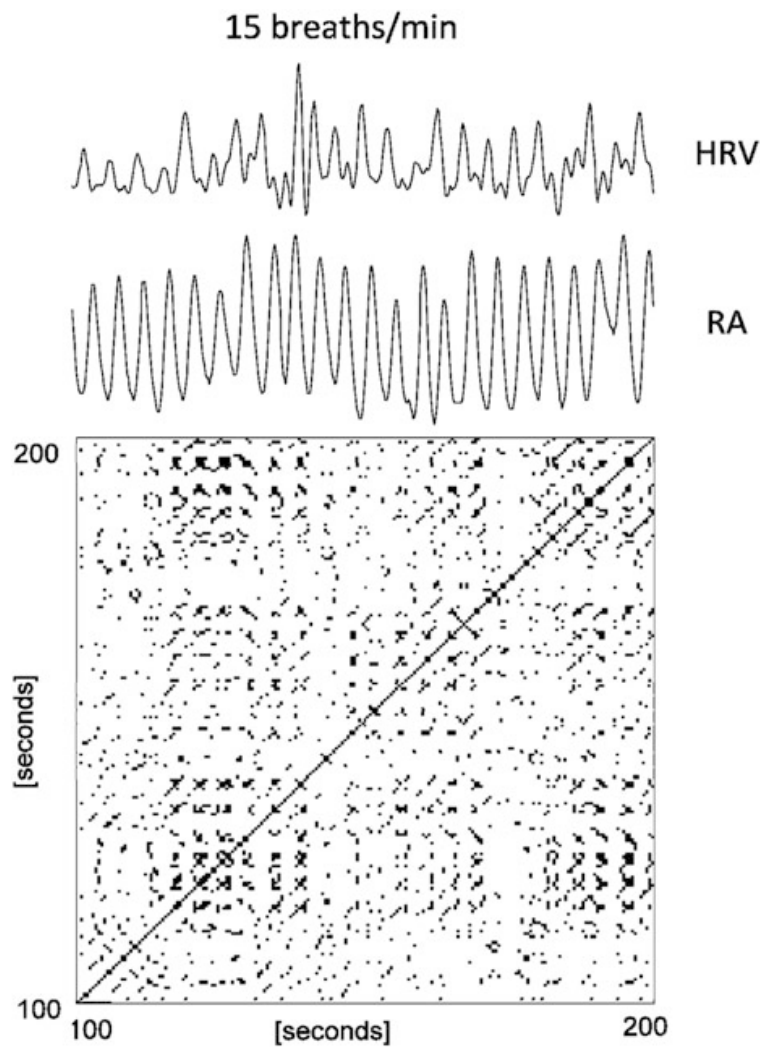


Fig. 6.5 Experimental results obtained during 15 breath/min. 100-second tracks of HRV and respiratory time series are reported, together with the corresponding RP

re-sampling at 2 Hz. According to this procedure, the HRV and RA series were expressed in arbitrary units (a.u., not indicated in the figures), and BPV in mmHg.

The protocol consisted of three stages, with the subject in paced breathing at various rates, lying supine in rest. Each stage lasted 8 min, followed by a 3-min recovery period to allow the periodic calibration of the Finapres. Subjects were instructed to initiate a breath with each tone of a series of auditory cues generated by a computer. Subjects were allowed to adjust their tidal volume to a comfortable level to preserve normal ventilation. Since the acquisitions were performed with the subjects in supine position, subjects were asked to lay on the bed at least 20 min before the acquisition beginning, to adapt themselves to the supine position. In order to avoid movement artifacts, subjects were invited to assume a comfortable position and to avoid muscular movements during the recording.

The selected breathing rates were 15, 12, and 8 breath/min corresponding to 0.25, 0.20 and 0.13 Hz, respectively. These breathing rates allowed to have the High Frequency (HF) component (except for 0.13 Hz) getting progressively closer to the spontaneous Low Frequency (LF) (0.04–0.15 Hz). Figures 6.5, 6.6 and 6.7

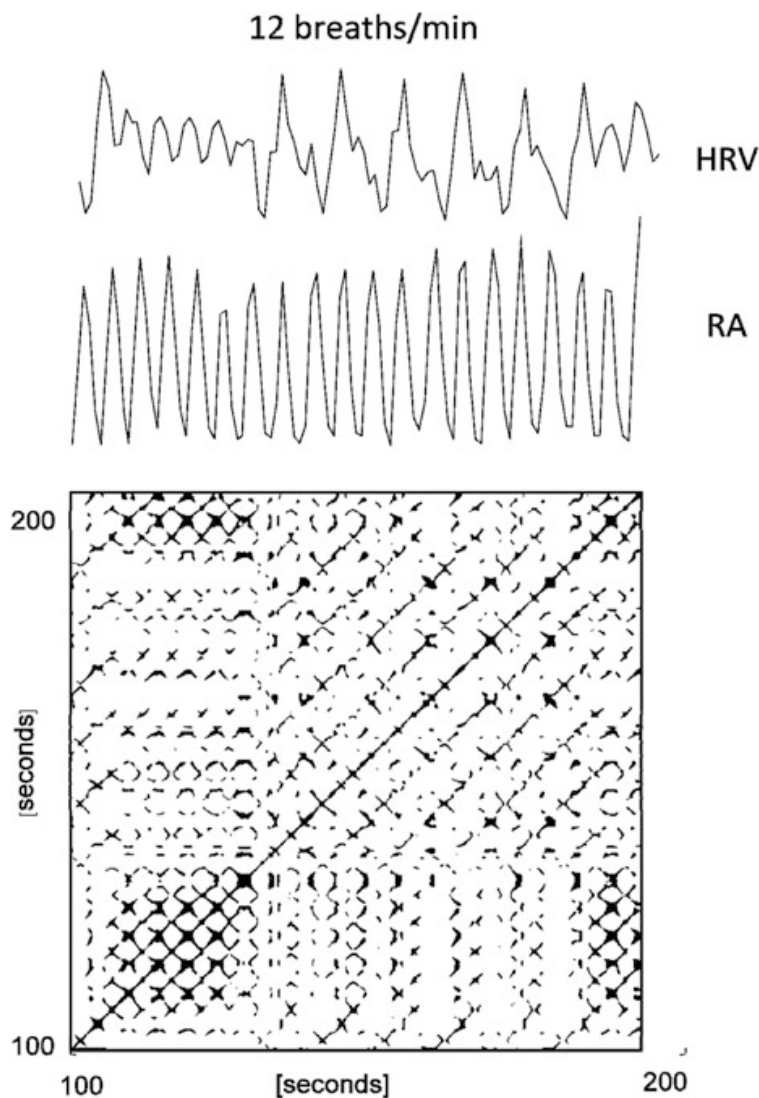


Fig. 6.6 Experimental results obtained during 12 breath/min. 100-second tracks of HRV and respiratory time series are reported, together with the corresponding RP

show the results. 100-second tracks of HRV and respiratory time series are reported, together with the corresponding RP. Note the increase of coupling between the two time series, from 15 to 8 breaths/min stage.

During the 15 breath/min stage the frequency of respiration was 0.25 Hz and the power spectral densities (PSD) of both HRV and BPV signals (not shown) are characterized by an HF component centered at the respiratory frequency and by an LF component which seems not to be influenced by the respiration (center frequency ranging 0.08–0.12 Hz). The RP does not show definite trajectories neither periodicity.

In the 12 breath/min recordings, the interaction between HRV and respiration appeared to have a more complex dynamics. The power spectrum of both HRV showed an LF component clearly centered at half HF frequency (0.10 Hz). Moreover, a more complex, transient interactions between LF and HF components

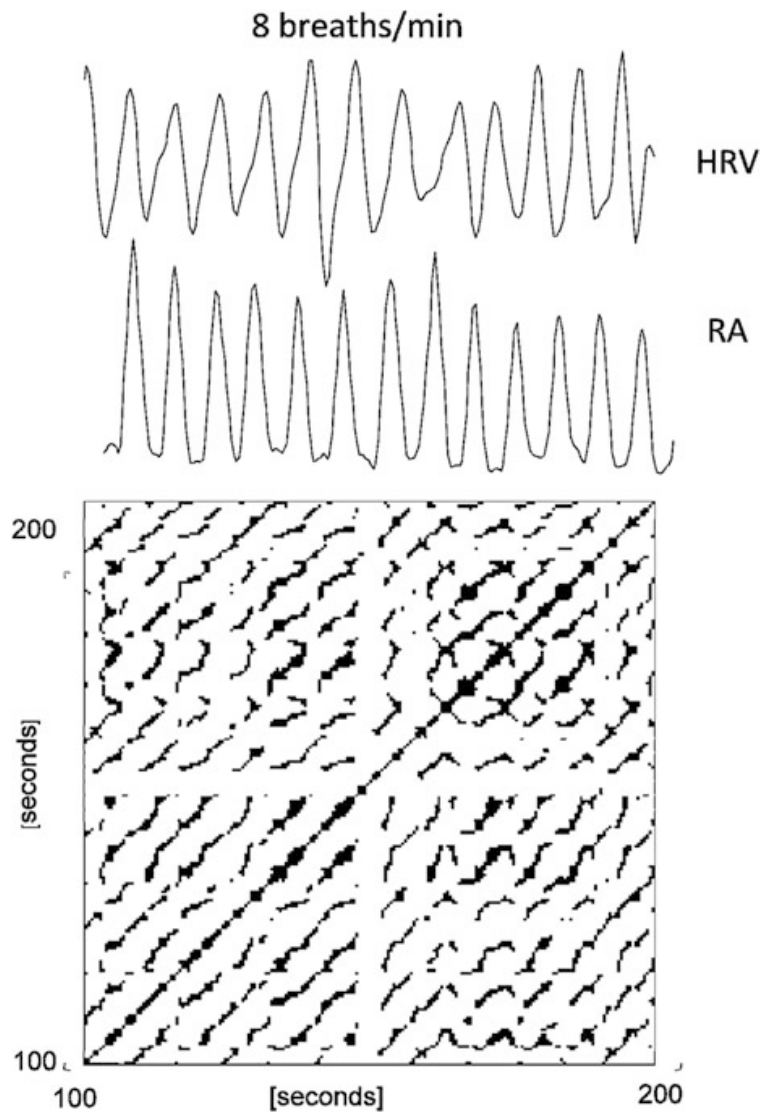


Fig. 6.7 Experimental results obtained during 8 breath/min. 100-second tracks of HRV and respiratory time series are reported, together with the corresponding RP

were observed, which appear to be a 1:2 PL phenomena. This phenomenon appears in the RPs display as diagonal lines recurring every LF period (about 10 s) and isolated points recurring every breath (5 s).

In the 8 breath/min stage only one component (0.13 Hz) and its higher harmonics are visible both in HRV signal: LF and HF components cannot be distinguished anymore and only a unique component centered at the respiratory frequency can be observed and diagonal lines characterize the RP.

Table 6.1 summarizes the results averaged all over the population (mean \pm standard deviation), also reported in Fig. 6.5, panel B.

PR, PD and ER obtained during the 12 breath/min stage are significantly different from those obtained during 15 breath/min stage ($p < 0.001$ for PD and $p < 0.01$ for PR and ER, Wilcoxon’s non-parametric test for paired data). PD and PR obtained during 12 breath/min stage are lower than those obtained during 8 breath/min stage,

Table 6.1 RQA results averaged all over the population (mean \pm standard deviation), for the coupling between HRV and respiration

Stage	HRV-respiration		
	PR (mean \pm standard deviation)	PD (mean \pm standard deviation)	ER (mean \pm standard deviation)
15 breath/min	14.8 \pm 4.2	24.5 \pm 4.8	4.2 \pm 0.13
12 breath/min	24.3 \pm 3.5	62.3 \pm 9.2	3.6 \pm 0.2
8 breath/min	27.2 \pm 3.5	68.7 \pm 6.2	2.3 \pm 0.2

whereas ER is higher; in addition the difference of PD and ER values reaches a statistical significance ($p < 0.01$ for PD and $p < 0.001$ for ER). For the three RQA variables, no overlapping is found between 8 and 15 breath/min stages.

6.5 Discussion

We hereby propose a quantitative assessment of the degree of coupling between respiration and cardiovascular variability signals, based on the RQA. We estimated the degree of coupling by quantifying the spatio-temporal recurrent patterns of the interaction between the respiratory rhythm and the HR and BP variability series, in different breathing conditions.

Although the autonomic nervous system (ANS) functionality has been extensively studied by means of spectral and cross-spectral analysis of cardiovascular variability signals, non-linear dynamics have been recently shown to provide a relevant description of operation of this control system [1–8, 31–33]. In addition, the interaction of breathing, heart rate and blood pressure are considered important for the understanding of the functional organization of the ANS.

The ANS functionality has been already investigated by the RQA of single embedded time series of HRV and/or BPV. Dabiré et al. [5] quantified the RP obtained from both HRV and BPV time series by using PR, PD and the longest diagonal, this latter inversely related to the Lyapunov exponent. They showed that these non-linear indexes better reflected the sympathetic tone of BP and the parasympathetic tone of HR, than linear indexes derived from the spectral analysis. The longest diagonal of a RP has been also demonstrated to be a useful index to survey the autonomic function in diabetic subjects. The variability of the cardiovascular signals is very complex and may reflect the non-linear interactions between sympathetic and parasympathetic systems and between the cardiovascular system and the respiratory one. Our approach could provide important physiological information for the understanding of the neurovegetative coordination underlying normal and pathophysiologically disturbed behaviors. Indeed, in normal subjects, the deeper the breaths, the more stretching of the lung, which induces a greater synchronization between lungs and heart both mechanically and neurally.

Obviously, the need to extend the method to a larger number of subjects as well as to the study of pathological situations is crucial. It could be interesting to test whether the non-linear mechanisms and the degrees of non-linear coupling obtained here can be reproduced in more patients, in different experimental conditions, in several patho-physiological conditions of clinical relevance, or in different variability signals (such as microvascular blood flow fluctuations and pupil diameter changes).

RPs have been extensively used to qualitatively detect transient phase locking patterns in the interaction between different oscillators. Following this approach, we used the RQA approach to quantify the repetitive patterns of interaction between respiration and cardiovascular variability signals in a time-to-time domain. An observed recurrent pattern indicates that the (temporal) periodicity of one series has almost the same features as the other series. The number and the length distribution of the diagonals in a RP, measured by PD and ER respectively, encode the deterministic features of the plot itself.

RQA turned out to be suitable for a quantitative evaluation of the observed coupling patterns among rhythms, both in simulated and real data, providing different degrees of coupling. The results from the simulated data showed that the increased degree of coupling between the signals was marked by the increase of PR and PD, and by the decrease of ER. When the RQA was applied to experimental data, PD and ER turned out to be the most significant variables, compared to PR; the better performance of these two variables can be explained by their low intrinsic sensitivity to spurious, isolated recurrent points, which may be introduced by random noise.

Coupling phenomena were investigated during a three-stage paced-breathing protocol, with the respiration rate getting progressively closer to the LF rhythm. Three main results have been obtained: no interaction between LF and HF rhythms at 15 breath/min; transient 1:2 PL phenomena during 12 breath/min stage; 1:1 PL between LF and respiration rhythm during 8 breath/min.

RQA detected different coupling patterns in the three stages of the experimental protocol: no interaction in the first stage (15 breath/min), a kind of sub-harmonic synchronization in the second stage (12 breath/min), and a full entrainment condition in the third one (8 breath/min).

The analysis of the frequency component of the signals by spectral analysis showed that during 15 breath/min the respiratory rhythm did not affect either HRV or BPV spontaneous fluctuations which are characterized by the LF component (but for the respiratory-induced fluctuations). During 12 breath/min, we observed typical sub-harmonic synchronization patterns between the respiration and the cardiovascular variability signals (LF rhythm locked to the half of the respiratory frequency). The LF rhythm resulted to be fully entrained with the respiratory one, during the 8 breath/min stage, in all the subjects examined. However, although the spectral analysis detects shifting of the LF rhythm to the HF frequency all over the recording period, this linear tool considers the cardiovascular variability signal as a linear superposition of oscillations at different frequencies and cannot quantify the degree of non-linear coupling between different rhythms.

The three RQA variables succeeded in discriminating the situation of no-interaction from the full-entrainment, for both the couples HRV-respiration and BPV-respiration. The sub-harmonic synchronization phenomena has been successfully distinguished from the situation of no-interaction but only to a minor extent from that of full synchronization. A certain degree of non-linear coupling exists during the sub-harmonic synchronization phenomenon, significantly higher than that occurring when the LF rhythm seems not to be influenced by the respiratory rhythm. However, the sub-harmonic synchronization is not easily distinguishable from the full-entrainment condition, although generally weaker. It can be speculated that the sub-harmonic synchronization phenomenon is due to a non-linear mechanism somehow similar to that characterizing a full synchronization, but featuring a lower degree of non-linear coupling. Still, it should be noted that ER succeeded in distinguishing between sub-harmonic and full synchronization for both HRV and BPV, while PD did not. This result suggests that the difference between sub-harmonic and full synchronization phenomena can be better detected from the diagonal length distribution (encoded by ER) rather than from the number of diagonals (encoded by PD). Further, since a diagonal length represents the time duration of a recurrence, the two interaction phenomena differ for the duration of stable recurrences and only to a minor extent for their number (quantified by PD). However, the strongest coupling between respiration and cardiovascular variability signals occurs during the last stage of the experimental protocol (8 breaths/min), which elicits a full synchronization between the LF and the HF rhythms in the cardiovascular variability series.

From a physiological viewpoint, it is difficult to precisely identify the mechanisms responsible for the observed non-linear interaction. Our knowledge of the functional organization of the ANS is indeed still incomplete, given its complexity and the large number of involved physiological variables. Several mechanisms can be hypothesized to explain the non-linear coupling between respiratory and spontaneous rhythms of cardiovascular variability signals, such as changes in venous return induced by lung volume, a central coupling of cardiovascular and respiratory centers at the brain stem level, or central non-linear oscillating interacting sources (generating LF, HF and respiratory rhythms) influenced by forcing actions arising from physiologic control loops.

A remarkable finding is the detection of transient 1:2 PL episodes between respiration and cardiovascular variability signals. When the respiration rate was about twice the frequency of the LF rhythm (0.2 Hz, 12 breaths/min stage), the power spectra of HRV and BPV showed a 1:2 frequency relationship between HF and LF rhythms. This phenomenon can be associated to a sub-harmonic synchronization between the two main rhythms of HR and BP variability series, similar to the effect observed in the Van Der Pol model, when the frequency of the external stimulus was about twice the frequency of the self-sustained oscillator. Again, whether this frequency entrainment is stable, or reflects an average behavior cannot be detected by spectral analysis. In four out of eight subjects, the time-domain analysis revealed short-time intervals of stable 1:2 frequency and phase synchronization between RA and HRV. The idea behind the use of the proposed methods is that the

cardiovascular control system features deterministic dynamical properties. This idea is also reflected in the simulated systems used for the verification of the methods. It has to be mentioned that different ideas of the physical nature of the studied signals lead to different mathematical approaches. Analyses based on symbolic dynamics or stochastic modeling might also be used to investigate cardiorespiratory synchronization data.

It can be speculated that phenomena such as 1:2 PL between LF and HF rhythms might be expression of a fine regulation of the ANS. Such synchronization phenomena can be evoked in experimental conditions (control breathing at 12 breaths/min) easily reproducible in clinical practice. In addition, the transient nature of the interaction phenomena could explain the observed amount of non-linear coupling.

Several factors such as the mean heart rate, the spontaneous LF frequency, the overall autonomic balance, and the actual respiration frequency jitter may produce the inter-individual variability observed in the 1:2 PL phenomena. Further investigation of this important point requires a larger number of subjects as well as the quantification of the ‘degree of coupling’ between LF and respiratory rhythms.

References

1. V. Sharma, Deterministic chaos and fractal complexity in the dynamics of cardiovascular behavior: perspectives on a new frontier. *Open Cardiovasc. Med. J.* **3**, 110–123 (2009)
2. D. Hoyer, R. Bauer, B. Walter, U. Zwiener, Estimation of nonlinear couplings on the basis of complexity and predictability – a new method applied to cardiorespiratory coordination. *IEEE Trans. Biomed. Eng.* **45**(5), 545–552 (1998)
3. D. Hoyer, O. Hoyer, U. Zwiener, A new approach to uncover dynamic phase coordination and synchronization. *IEEE Trans. Biomed. Eng.* **47**(1), 68–74 (2000)
4. H. Ding, S. Crozier, S. Wilson, A new heart rate variability analysis method by means of quantifying the variation of nonlinear dynamic patterns. *IEEE Trans. Biomed. Eng.* **54**(9), 1590–1597 (2007)
5. H. Dabiré, D. Mestivier, J. Jarnet, M.E. Safar, N.P. Chau, Quantification of sympathetic and parasympathetic tones by non-linear indexes in normotensive rats. *Am. J. Physiol.* **275**(4 Pt 2), H1290–H1297 (1998)
6. B. Pompe, P. Blindh, D. Hoyer, M. Eiselt, Using mutual information to measure coupling in the cardiorespiratory system. *IEEE Eng. Med. Biol.* **17**(6), 32–39 (1998)
7. J.S. Chang, K. Ha, I.Y. Yoon, C.S. Yoo, S.H. Yi, J.Y. Her, T.H. Ha, T. Park, Patterns of cardiorespiratory coordination in young women with recurrent major depressive disorder treated with escitalopram or venlafaxine. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **39**(1), 136–142 (2012)
8. E. Pereda, D.M. De la Cruz, L. De Vera, J.J. González, Comparing generalized and phase synchronization in cardiovascular and cardiorespiratory signals. *IEEE Trans. Biomed. Eng.* **52**(4), 578–583 (2005)
9. J.P. Eckmann, S.O. Kamphorst, D. Ruelle, Recurrence plots of dynamical systems. *Europhys. Lett.* **4**, 973–977 (1987)
10. C.L. Webber, J.P. Zbilut, Dynamical assessment of physiological systems and states using recurrence plot strategies. *J. Appl. Physiol.* **76**(2), 965–973 (1994)

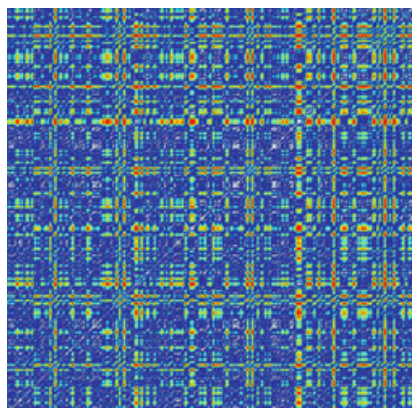
11. S. Ramdani, G. Tallon, P.L. Bernard, H. Blain, Recurrence quantification analysis of human postural fluctuations in older fallers and non-fallers. *Ann. Biomed. Eng.* **41**, 1713–1725 (2013)
12. C.D. Nguyen, S.J. Wilson, S. Crozier, Automated quantification of the synchrogram by recurrence plot analysis. *IEEE Trans. Biomed. Eng.* **59**(4), 946–955 (2012)
13. M. Mohebbi, H. Ghassemian, Prediction of paroxysmal atrial fibrillation using recurrence plot-based features of the RR-interval signal. *Physiol. Meas.* **32**(8), 1147–1162 (2011)
14. U.R. Acharya, S.V. Sree, S. Chattopadhyay, W. Yu, P.C. Ang, Application of recurrence quantification analysis for the automated identification of epileptic EEG signals. *Int. J. Neural Syst.* **21**(3), 199–211 (2011)
15. Y. Peng, Z. Sun, Characterization of QT and RR interval series during acute myocardial ischemia by means of recurrence quantification analysis. *Med. Biol. Eng. Comput.* **49**(1), 25–31 (2011)
16. M. Mazaheri, H. Negahban, M. Salavati, M.A. Sanjari, M. Parnianpour, Reliability of recurrence quantification analysis measures of the center of pressure during standing in individuals with musculoskeletal disorders. *Med. Eng. Phys.* **32**(7), 808–812 (2010)
17. P.I. Terrill, S.J. Wilson, S. Suresh, D.M. Cooper, C. Dakin, Attractor structure discriminates sleep states: recurrence plot analysis applied to infant breathing patterns. *IEEE Trans. Biomed. Eng.* **57**(5), 1108–1116 (2010)
18. S. Raiesdana, S.M. Golpayegani, S.M. Firoozabadi, J. Mehvari Habibabadi, On the discrimination of patho-physiological states in epilepsy by means of dynamical measures. *Comput. Biol. Med.* **39**(12), 1073–1082 (2009)
19. A. Porta, G. Baselli, N. Montano, T. Gneccchi-Ruscione, F. Lombardi, A. Malliani, S. Cerutti, Non-linear dynamics in the beat-to-beat variability of sympathetic activity in decerebrate cats. *Methods Inf. Med.* **33**, 89–93 (1994)
20. N. Marwan, J. Kurths, Nonlinear analysis of bivariate data with cross recurrence plots. *Phys. Lett. A* **302**, 299–307 (2002)
21. M. Carmen Romano, M. Thiel, J. Kurths, W. von Bloh, Multivariate recurrence plots. *Phys. Lett. A* **330**, 214–223 (2004)
22. A. Porta, G. Baselli, N. Montano, T. Gneccchi-Ruscione, F. Lombardi, A. Malliani, S. Cerutti, Classification of coupling patterns among spontaneous rhythms and ventilation in the sympathetic discharge of decerebrated cats. *Biol. Cybern.* **75**, 163–172 (1996)
23. F. Censi, V. Barbaro, P. Bartolini, G. Calcagnini, A. Michelucci, G.F. Gensini, S. Cerutti, Spatio-temporal recurrent patterns of atrial depolarization during atrial fibrillation assessed by recurrence plot quantification. *Ann. Biomed. Eng.* **28**(1), 61–70 (2000)
24. H. Ando, H. Suetani, J. Kurths, K. Aihara, Chaotic phase synchronization in bursting-neuron models driven by a weak periodic force. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **86**(1 Pt 2), 016205.25 (2012)
25. M.C. Mackey, L. Glass, Oscillation and chaos in physiological control systems. *Science* **197**, 287–289 (1977)
26. M.R. Guevara, L. Glass, A. Shrier, Phase locking, period-doubling bifurcations, and irregular dynamics in periodically stimulated cardiac cells. *Science* **214**, 1350–1352 (1981)
27. S.D. Wu, P.C. Lo, Cardiorespiratory phase synchronization during normal rest and inward-attention meditation. *Int. J. Cardiol.* **141**(3), 325–328 (2010)
28. M.C. Wu, C.K. Hu, Empirical mode decomposition and synchrogram approach to cardiorespiratory synchronization. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **73**(5 Pt 1), 051917 (2006)
29. V. Perltz, B. Cotuk, M. Lambertz, R. Grebe, G. Schiepek, E.R. Petzold, H. Schmid-Schönbein, G. Flatten, Coordination dynamics of circulatory and respiratory rhythms during psychomotor drive reduction. *Auton. Neurosci.* **115**(1–2), 82–93 (2004)
30. M.D. Prokhorov, V.I. Ponomarenko, V.I. Gridnev, M.B. Bodrov, A.B. Bespyatov, Synchronization between main rhythmic processes in the human cardiovascular system. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **68**(4 Pt 1), 041913 (2003)
31. G.M. Ramírez Ávila, A. Gapelyuk, N. Marwan, H. Stepan, J. Kurths, T. Walther, N. Wessel, Classifying healthy women and preeclamptic patients from cardiovascular data using recurrence and complex network methods. *Auton. Neurosci.* **178**, 103–110 (2013)

32. R. Guo, Y. Wang, J. Yan, H. Yan, Recurrence quantification analysis on pulse morphological changes in patients with coronary heart disease. *J. Tradit. Chin. Med.* **32**(4), 571–577 (2012)
33. M. Javorka, Z. Turianikova, I. Tonhajzerova, K. Javorka, M. Baumert, The effect of orthostasis on recurrence quantification analysis of heart rate and blood pressure dynamics. *Physiol. Meas.* **30**(1), 29–41 (2009)

Chapter 7

Analysis of Brain Recurrence

Clifton Frilot II, Paul Y. Kim, Simona Carrubba, David E. McCarty,
Andrew L. Chesson Jr., and Andrew A. Marino



Abstract Analysis of Brain Recurrence (ABR) is a method for extracting physiologically significant information from the electroencephalogram (EEG), a non-stationary electrical output of the brain, the ultimate complex dynamical system. ABR permits quantification of temporal patterns in the EEG produced by the non-autonomous differential laws that govern brain metabolism. In the context of appropriate experimental and statistical designs, ABR is ideally suited to the task of interpreting the EEG. Present applications of ABR include discovery of a human magnetic sense, increased mechanistic understanding of neuronal membrane processes, diagnosis of degenerative neurological disease, detection of changes in brain metabolism caused by weak environmental electromagnetic fields, objective characterization of the quality of human sleep, and evaluation of sleep disorders. ABR has important beneficial implications for the development of clinical and experimental neuroscience.

C. Frilot II

School of Allied Health Professions, LSU Health Sciences Center, Shreveport, LA, USA

P.Y. Kim • D.E. McCarty • A.L. Chesson Jr. • A.A. Marino (✉)

Department of Neurology, LSU Health Sciences Center, P.O. Box 33932, Shreveport,
LA 71130-3932, USA

e-mail: andrewamarino@gmail.com

S. Carrubba

Natural Sciences Department, Daemen College, Amherst, NY, USA

Abbreviations

ABR	Analysis of brain recurrence
AHI	Apnea–hypopnea index
ARA	Aperiodic rhythmic activity
EEG	Electroencephalogram
EEP	Electrosensory evoked potential
EMF	Electromagnetic field
EP	Evoked potential
FZAC	First zero of the autocorrelation function
MEP	Magnetosensory evoked potential
MS	Multiple sclerosis
REM	Rapid eye movement
WASO	Wake after sleep onset

7.1 Introduction

Human behaviors including somatic, cognitive, and emotional responses to stimuli, therapeutic reactions to drugs, and changes in signs or symptoms during progression of acute and chronic diseases are all mediated and controlled by the brain. Even though its structural complexity is daunting and its differential laws are mysterious, we accept the idea that the brain always acts lawfully. How it does so has been a perennial subject of scientific and philosophical interest.

In principle, the task of explaining the behaviors could be approached by attempting to deduce equations of motion, akin to the Hamiltonian formulation of mechanics, or by following a statistical approach similar to the determination of the pressure in a vessel by averaging over many atomic-level motions. But brain activity is such a complicated process that we have nil expectation it can be profitably analyzed by means of these traditional approaches. But even though we cannot deduce the brain’s governing differential laws, we are certain that they exist, because the behaviors governed by brain activity are lawful, not random.

Even though the task of understanding brain function as a necessary consequence of its differential laws is probably impossible, other goals regarding brain function can be achieved. An ability to make reliable predictions of human behaviors has enormous practical benefits in the fields of translational medicine and experimental neuroscience, even if the predictions are only probabilistic. Our primary goal here is to show how recurrence analysis, when employed in the context of an appropriate experimental and statistical framework, a technique we call *Analysis of Brain Recurrence* (ABR), yields such benefits.

In the next section we provide foundational information regarding the structure and function of the brain, with emphasis on the *complexity conjecture*, a fundamental principle that undergirds the unique applicability of recurrence analysis to

the study of brain activity. We shall see why recurrence analysis is extraordinarily well-suited for brain studies.

The following section describes the origin of recurrence analysis and what it is when looked at from the perspective of brain studies. Recurrence analysis is a remarkable invention that stemmed from the keen insight that biological systems are not crude versions of physical systems but rather irreducibly complex systems whose understanding for useful purposes requires novel analytical methods. The basic properties of recurrence analysis are discussed, and its ability to detect law-governed activity is demonstrated.

Next we present the results of published studies that encompass our applications of ABR. They include discovery of a human magnetic sense, the diagnosis of multiple sclerosis, detection of changes in brain metabolism caused by weak electromagnetic fields (EMFs), insights into the biophysical and physiological bases of EMF transduction, and applications to the objective characterization of sleep and the diagnosis of sleep disorders. Recurrence analysis must be employed in an appropriate statistical context to permit its results to go beyond geometric curiosities and reliably generalize to real-world problems. Consequently the true error rates for the hypotheses tested are listed, and references are provided to the original descriptions for the statistical details. The applications of ABR chosen for discussion should give the reader a good idea of its capabilities and limitations.

In the final section we attempt to place ABR in the general context of the scientific method so that the rock-solid foundation of ABR is clear. In contrast with attempts to explain the world in terms of time-reversible linear differential laws, ABR has different goals, makes opposite assumptions, and yields rewards not otherwise obtainable.

7.2 The Brain

7.2.1 *Physiological Role*

The brain is an open system that constantly interacts with the environment and incessantly changes, always existing in far-from-equilibrium states (Fig. 7.1). A continuous flow of mass, energy, and information crosses the skin and drives the brain's ceaseless activity, which occurs during the presence and absence of recognizable stimuli and ceases only in death. The brain generates information and also changes its internally-generated programs in response to information obtained from the environment. These processes embed the world in a unique context for each individual, determined by time and experience. The brain's governing laws are necessarily nonlinear and non-autonomous, because linear and/or autonomous systems don't move or change themselves or exhibit irreversible changes.

The physical appearance of the brain belies its abilities. The brain is small, about 0.34 kg at birth, 1.4 kg at maturity, 1.3 kg in advanced old age [1], and easily

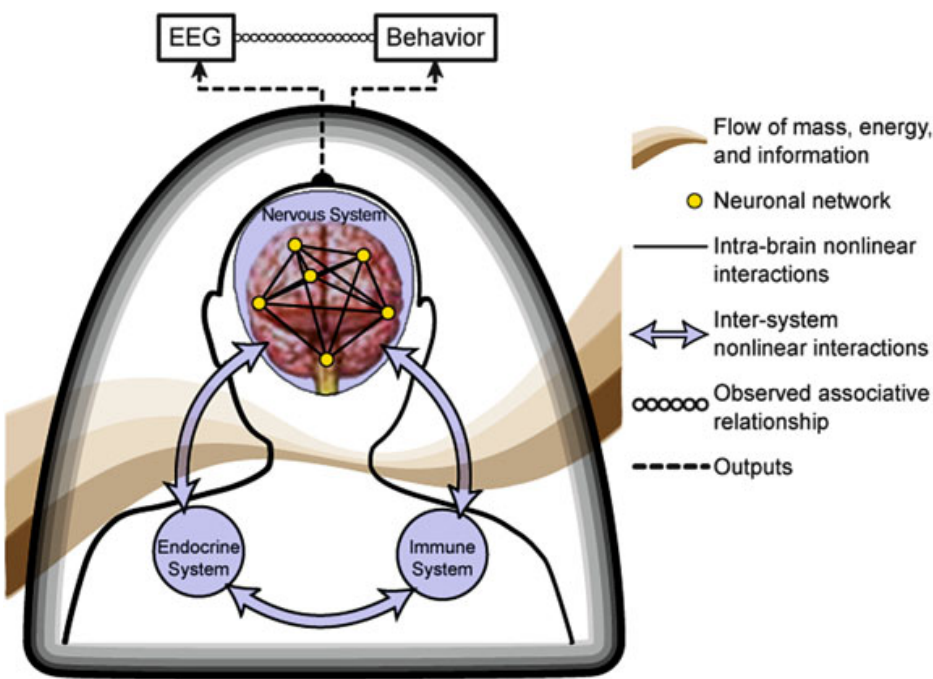
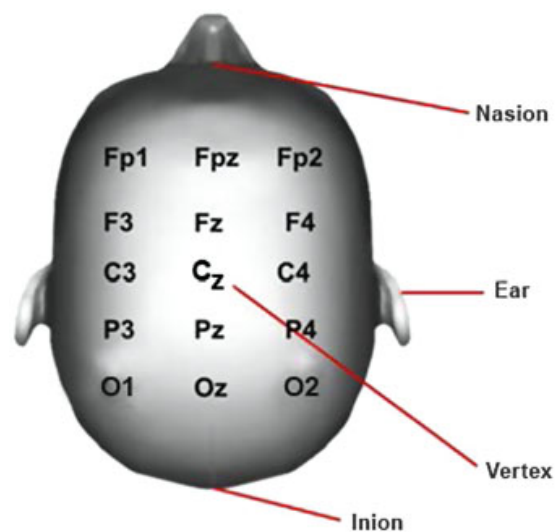


Fig. 7.1 Continuous flows of mass, energy, and information drive the brain’s master-control function, which is the governance of the body’s three main regulatory systems. Facilitated by interactions among a relatively small number of local neuronal networks, the brain receives numerous chemical and energetic inputs, and produces numerous outputs. The interactions determine all observable responses and behaviors, including the electroencephalogram (EEG). We regard observations and their contemporaneous EEGs as physiological doppelgangers that code for each other

deformed, like a gel. The brain’s abilities ultimately arise from the conjoint operation of the laws of mechanics, electricity, and thermodynamics in the context of what is probably the most complex structure in the universe. The brain’s cellularity and microscopic anatomy consists of $\sim 10^{11}$ distinct cells called neurons, with $\sim 10^{14}$ contacts between them [2], all of which are self-organized into $\sim 10^2$ structural and functional networks that are fed and nurtured by $\sim 10^{12}$ support cells of various types [3]. Information transfer at the neuronal level is reasonably well understood conceptually in terms of neurotransmitters, neuromodulators, cytokines, direct electrical communication between cell interiors via gap junctions, and electrotonic conduction through the extracellular fluid [4]. Some phenomenological descriptions of neuronal signaling have been developed, the Hodgkin–Huxley equations for example, but there are no direct explanations based on general laws.

At the organ level almost nothing is known about what governs brain activity, and we lack even phenomenological descriptions. Explaining human behaviors at the level of individuals is therefore problematical. Nevertheless each brain signs itself by producing an electroencephalogram (EEG), a time-dependent voltage signal conventionally measured on the scalp (Fig. 7.2) (measurable anywhere on the skin but with reduced sensitivity because of the passive electrical properties of tissue).

Fig. 7.2 Scalp locations and labels for EEG measurements (10–20 system). The EEG is measured at standardized locations on or symmetrically distributed about the mid-sagittal plane [5]. The locations are identified by rules based on proportional distance measurements made in relation to arcs between the nasion and inion, and between the ears



7.2.2 The Baseline EEG

The EEG (Fig. 7.2) is an emergent property of the dynamical activity of the brain, and consequently has no specific origin, like a smile. We conceptualize the EEG as a state property of the brain in the sense that some of the information it contains involves the cooperative activity of all its cells. In principle, an EEG measured during an arbitrary time interval would have been different if even one neuron had died immediately prior to the beginning of the interval.

The spectral energy in the EEG occurs almost entirely below 50 Hz. The signal therefore cannot directly reflect rapid electrophysiological events occurring at the neuronal membrane [4]. Attempts to interpret the EEG in relation to phenomena associated with higher organizational levels have traditionally been based on use of the root mean square of the signal (time averaging), or on its Fourier decomposition with subsequent determinations of relative power in standard frequency intervals (spectral analysis). The variables associated with both methods were assumed to be classical stochastic variables (described by fixed probability distributions). Unfortunately the EEG is dramatically non-stationary over the time scales pertinent to the problem of interpreting the EEG (Fig. 7.3). This salient fact conflicts with the logic of linear methods like time averaging and spectral analysis. Valid use of these methods requires that differences in means of dependent variables arise solely from changes in the independent variables, not from changes in the dynamical law of the system. But that is exactly what occurs in the brain, evidenced by the nonstationarity of the EEG (Fig. 7.3). In marked contrast, stochasticity is irrelevant to the validity of ABR. Indeed, its fundamental value is that it permits characterization of the brain’s non-autonomous activity, which is where human behaviors actually come from (see below).

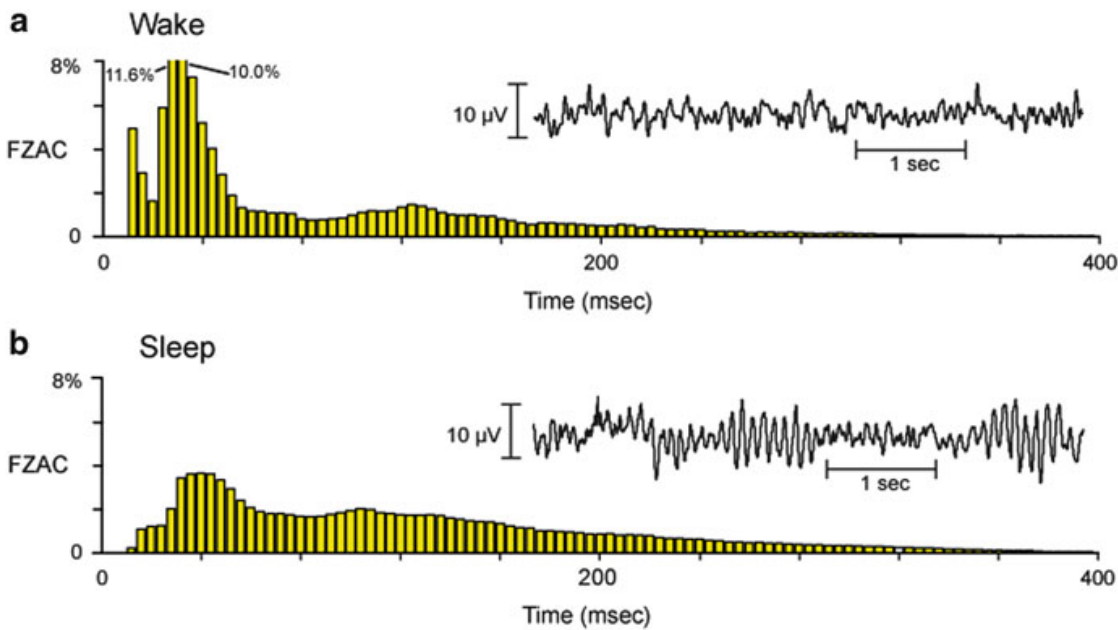


Fig. 7.3 Nonstationarity in the human EEG during wake and sleep. **(a)** EEGs were recorded for 5 min (O1, O2, P3, P4, C3, C4) from each of ten vigilant subjects, and the first zero of the autocorrelation function (FZAC) was determined second-by-second. The results ($\sim 1,800$ values/subject) were divided into 4-ms bins, averaged, and the resulting histogram was normalized. **(b)** EEGs (from C3) were recorded for 7–8 h from each of ten sleeping subjects, and the mean normalized FZAC histogram ($\sim 28,000$ values/subject) was determined. The histograms of the EEG during both wake and sleep indicate that the statistical properties of the EEG change drastically from second to second, which is the definition of a nonstationary time series. The nonstationary character of the EEG implies that brain activity is governed by non-autonomous differential laws. Sampling rate, 500 Hz

7.2.3 The Stimulated EEG

Somatic and/or cognitive stimuli produce transient stereotyped electrophysiological responses in the EEG called evoked potential (EPs), typically lasting 100–400 ms and occurring 100–400 ms after application of the stimulus [6,7]. Stimuli trigger both onset and offset responses, respectively caused by the onset and offset of the stimulus. For example a light pulse from a light-emitting diode has an onset and offset whose durations are the rise- and fall-times of the diode current, respectively, and both aspects of the stimulus trigger an EP. Different sets of neuronal networks are involved in attention to stimuli onset and offset [8–11]. The EPs are much weaker than the baseline EEG, with which they become convolved.

Evoked potentials have traditionally been detected by time averaging the EEG over intervals immediately following N application of the stimulus. The procedure results in an increase in signal (the EP) to noise (the EEG) $\propto \sqrt{N}$. Evoked potentials not time-locked to the stimulus are undetectable by time averaging, and hence were essentially unknown prior to the development of recurrence analysis (see below).

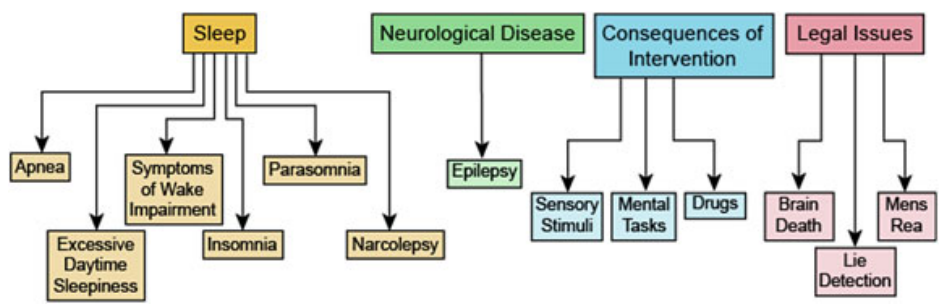


Fig. 7.4 Present uses of the EEG

7.2.4 *Uses of EEG*

The main clinical use of EEGs occurs in sleep medicine, where measurements are typically made from six derivations during overnight studies for use in characterizing the stages of sleep (typically, the first step in diagnosing sleep disorders) (Fig. 7.4). With the exception of epilepsy, EEG measurements play a minor role in diagnosis of neurological diseases, having largely been supplanted by magnetic resonance imaging. Evoked potentials are used clinically to evaluate the occurrence of lesions in the peripheral nervous system. Robust utilization of EPs occurs in experimental neuroscience because EPs are the only known method for objectively characterizing the functional activity of the brain over times on the order of milliseconds. In principle, the guilty mind (*mens rea*) and the lying witness can be detected based on EEG analysis, but efforts to do so have been unsuccessful. The absence of an EEG for a defined interval is a common basis for recognizing death.

Following the development of the theory of low-dimensional nonlinear dynamics, interest spiked in the possibility that the theory was directly applicable to the EEG. Reports describing calculations of fractal dimension and Lyapunov exponents were interpreted as evidence that the brain was a low-dimensional nonlinear system operating in the chaotic mode, like the Lorenz model of weather [12]. But the effort to show that EEGs were outputs of low-dimensional systems of the type for which the theory was created ultimately failed [13]. Nevertheless new initiatives developed, prompted by a dawning recognition that progress regarding understanding the lawfulness of the brain required approaches that were beyond linear and low-dimensional nonlinear approximations, and more faithfully mirrored the brain’s actual nonlinear dynamics, namely that it is governed by non-autonomous differential laws, and probably has dimensions closer to 10^{10} than to 10. As we shall see, ABR uniquely accomplished the task of revealing the footprints of the laws.

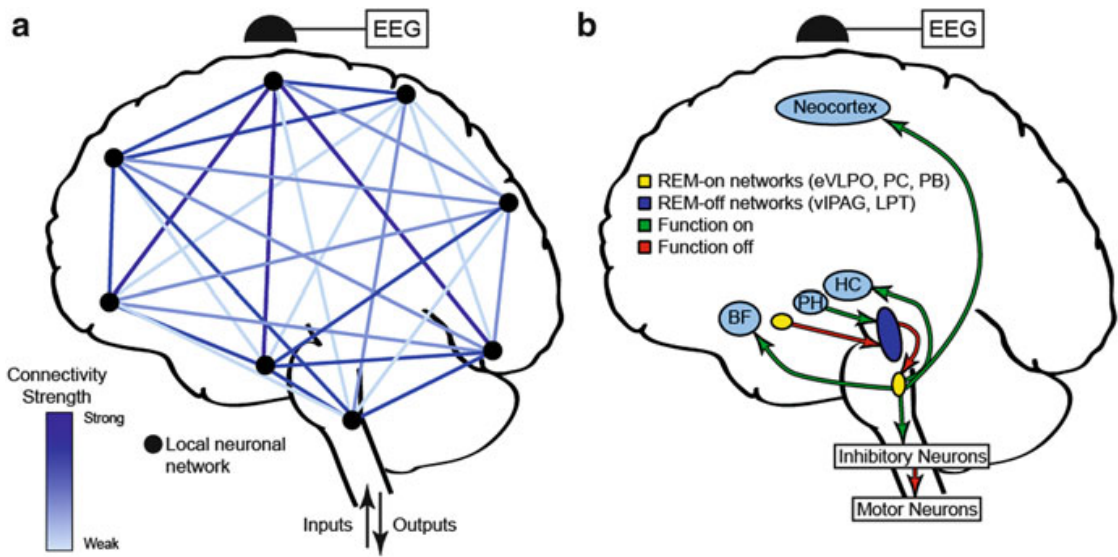


Fig. 7.5 The complexity conjecture regarding the origin of the human EEG. **(a)** Brain function is mediated by electrical activity in localized neuronal networks and their dynamic internetwork electrical connectivity, which result from synaptic and non-synaptic processes [4]. The instantaneous strength of the connectivity between local networks is represented by the color intensity of the line that joins them. The EEG measured for any Δt from any derivation characterizes the entire brain (brain state) during Δt . In practice, the time interval regarded as the characteristic duration of the state is chosen in relation to the behavior of interest. As examples, $\Delta t \approx 100$ ms for detecting evoked potentials, and $\Delta t = 1$ s for analyzing human sleep. **(b)** REM occurs when REM-on neurons in the extended ventrolateral pre-optic area (eVLPO) inhibit REM-off neurons in ventrolateral periaqueductal gray (vIPAG) and the lateral pontine tegmentum (LPT), whose function is to actively inhibit REM-on neurons in the pre-coeruleus (PC) and parabrachial (PB) networks (dorsolateral pons). The PC and PB networks send excitatory projections to the basal forebrain (BF), hippocampus (HC) and neocortex, and to the sublaterodorsal nucleus (not shown) from which neurons project to inhibitory interneurons which produce the muscle atonia that is characteristic of REM sleep

7.2.5 Complexity Conjecture

An explicit statement of our perspective regarding the functional framework of brain electrical activity will clarify why we expected ABR to be successful.

Cognition and physiological regulation are mediated by time-dependent electrical interactions among spatially distributed neuronal networks [14–17] (Fig. 7.5). In the absence of identifiable tasks or stimuli, network connectivity is highest in wakefulness, and lower during sleep, or in subjects with neurological disease. A decrease in connectivity corresponds to an increase in deterministic activity because the brain becomes more machine-like. Complexity is decreased (deterministic activity increased) during sleep and in association with neurological disorders.

The EEG during any particular time interval is a characterization of the global electrical activity of the brain that occurred in the interval as the brain processed inputs and generated outputs. If the inputs contain contributions from specific

stimuli external to the subject, the EEG baseline brain state will contain additional transient components called evoked potentials (EPs) that are convolved with the baseline EEG. Whether characteristic changes in the determinism of the baseline EEG or transient changes in the EEG actually occurred in association with changes in the subject's metabolism or environment are empirical questions.

The law-governed activity in the EEG can be detected by time embedding in low-dimensional space whose dimensions are the instantaneously dominant brain networks (Fig. 7.5). The dimensional variables are *not* stochastic variables and therefore not generally associated with fixed probability distributions.

Next, the origins of recurrence analysis and the bases of its applicability to brain activity are described.

7.3 Recurrence Analysis

7.3.1 *Historical Development*

Recurrence analysis developed as a means of exploiting the recurrence plot, a graphical tool introduced to facilitate visualization of hidden patterns in time series from low-dimensional nonlinear differential equations [18]. The plot was formed by embedding a time series in dimension D using the method of time delays to produce a sequence of N D -dimensional vectors. X_i , $i = 1, 2 \dots j \dots N$. The sequence was represented in two dimensions by plotting a point at the location addressed by (i, j) whenever X_i was near X_j . The plots were said to be useful for revealing departures from autonomous behavior in physical systems such as noisy Rayleigh–Bernard-convection data or Lorenz solutions with added drift [18]. Webber and Zbilut developed a method for quantifying the plots, and repurposed them for application to biological systems [19]. The insights and innovations they described are discussed below. Application of recurrence plots to physical systems is reviewed elsewhere [20].

Webber and Zbilut recognized that outputs of biological systems exhibited aperiodic rhythmicity which was intimately related to the physiological state of the organism, but that no method existed for quantifying the rhythmicity. Linear methods were sometimes used as first approximations, but they were formally unsuitable. Linear theory assumes that variables are stochastic, and that two groups of realizations of a stochastic variable are reliably different if and only if their means differ by more than that expected on the basis of chance. But for variables governed by nonlinear differential laws, a true difference between the groups may simply be averaged away—the more measurements made or the more nonlinear the laws, the less likely a true difference will be detected. Thus linear methods are prone to obscure rather than reveal reality in those instances where it is governed by nonlinear laws, which is typically the case for biological signals.

Low-dimensional nonlinear theory similarly was no answer to the problem of quantifying aperiodic rhythmic activity (ARA) in biological time series because biological systems are invariably high-dimensional. Moreover, actual biological systems such as the brain characteristically produced non-stationary time series, a statistical property of the data which invalidated general application of the theory and rendered meaningless the results of formal processes of calculating constants of the motion.

They further observed that recurrence plots for successive time-series intervals of common physiological signals exhibited patterns which changed in a more or less regular way, suggesting that the changes in ARA were somehow dependent on the state of the organism. These realizations led them to repurpose the recurrence plot from a device that was originally designed for the exclusive purpose of detecting non-autonomous deviations in autonomous systems, to a device that could directly quantify changes in ARA in non-autonomous systems. Quantification was accomplished by means of systematic procedures that mapped geometric features of the plot into numbers.

They conjectured that the number of points in the plot relative to the possible number (the $N \times N$ array) as well as various geometric features of the plot were objective measures of the dynamical activity exhibited by biological systems. Sequential application of the quantification step then yielded recurrence time series that could be analyzed statistically, like an ordinary random variable. Thus they made biological meaning of the quantitative temporal changes exhibited by sequences of recurrence plots. In this manner, they transformed their plot measures into *bona fide* dynamical variables, an entirely novel development in the history of biological quantification. It is important to recognize that they did not claim any specific physiological meaning for the variables. On the contrary, their examples made clear that physiological meaning must be inferred empirically [19].

They defined several plot-based variables. *Percent recurrence* (%R) quantified the amount of recurrence that existed in the system during the interval that the embedded time series was generated (given the choices for the various parameters related to embedding and plot formation). Graphically, %R corresponded to the fraction of the area of the plot that contains points, but %R had no direct relation to the structure of the plot. The other variables, in contrast, were measures of the arrangement of the points.

Percent determinism (%D) characterized the number of points in the plot that fell on diagonal lines. Various possible structures in the plots have differing dynamical correlates. For example, consider three points along a diagonal (Fig. 7.6a, left panel). By definition, the i th and j th points in phase space were near, as were the consecutive states of the system (i_1, j_1 and i_2, j_2). In other words, the system's dynamics repeated for at least two successive states. The longer such behavior continues (more adjacent points along the diagonal), the stronger is our conviction that the plot reflected true law-governed activity. Vertical and horizontal lines result from the nearness of a single state to a series of other states (Fig. 7.6b, c). The diagonal line is the centrally important plot structure, probably because it has the

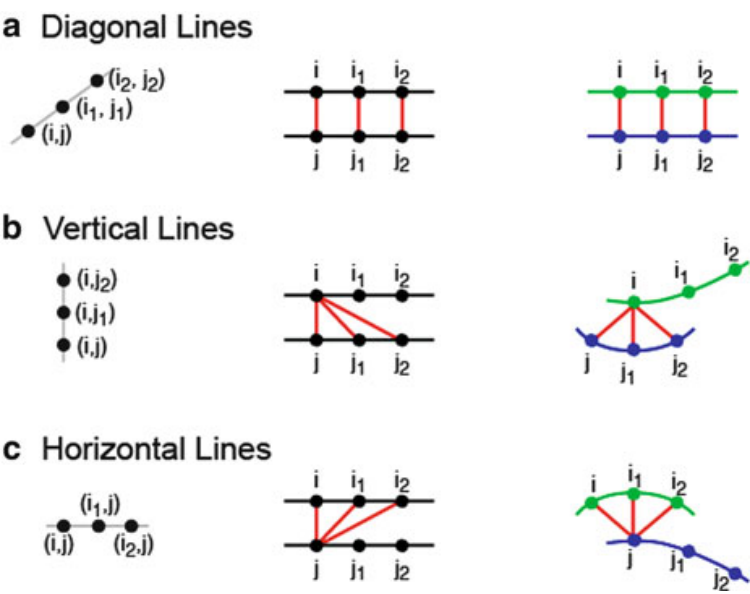


Fig. 7.6 Dynamical implications of lines in a recurrence plot. *Left column*, directions of lines in a plot. *Middle column*, schematic representation of the corresponding points in phase space. *Red lines* join points near another (based on the points shown and the definition of a plot). *Right column*, representation of the system’s trajectory in phase space that produces the lines. *Blue* and *green* points are located sequentially on two different segments of the trajectory

most direct and intuitive interpretation in terms of the system’s dynamical activity (Fig. 7.6a). Other variables related to the distribution of diagonal lines in a plot have also been defined [21].

Recurrence variables have no necessary relationship to dynamical theory, nor any specific meaning in relation to any standardized unit of measurement. There is no platinum bar as was accepted for the initial definition of a meter, and no acknowledged functional characterization similar to the manner in which an ohm was defined. Notwithstanding the absence of normative theories, definitions, or conventions, the variables have important and practical physiological significance that can be established directly by means of generally accepted scientific methods.

7.3.2 General Properties of Recurrence Analysis

Numerical calculations of recurrence variables from EEG signals depend on the values chosen for various parameters. There is no evidence that the principles commonly recommended for choosing parameters in connection with analysis of low-dimensional systems are optimal for analyzing the EEG. On the contrary the principles may be misleading, at least to the extent they imply that the EEG is

a stationary signal or the output of a low-dimensional system. Consequently the parameters used for ABR must be determined empirically, to maximize sensitivity for the effect sought.

The pertinent parameters include sampling frequency of the voltage time series, the window employed in the recurrence calculation, the step size used to shift the window along the time series (the shift parameter), the embedding dimension and time delay chosen for the phase-space calculations, the choice of the number of points regarded as necessary to constitute a line, and the definitions of nearness and distance (in phase space).

The sampling frequency and recurrence window used in ABR were chosen initially based on our estimates of the characteristic dynamical times in the baseline EEG and the transient response evoked in the baseline by external stimuli. Generally, %R and %D increased with increasing sampling frequency, but plateaued as a function of window size, which permitted choices of minimum window sizes that were optimal for capturing the determinism present in the signal.

Let $V_1(t)$ be an EEG interval; $\%R_1(t)$ and $\%D_1(t)$ are the corresponding percent recurrence and percent determinism time series computed from $V_1(t)$. $V_2(t)$ is a different EEG interval with corresponding recurrence times series $\%R_2(t)$ and $\%D_2(t)$. In general

$$\%R_{1,2} \neq \%R_1 + \%R_2 \quad (7.1)$$

$$\%D_{1,2} \neq \%D_1 + \%D_2 \quad (7.2)$$

where $\%R_{1,2}$ and $\%D_{1,2}$ are the recurrence time series of $V_1(t) + V_2(t)$, indicating that recurrence determinations do not follow the law of superposition. As a consequence of the non-applicability of the law of superposition, detection of the presence of a signal (an “effect” in an experiment in which a subject is exposed to a stimulus) can be manifested as any kind of a change in %R or %D relative to an appropriate control, not necessarily as a unidirectional change. In other words, in general, an effect due to a stimulus will be evidenced by a consistent change rather than by a consistent directional change.

Another important characteristic of recurrence analysis of the EEG involves the relationship between the recurrence time series and the EEG signal from which it was calculated. Assume that a signal was added to the EEG during the time window W , and that W was the window used for calculating each point in the recurrence time series (Fig. 7.7). The added signal will combine point-by-point with the baseline EEG. Some points in the recurrence time series will contain contributions from the baseline but not the added deterministic signal. In Fig. 7.7, for example, only one point in the recurrence time series includes all the determinism contributed by $V_2(t)$. Thus, in general, the time interval in the recurrence time series that contains a material contribution of the signal of interest will be shorter than the corresponding time interval in the EEG during which the signal was present.

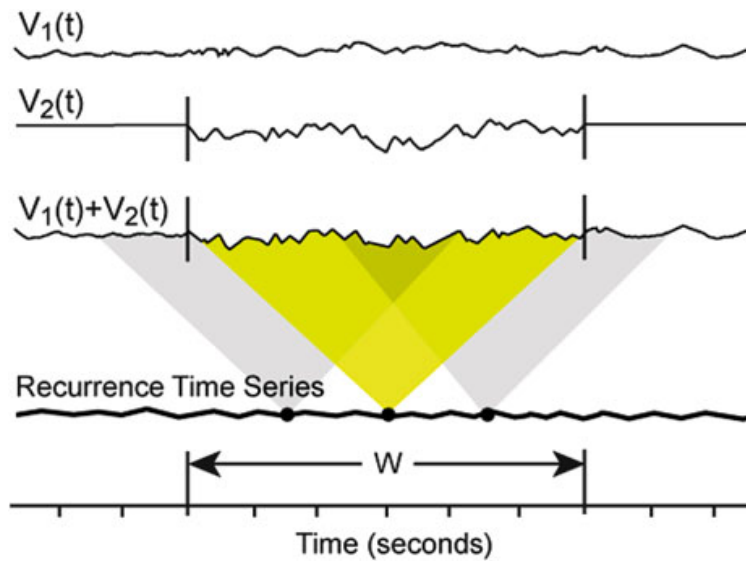


Fig. 7.7 The relationship between the EEG and its corresponding recurrence time series. $V_1(t)$ and $V_2(t)$ are continuous time series. $V_2(t)$ is non-zero during W . $V_1 + V_2$ is formed by point-wise addition of the two signals. If W is the window employed for recurrence calculations, only the center point in the recurrence time series will include all of the points in the composite that were contributed by V_2 (highlighted in yellow). All other points in the recurrence time series contain some points for which V_2 made no contribution, thereby lessening the overall ability of recurrence analysis to detect V_2 . Typically the EEG sampling rate is 300 Hz and W is 30 points (100 ms)

7.3.3 Recurrence Analysis of Model Systems

7.3.3.1 Model Systems

The capability of recurrence analysis to detect nonlinear determinism present in the EEG can be demonstrated by adding known deterministic signals to baseline EEG, and comparing the altered time series with appropriate controls. The process models the expected determinism associated with an EP, and affords the opportunity to study the same time series before and after the addition of determinism, an impossibility in actual EP experiments.

Baseline EEGs were recorded from standard derivations in normal human subjects (Fig. 7.2), band-passed at 0.5–35 Hz (standard band for recording EEGs), and sampled at 300 Hz. The recorded signals were divided into 2-s epochs; those containing motion artifacts were discarded and artifact-free epochs were randomly selected for use in the modeling procedures. To mimic transient deterministic signals occurring in the EEG in response to sensory stimuli, sine or Lorenz signals were added to the epochs. Three types of signals were considered: (1) a 10-Hz sine added at $t = 0.85$ – 1.15 s to each of 50 2-s epochs such that the phase of the added signal was identical in each epoch; (2) a 10-Hz sine with a phase that varied randomly from epoch to epoch; (3) a portion of the solution from the Lorenz equation ($a = 10.14$, $r = 28$, $b = 2.67$) [22]. In each case the rms signal-to-noise (S/N) ratio was 0.4, set on an epoch-by-epoch basis.

7.3.3.2 Detection of Nonlinear Determinism

When segments of sine waves were added to EEG epochs such that the phase of the sine was always the same (negative slope starting at 0), the presence of the sine was easily detected by time-averaging (Fig. 7.8b). But when the added sine had a random phase, the added signal was not seen in the time-averaged signal (Fig. 7.8c). The same result was found when the added signal consisted of segments of a solution to the Lorenz equation (Fig. 7.8d). Thus, as expected, time-averaging was unable to reveal the presence of known deterministic signals in the cases where the signals were inconsistent (random-phase sine) or aperiodic (Lorenz). In contrast, recurrence analysis detected the addition of inconsistent or aperiodic signals, as seen by comparing the recurrence time series in the presence and absence of the added segments (Fig. 7.9).

Localization of the added determinism was affected by the choice of the recurrence window (Fig. 7.7) as well as by random fluctuations in the baseline signals. For example, the random-phase sine wave yielded simulated effects at 0.90–0.92 s and 1.05–1.15 s, even though the added deterministic signal was present through the 300-ms interval centered at 1 s (Fig. 7.10a, bottom panel). Thus only the approximate temporal location of the added determinism could be inferred from an analysis of the recurrence result. Addition of Lorenz segments also illustrated the point (Fig. 7.10b, bottom panel); in this case, statistically reliable demonstration of the added signal was even more localized in time.

Detection of deterministic changes in the EEG was improved by averaging the recurrence time series over a sliding window (called the P window, to distinguish it from the window used for recurrence calculations) prior to statistical comparisons of the *E* and *C* epochs (Fig. 7.11).

A further demonstration of the ability of ABR to detect nonlinear determinism is shown in Fig. 7.12. The presence of a Van der Pol signal (Fig. 7.12a) was averaged away in the EEGs (Fig. 7.12b) but was clearly demonstrated in %R(t) (Fig. 7.12d).

7.3.4 Overview of Recurrence Analysis of Brain Electrical Activity

Recurrence analysis quantifies order in the EEG over any desired time interval, ranging from milliseconds, as in neuroelectrophysiological studies, to intervals lasting hours that are commonly employed in clinical diagnostic procedures. We presented a broad outline of the uses and limitations of recurrence analysis, and the reasons we anticipated that the method would be experimentally and clinically useful.

The EEG necessarily differs between the presence and absence of stimuli, and between the presence or absence of neurological disorders. In the next section the

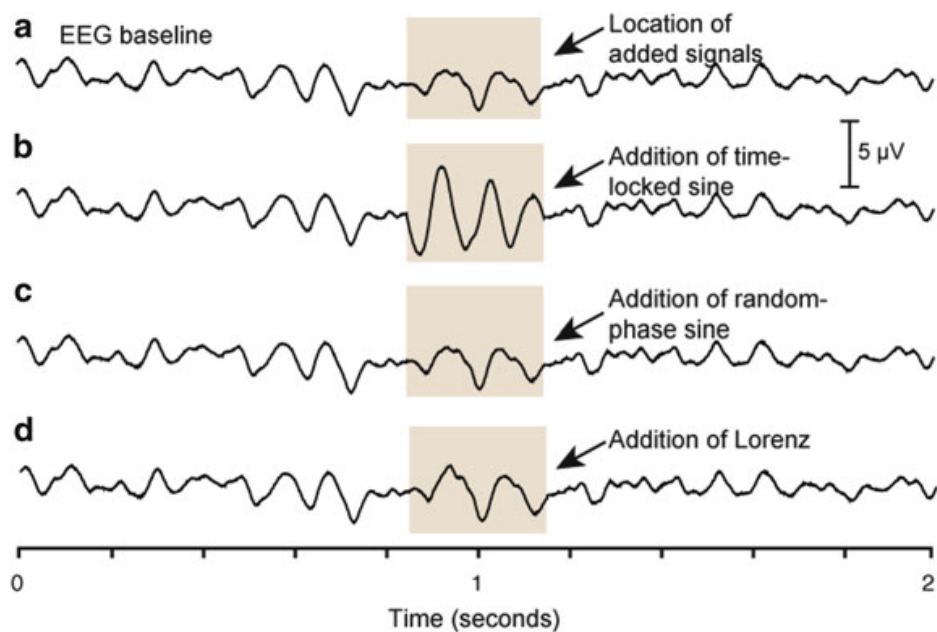


Fig. 7.8 Effect on EEG of addition of known signals. (a) Average of 50 2-s EEG epochs. (b) Effect of adding a constant-phase 10-Hz sine to each epoch, showing the effect of the averaging in revealing the presence of a phase-locked signal. (c) Effect of adding a random-phase sine to each epoch, showing the inability of time averaging to reveal a nonstationary signal. (d) Effect of adding segments of the Lorenz equation to each epoch, showing the inability of time averaging to reveal a nonlinear signal

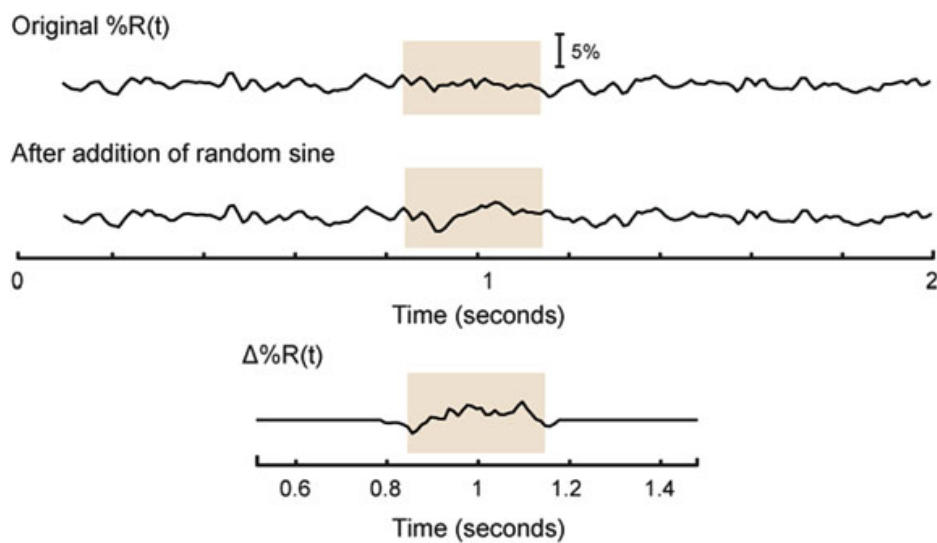


Fig. 7.9 Effect on $\%R(t)$ due to addition of nonstationary signals to the EEG. *Top*, average value of $\%R$ computed from 50 2-s EEG epochs. *Middle*, after addition of random-phase 10-Hz sine wave to each EEG epoch at $t = 0.85\text{--}1.15$ (shading). *Bottom*, time dependence of the point-wise differences of the signals before and after addition of the sine waves. Similar results were obtained following the addition of Lorenz segments. Sampling frequency 300 Hz. Embedding dimension 5. Delay five points. Recurrence window (see Fig. 7.7) 100 ms. Moving window step 1 point

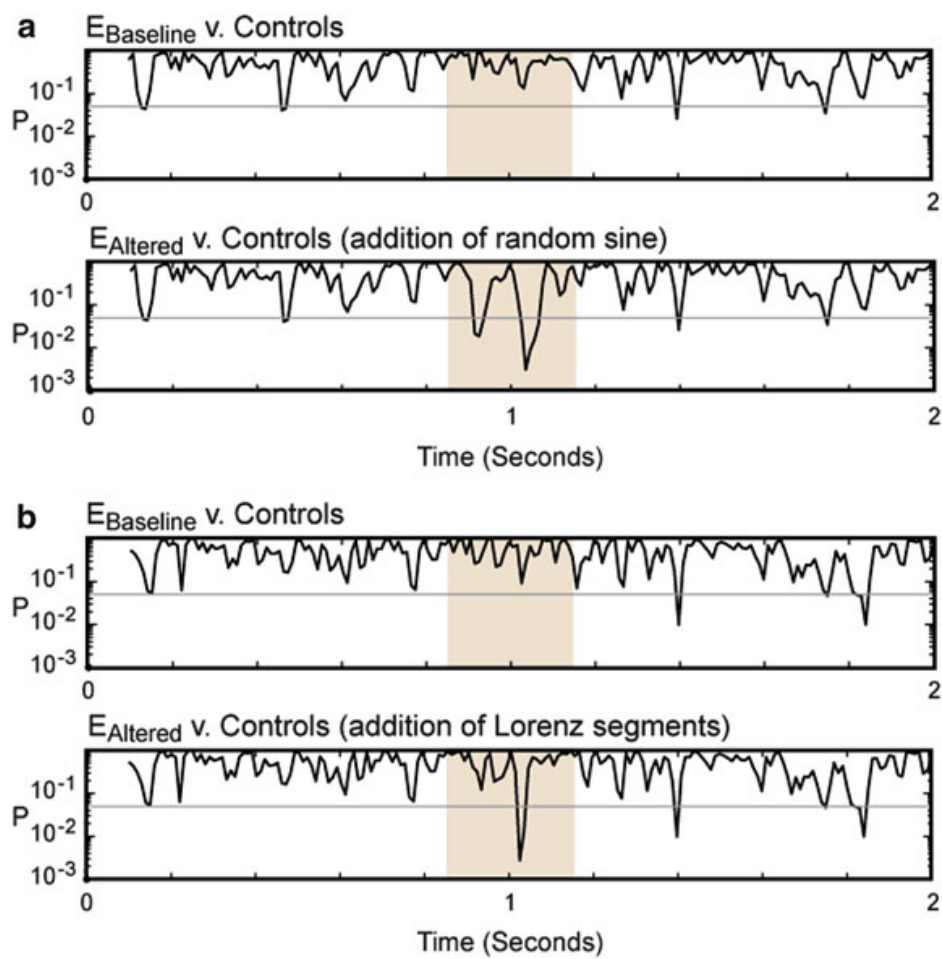


Fig. 7.10 Detection of addition of known signals to the EEG. $N = 50$ randomly-selected baseline epochs (E epochs) were compared point by point (t test) with $N = 50$ different randomly-selected control epochs (C epochs). Known determinism was added to the E epochs where shown (shading). (a) E versus C epochs in the absence of signals added to the E epochs (top), and after addition of random sine to the E epochs (bottom). (b) E versus C in the absence of signals added to the E epochs (top) and after addition of segments of the Lorenz signal to the E epochs (bottom). P , probability that the means at the indicated point in time were identical (total of approximately 600 paired t tests). Gray line, $P = 0.05$. The rms values of the added signals were adjusted epoch-by-epoch so that they were 40 % of the rms value of the EEG. The results show that recurrence analysis can detect the presence of a localized transient deterministic signal in the EEG with statistical certainty

goals, statistical structure, and results of experiments involving the capabilities of ABR are described. Examples are presented that evidence the ability of recurrence analysis to detect effects having profound implications for basic studies of brain activity, and that demonstrate applications of the method in translational medicine. The reader will see that recurrence analysis can produce knowledge of brain activity that is not otherwise obtainable.

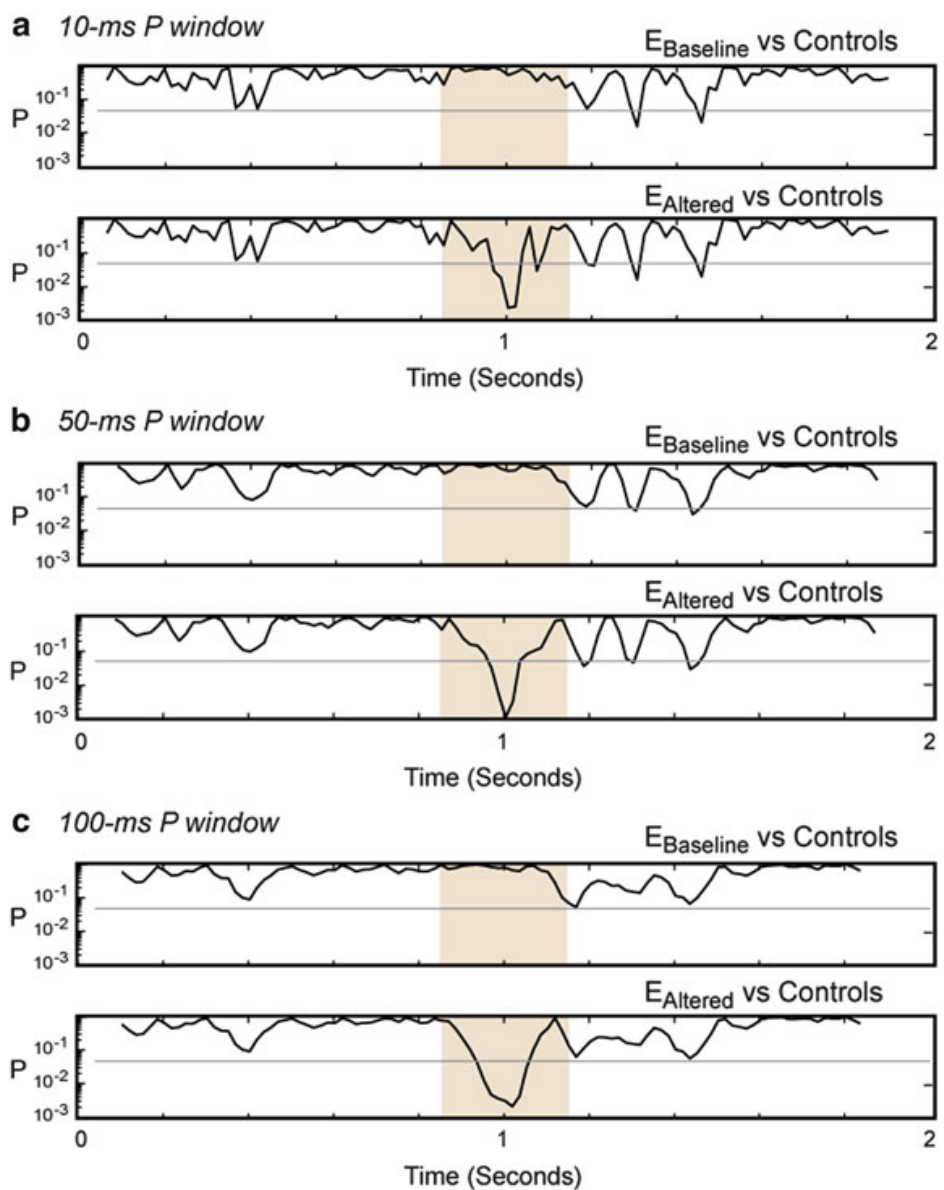


Fig. 7.11 Optimized conditions for detection of added Lorenz segments using %R. (a–c) Point-by-point statistical comparisons before and after addition of the signal to the baseline EEG, after employing the indicated P window. P , probability that the means at the indicated point in time were identical (total of approximately 600 paired t tests). Gray line, $P = 0.05$. The rms values of the added signals were adjusted epoch-by-epoch so that they were 40 % of the rms value of the EEG

7.4 Application of Recurrence Analysis

7.4.1 New Paradigm for Studying the Brain

A seminal problem in modern biology involves understanding how weak electromagnetic fields (EMFs) (fields whose strength is below the level of conscious detection) affect the growth and regulatory systems of the body, particularly the brain [23]. Early work suggested that EMFs impacted brain metabolism as

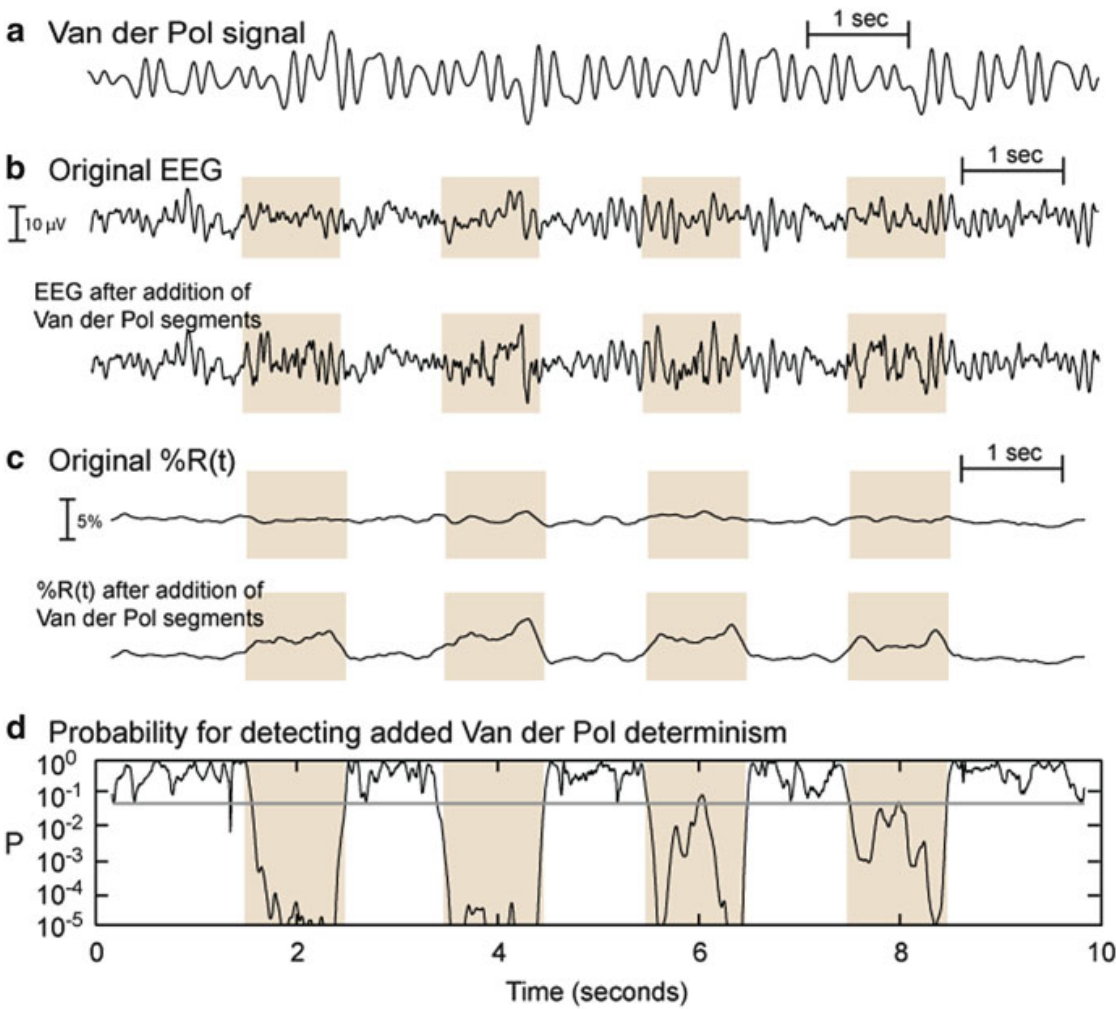


Fig. 7.12 Effects on %R(t) due to addition of Van der Pol segments to the EEG. **(a)** Representative segment of solution to Van der Pol equations. **(b, c)** Respectively, the EEG and %R averaged over 20 10-s EEG epochs before and after addition of 1-s Van der Pol segments where shown (shading) to each epoch. **(d)** Comparison of the original and altered epochs point-by-point (t-test). P, probability that the means at each indicated point in time were identical (total of 2,901 paired t tests). EEG sampling frequency 300 Hz; RQA window 50 pts; embedding dimension 5; time delay 5; radius 15 %; P value window 50 pts. Solution of the equation for the periodically driven Van der Pol oscillator $dy/dt = x - \beta x \sin(\omega t + \theta_2)$ and $dx/dt = -y - f(x) + \beta y \cos(\omega t + \theta_1)$, where $f(x) = (1/3)x^3 - \lambda x$, $\lambda = 3$, $\omega = 6.455$, $\beta = 45$, $\theta_1 = 24$, $\theta_2 = 26$. Gray line, P = 0.05

evidenced by changes in the EEG, but the traditional analytical methods of time-averaging and spectral analysis were inadequate in several respects [24].

Detection of familiar stimuli, light for example, occurs via a linear stimulus-response system. Consequently evoked potentials (EPs) triggered by common stimuli can be detected by time-averaging. But nonlinear responses cannot be efficiently detected by linear methods because they are not matched to the underlying dynamics. The first clear demonstration that EMFs consistently induced dynamical changes in brain activity was made possible by the invention of recurrence analysis, which permitted detection of nonlinear EPs [25]. Rabbits exposed to EMFs (2 G, 60 Hz magnetic field) exhibited a change in %D that was not seen directly in the

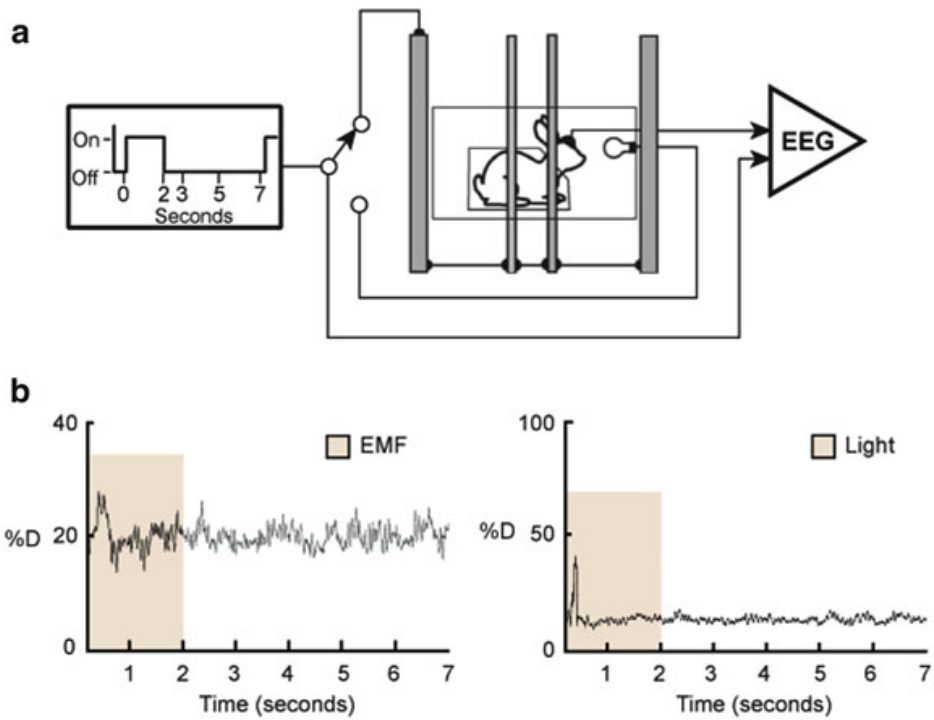


Fig. 7.13 A new paradigm for studying brain function. (a) Apparatus for exposing rabbits to EMF (2 G, 60 Hz magnetic field) or light (positive control stimulus). Stimuli on for 2 s during each 7-s trial. (b) Average %D(t) calculated from the EEG (50 trials) of a rabbit exposed to EMF (left) or light (right) [25]

EEG (Fig. 7.13). Even though the EMF was too weak to be consciously perceived, statistically significant changes were observed in all the rabbits, and the effects were manifested in both %R and %D (Fig. 7.14). Thus each of the recurrence variables captured EMF-caused law-governed activity in the EEG that could not be consistently detected by any other known method.

The generality of the phenomenon was established by showing that essentially the same results were obtained in rats [26] and human subjects [27]. In the human study, the EEG was measured from six derivations in each of eight subjects, and the effects were essentially identical regardless of the derivation. This was the expected result based on the complexity conjecture (Fig. 7.5), which assumed that each EEG was a sample measurement of the same brain state. The overall results [25–27] indicated that environmental strength EMFs could be transduced by the nervous system, resulting in subliminal changes in brain electrical activity that were consistently detectable by recurrence analysis, but not by linear methods of analysis.

7.4.2 Statistical Basis of Brain Recurrence Analysis

Application of recurrence analysis to biological time series requires an experimental and statistical framework to permit the meaning of the calculations to be ascertained.

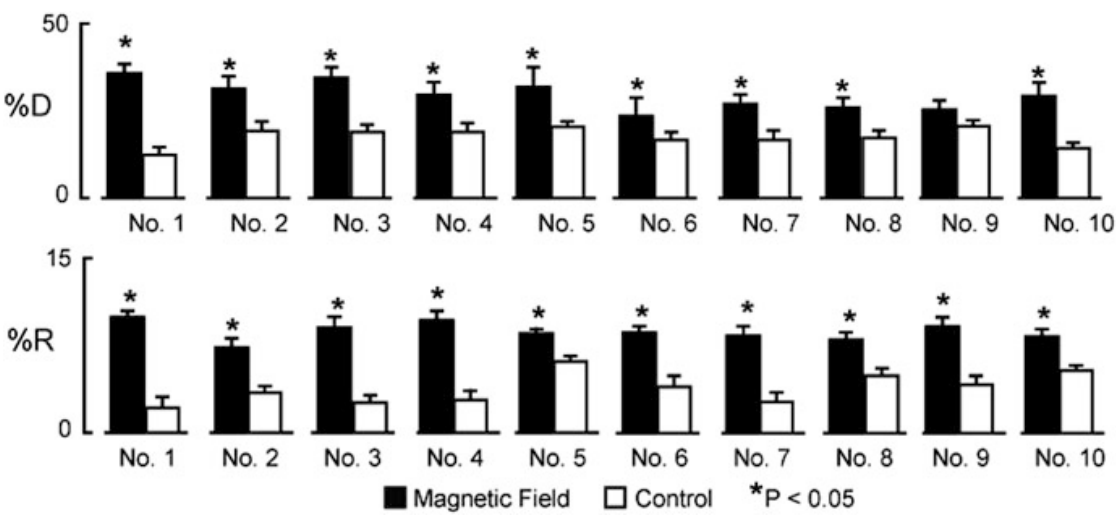


Fig. 7.14 Effect of 2 G, 60 Hz on each of 10 rabbits, 5 female (No. 1–5) and 5 male (No. 6–10) as assessed using percent determinism (%D) and percent recurrence (%R). For each rabbit and each quantifier, the difference between the exposed and control EEG epochs was evaluated using the Wilcoxon signed rank test. Window for recurrence analysis centered at 250 ms, with width of 250 ms. The average values of the quantifiers (\pm SD) are presented for each rabbit [25]. With one exception (%D, No. 9), statistical significance was obtained in each animal, for each recurrence variable

One approach involves comparing the experimental and control recurrence time series, point by point. For example consider a study in which EEGs were sampled at 300 Hz and analyzed to detect transient changes in brain electrical activity caused by the onset and/or offset of an auditory stimulus [28]. During independent trials, the stimulus was turned on and off and the inter-stimulus time period was used as the control for the immediately preceding stimulus period. %R(t) was extracted from the EEG and analyzed statistically to detect deterministic changes caused by the onset and offset of the stimulus.

Stimulus onset produced the expected linear EP in the time-averaged EEG, which peaked about 100 ms later (Fig. 7.15a, left panel). The EP was also seen in the probability curve, which displayed the P values point by point in the corresponding %R(t) relative to the controls (Fig. 7.15a, right panel). The EP was identified as a continuous series of tests that were pair-wise significant at $P < 0.05$. In addition, a nonlinear EP was detected (about 550 ms after stimulus onset) that was not resolved in the average EEG. Fig. 7.15b depicts an instance in which the subject did not exhibit a linear offset EP (left panel) but did exhibit a nonlinear EP detected by analysis of brain recurrence (ABR) (right panel).

Statistical considerations are critical if the ultimate goal is to make reliable, generalizable conclusions regarding the application of recurrence analysis. Simply varying embedding parameters or evaluating multiple recurrence variables yields only self-referential, non-generalizable results.

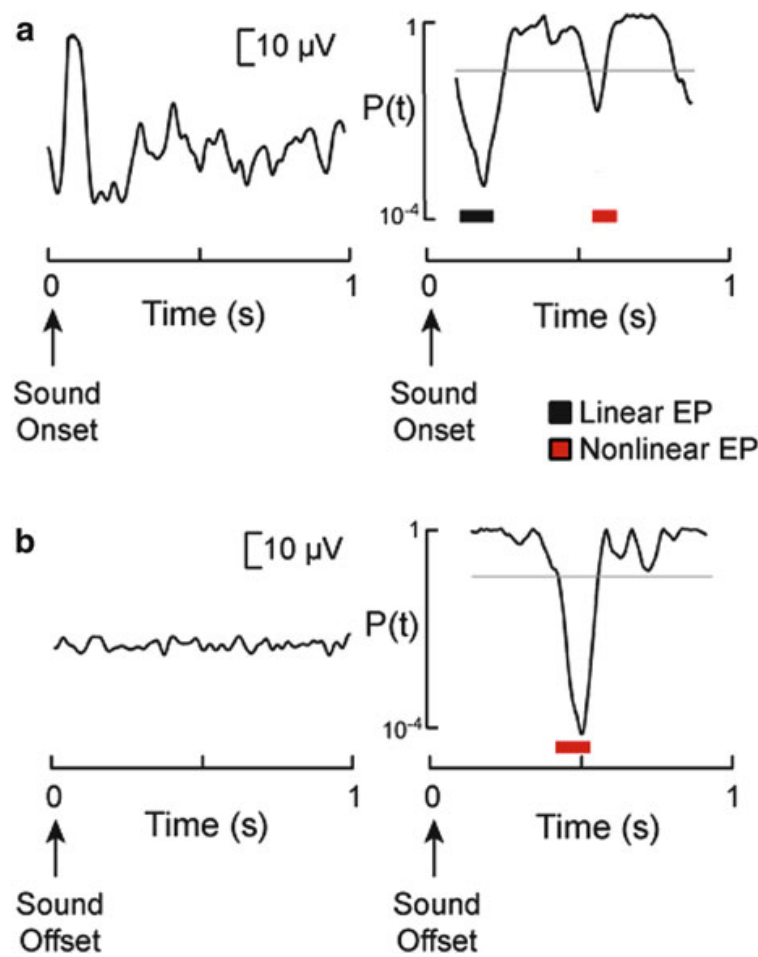


Fig. 7.15 Statistical basis for detection of auditory evoked potentials using recurrence analysis. *Left panels*, time-averaged EEG. *Right panels*, point by point probability (P) that $\%R(t)$ computed from the stimulated EEG was equal to the corresponding values computed from the controls. (**a**, **b**) EPs due to sound onset and offset, respectively [28]. Bars indicate EPs (>9 consecutive pair-wise significant t tests) ($N = 50$ trials). EP, evoked potential. Gray line, $P = 0.05$. Total of 300 t tests [28]

7.4.3 Discovery of Human Magnetic Sense

All known sensory modalities generate EPs in response to the onset and/or offset of the cognate stimulus. Consequently observations of EPs is good evidence of the existence of a sensory modality for the type of stimulus that elicited the EPs. We hypothesized the existence of a human sense capable of detecting weak magnetic fields, and used recurrence analysis to detect the EPs that we planned to interpret as evidence of the putative sensory capability.

Subjects were exposed to either sound (positive control) or a magnetic field for 2 s every 7 s (stimulus on at $t = 0$ and off at $t = 2$) (Fig. 7.16). Evoked potentials were expected within 100–400 ms of the stimuli. The control interval for both stimuli was the 1-s interval beginning at $t = 5$. In a study of 17 subjects, 16 subjects exhibited magnetosensory evoked potentials (MEPs) ($P < 0.05$ for *each* subject) [29]; the

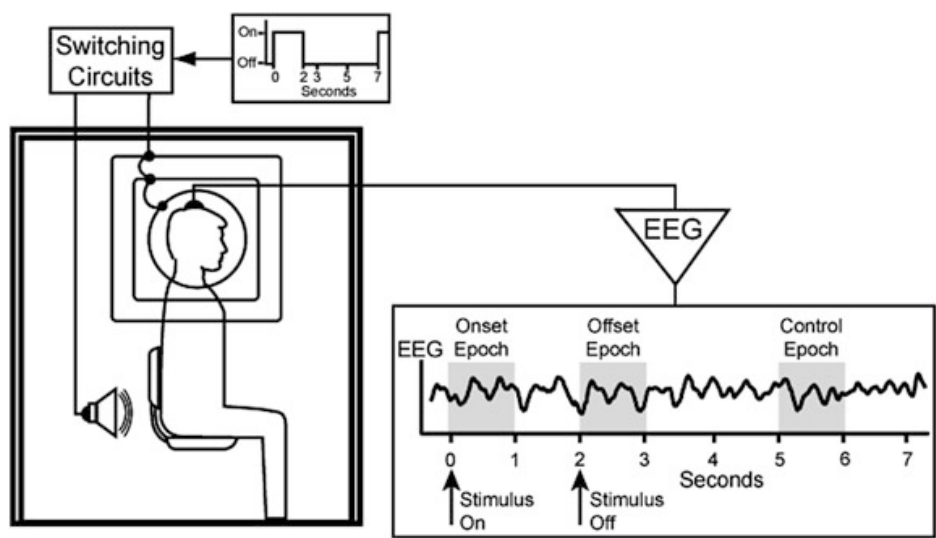


Fig. 7.16 Use of recurrence analysis to detect nonlinear magnetosensory evoked potentials. A computer-generated timing signal controlled application of the stimuli (on for 2 s, interstimulus period of 5 s). Location of onset, offset, and control epochs of a typical trial are shown. Magnetic field, generated by two sets of coils (separated by 65 cm), was homogeneous to within 5 %. Each subject had 80 trials of each stimulus [29]

latencies and durations of the EPs varied within the expected ranges, depending on the subject, and were triggered by onset and/or offset of the field. Onset results are shown in Fig. 7.17. No MEPs were detected directly in the EEG using the method of time averaging.

Because the ability to generate EPs is the hallmark of a sensory modality, the results of the study indicated that human beings can detect magnetic fields [29]. The existence of a human magnetic sense had not been discovered previously because there was no reliable method of observing nonlinear EPs prior to the development of recurrence analysis. When the new paradigm for studying the brain was employed in an appropriate statistical framework, the existence of phenomena not otherwise recognizable was validated. Both %R and %D were necessary to demonstrate the consistency of the effect. Their joint use permitted consistent detection of EPs from subject to subject, as would be expected if magnetodetection were a common human capability.

Because the study objective was fully achieved using only two recurrence variables, the other variables were not employed. In other words, %R and %D were necessary and sufficient.

7.4.4 Rationalizing Inconsistency

A question raised by the discovery of a human sensory capability for EMFs (Fig. 7.17) involved the nature of the dynamical relationship between the stimulus and the response. Field-induced changes were detected by recurrence analysis but

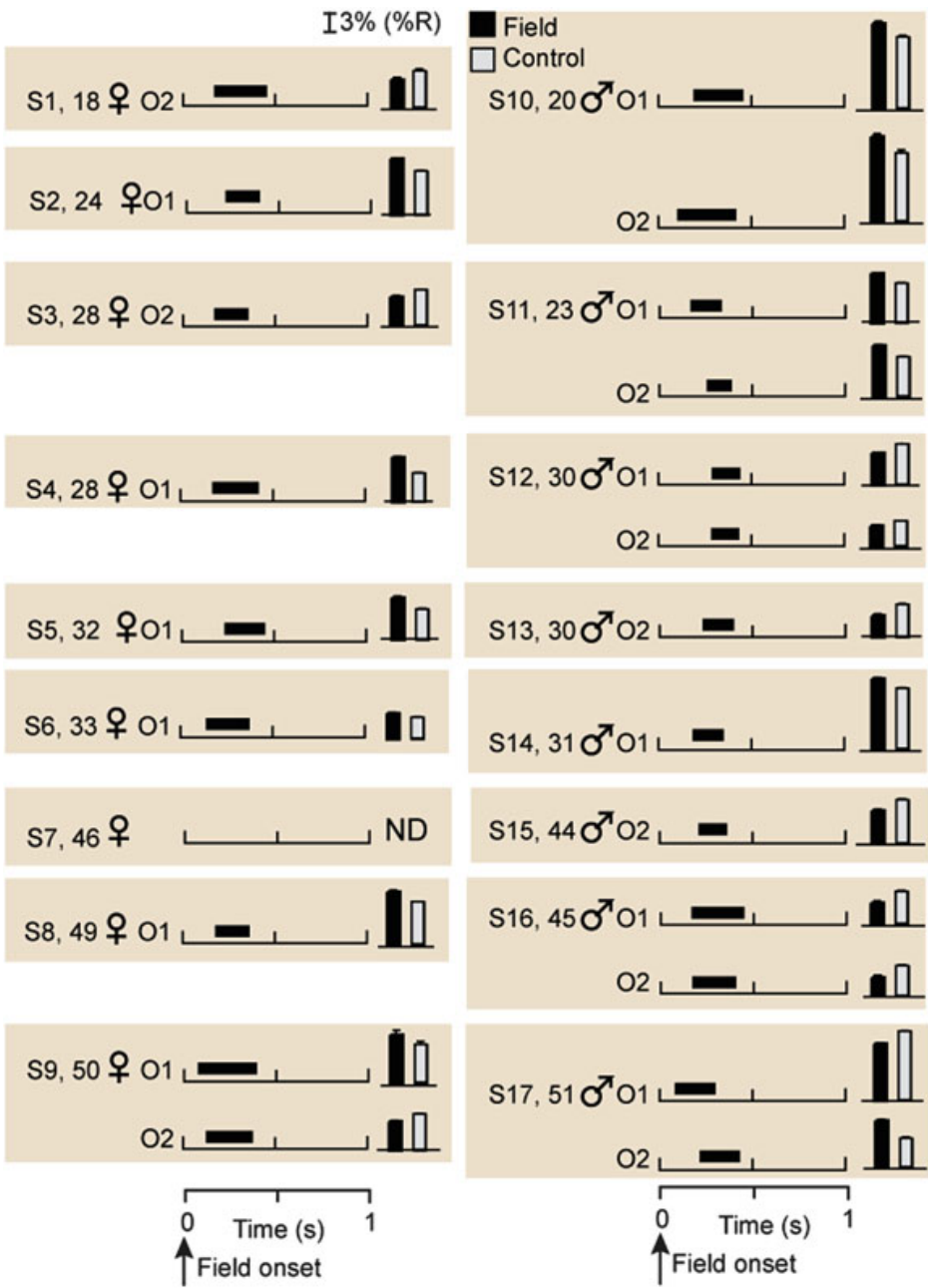


Fig. 7.17 Onset magnetosensory evoked potentials (MEPs) measured from occipital electrodes. Latency and duration in each subject are indicated by a bar over the time axis, which show the locations in the onset epochs within which individual points differed pair-wise ($P < 0.05$) from the corresponding control points. Bar graph, %R(t) (average of the significant points) (SD not resolved at scale presented). Alpha filtering was performed in nine subjects: S1, 3 (8–10 Hz), 5, 6, 9 (O1), 11, 12, 16, 17. ND, not detected [29]

not by linear methods, suggesting that responses of the subjects were governed by nonlinear laws. Nonlinear systems do not follow the law of superposition, and therefore the direction of their reactions to changes in external conditions cannot be predicted. Under the hypothesis that MEPs were nonlinearly related to the field, brain electrical responses exhibited by individual subjects would be expected to

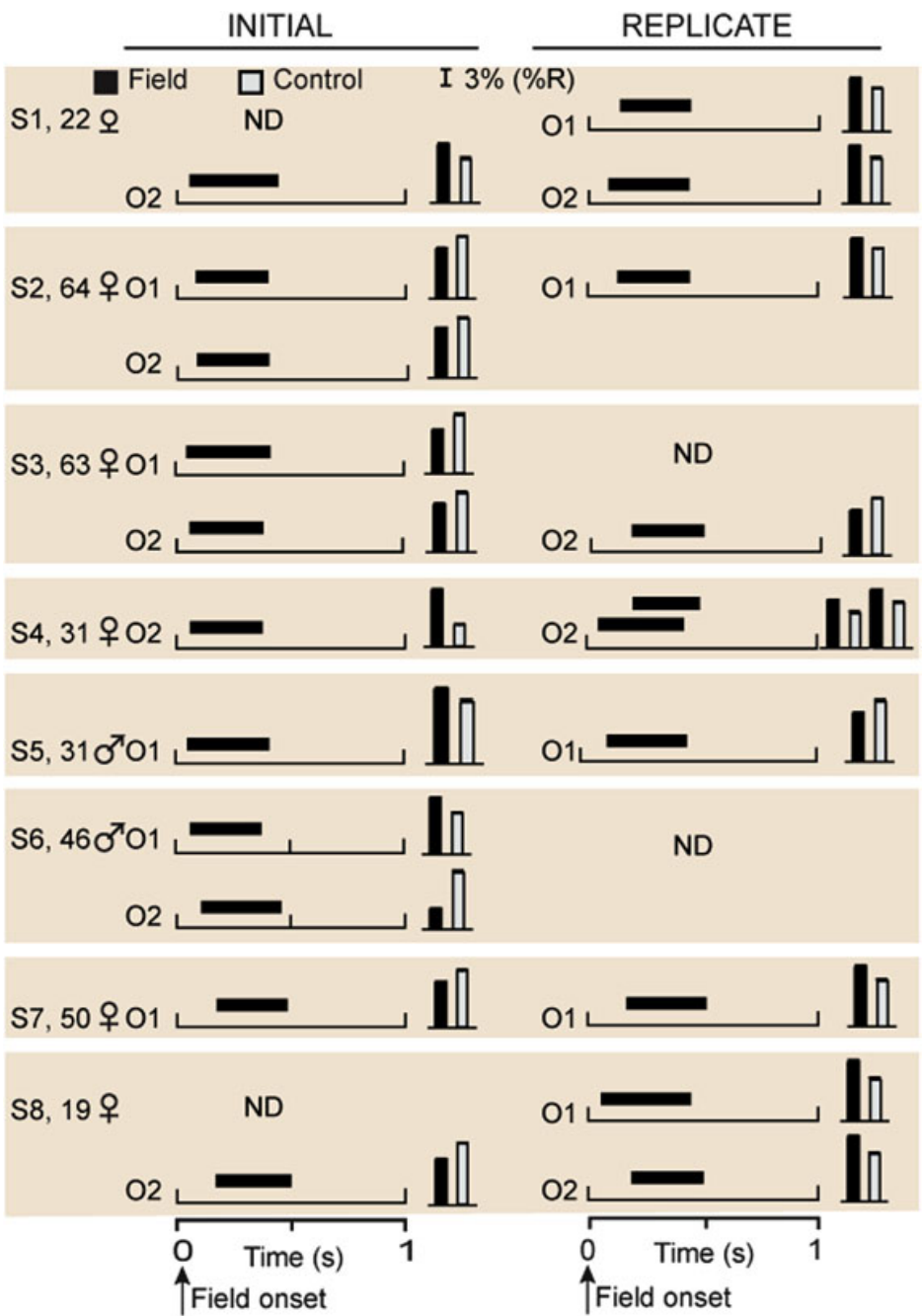


Fig. 7.18 Onset magnetosensory evoked potentials (MEP) in initial and replicate (performed at least 1 week later) studies, using recurrence analysis. Latency and duration in each subject are indicated by a bar over the time axis. Bar graphs, %R(t) (average of pair-wise significant points) (SD not resolved at scale presented). ND, not detected [30]

differ even when the experimental conditions were replicated. The hypothesis was tested in a group of eight subjects by comparing each subject's response to a specific magnetic stimulus at two times, at least 1 week apart. Intra-subject differences were observed that were of the type manifested only when the laws governing the relationship between the field and the responses were nonlinear (Fig. 7.18) [30].

Analysis of brain recurrence facilitated the discovery of new phenomena and new ways of characterizing brain activity, but ABR also had the unsettling effect of requiring a reconsideration of the process by which meaning was made from observations. Linear models of brain activity were so ubiquitous that the fundamental property of the models—adherence to the law of superposition—was often adopted as a criterion of the reliability of observations. It is not, of course, but that misconception commonly led to the false conclusion that inconsistent data (Fig. 7.17) was unreliable [31]. Reflection on the actual meaning of the reproducibility requirement of the scientific method reveals that the method requires *phenomena* be reproducible. In the present study, for example, that brain activity is reproducibly affected whenever the stimulus is applied, in accordance with the system's differential law. The reproducibility requirement applies to *data* only when the differential law is linear.

7.4.5 Canonical Conditions and ABR Variables

The optimal conditions for ABR, given the range of physiological and clinical problems considered thus far, were: sampling frequency 300–500 Hz; digital pass band 0.5–35 Hz; embedding dimension 5; delay 5; radius 15 %; scaling Euclidean; line parameter 2–20; recurrence window 30–500 points; P window 10–100 points. In all cases these canonical conditions were determined empirically, with the endpoint being the sensitivity with which the effect sought could be observed at a predetermined level of acceptability of the false-positive rate. The adopted parametric values were only bluntly optimal in the sense that reasonable departures in one or even a few choices generally had no material effect on the results. Whether future studies will require different parametric choices remains to be seen.

A further unresolved issue regarding ABR involves the potential usefulness of recurrence variables other than %R and %D. The approach we ultimately adopted regarding choices of variables was to begin with %R(t). When the putative effect sought could not be detected consistently, %D was added to the analysis, with appropriate care taken to ensure that the family-wise error rate for detecting an effect remained at the appropriate level. Use of only these variables proved sufficient to resolve all problems addressed thus far, and consequently the other suggested recurrence variables [21] have not been employed in any published studies involving ABR.

Despite the macroscale similarity of %R(t) and %D(t), use of both time series was absolutely necessary to demonstrate the consistency of the phenomena under investigation. An example is shown in Fig. 7.19 [32]. The subjects were exposed to a brief magnetic stimulus (50 ms, 60 or 30 Hz) such that the onset and offset evoked potentials (EPs) overlapped. The presence of the overlapped EPs was examined in each of six derivations from 15 subjects. For a given stimulus, say 60 Hz, six statistical tests were planned (one for each derivation) at a pair-wise significance level of 0.05. If at least three tests were significant, we planned to conclude that

the subject had exhibited an EP (family-wise error rate (PFW), 0.001). Otherwise, six additional tests were done using %D(t). Based on %R, four subjects met the a priori condition for exhibiting an EP (S1 (60 Hz), S11 (60 Hz), S12 (30 Hz), S13 (30 Hz)). When the results using %D were added, five additional cases of EP detection were revealed (S1 (30 Hz), S2 (60 Hz), S5 (30 Hz), S14 (30 Hz), S15 (60 Hz)), an inference that was justified based on consideration of the family-wise error rate. Analyses to reach the three-derivation threshold were conducted two additional times with %R and with %D, but after filtering the EEG to remove 8–10 Hz or 9–12 Hz energy (an interval that made no material contribution to the effect of the field, hence lowered the sensitivity of the t test), EPs (≥ 3 derivations) were found in 29 of 30 cases (Fig. 7.19, *All Effects*); S14 at 60 Hz was the exception. The results were family-wise significant at $P < 0.05$ (account taken for the number of tests performed) in 26 of the 29 cases (exceptions were S6 (60 Hz), $p = 0.14$; S12 (60 Hz), $p = 0.25$; S15 (30 Hz), $p = 0.15$).

Preconditioning the EEG time series (filtering at 8–10 Hz or 9–12 Hz) combined with use of %R and %D has been sufficient to successfully resolve all experimental questions regarding analysis of brain electrical activity considered thus far.

7.4.6 *Inferring Mechanisms*

We proposed that human detection of EMFs was mediated by an interaction of the field with charged glycoproteins attached to the gate of an ion channel, resulting in a force that tended to open the gate (Fig. 7.20a). The proposed model had been shown to be energetically consistent with physical and thermodynamic laws and sufficiently sensitive to respond to environmental-strength EMFs [33]. The question considered was whether the mechanism was sufficiently rapid to explain the observed field-induced evoked potentials. First we measured single-channel currents in a species of fish known to be electrically sensitive and determined that the EMF-sensitive membrane ion channels opened or closed in ~ 0.2 ms (Fig. 7.20b). We then applied a 0.2-ms magnetic stimulus to human subjects with the intent of interpreting observations of evoked potentials as evidence that the ion channel involved in field transduction was a force receptor (Fig. 7.20a).

A DC EMF that had a 10-ms rise-time and a 0.2-ms fall-time was used (Fig. 7.20c). Onset potentials were observed in all subjects, and offset potentials were observed in 60 % of the subjects [34]. These results were similar to those found earlier when both the rise- and fall-times were 10 ms [29]. We concluded therefore that the human EMF transduction system was capable of detecting fields that changed at least as rapidly as 0.2 ms (on-to-off), thus supporting the theory that transduction was directly initiated by a force receptor (Fig. 7.20a).

Time-varying magnetic fields induce electric fields in accordance with Faraday's law. Consequently *both* fields were present simultaneously in the brain in the studies involving application of magnetic fields (Figs. 7.16 and 7.20). To determine whether the electric field *alone* could explain the occurrence of the brain potentials

Subject	Stimulus (Hz)	%R	%D	%R (8-10Hz)	%D (8-10Hz)	%R (9-12Hz)	%D (9-12Hz)	All Effects	No. Tests	PFW
S1 (30F)	60	O1 C4 P4	—	—	—	—	—	O1 C4 P4	6	0.001
	30	O2 C3	O2 C3	—	—	—	—	O2 O2 C3 C3	12	0.001
S2 (54M)	60	O2	O2 P3	—	—	—	—	O2 O2 P3	12	0.004
	30	O1	X	C4	O1	—	—	O1 O1 C4	23	0.022
S3 (23M)	60	X	C4 P3	X	O1	—	—	O1 C4 P3	22	0.047
	30	P3	P3	O2, C4	—	—	—	O2 C4 P3 P3	17	0.004
S4 (22M)	60	O1	O1	C4	—	—	—	O1 O1 C4	17	0.009
	30	C3	C3	O1	—	—	—	O1 C3 C3	17	0.025
S5 (51F)	60	X	X	O1	O1	C3	—	O1 O1 C3	29	0.042
	30	O1 P3	P3	—	—	—	—	O1 P3 P3	12	0.01
S6 (23M)	60	C4	C4	X	X	P4	—	C4 C4 P4	27	0.14
	30	X	O1	X	X	O2 P3 P4	—	O1 O2 P3 P4	29	0.017
S7 (29F)	60	X	X	O1 O2 C4 P3 P4	—	—	—	O1 O2 C4 P3 P4	18	0.001
	30	C4	C4	C3	—	—	—	C3 C4 C4	17	0.046
S8 (45F)	60	X	O1	O2	O2	—	—	O1 O2 O2	23	0.003
	30	X	X	P4	P4	O2	—	O2 P4 P4	29	0.09
S9 (25M)	60	C4	X	O1	O1	—	—	O1 O1 C4	22	0.028
	30	X	X	O1	O1 C4	—	—	O1 O1 C4	24	0.03
S10 (21M)	60	O2	P3	C4	—	—	—	O2 C4 P3	17	0.025
	30	O2	O2	P3	—	—	—	O2 O2 P3	17	0.009
S11 (31F)	60	C3 C4 P4	—	—	—	—	—	C3 C4 P4	6	0.002
	30	O2	O2	X	C4	—	—	O2 O2 C4	22	0.017
S12 (51F)	60	X	X	X	P4	C4 P4	—	C4 P4 P4	26	0.251
	30	O2 P3 P4	—	—	—	—	—	O2 P3 P4	6	0.001
S13 (31F)	60	X	X	O1 O2	O2 C3	—	—	O1 O2 O2 C3	24	0.003
	30	O1 O2 P3	—	—	—	—	—	O1 O2 P3	6	0
S14 (24F)	60	X	X	O2 P3	X	X	X	O2 P3	24	0.286
	30	P3	C3 P3	—	—	—	—	C3 P3 P3	12	0.018
S15 (36M)	60	O2	O2 C3	—	—	—	—	O2 O2 C3	12	0.004
	30	X	X	X	P3	O2	O2	O2 O2 P3	25	0.15

Fig. 7.19 Magnetosensory evoked potentials from indicated electrode derivations in 15 subjects. Column heads indicate conditions of analysis. Effects in R(t) and D(t) are shown in *red* and *green*, respectively. X, MEPs not detected. Dashes indicate conditions not analyzed (unnecessary because the EPs had already been detected). PFW, family-wise error. The stimulus was applied coronally to subjects S1–S10, and sagittally to subjects S11–S15 [32]

detected using ABR, we repeated the experiments, using an external electric field that produced an internal electric field but no magnetic field. The results obtained using electric fields of 27–430 V/m were identical to the results seen when magnetic fields were applied [35].

Our observation that electric fields as low as 27 V/m triggered EP that could be detected by ABR was particularly surprising because 27 V/m was roughly equivalent to the electric field induced by a magnetic field of about 10 mG, which is ubiquitous in the general environment. We confirmed this result by performing additional studies using 10 and 50 mG, and EPs were again observed (Fig. 7.21)

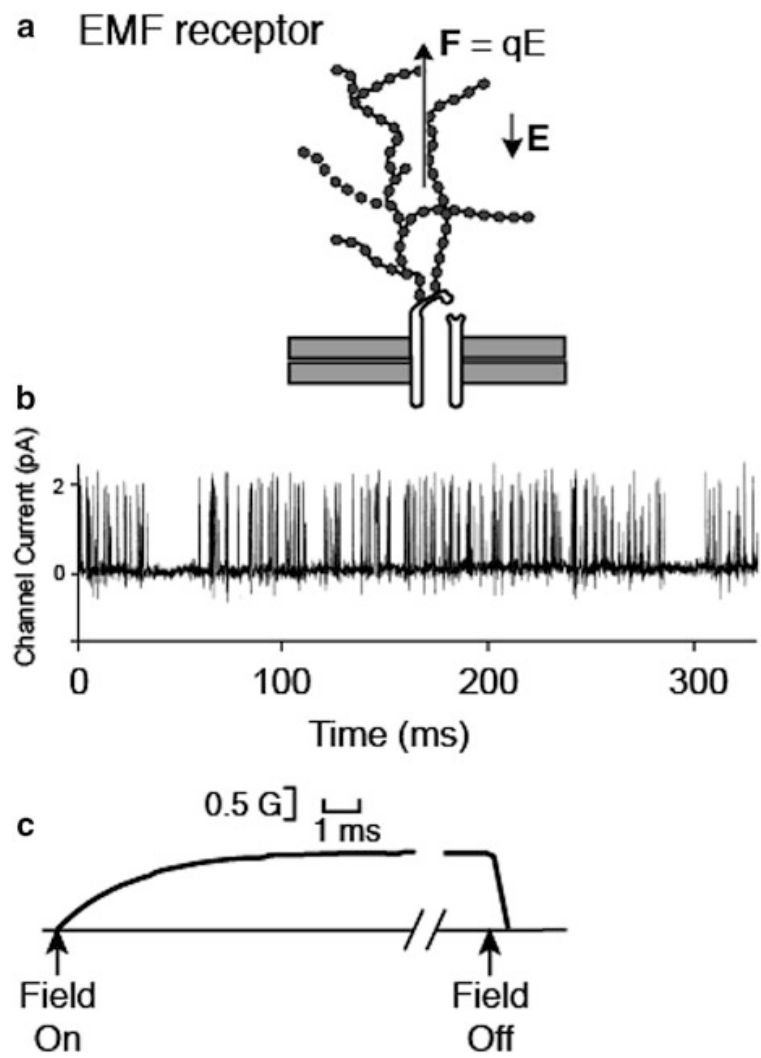


Fig. 7.20 Transduction of weak electromagnetic fields (EMFs) [33]. (a) Proposed EMF receptor; the rate of change of the applied magnetic field induces an electric field (E) that produces a force (F) on charged glycoproteins attached to a channel gate. (b) Single channel current from a voltage-sensitive channel in an electroreceptor cell in *Kryptopterus bicirrhys* (an electrosensitive species of catfish), indicating that the channel can open or close in about 0.2 ms. (c) Rise- and fall-times of the DC magnetic stimulus. Experimental set-up shown in Fig. 7.16

[36]. As expected, the EPs were manifested in $\%R(t)$ and $\%D(t)$ as both increases and decreases in determinism. No known method of EEG analysis other than ABR can detect this kind of dynamical change.

ABR was also capable of evidencing specific neurotransmission processes in the brain (Fig. 7.22). The hypothesis considered was that EMF detection involved synapses in the trigeminal nucleus that projected to the thalamus via glutamate-dependent pathways. If so, an anesthetic agent that antagonized glutamate neurotransmission would be expected to degrade EMF-evoked potentials (EEPs). We tested the hypothesis using ketamine which blocks glutamate receptors and xylazine which does not do so. EEGs of rats were examined using ABR to observe EEPs in the presence and absence of ketamine and xylazine anesthesia. EEPs were observed

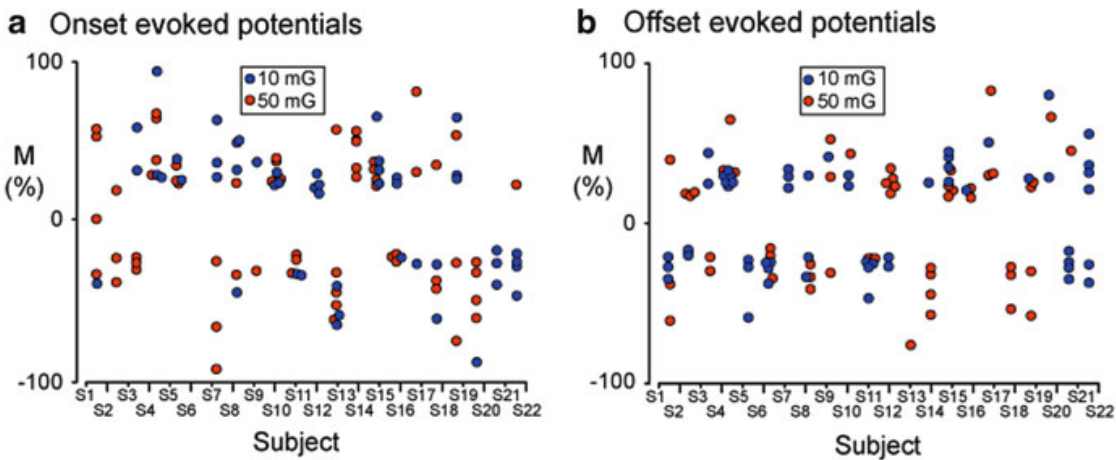


Fig. 7.21 Relative magnitude (M) of each evoked potential (expressed in percent) from each subject as determined by recurrence analysis. (a, b) Onset and offset responses, respectively. For each potential, $M = 100(E - C)/0.5(E + C)$, where E was the average of the recurrence variable (%R or %D) over the statistically significant latency interval, and C was the corresponding average in the control epoch. Where necessary, points were jittered to facilitate resolution. Values greater than 100 % are shown as 100 % [36]

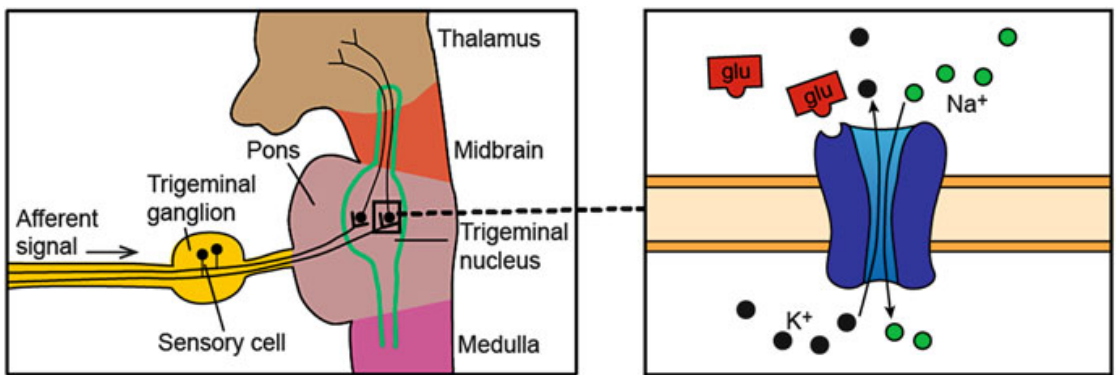


Fig. 7.22 Hypothesis regarding synaptic processes responsible for the afferent signal produced by EMF detection [37]. The afferent signal generated by EMF transduction triggers glutamate-mediated (glu) neurotransmission in the trigeminal nucleus leading to thalamic projection of the signal

in rats under xylazine anesthesia, but not when ketamine was used, indicating that the afferent signal triggered by transduction of EMFs was likely mediated by glutamate [37].

7.4.7 Cell-Phone Effects on the Brain

Questions have been raised concerning whether the electromagnetic fields emitted by cell phones could affect brain electrical activity, thereby raising the possibility that the field could ultimately lead to brain cancer or other diseases. Cell phones

emit a complicated temporal array of electromagnetic, acoustic, thermal, and tactile stimuli, any one of which might be responsible for effects on brain electrical activity that might be observed in association with cell-phone use. Thus the problem of determining whether cell-phone EMFs affect the brain requires a showing that brain activity is affected when *only* an EMF stimulus is presented, which obviates use of an actual cell phone in controlled studies.

Cell phones generate narrow pulses of high-frequency radiation that are formed by current pulses from the phone's battery. We knew that the rise- and fall-times of the pulses were within the range detectable by the nervous system [34], and that the physical process which coupled the applied magnetic field to the brain was the induced electric field generated by the rate of change of the magnetic field (Faraday's law) [35]. There was a high a priori likelihood that the brain detected the cell-phone magnetic pulses because the rate of change of the pulses produced an extraordinarily high electric field, far greater than that induced by the 60-Hz fields [36].

The question whether a *single* cell-phone pulse could be detected by the brain was addressed by applying simulated pulses in a series of independent trials, and using ABR to evaluate whether the pulses produced EPs (Fig. 7.22). Evoked potentials due to just a single cell-phone pulse were demonstrated in 18 of 20 subjects [38]. The pulse rate of a typical cell phone is 216 Hz. Consequently, the results of the study implied that 216 EPs were produced each second in the brain of a typical cell-phone user. The potential public-health implications remain to be assessed, but the salient point here is that the potential health problem could not have been detected except for ABR.

A critically important property of recurrence analysis is its applicability to *any* time series. A broad range of data manipulations may therefore be performed on the original time series or even on the computed recurrence time series, assuming that proper controls are included in the process. The possible changes include any systematic modification of the data such that features which do not contribute to recurrence-based discrimination between groups being compared are minimized, thereby increasing the sensitivity of the analysis (more likely that the effects sought will be resolved), alpha filtering for example (Fig. 7.19).

As an example of increased sensitivity based on changes in both the EEG and recurrence time series, consider a study of the effect of the high-frequency field produced by cell phones on brain electrical activity [39]. The internal antenna in a cell phone was disconnected and replaced with an external antenna positioned directly above the head of a rabbit. The antenna was energized for 2 s, with a 5-s inter-stimulus period ($N \geq 60$ 7-s trials); the last 2 s of each trial served as the control (Fig. 7.23). Both %R and %D were computed over a 300-ms window, centered at 250 ms from the beginning of the stimulus epoch. During preliminary studies involving one rabbit, the EEG and the recurrence time series were systematically altered with the goal of finding conditions of analysis that yielded statistically significant differences between the *E* and *C* epochs (Fig. 7.23). When the EEG was digitally filtered to remove 3, 4, and 8–12 Hz, and only 85 % of the attractor volume was included in the *E–C* comparisons, both %R and %D were significantly different

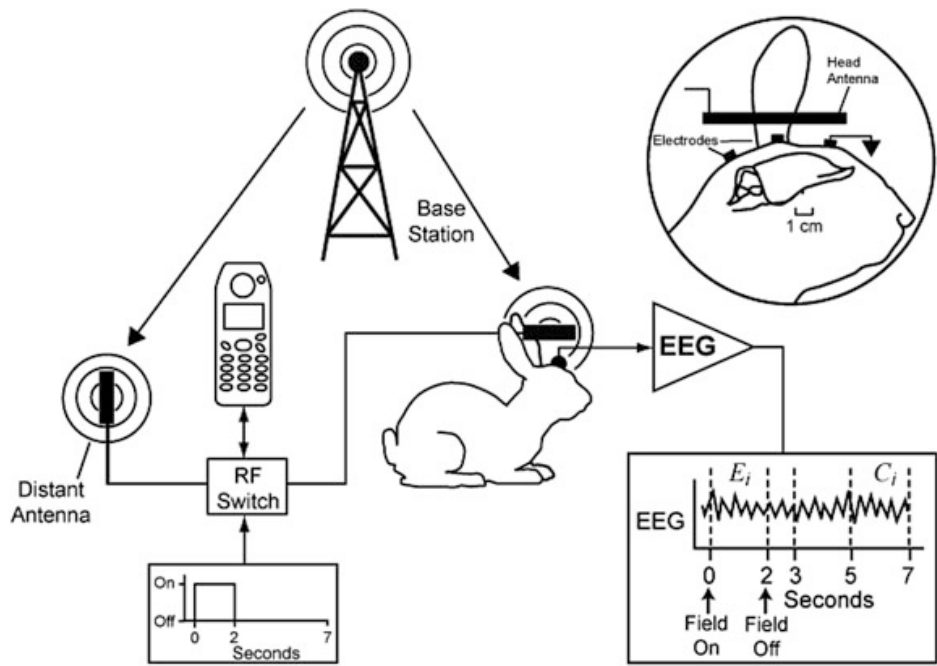


Fig. 7.23 Schematic representation of the experimental system. The detail shows the location of the electroencephalogram (EEG) electrodes relative to the head antenna [39]

($P < 0.05$). The data-conditioning procedures were then prospectively applied to nine additional rabbits and both recurrence variables were shown to be consistently altered during the exposure epochs [39].

The study demonstrated that changes in brain electrical activity associated with specific external conditions were detected more efficiently when preliminary steps were taken to precondition both the original time series and the recurrence time series to minimize signal characteristics that did not contribute to discrimination between the experimental and control data.

7.4.8 Detecting the Presence Effect

Another experimental question that can be addressed using ABR that previously could not be considered because of the absence of a suitable method for quantifying aperiodicity in the EEG is depicted in Fig. 7.24. When a stimulus is applied at $t = 0$ and removed at $t = 2$ s, nonlinear onset and offset EPs are produced that can be detected by recurrence analysis. With regard to the interval after the onset EP has decayed but while the stimulus is still present (E epoch), we can ask whether the *presence* of the stimulus (which entails a change in brain electrical activity) can be detected. Surprisingly, for typical stimuli such as light or sound, there is no objective method by which the subjective sense of detection (perception) can be objectively verified. By calculating %R and %D during the on-time of the stimulus but after the onset EP had decayed, objective evidence was found that confirmed the subject's subjective report of stimulus perception (hence confirming the change

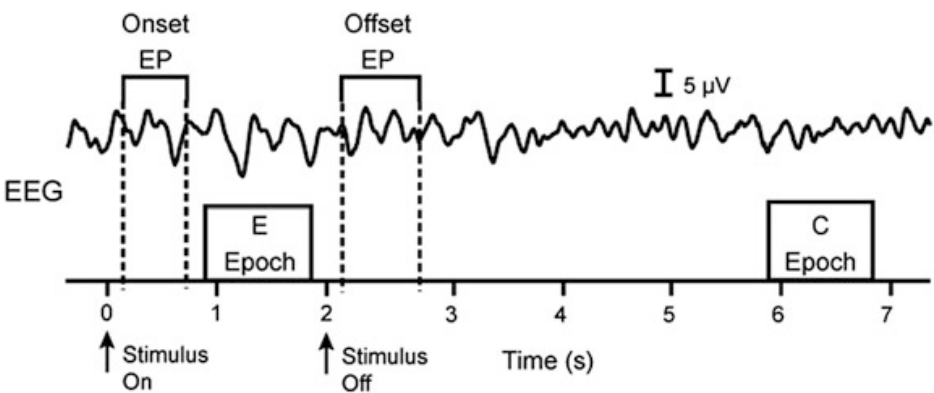


Fig. 7.24 Detection of changes in the EEG induced by the *presence* of a stimulus. EEG trial showing the locations of the epochs used to study the physiological effect of the presence of a stimulus (in distinction to the transient effect caused by its onset and offset). *EP*, evoked potential; *E*, exposed; *C*, control [40]

a) Sound Stimulus				
Subject	Age/Gender	%D	%R	V _{rms}
S1	30/M	O1 C3 C4 P3	O1 C3 C4 P3	ND
S2	45/M	O1 O2 C3 C4 P3 P4	O2 C3 C4 P3 P4	ND
S3	32/F	O1 O2 C4 P4	ND	ND
S4	29/F	ND	C3 C4 P4	ND
S5	28/F	O1 C3 C4 P3	O1 P4	O1 C3 C4 P3
b) Field Stimulus				
Subject	Age/Gender	%D	%R	V _{rms}
S6	18/F	O1 C3 C4 P3 P4	ND	ND
S7	30/M	O1 O2 C3 C4 P3 P4	O1 P3 P4	ND
S8	50/F	O1 O2 C3 C4 P3 P4	ND	ND
S9	49/F	O1 O2 C3 C4 P3 P4	O1 O2 P3 P4	ND
S10	46/F	O1 C3 C4 P3	ND	ND
c) Light Stimulus				
Subject	Age/Gender	%D	%R	V _{rms}
S11	51/F	C3 P4	ND	ND
S12	29/M	ND	ND	ND
S13	50/M	ND	O1 C4	C4 P4
S14	46/F	O1 O2 P3 P4	ND	ND
S15	31/F	O1 C4	ND	ND

Fig. 7.25 Detection of the presence effect. Light onset and offset were at $t=0$ and $t=2$ s, respectively. Using nonlinear (%D, %R) and linear (V_{rms}) analysis, brain electrical activity at 0.7–1.7 s was compared with that at the inter-stimulus epoch (3.7–4.7 s) for each of the six derivations from each subject. Effects in %R(t), %D(t) are shown in *red* and *green*, respectively. The derivations for which the comparisons differed significantly ($p < 0.05$) are listed. The presence effect was not detected by linear analysis (V_{rms}) except in S5 and S13 [40]

in brain electrical activity). The same conditions of analysis permitted an objective verification of a change in brain electrical activity when an EMF was applied (a subliminal stimulus for which the subjects cannot report the subjective sensation of perception) (Fig. 7.25). Thus ABR offers the possibility to study the dynamics of continuous detection (perceptual and non-perceptual) of external stimuli (the *presence effect*).

7.4.9 *Diagnosing Multiple Sclerosis*

Multiple sclerosis (MS) is an immune-system-mediated disease that degrades brain structure, leading to serious clinical consequences. We hypothesized that the neuronal networks involved in attention to stimuli onset and/or offset would be affected by the presence of MS, and that the effect could be detected by ABR. An onset response to an EMF stimulus occurred in only 27 % of a group of MS patients studied, compared with 90 % in the control group (Fig. 7.26). Further, ABR analysis of the baseline EEG in both groups provided additional evidence that the two groups could be distinguished; mean %R was 30 % greater in the MS group, and %D was 15 % greater [42]. In principle, EEG analysis could provide the basis of a functional method for detecting organ-level changes in brain activity associated with MS even before the changes were detected by imaging brain structure.

7.4.10 *Applications in Sleep Medicine*

Human sleep is commonly studied by analyzing simultaneously digitized signals from the brain, heart, skeletal muscle, and other physiological systems [43]. Sleep macroarchitecture is characterized by examining the signals in 30-s epochs and classifying them on the basis of standardized rules into one of four mutually exclusive stages, either rapid-eye-movement (REM) sleep or progressively deeper stages of non-REM sleep respectively termed N1, N2, and N3; the other recognized stage is wake after sleep onset (WASO) [44]. A graphical record of the distribution of sleep-stage changes during overnight sleep, termed a hypnogram (Fig. 7.27, bottom panel) provides sleep-medicine specialists an overview of the macroarchitecture of the night's sleep. Knowledge of sleep-stage distributions permits normal and pathological sleep to be distinguished [46].

The concept of sleep stage is fundamental to an understanding of sleep physiology, but has several limitations. Staging emphasizes a discontinuity of sleep, leading to its representation as a discrete process rather than a continuous process which is actually the case. Second, sleep staging is rule-determined [44], and application of the rules depends on observer judgment; typical average inter-rater agreement among experts is 80–85 %. Finally, and probably most importantly, the non-REM stages are defined in terms of the behavior of the EEG, whereas REM sleep is defined in terms of the coordinated behavior of three signals, only one of which is the EEG. This fundamental difference prevented characterization of sleep stages in terms of a single, continuous, objective, physiological variable. ABR provides a complementary perspective on sleep characterization that is based solely on EEG metrics [47].

For application of ABR to the sleep EEG, %R and %D were calculated second-by-second and then averaged over 30-s epochs, resulting in approximately 900 values for a typical 8-h sleep study. The results, when viewed graphically, reveal

a) Participants with multiple sclerosis		
Participant	Onset Stimulus	PFW
1 (40)	NE	—
2 (34)	NE	—
3 (52)	NE	—
4 (32)	O1 O2 C3 C3 C4	0.003
5 (19)	NE	—
6 (30)	O2 O2 C3	0.029
7 (18)	NE	—
8 (27)	C3 C4 P4	0.029
9 (50)	NE	—
10 (31)	NE	—
11 (38)	NE	—
b) Participants who had no medical complaints		
Participant	Onset Stimulus	PFW
1 (51)	O2 O2 C3	0.031
2 (66)	O2 C3 C3 P4	0.001
3 (22)	NE	—
4 (26)	C3 C4 C4 P3	0.001
5 (23)	C3 C4 P4	0.001
6 (23)	C3 C3 C4 C4	0.001
7 (23)	O1 C3 C3 P3	0.004
8 (46)	O1 O1 C3	0.005
9 (23)	O1 O2 C4 C4 P3 P4	0.000
10 (25)	P3 P3 P4	0.084

Fig. 7.26 Changes in brain electrical activity induced by EMF stimuli. NE, No effect. PFW, family-wise error. Age (years) in parentheses. Effects in %R(t), %D(t) are shown in *red* and *green*, respectively. Stimulus was a subliminal electric field [41]

the expected ultradian macroarchitecture (2–5 relative maxima), and associated fine structure not determinable in the standard hypnogram (Fig. 7.27, top panel). When the individual epoch values were color coded based on the clinically-assessed sleep stage, both %R and %D had their highest values during deep sleep (N3), lowest values during wake, with the other stages exhibiting intermediate values (Fig. 7.27). Thus %R and %D provided a continuous measure of sleep depth, which is a crucially important variable in determining sleep quality. The generality of the phenomenon was established by evaluating values of %R and %D as a function of sleep stage in 20 patients (Fig. 7.28).

The results (Fig. 7.28) suggested a possible basis for the use of ABR for diagnosing sleep disorders. Obstructive sleep apnea (OSA) is a disorder characterized

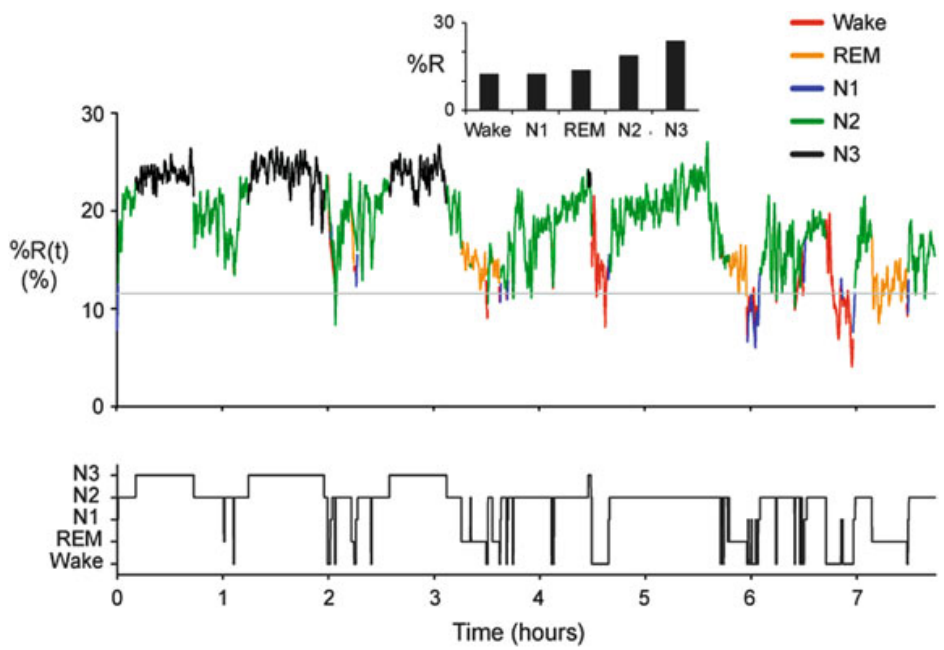


Fig. 7.27 Typical percent recurrence $\%R(t)$ in the EEG (C3) from an overnight sleep study of a patient with OSA (AHI = 9.4). Percent recurrence was calculated every second, averaged epoch-by-epoch, and color-coded by sleep stage (hypnogram, *lower panel*). For clarity in presentation, the curve was smoothed using a Savitzky–Golay filter. Gray line ($\%R = 11.6$) indicates average value of percent recurrence from clinically normal subjects during wake [45]. Insert shows sleep-stage-specific average values of percent recurrence. OSA, obstructive sleep apnea. AHI, apnea–hypopnea index (accepted clinical measure of OSA severity)

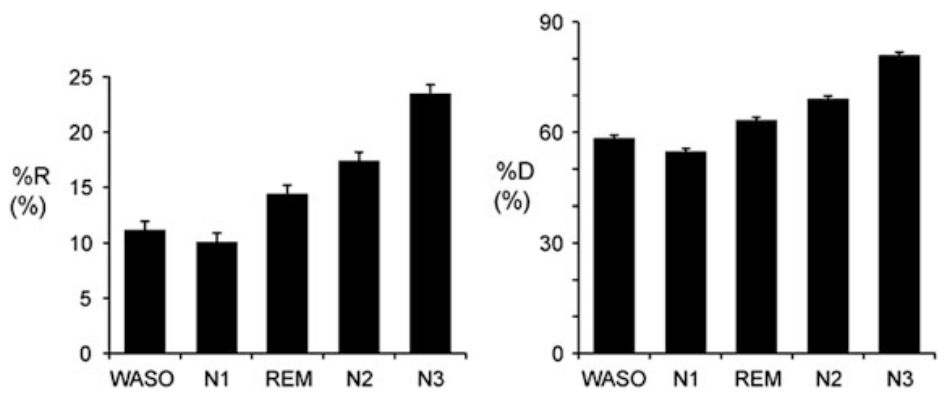


Fig. 7.28 Sleep-stage-specific percent recurrence ($\%R$) and percent determinism ($\%D$) in the EEGs (C3) from 20 patients with OSA (AHI 5–30). The recurrence values were computed second-by-second from each patient and averaged across sleep stage. Grand averages (\pm SE) [45]

by periodic collapse of the upper airway during sleep, resulting in intermediate hypoxemia, hypercarbia, and variable degrees of sleep disruption. Blood oxygen levels are affected almost immediately by apneic events, resulting in afferent signals to the brain that initiate appropriate compensatory responses. These autonomic

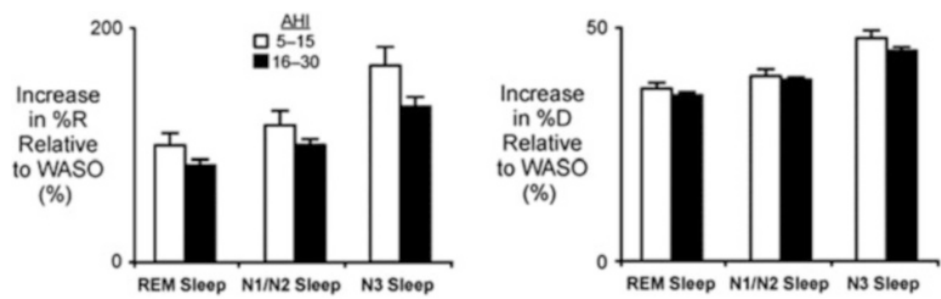


Fig. 7.29 Effect of OSA severity on sleep-stage-specific changes in %R and %D. Both recurrence variables were computed second-by-second from the EEG (C3), normalized by the patient’s value during WASO, and averaged over ten patients with mild OSA (AHI 5–15) and ten patients with moderate OSA (AHI 16–30). The N1 and N2 stages were combined. Sleep-stage-specific means \pm SE [45]

processes are necessarily accompanied by changes in the functional state of the brain, compared with what the state would otherwise have been.

We expected that the changes could be detected in patients with more severe OSA, and Fig. 7.29 demonstrates that this was indeed the case; the increases in %R and %D during sleep (associated with deeper and hence more restful sleep) were less in all sleep stages in patients with more severe OSA.

Individual sleep-stage-specific ABR markers for sleep depth (Fig. 7.28) cannot presently sustain a diagnosis of a sleep disorder because of inter-patient variations. But when the diagnostic/prognostic power associated with individual sleep-stage-based recurrence markers is combined in the context of a suitable statistical design, successful prediction becomes possible. For example, ABR was used to construct four time series from an overnight EEG, namely %R(t), %D(t), and two additional time series based on measures of their time-dependent variances. Combining the four time series with the five sleep stages yielded a total of 20 markers. When the markers were analyzed using linear discriminant analysis, the resulting biomarker function was 100 % accurate in diagnosing patients with either mild or moderate OSA [45].

7.5 Summary

The aperiodic rhythmic activity (ARA) displayed in the EEG has long been suspected to code for specific physiological processes. A plethora of methods for discerning the latent meaning of the EEG have been proposed, but ultimately they were unsuccessful in yielding a general approach to the problem of interpreting the EEG. A novel solution became possible following the development of recurrence analysis. The techniques developed to study the dynamics of low-dimensional systems were applied in a model-free fashion to the study of time series outputs from high-dimensional systems, the premier example of which is the EEG generated by the brain. Using recurrence analysis, we developed ABR, an approach that

quantitates ARA without explicit resort to specific dynamical models, in distinction to traditional approaches such as Fourier decomposition which assumes that the ARA are composed of parts (individual frequencies), and in distinction to chaos theory which assumes that the ARA arises from low-dimensional dynamical activity. In the context of controlled statistically-based experiments that addressed specific hypotheses, ABR was consistently useful in helping to understand brain function. Examples of such studies were described in the preceding section.

The usefulness of ABR ultimately stems from its incorporation of an additional analytical step in the process of analyzing the EEG. The nature of this additional step merits reflection. Two fundamental aspects of the scientific method for determining true cause–effect relationships (and studies of brain metabolism are no exception) involve the concepts of *averaging* and *replication*. Identification of a cause–effect relationship requires multiple independent observations of a phenomenon under controlled circumstances, and the combining of the observations by means of averaging. The point in the analysis where the averaging should be performed is a critically important but underappreciated issue. In experiments involving linear systems, the dependent variable is stochastic and consequently its values can be averaged directly to form a mean that may properly be regarded as the true characteristic behavior of the system. The *replication* requirement is satisfied by showing that the mean value can be reproduced (more or less) at will. But when the system under consideration is nonlinear, the dependent variable is *not* a stochastic variable. In this case repeated measurements under the same conditions do not yield necessarily similar values because the law of superposition does not generally apply to nonlinear systems. As a consequence of this common behavior of nonlinear systems, successively determined means *in principle* are not replicates. Consequently direct resort to averaging for purposes of evaluating the relation between the independent and dependent variables can have the opposite effect of that intended—obscuring knowledge of the system’s determinism rather than revealing it (Fig. 7.8). ABR obviates this problem by interposing a second dependent variable (derived algorithmically from the directly-measured variable) that captures and quantifies the experimentally-induced determinism (the difference between the E and C groups in a controlled study) *prior* to averaging. Consequently each independent trial can contribute to the requisite sensitivity needed to rationalize the existence of an effect, because each trial adds to the ability to distinguish the determinism of interest from background signals (signals that do not depend on the presence of the independent variables) (Fig. 7.12). In many cases of experimentally-induced nonlinear determinism, the scientific requirement of replication is fulfilled by a showing that the *phenomenon* (as opposed to the magnitude and/or sign of a mean) is replicable. Examples of this behavior have been discussed (Figs. 7.18 and 7.21).

The studies discussed in the previous section demonstrated ABR can be exploited to yield a broad range of useful results in the areas of basic and translational neuroscience.

References

1. S.M. Blinkov, I.I. Glezer, *Human Brain in Figures and Tables*, 1st edn. (Basic Books, New York, 1968)
2. R.W. Williams, K. Herrup, The control of neuron number. *Annu. Rev. Neurosci.* **11**, 423–453 (1988)
3. E. Bullmore, O. Sporns, Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* **10**(3), 186–198 (2009)
4. O.V. Kolomytkin, A.A. Marino, Neurobiophysics, in *Handbook of Molecular Biophysics: Methods and Applications*, ed. by H.G. Bohr (Wiley-VCH, Hoboken, 2009), pp. 523–556
5. E.L. Reilly, EEG recording and operation of the apparatus, in *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, ed. by E. Niedermeyer, F. Lopes da Silva (Lippincott Williams & Wilkins, Philadelphia, 2004)
6. S. Luck, E.S. Kappenman, *The Oxford Handbook of Event-Related Potential Components* (Oxford University Press, New York, 2013)
7. T. Yamada, E. Meng, *Practical Guide for Clinical Neurophysiologic Testing* (Lippincott Williams & Wilkins, New York, 2011)
8. H. Pratt, A. Starr, H.J. Michalewski, N. Bleich, N. Mittelman, The auditory P50 component to onset and offset of sound. *Clin. Neurophysiol.* **119**(2), 376–387 (2008)
9. F. Bandini, M. Pierantozzi, I. Bodis-Wollner, Parkinson's disease changes the balance of onset and offset visual responses: an evoked potential study. *Clin. Neurophysiol.* **112**, 976–983 (2001)
10. B.A. Clementz, A. Keil, J. Kissler, Aberrant brain dynamics in schizophrenia: delayed buildup and prolonged decay of the visual steady-state response. *Cogn. Brain Res.* **18**, 121–129 (2004)
11. E. Tanaka, K. Inui, T. Kida, R. Kakigi, Common cortical responses evoked by appearance, disappearance and change of the human face. *BMC Neurosci.* **10**, 38–46 (2009)
12. J. Gleick, *Chaos* (Penguin, New York, 2008)
13. W.S. Pritchard, D.W. Duke, Measuring chaos in the brain: a tutorial review of nonlinear dynamical EEG analysis. *Int. J. Neurosci.* **67**, 31–80 (1992)
14. C. Büchel, K.J. Friston, Dynamic changes in effective connectivity characterized by variable parameter regression and Kalman filtering. *Hum. Brain Mapp.* **6**(5–6), 403–408 (1998)
15. A.M. Kelly, L.Q. Uddin, B.B. Biswal, F.X. Castellanos, M.P. Milham, Competition between functional brain networks mediates behavioral variability. *Neuroimage* **39**(1), 527–537 (2008)
16. C.M. Lewis, A. Baldassarre, G. Committeri, G.L. Romani, M. Corbetta, Learning sculpts the spontaneous activity of the resting human brain. *Proc. Natl. Acad. Sci. U. S. A.* **106**(41), 17558–17563 (2009)
17. Y. Liu, J.H. Gao, M. Liotti, Y. Pu, P.T. Fox, Temporal dissociation of parallel processing in the human subcortical outputs. *Nature* **400**(6742), 364–367 (1999)
18. J.-P. Eckmann, S.O. Kamphorst, D. Ruelle, Recurrence plots of dynamical systems. *Europhys. Lett.* **4**, 973–979 (1987)
19. C.L. Webber Jr., J.P. Zbilut, Dynamical assessment of physiological systems and states using recurrence plot strategies. *J. Appl. Physiol.* **76**, 965–973 (1994)
20. M.C. Casdagli, Recurrence plots revisited. *Physica D* **108**(1–2), 12–44 (1997)
21. J.P. Zbilut, C.L. Webber Jr., Recurrence quantification analysis, in *Wiley Encyclopedia of Biomedical Engineering*, ed. by M. Akay (Wiley, Hoboken, 2006), pp. 2979–2986
22. H.D. Abarbanel, Nonlinear systems, in *Encyclopedia of Applied Physics*, ed. by G.L. Trigg (VCH Publishers, New York, 1994), pp. 417–439
23. R.O. Becker, A.A. Marino, *Electromagnetism & Life* (State University of New York Press, Albany, 1982)
24. G.B. Bell, A.A. Marino, A.L. Chesson, Alterations in brain electrical activity caused by magnetic fields: detecting the detection process. *Electroencephalogr. Clin. Neurophysiol.* **83**, 389–397 (1992)
25. A.A. Marino, E. Nilsen, C. Frilot, Consistent magnetic-field induced changes in brain activity detected by recurrence quantification analysis. *Brain Res.* **951**, 301–310 (2002)

26. C. Frilot II, S. Carrubba, A.A. Marino, Magnetosensory function in rats: localization using positron emission tomography. *Synapse* **63**, 421–428 (2009)
27. A.A. Marino, E. Nilsen, A.L. Chesson Jr., C. Frilot, Effect of low-frequency magnetic fields on brain electrical activity in human subjects. *Clin. Neurophysiol.* **115**, 1195–1201 (2004)
28. S. Carrubba, C. Frilot, A. Chesson, A. Marino, Detection of nonlinear event-related potentials. *J. Neurosci. Methods* **157**, 39–47 (2006)
29. S. Carrubba, C. Frilot, A.L. Chesson Jr., A.A. Marino, Evidence of a nonlinear human magnetic sense. *Neuroscience* **144**, 356–367 (2007)
30. S. Carrubba, C. Frilot, A.L. Chesson Jr., A.A. Marino, Nonlinear EEG activation by low-strength low-frequency magnetic fields. *Neurosci. Lett.* **417**, 212–216 (2007)
31. A.A. Marino, C. Frilot, Comment on “Proposed test for detection of nonlinear responses in biological preparations exposed to RF energy”. *Bioelectromagnetics* **24**, 70–72 (2003)
32. S. Carrubba, C. Frilot, A.L. Chesson Jr., C.L. Webber Jr., J.P. Zbilut, A.A. Marino, Magnetosensory evoked potentials: consistent nonlinear phenomena. *Neurosci. Res.* **60**, 95–105 (2008)
33. O.V. Kolomytkin, S. Dunn, F.X. Hart, C. Frilot, D. Kolomytkin, A.A. Marino, Glycoproteins bound to ion channels mediate detection of electric fields: a proposed mechanism and supporting evidence. *Bioelectromagnetics* **28**, 379–385 (2007)
34. A.A. Marino, S. Carrubba, C. Frilot, A.L. Chesson Jr., Evidence that transduction of electromagnetic field is mediated by a force receptor. *Neurosci. Lett.* **452**, 119–123 (2009)
35. S. Carrubba, C. Frilot II, F.X. Hart, A.L. Chesson Jr., A.A. Marino, The electric field is a sufficient physical determinant of the human magnetic sense. *Int. J. Radiat. Biol.* **85**, 622–632 (2009)
36. S. Carrubba, C. Frilot II, A.L. Chesson Jr., A.A. Marino, Numerical analysis of recurrence plots to detect effect of environmental-strength magnetic fields on human brain electrical activity. *Med. Eng. Phys.* **32**(8), 898–907 (2010)
37. C. Frilot II, S. Carrubba, A.A. Marino, Sensory transduction of weak electromagnetic fields: role of glutamate neurotransmission by NMDA receptors. *Neuroscience* **258**, 184–191 (2014). <http://dx.doi.org/10.1016/j.neuroscience.2013.11.009>
38. S. Carrubba, C. Frilot II, A.L. Chesson Jr., A.A. Marino, Mobile-phone pulse triggers evoked potentials. *Neurosci. Lett.* **469**, 164–168 (2010)
39. A.A. Marino, E. Nilsen, C. Frilot, Nonlinear changes in brain electrical activity due to cell-phone radiation. *Bioelectromagnetics* **24**, 339–346 (2003)
40. S. Carrubba, C. Frilot II, A.L. Chesson Jr., A.A. Marino, Method for detection of changes in the EEG induced by the presence of sensory stimuli. *J. Neurosci. Methods* **173**, 41–46 (2008)
41. S. Carrubba, A. Minagar, E. Gonzalez-Toledo, A.L. Chesson Jr., C. Frilot II, A.A. Marino, Multiple sclerosis impairs ability to detect abrupt appearance of a subliminal stimulus. *Neurol. Res.* **32**, 297–302 (2010)
42. S. Carrubba, A. Minagar, A.L. Chesson Jr., C. Frilot II, A.A. Marino, Increased determinism in brain electrical activity occurs in association with multiple sclerosis. *Neurol. Res.* **34**(3), 286–290 (2012)
43. M.H. Kryger, T. Roth, W.C. Dement, *Principles and Practice of Sleep Medicine* (Saunders, Philadelphia, 2010)
44. American Academy of Sleep Medicine (2007) *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*. American Academy of Sleep Medicine
45. P.Y. Kim, D.E. McCarty, L. Wang, C. Frilot II, A.L. Chesson Jr., A.A. Marino, Two-group classification of patients with obstructive sleep apnea based on analysis of brain recurrence. *Clin. Neurophysiol.* **125**, 1174–1181 (2014). doi:[10.1016/j.clinph.2013.11.002](https://doi.org/10.1016/j.clinph.2013.11.002)
46. S. Chokroverty, R.J. Thomas, M. Bhatt, *Atlas of Sleep Medicine*, 1st edn. (Elsevier, Philadelphia, 2005)
47. L. Wang, P.Y. Kim, D.E. McCarty, C. Frilot II, A.L. Chesson Jr., S. Carrubba, A.A. Marino, EEG recurrence markers and sleep quality. *J. Neurol. Sci.* **331**, 26–30 (2013)

Chapter 8

Recurrence Analysis of Otoacoustic Emissions

Giovanna Zimatore and Marta Cavagnaro



Abstract Otoacoustic emissions are sounds generated inside the inner ear. Recurrence Quantification Analysis (RQA) has proven to be particularly suited for studying such signals, being able to evidence their essential dynamical characteristics. In this chapter the fundamental features of the auditory system will be briefly reviewed, then the results obtained in the literature, and linked to the application of RQA to otoacoustic emissions, will be reported and discussed.

8.1 Introduction

In this chapter the potentialities of the RQA technique are described with reference to otoacoustic signals, which are signals recorded in the ear canal and which are used clinically for the diagnosis of hearing losses. RQA is extremely fit to study non-stationary signals, as otoacoustic signals, and represents a valid alternative to wavelet analysis. In fact, the embedding procedure allows to expand the otoacoustic mono-dimensional signal into a multidimensional space, thus letting the identification of subtle peculiarities of the sampled series that in turn are described by few

G. Zimatore (✉)

CNR-IDASC – Institute of Acoustics and Sensor “O.M. Corbino”, Rome, Italy

e-mail: giovanna.zimatore@uniroma1.it

M. Cavagnaro

DIET – Sapienza University of Rome, Rome, Italy

e-mail: cavagnaro@diet.uniroma1.it

global parameters. In this way a synthetic description of the characteristics of very complex signals is allowed.

RQA has proven to be able to assess the fine structure of otoacoustic signals without any a priori hypothesis (i.e. stationarity) and any data manipulation; moreover it has proven to be independent from differences in signal amplitude.

The chapter is introduced with a brief description of ear morphology and physiology. This description is not intended to be exhaustive but it allows understanding the origin of the different hearing losses and the existence of sounds generated by the inner structures of the ear, named otoacoustic emissions. The hearing losses as well as the ear-generated sounds are reviewed in the successive paragraphs. Then, the different instrumentation used in clinical settings to diagnose hearing losses is briefly depicted. Successively, three paragraphs are devoted to the possible applications of RQA to study the different features of otoacoustic emissions, reporting promising results of research studies. It is the Authors' hope that at the end of the chapter the reader will be able to understand the fundamental characteristics of otoacoustic emissions, and to appreciate the tremendous potentialities that the RQA technique has on evaluating such characteristics.

8.2 Ear Morphology and Physiology

The auditory system can be subdivided into a peripheral and a central auditory system. The peripheral system converts the sound energy, represented by condensations and rarefactions of air molecules, into neural codes that are interpreted by the central auditory system as specific sound tokens [1].

The peripheral auditory system is in turn divided into the external ear, made by the auricle and the external ear canal, the middle ear, which stands between the tympanic membrane and the oval window, and the inner ear corresponding to the cochlea and semicircular canals of the vestibular system (Fig. 8.1).

The external ear collects the sound waves and converts them into mechanical motion at the eardrum. This motion is transferred through the middle ear ossicles (malleus, incus, stapes) up to the oval window, where it is converted into a pressure wave and transferred to the inner ear organs (Fig. 8.2). Here, it is finally analyzed and converted into neural codes that are carried by the auditory nerve to the central auditory system (acoustic cortex).

The cochlea is a shell shaped organ, filled with incompressible fluids. In the human ear it is made by 2 and 3/4 turns for a total unrolled length of about 35 mm. The cochlea is divided into three canals: the scala vestibuli and scala tympani, which are filled with the perilymph fluid and linked through the helicotrema at the cochlea apex (Fig. 8.2), and the scala media, which is filled with endolymph and it is isolated from the other two scales by the basilar membrane and the Reissner's membrane (Fig. 8.3).

Fig. 8.1 Schematic view of the peripheral auditory system showing the ear canal, tympanic membrane, middle ear ossicles, and the inner ear made by the cochlea and semicircular canals of the vestibular system

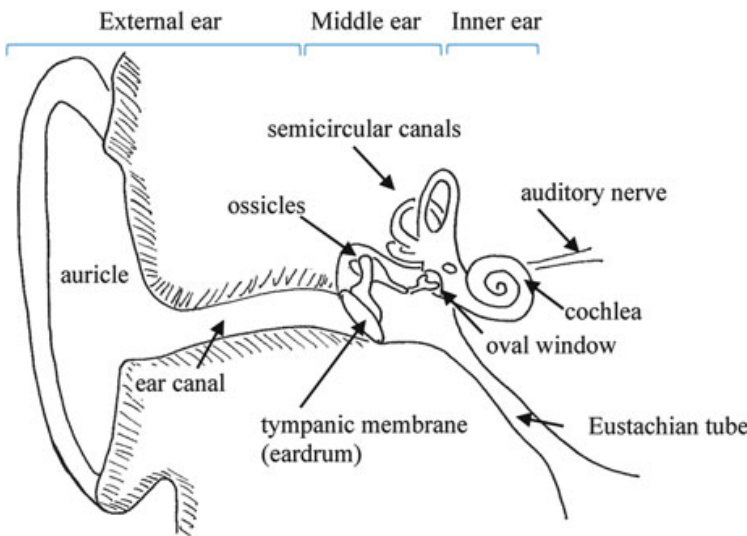


Fig. 8.2 Schematic view of a longitudinal section of uncoiled cochlea showing the oval window, and the scala vestibuli (SV), scala media (SM) and scala tympani (ST) within the cochlea

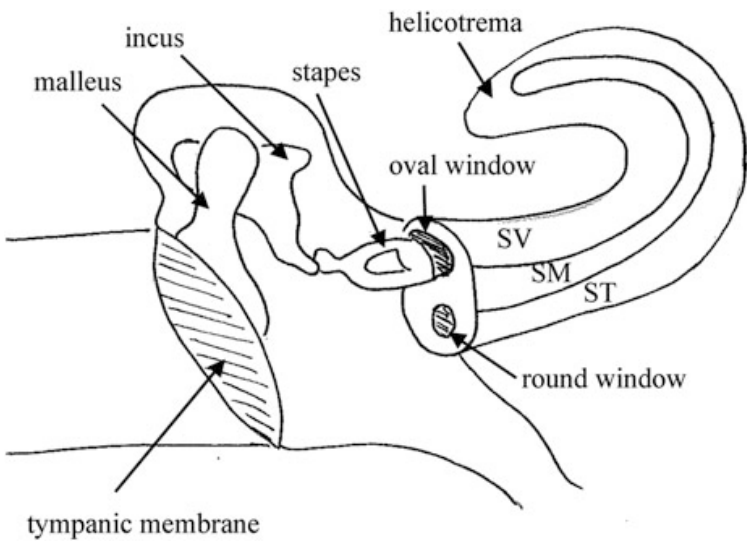


Fig. 8.3 Schematic view of a cross section of the cochlea showing the scala vestibuli (SV), scala media (SM) and scala tympani (ST). The organ of Corti is also evidenced

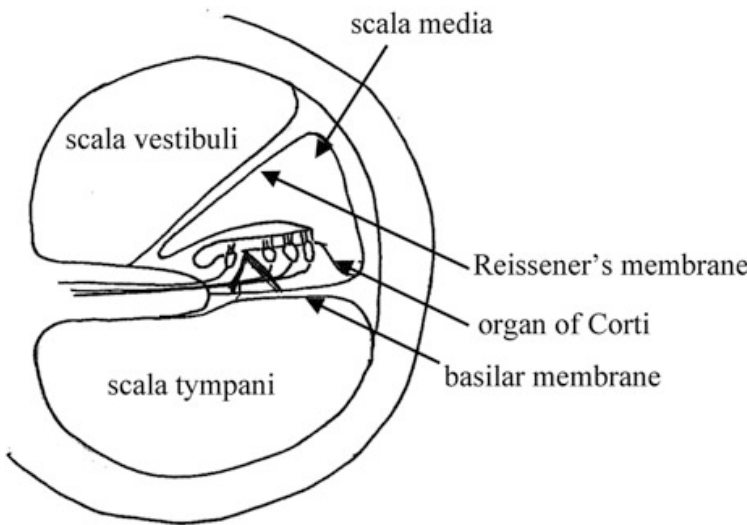
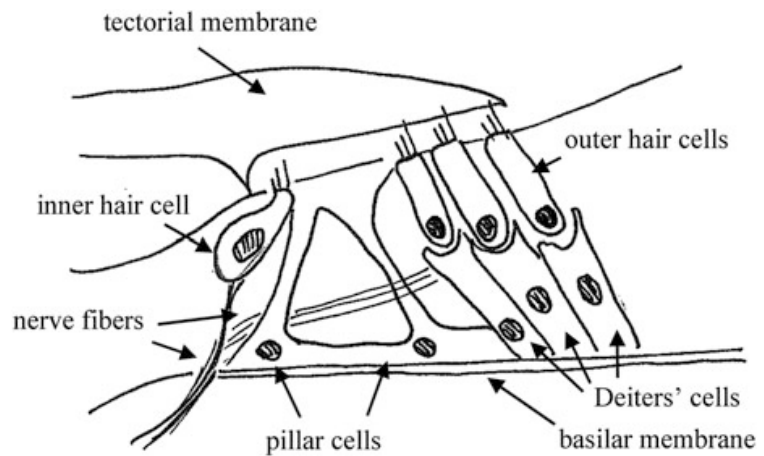


Fig. 8.4 Schematic view of a section of the organ of Corti. The pillar cells, Deiters' cells, outer and inner hair cells are shown



On the basilar membrane stands the organ of Corti, which is responsible for the transduction of the pressure waves into neural signals. The Corti's organ extends between the basilar and the tectorial membrane within the scala media (Fig. 8.4). It is made by a series of structures that repeat themselves all along the cochlea (in the human ear their number is about 3,500).

These structures are constituted by several types of cells each with a specific role for the sound transduction (Fig. 8.4): the pillar cells which contribute to the mechanical stability of the organ of Corti and shape a tunnel where nerve fibers run; the Deiters' cells which form the base of the Outer Hair Cells (OHC) and help amplifying or dumping the incoming pressure wave; the OHC, which are equipped with cilia terminating into the tectorial membrane and are organized in three lines, and finally the Inner Hair Cells (IHC), one every three OHC, which have cilia not terminating into the tectorial membrane.

Fluid and tissue displacements travel from the oval window all along the cochlea from the base to the apex. Pressure relief is provided for the incompressible fluids of the inner ear by the round window membrane (Fig. 8.2). Under the action of the fluid pressure, the basilar membrane and the OHC, which are in-built on it, oscillate. Due to these oscillations the OHC cilia bend, provoking the opening of ion channels and the depolarization of the cells that, as a consequence, shrink.

The OHC contraction is selectively amplified by the Deiters' cells. Moreover, the OHC organization on the basilar membrane is such that, when a pressure wave stimulates the OHCs, the stimulus is amplified only by the cells located in a particular position of the basilar membrane, corresponding to the frequency of the incoming sound. This organization is defined as tonotopic and helps de-codifying the different frequencies composing a sound: higher frequencies are detected close to the base of the cochlea, while lower frequencies travel longer along it. The OHC action causes the movement of the tectorial membrane, which drags along the movement of the IHC cilia. Since IHCs are contacted by nerve fibres that link the organ of Corti with the cochlear nuclei in the brainstem, the movement of IHC cilia finally stimulates the auditory nerve.

According to the previous description, the position of the IHCs that stimulate the nerve fibers carries on the information on the frequency of the incoming stimulus. On the contrary, the number of nerve fibers activated is proportional to the intensity of the acoustic stimuli. In this way the acoustic cortex can elaborate real time information about frequency and intensity of incoming sounds.

8.3 Otoacoustic Emissions (OAE)

OAEs are low-level sounds generated in the inner ear and measurable in the external auditory canal [2]. They can be recorded in a normal ear when the auditory periphery is solicited by an external acoustic stimulus, such as a transient sound (Transient evoked OAE, TEOAE) or two pure tones (Distortion Product OAE, DPOAEs), and even in the absence of stimulation (Spontaneous OAE, SOAE) [3].

The actual origin of OAE is still unclear. Nowadays, their generation has been explained as the combination of two mechanisms: a linear reflection of the incident traveling wave within the inner ear, due to irregular discontinuities of the basilar membrane [4, 5], and a nonlinear distortion process based on the active contractile mechanisms of the OHC in the Corti's organ [6–9]. Even if a general consensus on the presence of the two mechanisms is lacking, it is generally agreed that OHC active processes are fundamental in OAE generation.

OAE represent an objective, non-invasive test that can assess with great accuracy OHC activity, as demonstrated by experimental and clinical studies [10–12]. Their evaluation does not need the collaboration of the patient, and they show a great reproducibility and stability in time. Accordingly, OAE have been increasingly used in clinical applications, e.g. for the assessment of ear functionality in newborn babies. Moreover, they have been used to monitor the cochlear effects of exogenous factors, such as ototoxic drugs, solvents, and high-level sound exposure [13, 14]. Recently TEOAE were used to study tinnitus subjects with normal hearing to assess whether a minor cochlear or efferent dysfunction might play a role in tinnitus [15].

8.4 Hearing Losses

The lowest sound that an average person is able to hear (hearing threshold) has an amplitude of about 20 decibel. The decibel (dB, or dB HL) is a logarithmic scale normalized to the maximum value. A person with a hearing threshold higher than that of an average person is said to have hearing losses. At present, about 360 million people worldwide have disabling hearing losses [16].

Hearing losses can be classified as conductive, sensory-neural, and retrocochlear or central [17]. In the first case the damage responsible for the hearing loss is located in the external or in the middle ear, so that the sound path is interrupted; in the second case the damage is located in the inner ear, while in central hearing loss the damage is located along the pathway from the cochlea to the cortex.

Several external as well as internal agents can induce hearing loss, with different consequences both on the ear physiology and functionality. Conductive hearing losses can be due to trauma, that can damage the tympanic membrane or disrupt the ossicle chain, or by otosclerosis. Sensory-neural hearing losses are linked to the death of OHC. This can be the result of ageing, of exposure to drug or to high level sounds or noise.

In the following, the age-related hearing losses and those associated to noise exposure will be briefly reviewed.

8.4.1 Age-Related Hearing Loss (ARHL)

Age Related Hearing Loss (ARHL), also referred to as presbycusis, is the progressive deterioration of hearing ability that occurs with normal aging and it is one of the most chronic conditions in elderly people [18]. ARHL affects 23 % of the population between 65 and 75 years of age, and 40 % of the population older than 75 years in the USA. In Europe, it affects 2.3 % of the population between 40 and 50 years of age, and over than 30 % of the population above 70 years [19]. Presbycusis is common in industrialized societies and is less pronounced in other societies. This discrepancy has been attributed to many factors including genetics, diet, and socioeconomic and environmental factors.

The pioneering work in the investigation of age-related hearing loss (ARHL) was by Schuknecht et al. [20, 21], which defined four types of ARHL according to the origin of the lesion: sensory, strial, conductive and neural. Sensory loss is characterized by the death of the outer hair cells (OHCs) in the Corti's organ, usually beginning with the OHC located at the cochlea's base where the higher frequencies of the sounds are detected, and then progressing toward the apex where the lower frequencies are detected. Strial loss, also referred to as metabolic loss, is linked to the atrophy of the capillary and blood vessels, which help maintaining the metabolic processes of the cochlea. Strial loss is a process that involves the whole cochlea. Conductive loss is due to the loss of motility of the basilar membrane due to its thickening and stiffening, while neural loss corresponds to the dead of nerve cells. Several studies have identified a complex interaction between intrinsic (genetic) and extrinsic (environmental) factors at the basis of ARHL [22–25], so that nowadays there is a general consensus on considering this inevitable deterioration in hearing ability a multifactorial process [23, 24, 26, 27].

8.4.2 *Noise Induced Hearing Loss (NIHL)*

Noise Induced Hearing Loss (NIHL) is a sensory-neural hearing loss with great public health relevance. Every year, approximately 30 million people in the United States and 30 million people in Europe are exposed to occupational noise [19, 28, 29].

Prolonged exposure to intense noise can induce a transitory fatigue of the organ of Corti, which in turn leads to a transitory rise of the hearing threshold levels, i.e. a rise of the intensity of the minimum audible incoming sound, or can induce a permanent damage with particular reference to the higher frequencies.

Depending on the intensity of the noise, multiple anatomical changes can be found in the inner ear after the acoustic trauma, ranging from distorted stereocilia of inner and outer hair cells, to a complete destruction of OHCs followed by IHCs death and rupture of intracochlear membranes. Death of hair cells may also result in a partial degeneration of spiral ganglion neurons [30]. Many studies have shown that the generation of reactive oxygen species (ROS) and free radicals is involved in the cascade of cochlear events induced by the acoustic trauma [31]. A synaptic repair mechanism can occur within the first few days post-exposure, partially helping in the recovery of acute hearing loss (temporary threshold shifts, TTS) [32]. However, whereas most epithelial tissues present a high turnover rate and tend to regenerate after a traumatic lesion, the mammalian cochlea does not have regenerative properties leading to irreversible hearing loss. Different approaches have been used to prevent noise induced hearing loss such as antioxidant drugs [14, 33–36].

8.5 Clinical Practice

Tests performed to discern hearing losses include: tympanometry, electrocochleography, evoked potential, pure tone audiometry, and the ILO test. In particular, pure tone audiometry and the ILO test are those more widely used in the clinical practice; they will be briefly reviewed in the following.

8.5.1 *Pure Tone Audiometry*

The pure tone audiometry (PTA) is the more widespread method used in the assessment of hearing. It consists in the evaluation of the hearing threshold at different frequencies: at each considered frequency, pure tones are presented into the aural canal with decreasing intensities until the patient is no longer able to hear them. Although the human auditory range is from 20 to 20,000 Hz, PTA covers in a strict sense the speech spectrum: 250, 500, 1,000, 2,000, 3,000, 4,000, 6,000 and

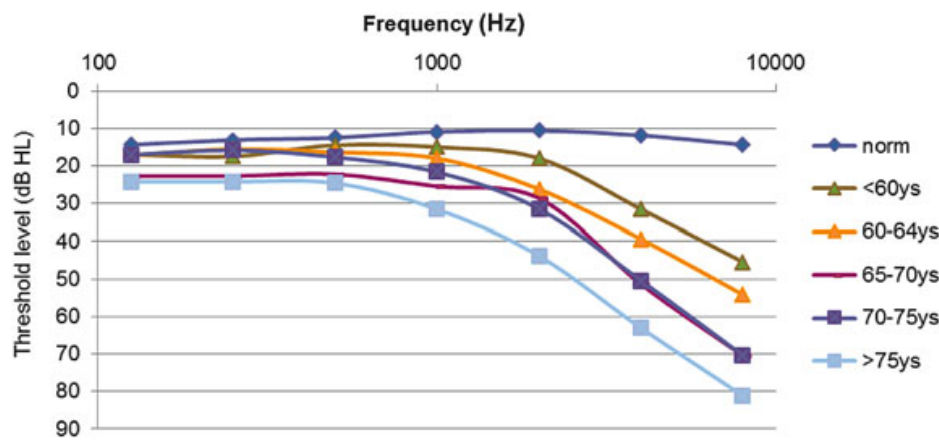


Fig. 8.5 Audiogram with the mean threshold values for five age-groups

8,000 Hz. The PTA output is a graphical representation of the obtained thresholds (audiogram) with the frequency in hertz (Hz) on the horizontal axis and the hearing level (HL) measured in dB on the vertical axis. On the PTA graphs the hearing thresholds are reported in an inverted way: the higher the threshold, the lower the position within the vertical scale. Starting from a normal threshold from 0 to 20 dB at all frequencies, a 6 dB increase in the hearing threshold represents a doubling of sound pressure level. Accordingly, hearing loss is classified according to how far the marks go down on the audiogram, and at what frequencies that occurs. PTA is a subjective test, which needs the active cooperation of the patient. As a consequence it can neither be performed on young babies nor on mentally retarded or elderly people with brain disorders.

Figure 8.5 shows the characteristic differences in threshold levels due to the increasing age. From the figure it can be noted a progressive loss of hearing sensitivity at higher frequencies. However, as noted before, ARHL is a multifactorial process so that it could be the result of a combination of different loss types and the estimate of the dominant contribution to the loss cannot be made on the basis of an audiogram alone.

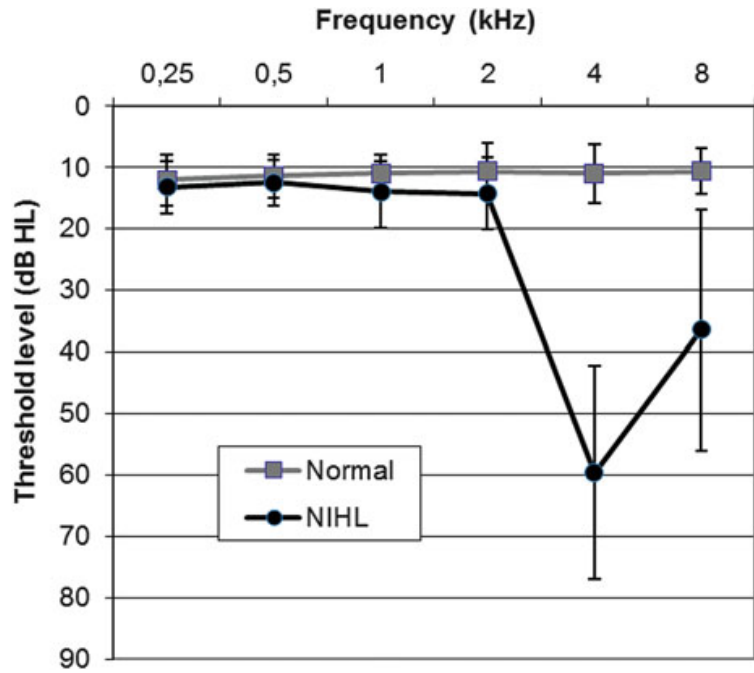
When the hearing loss is due to noise exposure, the audiogram shape shows a characteristic notch at the frequency of 4 kHz. As an example, Fig. 8.6 shows the audiogram obtained from normal ears (grey line), and NIHL ears (black line).

The peculiar shape of NIHL audiogram is due to the fact that the human ear is particularly sensible at this frequency, so that the OHC located at the tonotopic position of 4 kHz are the first to die when the ear collects sounds too high. Some authors identified in 6 kHz another frequency sensible to sound damage [37, 38].

8.5.2 ILO Test

The ILO test is performed through the Otodynamic Analyzer (ILO92, Otodynamics Ltd., Hatfield, UK) developed by Kemp after his discovering of OAE [2]. The

Fig. 8.6 Mean pure-tone hearing thresholds for normal ears (*square symbols and grey line*) and NIHL (*circle and black line*). Bars represent the standard deviation from the mean. Reproduced with permission [74]



ILO test is based on the measurements of OAEs through commercial devices (the ILO88 or ILO92; Otodynamics Ltd., Hatfield, UK) and on the evaluation of the reproducibility and signal to noise ratio of the measured waveforms. In the clinical practice protocols based on TEOAEs are widely used for identifying hearing losses [3, 39–41] and, as in DPOAEs, may provide an early indication of cochlear damage before evidence of NIHL appears in pure-tone audiometry.

The TEOAE signal is recorded at the ear canal in response to a stimulus. It includes a passive ringing of the speaker, ear canal, and middle ear, substantially linear with the incident stimulus and arriving with short latencies. This ringing is followed by a smaller, long-latency and long-duration oscillation constituting the non-linear TEOAE [42]. Figure 8.7 shows a TEOAE waveform recorded from a normal ear, after removing the first 2.5 ms to rule out the initial ringing. According to the tonotopic organization of the OHCs within the cochlea, the TEOAE signal presents a characteristic latency of the different frequencies composing the signal, with the higher frequencies coming first, followed by the lower frequencies (Fig. 8.7).

To better discriminate the TEAOE signal with respect to the initial ringing as well as to the background noise, a non-linear estimation method has been developed [43]. In this method the stimulus is made of a train of three rectangular pulses of equal amplitude and polarity (75–85 dB sound pressure level—SPL) lasting 80 μ s followed by a fourth pulse three times higher than the previous ones and with inverted polarity [43]. The clinical signal is obtained sending 520 stimuli trains, for a total of 2,080 clicks, recording the obtained TEOAE in two different buffers, and evaluating the average signal in each buffer. Then the correlation between the two averaged waveforms is evaluated (Pearson correlation coefficient \times 100). According to the clinical settings, when the whole waveform reproducibility (REPRO or

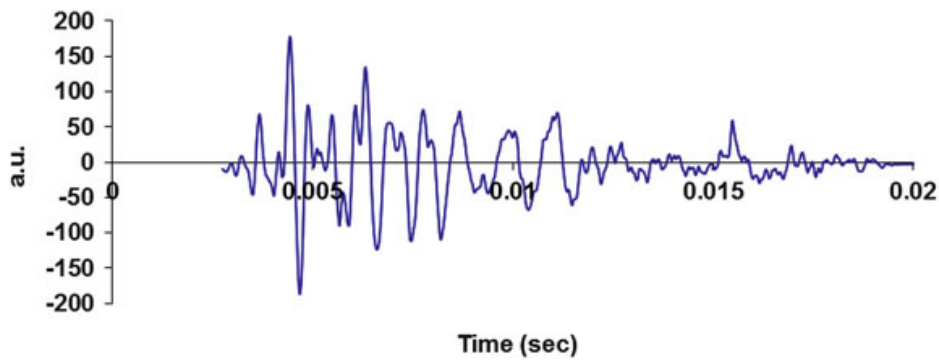


Fig. 8.7 TEOAE signal recorded from a normal ear

WWR) value is higher than 70 and the SNR (Signal to Noise Ratio) is higher than 3 dB, the signal is considered as physiological (ILO Test is Passed), whereas a WWR value lower than 70 indicates the presence of a possible hearing malfunction (ILO Test is Failed) [44].

OAEs tests are non-invasive and do not require co-operation from the patient, as does audiometry instead; still, 30 years after their discovery, clinical applications of OAEs are largely limited to qualitative pass/fail testing [2, 10, 45].

Researches have suggested that OAEs may provide early indication of cochlear damage before evidence for hearing loss appears in pure-tone audiometry [46, 47].

In particular, Lucertini et al. [48] compared TEOAE variables with pure tone audiograms, showing that the firsts could discriminate between non-exposed normal subjects and subjects with unilateral noise-induced hearing loss. Accordingly, OAE can be used to monitor subjects at risk of hearing impairment, as workers in noisy environment or patients receiving lifesaving ototoxic drugs. Moreover, since OAE are generated in the inner ear and propagate toward the middle to the external ear, they can be used to discriminate among different causes of hearing loss [42]. However, the validity of TEOAE measurements is hampered by their subject-dependent variations [45].

8.6 RQA to Study the Characteristics of Otoacoustic Emissions

In this section results obtained studying the TEOAE signal by way of RQA will be reviewed in order to emphasize the great potentiality both of the TEOAE signal and of the RQA analysis for improving the knowledge of the human ear physiology and building the basis for developing new methods of diagnosis. RQA is particularly suited for studying OAE being not limited by specific requirements on the studied signal, like stationary or length of data series, and being able to quantitatively evidence the amount of deterministic structure of a signal [49, 50].

Fig. 8.8 Recurrence plot of a typical TEOAE recorded in a normal ear (DET = 86.01)

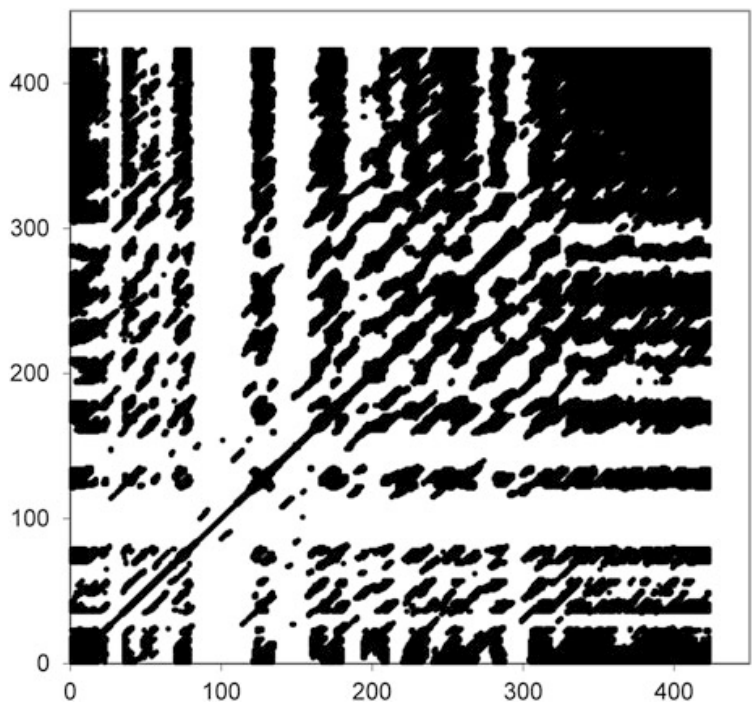
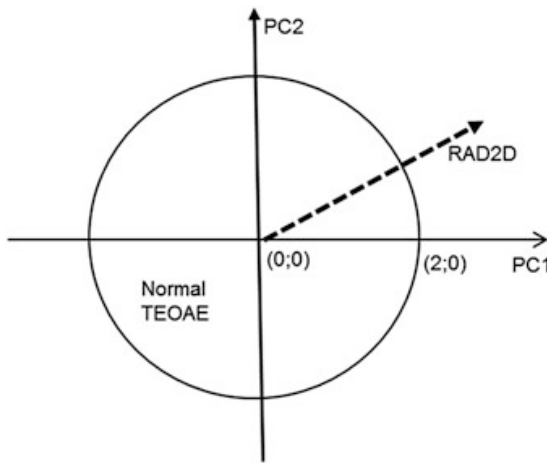


Figure 8.8 shows the Recurrence plot of a typical TEOAE recorded in a normal ear. Three RQA descriptors proved to be the most suitable in the study of TEOAE signals through RQA [51]: Percent Recurrence (REC), which represents the fraction of the recurrence plot occupied by recurrent points, measuring the amount of recurrent (both periodic and auto-similar) behavior of the signal; Percent Determinism (DET), that is the fraction of recurrent points aligned parallel to the main diagonal (deterministic lines) and indicates the degree of deterministic structures of the signal; Entropy (ENT), which is a Shannon entropy estimated over the distribution of the length of deterministic lines, and it is linked to the richness of deterministic structure. The RQA descriptors REC, DET, ENT were calculated with the following choice of working parameters: time course of the original signal described by 512 points; delay in the embedding procedure (lag) equal to 1; 10 columns included in the embedding matrix (embedding dimension); cut-off distance (radius) equal to 15 [51]. The ability of the RQA technique to represent the main features of TEOAE signals was evidenced in a study which compared the two recurrence plots (not shown) obtained from a TEOAE signal as well as from its shuffled version [51]. The comparison showed that while the TEOAE signal exhibited a rich deterministic structure, this was no longer present after shuffling. Moreover, all three RQA descriptors REC, DET, and ENT evaluated on the TEOAE signal showed a correlation with the ILO REPRO value. In particular, studying TEOAE signals as a function of the intensity of the stimulus, it was found that the DET descriptor changed, starting from a sub-threshold state, going through a phase of linear increase with the stimulus, and finally coming to a saturation phase in which it did not change with the intensity of the stimulus any more. These

Fig. 8.9 NA circle in the PC1/PC2 plane



changes of the DET value for different stimulus intensities correspond to a higher organization of the TEOAE signal with the increasing stimulus [52].

The three RQA descriptors were analyzed through Principal Component Analysis (PCA), a statistical technique which projects a multidimensional data set into a space of orthogonal axes, called principal components (PC) [53]. PC are selected, one after the other (PC1, PC2, etc.), on the basis of the maximal variance explained in the space of the original variables. The presence of correlations between the original variables allows reducing the data set dimension in the new space without noticeable loss of information. Zimatore et al. [51, 52] showed that, choosing a set of at least 70 TEOAE signals as a reference (training set) for the RQA–PCA analysis, the first two principal components (PC1, PC2) can explain more than 90 % of the observed variability. Moreover, since PC1 and PC2 have, by construction, zero mean and standard deviation equal to 1.96 % of real signals, if taken from a homogeneous population, should fall within a circle centered in the origin of the PC1/PC2 plane with radius equal to 2 (NA circle), thus allowing for a straightforward test of ‘normality’ [51]. Figure 8.9 shows the PC1/PC2 plane with the NA circle evidenced.

The RQA/PCA analysis was applied to the study of TEOAE signals to find out subject-dependent characteristics, which could be used to evidence TEOAE features, and subsequently correlated to changes in ear physiology. The first observation that could be performed was that representing on the PC plane different TEOAE signals measured from different subjects, signals recorded from different ears located in different positions of the PC plane, while signals recorded from the same ear of the same subject located close each other [51]. However, it was noted that if the stimulus used to record the TEOAE was of medium intensity, the clustering of signals coming from the same ear was not present. On the other hand, the clustering was present when high-level stimulus intensities were considered [52].

Since TEOAE can be recorded without the need of an active cooperation from the subject, they are extensively used in neonatal screening of the ear functionality. It is well known that the neonatal TEOAE responses are characterized by a large

intra-subject variability (up to 6 year of age) [54–56], by larger signal amplitudes with respect to the adult ones [57, 10] and by a wider and more uniform spectrum, sometimes shifted toward the higher frequencies [58]. Comparing the RQA/PCA descriptors obtained by TEOAE recorded in newborn babies and in adults, DET values in the neonatal group were lower than those from the adult subjects, indicating a minor presence of determinism in the neonatal signals [59]. Moreover, both the inter-subject variability and the intra-subject variability were larger in newborns than in adults, showing a weakness of the REPRO value of the ILO test in the evaluation of newborn acoustic performances. Finally, the ratio of inter-subject to intra-subject variability had similar values in the two groups, pointing to the presence of the same amount of individual features, i.e. the individual characteristics of the TEOAE signal are present since the very beginning of life [59]. However, it is interesting to note that no correlations were obtained between signals from the two ears of the same subject. This implies that origin of TEOAE individual features should be searched on a micro-scale level, e.g. as a different distribution of OHCs, which could be different in the two ears even for newborns.

8.7 RQA and Models of the Ear to Study the Origin of Otoacoustic Emissions

Many researchers are interested in modeling the dynamic behavior of the cochlea or in physically interpreting the results of cochlear measurements. The objectives are to produce tools that can be used to predict the response of the cochlea to various inputs thus helping in developing devices to overcome hearing losses.

Several models of the ear have been developed with the aim of studying the ear functionality, with reference to speech, hearing, and OAE. These models can be broadly divided into two categories: functional and physical models [60, 61]. Functional models, such as the gammatone [59, 62] or the electronic cochlea [63], aim at reproducing a specific behavior of the ear, as e.g. the general shape of TEOAE, without looking for a correspondence between model elements and anatomical structures. On the contrary, physical models are based on a strict correspondence between the two. Within the latter models, those based on the electroacoustic analogy substitute acoustic quantities, as sound pressure and volume velocity, with electric quantities like voltage and current, respectively [64].

In the gammatone model it is supposed that the inner ear behaves as a bank of narrowband filters, each centered on a specific frequency (f_c). In the time domain, the impulse response of a filter is given by:

$$\gamma(t) = at^3 e^{-\beta\omega_c t} \cos(\omega_c t) \quad (8.1)$$

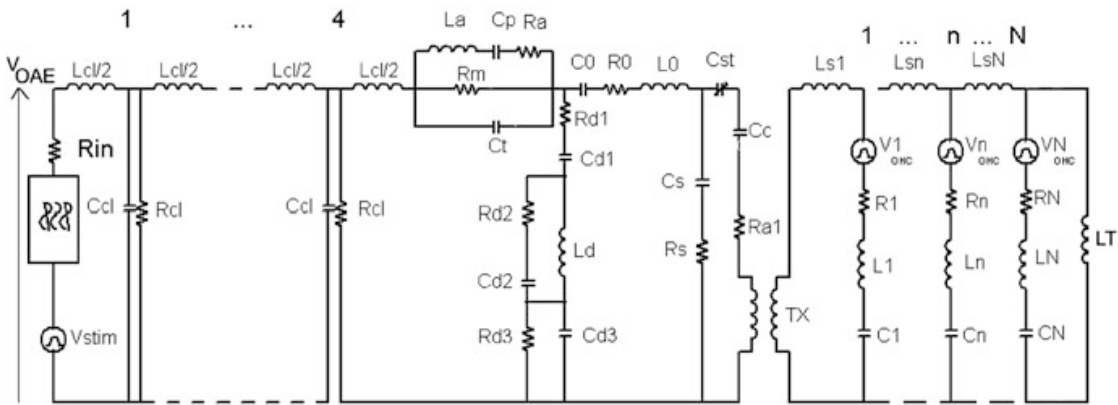


Fig. 8.10 Ear model based on the electro-acoustic analogy. Reproduced with permission [70]

where t is the time, $\omega_c = 2\pi f_c$ is the angular frequency, $a = (\omega_c)^{3.5}$ is an amplitude factor used to obtain a gammatone power independent from the frequency, and β represents the signal dumping [62].

In this model, it is supposed that the TEOAE signal is the sum of the impulse responses of each filter. In effect, if a set of five gammatones is used, centered at the frequencies of 1.0, 1.5, 2.2, 3.3, and 5.0 kHz and with a dumping factor β of 0.11 [62], a TEOAE waveform with a shape close to the real one is obtained [59]. However, if the thus obtained TEOAE waveform is studied with a single value decomposition (SVD) analysis, which aims at evaluating the number of independent modes (eigenvalues) excited in a system from the percentage of variance explained by each eigenvalue, it is found that the gammatone model generates a TEOAE signal more complex than the real one. In fact, more modes were needed to explain the same percentage of variability than the real signal, probably due to the lack of coupling between the different gammatone filters [59].

More interesting results have been obtained by using the model of the ear developed by Giguère and Woodland [65, 66], and the RQA/PCA technique [67]. The Giguère and Woodland [65, 66] model is based on the electroacoustic analogy, which, as already cited, represents the sound pressure with the electric voltage and the volume velocity with electric current. In this model, shown in Fig. 8.10, the outer ear is represented with a cascade of four T-sections, corresponding to the segmented form of an uniform transmission line [68], and the middle ear is modeled as a complex electrical network based on its functional anatomy [69]. An ideal transformer connects the middle ear to the cochlea, to represent the acoustic transformer ratio between the eardrum and the oval window [65]. The cochlea is modeled as an active non-uniform and non-linear transmission line, divided into several sections (named partitions) from its base to the apex.

Each partition consists of a series inductor, a shunt resonant circuit (composed of a resistor, an inductor, and a capacitor), and a non-linear voltage source. In the electro-acoustic analogy, the series inductors represent the acoustic mass of the cochlear fluids; the resistors, inductors and capacitors forming the shunt resonant circuits represent the acoustic resistance, mass and stiffness of the basilar

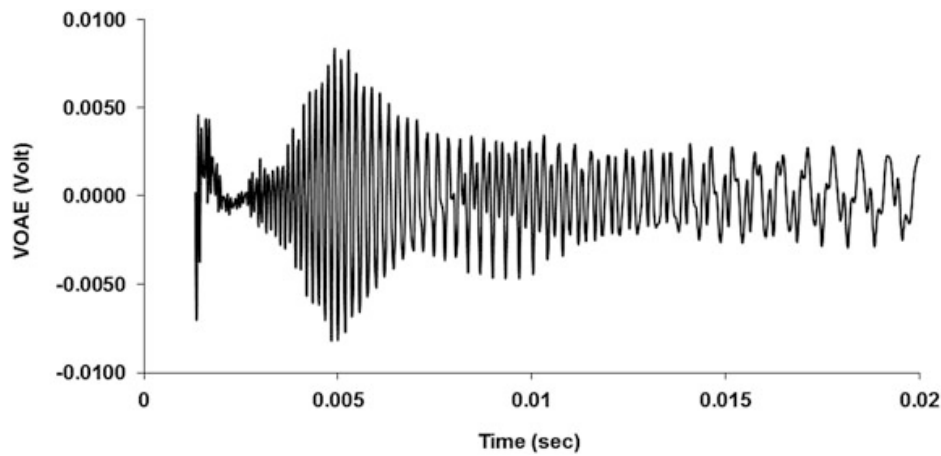


Fig. 8.11 TEOAE simulated signal

membrane, respectively, and the non-linear voltage sources represent the OHC active processes. Finally, an inductor models the helicotrema (L_T). Spontaneous OAE can be simulated inserting random noise sources into the cochlea partitions. To simulate TEOAE, the ear model is fed with a voltage source modeling the clinical set-up. The input voltage is a rectangular pulse $80\text{ }\mu\text{s}$ wide, whose amplitude corresponds to a stimulus of 80 dB SPL. The loudspeaker influence on the input signal is considered by inserting a band-pass filter ($f_{\text{LOW}} = 6\text{ kHz}$, $f_{\text{HIGH}} = 20\text{ kHz}$, ripple 3 dB, stop-band 5 dB) in series with the input voltage source (Fig. 8.10).

The Giguère and Woodland’s model was adapted to be solved through the commercial software PSpice® [67] by using the parameter values proposed in [65]. In particular, the cochlea has been initially modeled by using 128 sections according to what done by Giguère and Woodland [65].

The TEOAE simulated signal is reported in Fig. 8.11. From the figure it can be noted that the simulated signal is close to the signal recorded from a normoacoustic subject [67]. In particular, the simulated signal, after removing the firsts 2.5 ms, to eliminate the initial ringing, shows an oscillating behavior with time lasting up to 20 ms after the initial external excitation, and the typical latency-frequency TEOAE behavior, that is to say the higher frequencies have a shorter latency, and the lower frequencies have a longer one.

Figure 8.12 shows the latency of the simulated TEOAE signal compared with the latency of the real signal. The data are obtained evaluating, through a wavelet analysis, the time needed for each frequency that compose the TEOAE signal to cover the round trip from the oval window to the corresponding resonant partition and back. From the figure a good agreement can be obtained between the different curves. However, if the slope of the latency–frequency curve is evaluated, a value of -0.8 is obtained for the simulated signal, higher than that usually obtained from normoacoustic subjects (about -0.48), but in agreement with the outputs of models which use the transmission line formalism to represent the cochlea [71]. In fact, in [72] it is noted that an inverse proportionality relationship between latency and frequency (i.e. a slope of -1) is inherent to scale invariant cochlear models, which

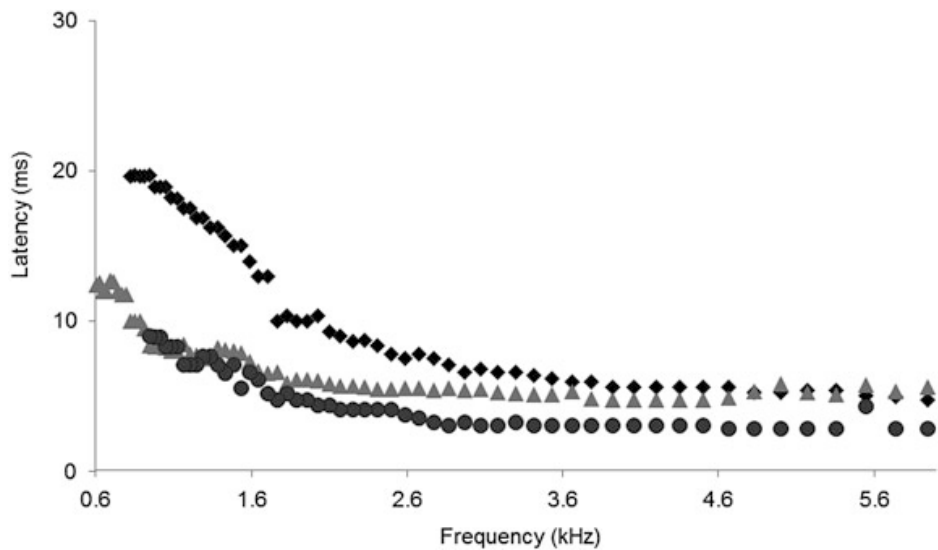


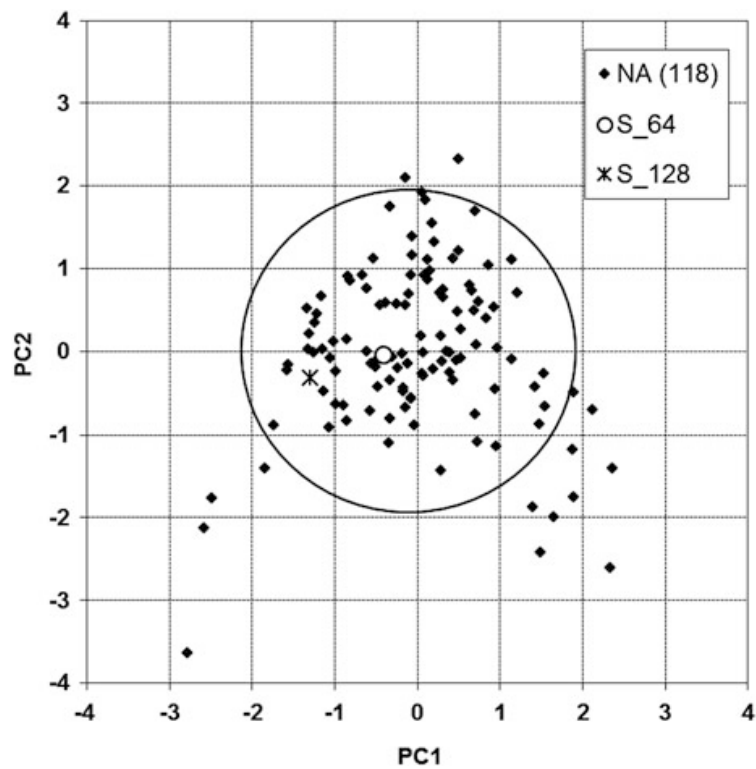
Fig. 8.12 Latency of TEOAE signals as a function of frequency. *Grey triangles*: measured TEOAE, *black diamonds*: TEOAE simulated using 128 partitions to model the cochlea; *black circles*: TEOAE simulated using 64 partitions to model the cochlea

are, in the transmission line formalism, the models with a constant quality factor (Q) in all the cochlea sections, as the electronic model [65].

To ameliorate the latency–frequency behavior of the simulated signal without changing the ear model, the characteristics of the transmission line which models the cochlea can be exploited. In fact, each section modeling the cochlea induces a time delay on the traveling signal, as a function of both position and frequency. Consequently, lowering the number of cochlear sections could reduce the time delay in the simulated signal; moreover, since the higher frequencies travel along the transmission line for a lower number of sections with respect to the lower frequencies (each frequency travels up to its tonotopic site, where the wave is reflected back on the line), this time delay reduction should be more pronounced for the lower frequency components of the signal. This, in turn, should lower the latency–frequency slope of the simulated signal. This strategy, of decreasing the number of cochlear sections used in the model, has to be intended as a best fitting procedure, justified by the typical redundancy of biology systems, void of any one-to-one mapping of the biological reality.

Following these considerations, a circuit with 64 partitions representing the cochlea has been developed, and the obtained signal has been compared with a real signal and with the one obtained from the 128-partitions model. While the time waveform of the signal is as close to the real one as the 128-partition model signal, the slope of the latency–frequency curve (see Fig. 8.12) for the 64-partitions signal was evaluated as equal to -0.60 , i.e. closer to the real one. Other characteristics have been evaluated as e.g. the basilar membrane volume velocity, reflections from the different sections modeling the cochlea, all showing a better agreement of the 64-partitions model to the real signal than the 128-partitions model.

Fig. 8.13 TEOAE signals in the PC1/PC2 plane after RQA/PCA analysis. *NA*: normoacoustic ear, *S_64*: simulated TEOAE with the 64-partitions model, *S_128*: simulated TEOAE with the 128-partitions model



Coming to the RQA/PCA technique, it was found that the simulated signal falls well within the NA circle in the PC1–PC2 plane. Figure 8.13 shows the NA circle obtained from 104 TEOAE signals recorded from normoacoustic ears. In the figure the diamonds represent the 104 signals, confirming that more than 96 % of real signal, if taken from a homogeneous population, fall within the NA circle. Moreover, in the figure the signal obtained from the 128-partitions model (square), and that obtained from the 64-partition model (circle) are also reported. From the figure it can be obtained once again that the 64-partitions model gives a simulated signal close to a real one, falling well within the NA circle.

The electric circuit modeling the ear can be used to study the influence of individual variability on the TEOAE signal. To this end, some middle ear elements were properly changed according to the experimental findings of Avan et al. [11], as reported in Table 8.1. In the table, C_{st} represents a stapes capacitor (see Fig. 8.10). When C_{st} has a large value, its impedance is small, corresponding to small tension in the stapedius muscle (C_{st} equal to infinity corresponds to no stiffness in the resting condition); conversely, when C_{st} is small, its impedance is large, corresponding to muscle high tension. C_{st} was equal to infinity in the model proposed by Giguère and Woodland [65], so that this condition is considered as the reference one. C_0 represents changes in the tympanic membrane stiffness, to account for changes in the middle ear pressure, while L_0 represents changes in the tympanic membrane mass, to simulate an additive mass. All the parameters' values reported in the table correspond to physiologic conditions [11].

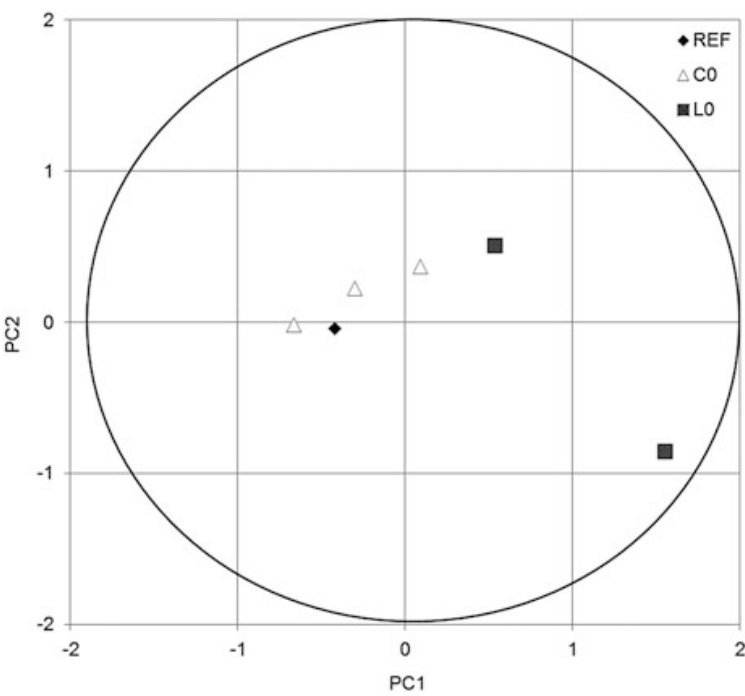
Figure 8.14 reports the simulated TEOAE signals in the PC1/PC2 plane. As can be noted from the figure, the position of the points in the PC1/PC2 plane

Table 8.1 Values of the middle ear parameters used to simulate individual variability [11]

Set	C_{st} (μF)	C_0 (μF)	L_0 (mH)
Ref	∞	1.40	40
1	0.25	1.40	40
2	0.25	0.700	40
3	0.25	0.233	40
4	0.25	1.40	80
5	0.25	1.40	160

The first row reports the starting values (reference set)

Fig. 8.14 Simulated TEOAE signals in the PC1/PC2 plane after RQA/PCA analysis. The different symbols represent different values of some parameters of the middle ear (see text)



changes when the value of a parameter of the middle ear section is changed, but the signal remains within the NA circle. As a consequence, it could be supposed that differences in the middle ear anatomy contribute to individual variability of TEOAE signals.

8.8 RQA in the Diagnosis of Hearing Losses

The complexity and fragility of the inner ear makes it extremely susceptible to damage. Research has shown that the most common cause of deafness is the loss of hair cells and not the loss of auditory neurons. In fact, OHCs represent the most sensitive and consequently the most vulnerable part of the ear.

Being the TEOAEs generated by OHC and transmitted through the middle ear and the ear canal, they reflect functions of outer hair cells (OHCs), and in some extent of the middle ear and ear canal also. However, if the damage is of limited

entity, it may not be detected by looking at the general characteristics of TEOAE. On the contrary, by studying the properties of TEOAEs through RQA, it is possible to discriminate different hearing losses and even to evidence inner ear damages before they can be evidenced by standard audiology tests. In the following lines the application of the RQA/PCA technique for the detection and discrimination of different hearing losses will be shown.

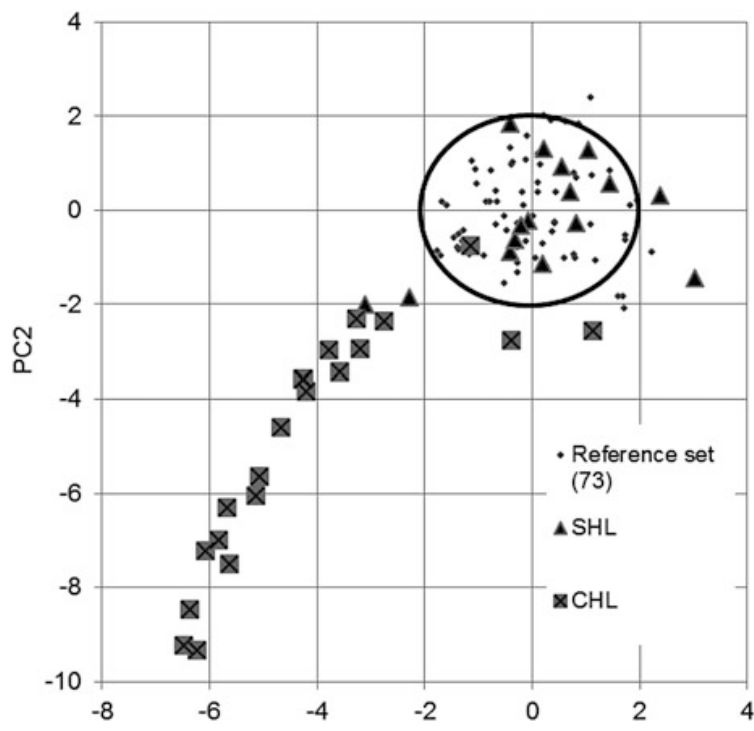
8.8.1 RQA Applied to the Study of CHL and SHL

In clinical applications the main problem of OAE consists in the large observed inter-subject variability, resulting in a different relationship between reproducibility and audiometric thresholds for different ears. In fact, it was found that the same audiometric threshold was associated with different REPRO values [73]. Moreover, subject-dependent variations in middle-ear transmittance and reflectance influence OAEs, and should be controlled [45]. The validity of OAE measurements can be increased by reducing their variance. This can be easily performed through the RQA technique. In Zimatore et al. [52], RQA was applied to TEOAE signals recorded from CHL and SHL ears at four intensities of the incoming stimulus (50, 65, 70, and 80 dB). In that work it was shown that the DET value was significantly different between the two groups of pathological signals, as well as between each of the two pathologies and the control group. This result points to the suitability of RQA in discriminating among the different pathologies. Figure 8.15 shows the TEOAE signals recorded from normal, CHL and SHL ears projected in the PC1/PC2 plane after the RQA/PCA analysis. From the figure it can be noted that while CHL signals move towards the negative axes, going away from the plane center and from the NA circle, SHL signals tend to superimpose to the normal ones, indicating still the presence of TEOAE essential features. Nevertheless, as said, the DET variable was able to discriminate among the three different groups [52].

8.8.2 RQA Applied to the Study of ARHL and NIHL

The main difference that can be foreseen between ARHL and NIHL is the location within the cochlea of the damaged OHCs: while in NIHL these are mainly concentrated on the cochlea segment corresponding to the 4 kHz frequency, in ARHL they are more uniformly distributed all along the cochlea, with a preferential location close to the higher frequencies. The ARHL and NIHL audiograms are very different (see Figs. 8.5 and 8.6) but the TEOAE signals can be indistinguishable in terms of REPRO and SNR. Sometimes, particularly for ARHL, the TEOAE signal analyzed through REPRO and SNR could be indistinguishable even from a normal TEOAE.

Fig. 8.15 TEOAE signals in the PC1/PC2 plane. *SHL*: signals from ears with sensorineural hearing losses; *CHL*: signals from ears with conductive hearing losses. Reproduced with permission [70]



RQA and RQA/PCA techniques were applied for studying noise-related hearing losses as well as age-related hearing losses, to look for possible early indication of the two types of damage. Figure 8.16 shows the recurrence plot of a TEOAE recorded from an NIHL (Fig. 8.16a) and from an ARHL subject (Fig. 8.16b). It has to be noted here that the chosen ARHL signal does not represent a severe condition ($REPRO = 78$). By comparing Fig. 8.16a with Fig. 8.8, it is clear that recurrence plot distinguishes between normal hearing and hearing losses, especially in terms of a reduction in the deterministic structure. On the other hand, Fig. 8.16b shows that the ARHL condition at the beginning stage could be indistinguishable even in the RQA plot from a normal condition. However, the reduction in determinism is confirmed from the mean values of the DET reckoned in different groups. In the three recurrence plots (Fig. 8.8 and in both panels of Fig. 8.16) the vertical and horizontal white bands reveal that transitions may have occurred; however the absence of fading to the upper left and lower right corners reveals that no trend or shift occurs.

Figure 8.17 reports the TEOAE signals recorded from NIHL ears as well as from normal ears in the PC1/PC2 plane. From the figure it can be noted that the NIHL signals fall in part within the normality circle and in part outside it, evidencing different levels of damage.

To quantify the amount of damage, a parameter, named RAD2D (two-dimensional radius), has been introduced. RAD2D measures the Euclidean distance between the origin of the PC1/PC2 plane and the position of the point representing a TEOAE signal in that plane [74, 75].

To evaluate the RAD2D values, the electronic model discussed in Sect. 8.7 was used, simulating increasing damages of the OHCs in the cochlea. A linear increase

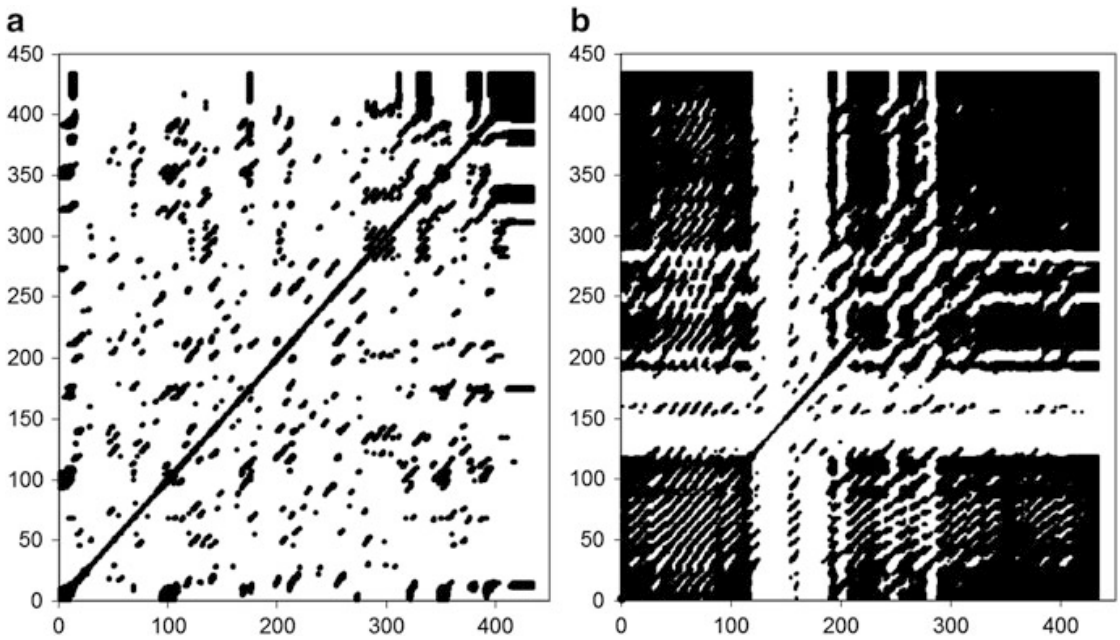


Fig. 8.16 Recurrence plot of a representative (a) NIHL (DET = 49.81) (b) ARHL TEOAE waveform (DET = 90.95)

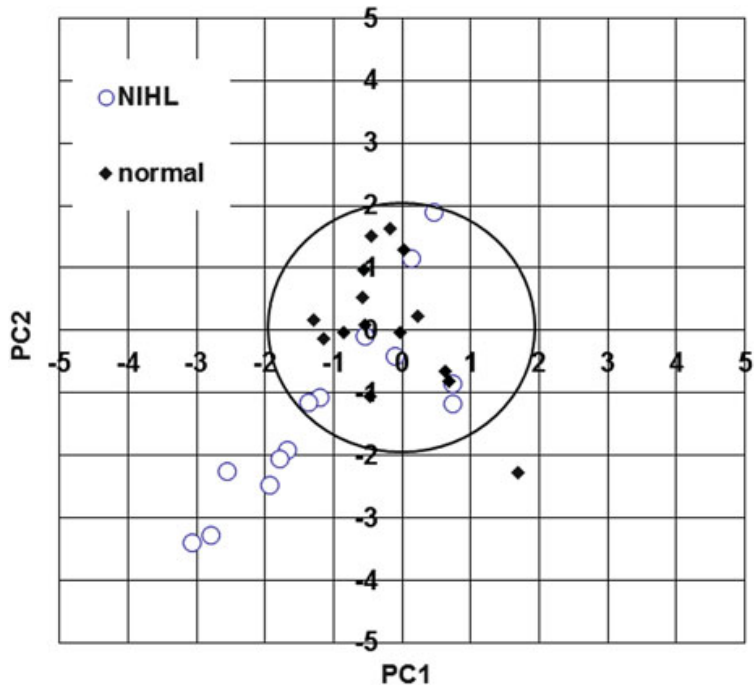


Fig. 8.17 TEOAE signals in the PC1/PC2 plane. *real_NIHL* represent ears with noise induced hearing losses; *real_Normal* represent normoacoustic ears

of the RAD2D with the amount of simulated damage was found [74]. Moreover, it is interesting to note that the RAD2D value obtained by the simulation of a normal ear (equal to 0.20) was close to the value of 0.21 obtained as the mean value from 15 TEOAE signals recorded in normal ears [74].

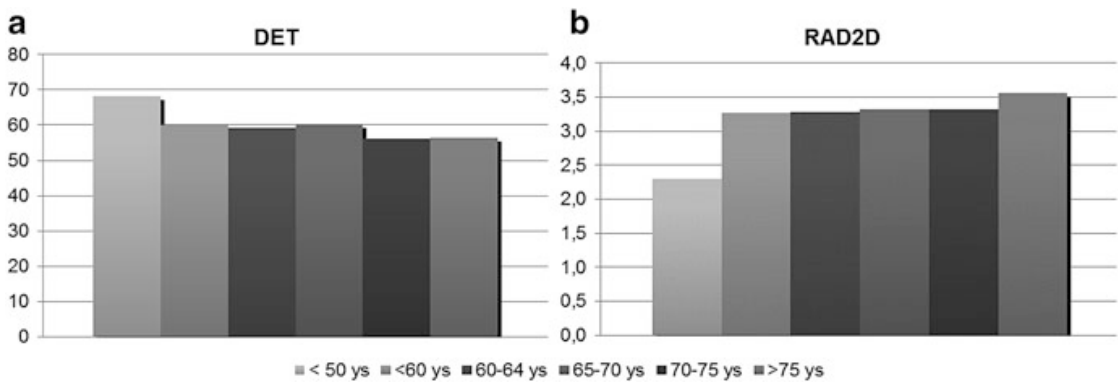


Fig. 8.18 (a) DET and (b) RAD2D values for five age-groups of ARHL ears. Adapted from [76]

A comparison was conducted among TEOAE signals recorded from three different groups: normoacoustic ears (Norm), ears with ARHL, and ears with NIHL centered at 4 kHz [76]. ARHL was defined if the subject was older than 50 years-old and presented at the PTA examination a threshold at 8 kHz higher than that at 4 kHz, and the threshold at 4 kHz higher than that at 2 kHz. From the RQA study it came out that while REC was not statistically different among the three groups, DET and RAD2D were statistically different between the ARHL and NIHL groups, as well as between each of the two impaired-ear groups and the normal group [76]. It is interesting to note here that 36 signals out of 115 among the pathological ones scored a REPRO value higher than 70, i.e. they would have passed the standard clinical ILO test.

Sub-dividing the ARHL group according to the increasing age, five groups were defined: younger than 60, between 60 and 64, 65–70, 70–75, and older than 75 years-old (audiograms shown in Fig. 8.5). Figure 8.18 reports the DET values and the RAD2D values obtained in the five age groups, compared with the values obtained from the normal group. From the figure it can be supposed that DET decreases and RAD2D increases with age, showing a slow age-related effect. However, the small numbers of used signals do not allow obtaining statistically significant differences among the different groups [76].

8.9 Conclusions and Future Directions

In this chapter RQA was presented as a powerful method to study the characteristics of OAEs. While OAEs are essential in the diagnosis of hearing losses, the information they carry on are only partially exploited by the actual diagnosis tools.

RQA has been applied to TEOAE signal sometimes in combination with the PCA technique. In this case the RQA descriptors were projected into the PCA plane, identifying the different signals with a simple dot in that space. This very simple representation of the TEOAE waveform proved to be able to discriminate among different signals coming from different ears, to cluster signals coming from the same

ear but recorded at different times, to discriminate among adults and newborns and among different types of hearing losses, as conductive versus sensorineural or age-related versus noise-induced hearing losses.

The reported research shows that the proposed approach can be useful in screening of adults, in longitudinal studies, in studies devoted to the evaluation of the efficacy of new pharmacological treatments, in conservation programs devoted to presbycusis, and in protection programs devoted to noise induced hearing losses. Moreover, it shows that RQA and RQA/PCA can be successfully used to study TEOAE signals and could be the basis for the definition of new clinical protocols, which could work side by side with existing ones, improving the actual diagnosis capabilities. However, it must be said that the reported results are just examples of the RQA potentiality in this field of research, and that further work, as well as a broader confirmation of the reported results needs to be carried out.

Software

The software used for the RQA analysis, that generates Recurrence plots and other RQA utilities, was downloaded from <http://homepages.luc.edu/~cwebber/>.

References

1. J. Bronzino (ed.), *The Biomedical Engineering Handbook* (CRC Press, Boca Raton, FL, USA 1995)
2. D.T. Kemp, Stimulated acoustic emissions from within the human auditory system. *J. Acoust. Soc. Am.* **64**(5), 1386–1391 (1978)
3. S. Hatzopoulos, J. Petrucelli, T. Morlet, A. Martini, Otoacoustic emission protocols revised. Data from adult subjects. *Int. J. Audiol.* **42**(6), 339–347 (2003)
4. C.A. Shera, J.J. Guinan Jr., Evoked otoacoustic emissions arise by two fundamentally different mechanisms: a taxonomy for mammalian OAEs. *J. Acoust. Soc. Am.* **105**, 782–798 (1999)
5. D. Konrad-Martin, D.H. Keefe, Time–frequency analysis of transient-evoked stimulus-frequency and distortion-product otoacoustic emissions: testing cochlear model predictions. *J. Acoust. Soc. Am.* **114**(4), 2021–2043 (2003)
6. W.E. Brownell, Outer hair cell electromotility and otoacoustic emissions. *Ear Hear.* **11**(2), 82–92 (1990)
7. R. Sisto, A. Moleti, Otoacoustic emissions and cochlear reflectivity. *J. Acoust. Soc. Am.* **124**(5), 2995–3008 (2008)
8. R.H. Withnell, C. Hazlewood, A. Knowlton, Reconciling the origin of the transient evoked otoacoustic emission in humans. *J. Acoust. Soc. Am.* **123**(1), 212–221 (2008)
9. S. Verhulst, J.M. Harte, T. Dau, Temporal suppression and augmentation of click-evoked otoacoustic emissions. *Hear. Res.* **246**, 23–35 (2008)
10. R. Probst, B.L. Lonsbury-Martin, G.K. Martin, A review of otoacoustic emissions. *J. Acoust. Soc. Am.* **89**(5), 2027–2067 (1991)
11. P. Avan, B. Buki, B. Maat, M. Dordain, H.P. Wit, Middle ear influence on otoacoustic emissions. I: non invasive investigation of the human transmission apparatus and comparison with model results. *Hear. Res.* **140**, 189–201 (2000)
12. P. Avan, P. Bonfils, Distortion-product otoacoustic emission spectra and high-resolution audiometry in noise-induced hearing loss. *Hear. Res.* **209**(1–2), 68–75 (2005)

13. A.L. Hamdan, K.S. Abouchacra, A.G. Al Hazzouri, G. Zaytoun, Transient-evoked otoacoustic emissions in a group of professional singers who have normal pure-tone hearing thresholds. *Ear Hear.* **29**(3), 360–377 (2008)
14. A.R. Fetoni, M. Garzaro, M. Ralli, V. Landolfo, M. Sensini, G. Pecorari, A. Mordente, G. Paludetti, C. Giordano, The monitoring role of otoacoustic emissions and oxidative stress markers in the protective effects of antioxidant administration in noise-exposed subjects: a pilot study. *Med. Sci. Monit.* **15**(11), PR 1–PR 8 (2009)
15. A. Paglialonga, S. Fiocchi, L. Del Bo, P. Ravazzani, G. Tognola, Quantitative analysis of cochlear active mechanisms in tinnitus subjects with normal hearing sensitivity: time–frequency analysis of transient evoked otoacoustic emissions and contralateral suppression. *Auris Nasus Larynx* **38**(1), 33–40 (2011)
16. World Health Organization (WHO). Deafness and hearing loss. Fact sheet N°300 (2013). <http://www.who.int/mediacentre/factsheets/fs300/en/index.html>. Accessed 21 July 2013
17. S. Stenfelt, Towards understanding the specifics of cochlear hearing loss: a modelling approach. *Int. J. Audiol.* **47**(S2), S10–S15 (2008)
18. G.A. Gates, J.H. Mills, Presbycusis. *Lancet* **366**(9491), 1111–1120 (2005)
19. EU-OSHA – European Agency for Safety and Health at Work. Combined exposure to Noise and Ototoxic Substances, 2009
20. H.F. Schuknecht, Presbycusis, in *Pathology of the Ear*, ed. by H.F. Schuknecht, M.R. Gacek (Harvard University Press, Cambridge, 1974)
21. H.F. Schuknecht, M.R. Gacek, Cochlear pathology in presbycusis. *Ann. Otol. Rhinol. Laryngol.* **102**(1 Pt 2), 1–16 (1993)
22. C.R. Jennings, N.S. Jones, Presbycusis. *J. Laryngol. Otol.* **115**(3), 171–178 (2001)
23. X.Z. Liu, D. Yan, Ageing and hearing loss. *J. Pathol.* **211**(2), 188–197 (2007)
24. K.K. Ohlemiller, M.E. Rybak Rice, J.M. Lett, P.M. Gagnon, Absence of strial melanin coincides with age-associated marginal cell loss and endocochlear potential decline. *Hear. Res.* **249**(1–2), 1–14 (2009)
25. E.C. Bielefeld, C. Tanaka, G.D. Chen, D. Henderson, Age-related hearing loss: is it a preventable condition? *Hear. Res.* **264**, 98–107 (2010)
26. Q. Huang, J. Tang, Age-related hearing loss or presbycusis. *Eur. Arch. Otorhinolaryngol.* **267**, 1179–1191 (2010)
27. A.R. Fetoni, P.M. Picciotti, G. Paludetti, D. Troiani, Pathogenesis of presbycusis in animal models: a review. *Exp. Gerontol.* **46**(6), 413–425 (2011)
28. OSHA United States Occupational Safety and Health Administration, 1910.95 CFR occupational noise exposure: hearing conservation amendment (final rule). *Fed. Regist.* **48**, 9738–9785 (1983)
29. EU-OSHA – European Agency for Safety and Health at Work. Noise in Figures, Risk Observatory Thematic Report, 2006
30. D. Henderson, E.C. Bielefeld, K.C. Harris, B.H. Hu, The role of oxidative stress in noise-induced hearing loss. *Ear Hear.* **27**(1), 1–19 (2006)
31. C.G. Le Prell, D. Yamashita, S.B. Minami, T. Yamasoba, J.M. Miller, Mechanisms of noise-induced hearing loss indicate multiple methods of prevention. *Hear. Res.* **226**, 22–43 (2007)
32. J.L. Puel, S. Saffiedine, C. Gervais d’Aldin, M. Eybalin, R. Pujol, Synaptic regeneration and functional recovery after excitotoxic injury in the guinea pig cochlea. *C. R. Acad. Sci. III* **318**(1), 67–75 (1995)
33. A.R. Fetoni, A. Ferraresi, C.L. Greca, D. Rizzo, B. Sergi, G. Tringali, R. Piacentini, D. Troiani, Antioxidant protection against acoustic trauma by coadministration of idebenone and vitamin E. *Neuroreport* **19**(3), 277–281 (2008)
34. A.R. Fetoni, R. Piacentini, A. Fiorita, G. Paludetti, D. Troiani, Water-soluble coenzyme Q(10) formulation (Q-ter) promotes outer hair cell survival in a guinea pig model of noise induced hearing loss (NIHL). *Brain Res.* **1257**, 108–116 (2008)
35. A.R. Fetoni, M. Ralli, B. Sergi, C. Parrilla, D. Troiani, G. Paludetti, Protective effects of *N*-acetylcysteine on noise induced hearing loss in guinea pigs. *Acta Otorhinolaryngol. Ital.* **29**(2), 70–75 (2009)

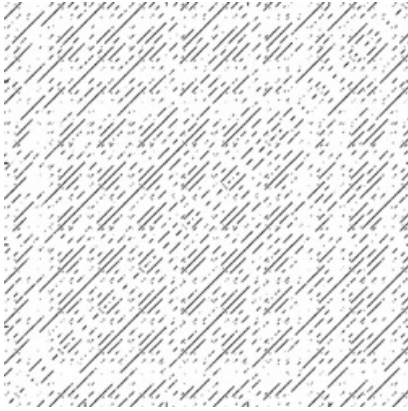
36. A.R. Fetoni, M. Ralli, B. Sergi, C. Parrilla, D. Troiani, G. Paludetti, Protective properties of antioxidant drugs in noise-induced hearing loss in the guinea pig. *Audiol. Med.* **6**(4), 271–277 (2009)
37. B.M. Vinck, P.B. van Cauwenberge, L. Leroy, P. Corthals, Sensitivity of transient evoked and distortion product otoacoustic emissions to the direct effects of noise on the human cochlea. *Audiology* **38**, 44–52 (1999)
38. Y. Uchida, T. Nakashimat, F. Ando, N. Niino, H. Shimokata, Is there a relevant effect of noise and smoking on hearing? A population-based aging study. *Int. J. Audiol.* **44**, 86–91 (2005)
39. D. Balatsouras, A. Kaberos, E. Karapantzios, E. Homsoglou, N.C. Economou, S. Korres, Correlation of transiently evoked otoacoustic emission measures to auditory thresholds. *Med. Sci. Monit.* **10**(2), MT24–MT30 (2004)
40. D. Davilis, S.G. Korres, D.G. Balatsouras, E. Gkoritsa, G. Stivaktakis, E. Ferekidis, The efficacy of transiently evoked otoacoustic emissions in the detection of middle-ear pathology. *Med. Sci. Monit.* **11**(12), MT75–MT78 (2005)
41. S. Hatzopoulos, A. Grzanka, A. Martini, W. Konopka, New clinical insights for transiently evoked otoacoustic emission protocols. *Med. Sci. Monit.* **15**(8), CR403–CR408 (2009)
42. M.L. Whitehead, B.B. Stagner, B. Lonsbury-Martin, G.K. Martin, Measurement of otoacoustic emissions for hearing assessment. *IEEE Eng. Med. Biol. Mag.* **13**, 210–226 (1994)
43. P. Ravazzani, F. Grandori, Evoked otoacoustic emissions: nonlinearities and response interpretation. *IEEE Trans. Biomed. Eng.* **40**(5), 500–504 (1993)
44. Otodynamic. ILO OAE Instrument User Manual. Issue 5a, 1997
45. C.A. Spera, Mechanisms of mammalian otoacoustic emission and their implications for the clinical utility of otoacoustic emissions. *Ear Hear.* **25**(2), 86–97 (2004)
46. J. Attias, M. Furst, V. Furman, I. Reshef, G. Horowitz, I. Bresloff, Noise-induced otoacoustic emission loss with or without hearing loss. *Ear Hear.* **16**(6), 612–618 (1995)
47. A. Shupak, D. Tal, Z. Sharoni, M. Oren, A. Ravid, H. Pratt, Otoacoustic emissions in early noise-induced hearing loss. *Otol. Neurotol.* **28**(6), 745–752 (2007)
48. M. Lucertini, A. Moleti, R. Sisto, On the detection of early cochlear damage by otoacoustic emission analysis. *J. Acoust. Soc. Am.* **111**(1–2), 972–978 (2002)
49. J.P. Zbilut, A. Giuliani, C.L. Webber, Recurrence quantification analysis and principal components in detection of short complex signals. *Phys. Lett. A* **237**, 131–135 (1998)
50. C. Manetti, M.A. Ceruso, A. Giuliani, C.L. Webber, J.P. Zbilut, Recurrence quantification analysis as a tool for characterization of molecular dynamics simulation. *Phys. Rev. E* **59**, 992–998 (1999)
51. G. Zimatore, A. Giuliani, C. Parlapiano, G. Grisanti, A. Colosimo, Revealing deterministic structures in click-evoked otoacoustic emissions. *J. Appl. Phys.* **88**(4), 1431–1437 (2000)
52. G. Zimatore, A. Giuliani, S. Hatzopoulos, A. Martini, A. Colosimo, Otoacoustic emissions at different click intensities: invariant and subject dependent features. *J. Appl. Phys.* **95**(6), 2299–2305 (2003)
53. P. Ravazzani, G. Tognola, M. Parazzini, F. Grandori, Principal component analysis as a method to facilitate fast detection of transient-evoked otoacoustic emissions. *IEEE Trans. Biomed. Eng.* **50**(2), 249–252 (2003)
54. W.A. Harrison, S.J. Norton, Characteristics of transient evoked otoacoustic emissions in normal-hearing and hearing impaired children. *Ear Hear.* **20**, 75–86 (1999)
55. A.B. Maxon, B.R. Vohr, K.R. White, Newborn hearing screening: comparison of a simplified otoacoustic emissions device (ILO1088) with the ILO88. *Early Hum. Dev.* **45**, 171–178 (1996)
56. T. Morlet, L. Goforth, L.J. Hood, C. Ferber, R. Duclaux, C.I. Berlin, Development of human cochlear active mechanism asymmetry: involvement of the medial olivocochlear system? *Hear. Res.* **134**, 153–162 (1999)
57. S.J. Norton, M.P. Gorga, J.E. Widen, R.C. Folsom, Y. Sininger, B. Cone-Wesson, B.R. Vohr, K. Mascher, K. Fletcher, Identification of neonatal hearing impairment: evaluation of transient evoked otoacoustic emission, distortion product otoacoustic emission, and auditory brain stem response test performance. *Ear Hear.* **21**(5), 508–528 (2000)

58. S.J. Norton, Application of transient evoked otoacoustic emissions to pediatric populations. *Ear Hear.* **14**, 64–73 (1993)
59. G. Zimatore, S. Hatzopoulos, A. Giuliani, A. Martini, A. Colosimo, Comparison of transient otoacoustic emission responses from neonatal and adult ears. *J. Appl. Physiol.* **92**(6), 2521–2528 (2002)
60. J.B. Allen, Nonlinear cochlear signal processing, in *Physiology of the Ear*, ed. by A.F. Jahn, J. Santos-Sacchi, 2nd edn. (Singular Thompson, San Diego, 2001), pp. 393–442
61. L. Robles, M.A. Ruggero, Mechanics of the mammalian cochlea. *Physiol. Rev.* **81**, 1305–1352 (2001)
62. H.P. Wit, P. van Dijk, P. Avan, Wavelet analysis of real ear and synthesized click-evoked otoacoustic emissions. *Hear. Res.* **73**, 141–147 (1994)
63. R.F. Lyon, C. Mead, An analog electronic cochlea. *IEEE Trans. Acoust. Speech Signal Process.* **36**(7), 1119–1134 (1988)
64. J. Merhaud, *Theory of Electroacoustics* (McGraw-Hill, New York, 1981)
65. C. Giguère, P.C. Woodland, A computational model of the auditory periphery for speech and hearing research. I. Ascending path. *J. Acoust. Soc. Am.* **95**(1), 331–342 (1994)
66. C. Giguère, P.C. Woodland, A computational model of the auditory periphery for speech and hearing research. II. Descending paths. *J. Acoust. Soc. Am.* **95**(1), 343–349 (1994)
67. G. Zimatore, M. Cavagnaro, A. Giuliani, A. Colosimo, Reproducing cochlear signals by a minimal electroacoustic model. *Open J. Biophys.* **2**, 33–39 (2012)
68. M.B. Gardner, M.S. Hawley, Network representations of the external ear. *J. Acoust. Soc. Am.* **52**, 1620–1628 (1972)
69. M.E. Lutman, A.M. Martin, Development of an electroacoustic analogue model of the middle ear and acoustic reflex. *J. Sound Vib.* **64**(1), 133–157 (1979)
70. G. Zimatore, M. Cavagnaro, A. Giuliani, A. Colosimo, Human acoustic fingerprints. *Biophys. Bioeng. Lett.* **1**(2) (2008)
71. L. Zheng, Y.T. Zhang, F.S. Yang, D.T. Ye, Synthesis and decomposition of transient-evoked otoacoustic emissions based on an active auditory model. *IEEE Trans. Biomed. Eng.* **46**(9), 1098–1106 (1999)
72. R. Sisto, A. Moleti, On the frequency dependence of the otoacoustic emission latency in hypoacoustic and normal ears. *J. Acoust. Soc. Am.* **111**, 297–308 (2002)
73. G. Tognola, F. Grandori, P. Avan, P. Ravazzani, P. Bonfils, Frequency-specific Information from click evoked otoacoustic emissions in noise-induced hearing loss. *Audiology* **38**(5), 243–250 (1999)
74. G. Zimatore, A.R. Fetoni, G. Paludetti, M. Cavagnaro, M.V. Podda, D. Troiani, Post-processing analysis of transient-evoked otoacoustic emissions to detect 4 kHz-notch hearing impairment – a pilot study. *Med. Sci. Monit.* **17**(6), MT41–MT49 (2011)
75. G. Zimatore, D. Stanzial, M.P. Orlando, Otoacoustic emissions, in *Acoustic Emission – Research and Applications*, ed. by W. Sikorski (InTech, Rijeka, 2013), pp. 203–223
76. G. Zimatore, Noise and aging effects in otoacoustic emissions. Ph.D. Dissertation, Catholic University Medical School “A. Gemelli” of Rome Italy, 2012

Chapter 9

Vibration Analysis in Cutting Materials

Arkadiusz Syta and Grzegorz Litak



Abstract The orthogonal cutting process is a highly nonlinear phenomenon, where dry friction is combined with possible contact loss and time delay effects. Unwanted and harmful vibrations may appear during cutting; the vibrations may have a regular or chaotic nature depending on the system vibrations. These vibrations disturb the process and affect the final surface quality of the workpiece. We investigate the dynamics of the cutting process by modeling it as a two-degrees of freedom nonlinear system. Using nonlinear analysis methods, such as nonlinear time series embedding approach, recurrence plots (RP) and recurrence quantification analysis (RQA), we try to identify the system variables and cutting forces that lead to chaotic motion.

9.1 Introduction and the Dynamical Model

Metal cutting is a strongly nonlinear process [1, 2] where the chaotic motion can occur due to frictional effects caused by various phenomena: the interaction of the tool and cutting chips [1–9], delay dynamics induced by sequential passes of the cutting tool [10–13] other structural nonlinearities caused by geometrical and material properties [14, 15] or tool-workpiece contact phenomena (Fig. 9.1) [16–18]. To confirm theoretical predictions, various experimental test have been performed

A. Syta (✉) • G. Litak

Faculty of Mechanical Engineering, Lublin University of Technology, Nadbystrzycka 36,
20-618 Lublin, Poland

e-mail: a.syta@pollub.pl; g.litak@pollub.pl

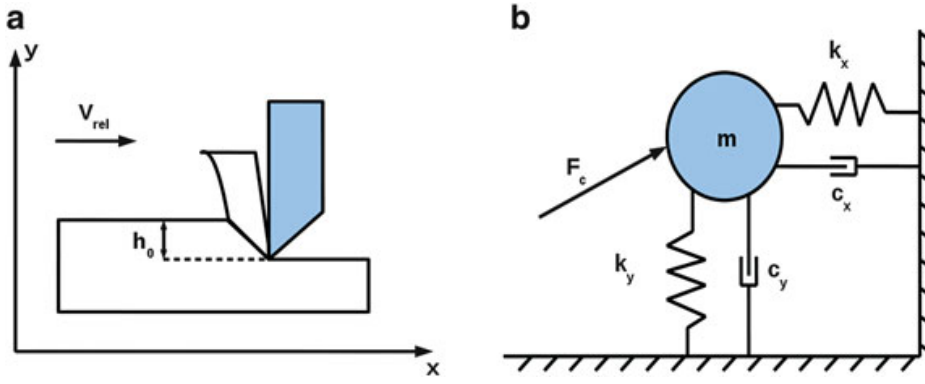


Fig. 9.1 Schematic of the cutting process (a). Effective model of vibrations (b)

[19–25]. In this study, we examined vibrations within a model with two degrees of freedom introduced by Grabec [1,2]. In this model the cutting edge is placed parallel to the material and perpendicular to the direction of cutting (Fig. 9.1a). Vibrations are caused by the cutting forces that act on the tool and may lead to its deformation (we assumed that the depth of cut is significantly smaller than the width of the material, so the tool operates only in a two-dimensional plane).

The simplified model can be considered as two coupled oscillators with appropriate equations of motion [1, 3, 26] (Fig. 9.1b):

$$\begin{aligned}\ddot{x} + D_x \dot{x} + Ax &= F_x, \\ \ddot{y} + D_y \dot{y} + By &= F_y.\end{aligned}\quad (9.1)$$

By neglecting linear damping coefficients and using dimensionless coordinates and time $(x, y, t) \rightarrow (X, Y, T)$:

$$x \rightarrow X = \frac{x}{h_0}, \quad y \rightarrow Y = \frac{y}{h_0}, \quad t \rightarrow T = \frac{v_0}{h_0} t \quad (9.2)$$

(h_0, v_0 are the assumed initial cutting depth and cutting velocity, respectively) as well as all system variables, the equations of motion can be written:

$$\begin{aligned}\ddot{X} &= F - AX, \\ \ddot{Y} &= K\mu - BY,\end{aligned}\quad (9.3)$$

where the cutting force (for a wide class of materials [27]) can be given by:

$$\begin{aligned}F &= f_0 H(C_1(V - 1)^2 + 1) \Theta(H) \Theta(V), \\ \mu &= \mu_0 (C_2(|V_p| - 1)^2 + 1) (C_3(H - 1)^2 + 1) \Theta(F) \operatorname{sgn}(V_p), \\ R &= R_0 (C_4(V - 1)^2 + 1).\end{aligned}\quad (9.4)$$

In the above equations $C_1 = 0.3$, $C_2 = 0.7$, $C_3 = 1.5$, $C_4 = 1.2$, $\mu_0 = 0.35$, $h_0 = 0.5$ and $R_0 = 2.2$ are certain system variables which model the cutting force and the material cut resistance. A , B constants denote the effective stiffness in X and Y directions (Fig. 9.1b):

$$\begin{aligned} A &= k_x / m\omega_D^2, \\ B &= k_y / m\omega_D^2, \end{aligned} \quad (9.5)$$

$\omega_D = 2.7 \cdot 10^4 \text{ s}^{-1}$ is the natural frequency of free tool vibration, $\Theta(x)$ denotes the Heaviside step function, $V = \dot{X}$ is the tool velocity in the x direction. Other variables include: F , the vertical cutting force component, R , the material deformation resistant constant while μ , the friction coefficient, and V_p , the chip flow relative velocity defined as:

$$V_p = \frac{v}{v_0 R}. \quad (9.6)$$

The actual cutting depth $h(T)$ is expressed as the uncut chip thickness, h_0 , and the tool displacement $Y(T)$ as:

$$h(T) = H(T)h_0 = (1 - Y(T))h_0. \quad (9.7)$$

9.2 Simulation Results

The dynamical system [Eq. (9.4)] has been solved numerically using the fourth order Runge–Kutta solver with the integration step set to $\Delta T = 0.001$. As mentioned before the chatter vibration can occur in certain regions of cutting force values. To observe differences in the system response we analyze plots of the cutting forces versus the relative velocity V_p obtained for two different values of the cutting force coefficient $f_0 = 0.25, 0.50$ (Fig. 9.2). Note, we consider the two different cases already discussed in the scientific literature [1, 2, 9] to show the relative differences between corresponding solutions.

As can be seen in Fig. 9.2, the force function F_x is nonsmooth for the relative velocity $V_p = 0$. Furthermore, in the case of regular response (for $f_0 = 0.25$), the force vanishes ($F_x = 0$) for $V_p < 0$. However, in the case of irregular response (for $f_0 = 0.5$) this condition is not satisfied. The reason for that is the cutting depth $h < 0$. This means loss of contact between the tool and workpiece material (which is quite undesirable). Moreover, a change in one of the system variables, such as the material dependent cutting-force coefficient f_0 , influences the cutting depth h . Figure 9.3 presents the cutting depth time evolution $h(t)$ depending on different values of the cutting-force coefficient f_0 . Note that the time dependence of cutting depth $h(t)$ with respect to the cutting velocity defines the shape of the machined

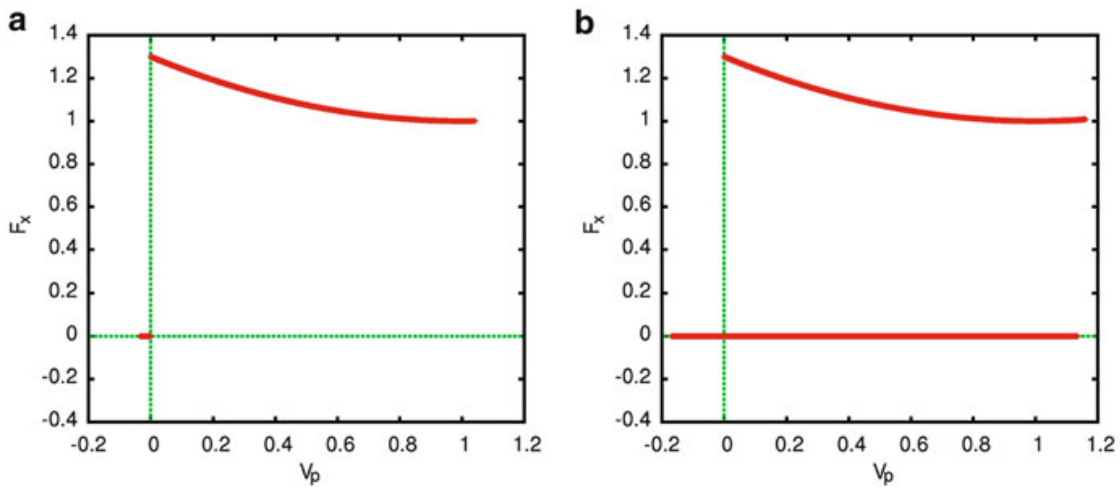


Fig. 9.2 The simulated cutting force $F_x(V_p)$ obtained for two cases $f_0 = 0.25$ (a) and $f_0 = 0.50$ (b). Note that the case (b) corresponds to contact loss between the tool and workpiece

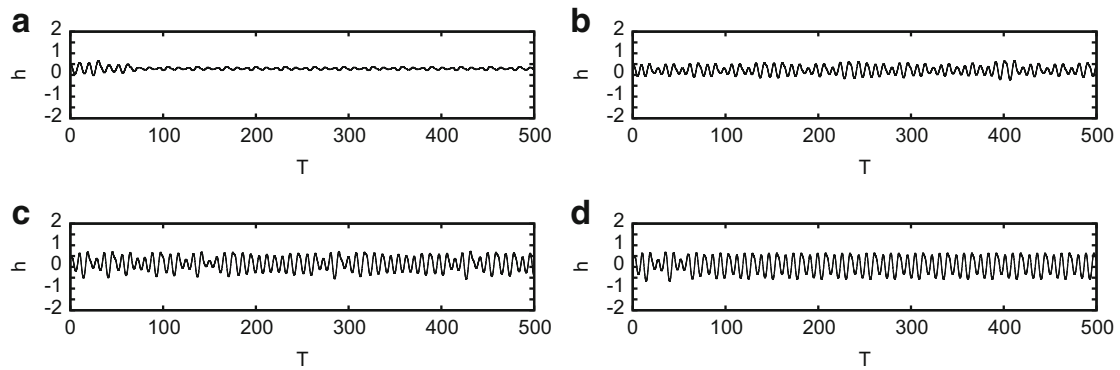


Fig. 9.3 Simulated cutting depth histories for different values of the cutting depth coefficient f_0 . f_0 was set to 0.25, 0.50, 0.75, 1.00 for (a), (b), (c), and (d), respectively

surface. Figure 9.3a–d show results for various f_0 . It is clear that increasing f_0 implies larger amplitudes in resulting oscillations in the cutting depths. Concluding these results, we note that simulated response for cutting forces and corresponding depths (see Figs. 9.2 and 9.3), one can say that the system response is very sensitive to the system parameter changes. This is typical of nonlinear systems.

By a more careful examination of Fig. 9.3, one can also notice that for a large enough f_0 ($f_0 = 0.50, 0.75$ and 1.00), h can reach negative values, which implies contact loss events. The values of f_0 were taken from the studies by Grabec et al. [1], where various responses of the dynamical system of the cutting process were identified based on their frequency spectra. For more clarity, we present bifurcation diagrams with respect to displacement and its velocity (x and \dot{x}) (Fig. 9.4) plotted against the parameter f_0 . The points on the bifurcation diagram match the corresponding Poincaré section projections.

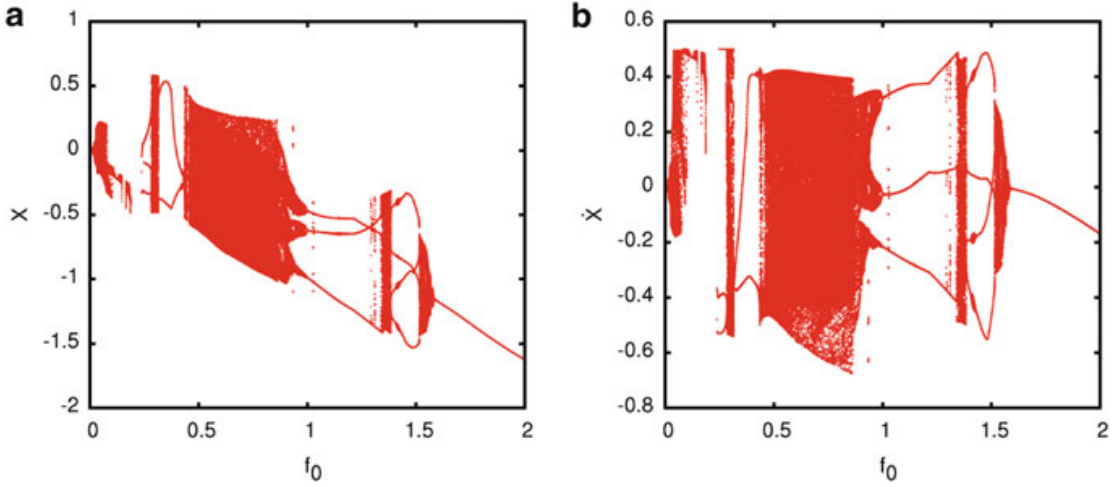


Fig. 9.4 Bifurcation diagrams X (a) and \dot{X} (b) versus f_0 . All the system parameter as in Fig. 9.2a, b, respectively

It is worth recalling that our vibrating dynamic system has no external excitation but is driven by a self-excited mechanism. This complicates the construction of the Poincaré map which could not be easily defined via stroboscopic properties. For contraction purposes, in the present case we assumed:

$$Y > 0.25 \text{ and } |\dot{Y}| < 0.0001, \quad (9.8)$$

and registered 80,000 points in the steady-state time evolution to show the differences.

More light can be shed by examining the corresponding phase portraits and Poincaré maps (see Fig. 9.5a–l). Figure 9.5a, j show the periodic response of the system for $f_0 = 0.25$ and $f_0 = 1.0$. Note that multiperiod character of solutions in Fig. 9.6a, d is also clearly expressed in terms the Poincaré sections (Fig. 9.5).

In these cases, the Poincaré maps (Fig. 9.5b, k) consist of a collection of singular points. Together with phase portraits, one can also find the closed lines (Fig. 9.5a, c, j, l).

On the other hand, for the coefficient f_0 set to 0.50 and 0.75 the Poincaré sections consist of many points forming strange attractors of the fractal topology (see Fig. 9.5c, h, respectively). Note also that the corresponding phase portraits do not form the closed loops any more (see Fig. 9.5d, f, g, i).

9.3 Recurrence Plot Analysis

The nonsmooth dynamical systems are difficult to study analytically and numerically. For more detailed analysis, the Recurrence Plots (RPs) approach can be applied. With this method, we need to optimize the data sampling. As the simulation

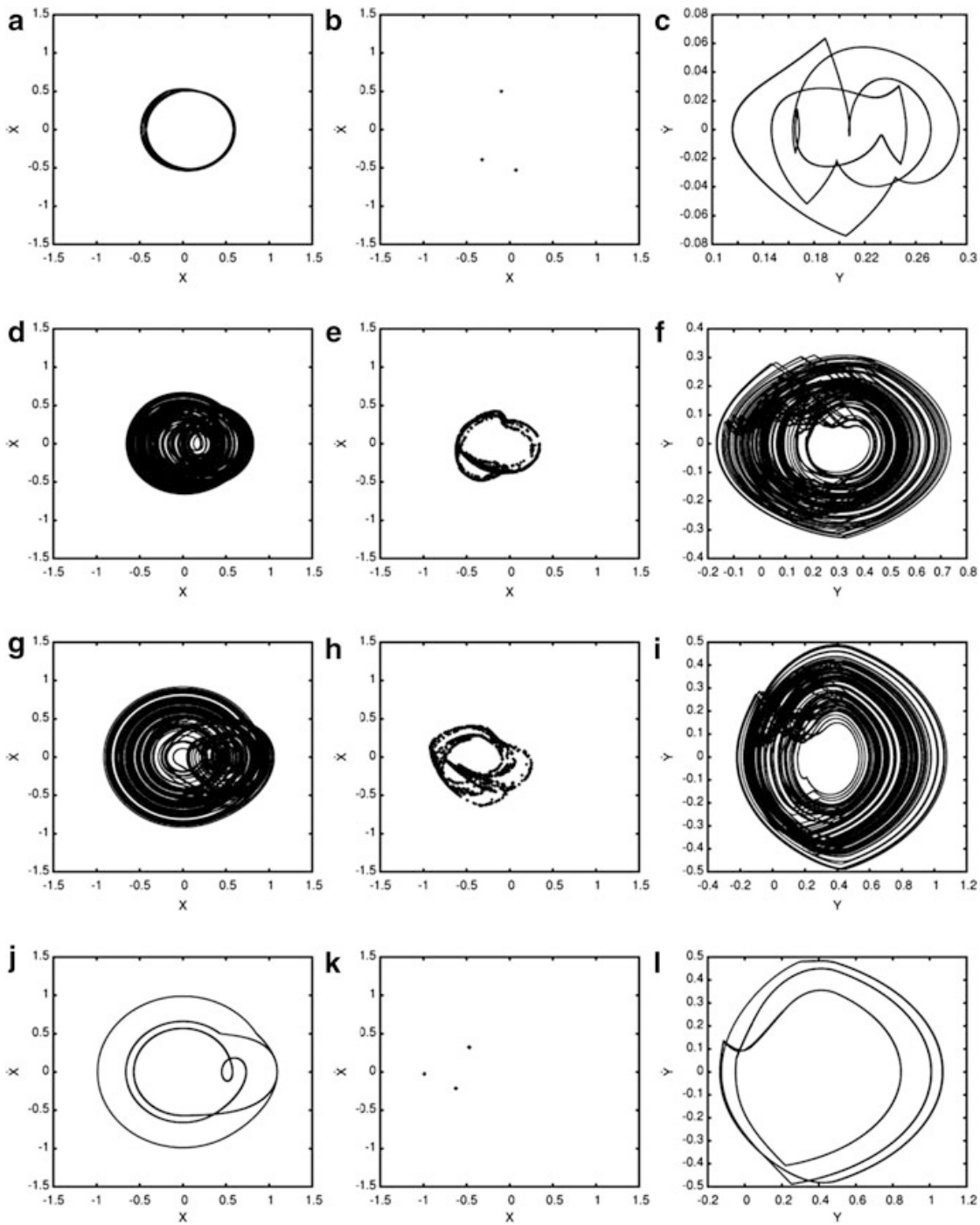


Fig. 9.5 Phase portraits (x, \dot{x}) and (y, \dot{y}) (a, c) for $f_0 = 0.25$ (d, f) $f_0 = 0.50$ (g, i) $f_0 = 0.75$, and (j, l) $f_0 = 1.00$. Poincaré maps (b, e, h, k) for $f_0 = 0.25, 0.50, 0.75$ and 1.00 , respectively. Note the different axis scale in (y, \dot{y}) (c)

time step ΔT was chosen relatively small, we reduced the system state representation using the Average Mutual Information (AMI) method. After Fraser [28–30] one can use the AMI as a kind of correlation function. The corresponding functional AMI is defined in terms of the multiple sampling intervals δi :

$$\text{AMI}(\delta i) = - \sum_{k,l} p_{kl}(\delta i) \ln \frac{p_{kl}(\delta i)}{p_k p_l}, \quad (9.9)$$

where for some partition the horizontal displacement band ($X \in [X_{\min}, X_{\max}]$), p_k is the probability of finding the displacement X in the k -th interval of $[X_{\min}, X_{\max}]$ divided into 16 intervals, and p_{kl} is the joint probability that displacement X , which was initially in k -th interval falls later into the l -th one within the time interval δi . Here i represents the instantaneous time ($T_i = i \Delta T$), while the characteristic time delay $\delta T = \Delta T \tau$ was estimated as the first minimum of AMI $\tau = \delta i$ for $f_0 = 0.25, 0.50, 0.75$ and 1.00 , τ was estimated as 180, 177, 182, 169 in the units of the time step ΔT , respectively.

The Recurrence Plot (RP) is given by [31]:

$$R_{ij}^\epsilon = \Theta(\epsilon - \|\mathbf{X}_i - \mathbf{X}_j\|), \quad (9.10)$$

where \mathbf{R} is the recurrence matrix and ϵ is the threshold value. In an RP, the elements 0 and 1 of the matrix \mathbf{R} are represented by an empty space and a coloured dot, respectively. Here, instead of building the embedding space, we used the simulated 4d space

$$\mathbf{X} = \{X, \dot{X}, Y, \dot{Y}\}, \quad (9.11)$$

with the Euclidean distance measure.

Webber and Zbilut [32] and later Marwan et al. [33–35] developed on Recurrence Quantification Analysis (RQA) for recurrence plots. First of all RQA analysis includes the recurrence rate variable RR . The RR expresses the measure of ability for the system to return in the neighbourhood of previous states and has the following definition,

$$RR = \frac{1}{N(N-1)} \sum_{i,j \neq i}^N R_{ij}^\epsilon. \quad (9.12)$$

Furthermore, the RQA can be used to identify maximal lengths L_{MAX} , V_{MAX} for diagonal and vertical lines, respectively. In its frame the RQA provides us with the probability $p(l)$ or $p(v)$ of line distribution according to their lengths l or v (for diagonal and vertical lines). Practically, they are calculated

$$p(z) = \frac{P^\epsilon(z)}{\sum_{z=z_{\min}}^N P^\epsilon(z)}, \quad (9.13)$$

where $z = l$ or v depending on diagonal or vertical structures in the specific recurrence plot. $P^\epsilon(z)$ denotes the histogram of z lengths and a fixed value of ϵ .

Based on the distributions $p(x)$ the Shannon information entropies (LE and VE) can be given by

$$LE = - \sum_{l=l_{min}}^N p(l) \ln p(l), \quad (9.14)$$

$$VE = - \sum_{v=v_{min}}^N p(v) \ln p(v) \quad (9.15)$$

Higher LE or VE would signal an increase in the complexity of the considered time series. Other properties, such as: determinism DET , laminarity LAM , trapping time TT and average length L , are also based on histograms $P^\epsilon(z)$

$$DET = \frac{\sum_{l=l_{min}}^N l P^\epsilon(l)}{\sum_{l=1}^N l P^\epsilon(l)}, \quad (9.16)$$

$$LAM = \frac{\sum_{v=v_{min}}^N v P^\epsilon(v)}{\sum_{v=1}^N v P^\epsilon(v)}, \quad (9.17)$$

$$L = \frac{\sum_{l=l_{min}}^N l P^\epsilon(l)}{\sum_{l=l_{min}}^N P^\epsilon(l)}, \quad (9.18)$$

$$TT = \frac{\sum_{v=v_{min}}^N v P^\epsilon(v)}{\sum_{v=v_{min}}^N P^\epsilon(v)}, \quad (9.19)$$

where l_{min} and v_{min} denote minimal value which should be chosen for a specific dynamical system. To exclude the isolated points, in our calculations, we have used $l_{min} = 2$ and $v_{min} = 5$. The determinism DET is the measure of predictability of the examined time series and gives the ratio of the recurrence points formed in diagonals to all recurrent points. Note in a periodic system all points would inscribe lines. On the other hand, the laminarity LAM is a similar measure which corresponds to points formed in vertical lines. This measure indicates the periodic dynamics behind sampling point changes. For small point-to-point changes the consecutive points form a vertical line. Finally, the average diagonal length L refers to the predictability time of the dynamical system while the trapping time TT refers to the average length of vertical lines measuring the time scale (in terms of sampling intervals) of small changes in the examined time history.

In the case of the RP and RQA, ϵ is an important measure of the threshold distance between recurrence points (corresponding to states in the space \mathbf{X}). This distance cannot be too small as the recurrences tend to 0 in this limit. On the other hand, for a fairly large ϵ most of points (states) become recurrence points. One of the optimal choices for ϵ is based on the observation of the RR variable. It should remain in the interval $[0.01, 0.05]$ [34] to stay in the range of linear growth $RR(\epsilon)$.

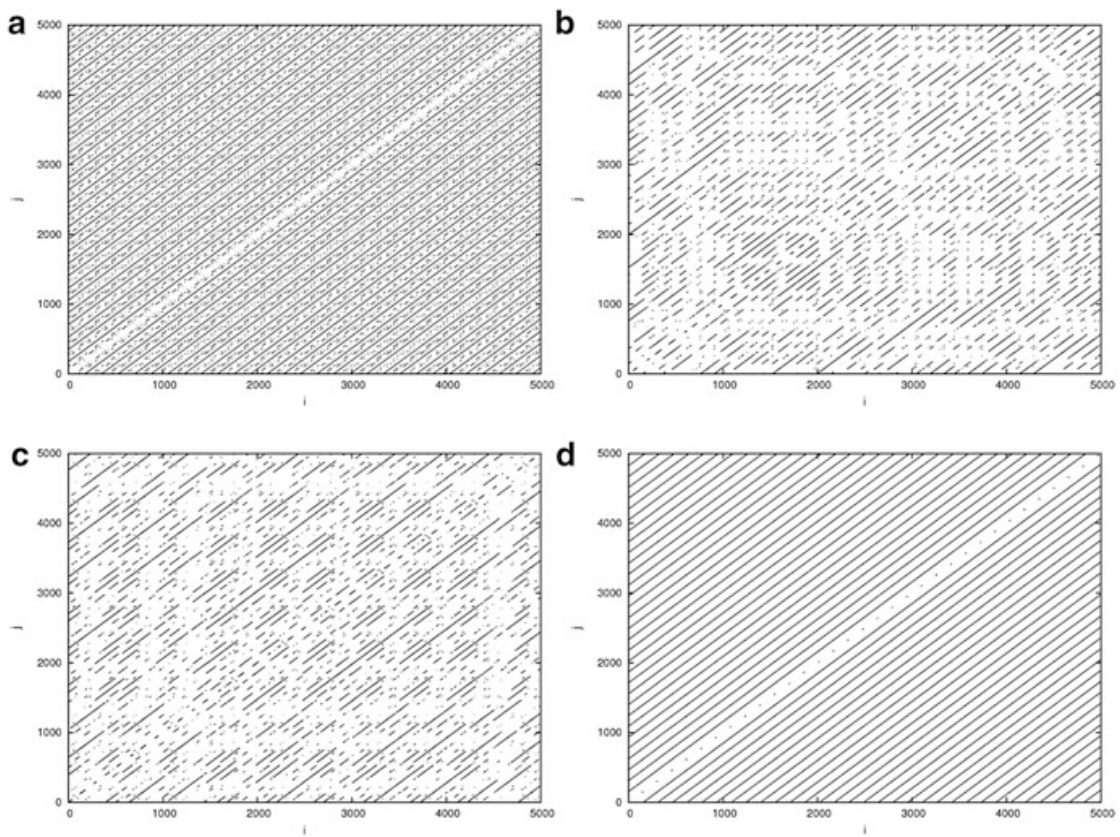


Fig. 9.6 Recurrence plots for $f_0 = 0.25$ (a), 0.50 (b), 0.75 (c), 1.00 (d). For better clarity, we used different RR s

Table 9.1 Recurrence variable based on diagonal lines for various f_0

f_0	ϵ	DET	$LMAX$	L	LE
0.25	0.055	0.9795	4825	47.766	3.4259
0.50	0.14	0.9851	638	33.867	4.1137
0.75	0.175	0.9746	903	30.040	3.9176
1.0	0.09	0.9870	4689	48.731	3.1923

For each case $RR = 0.01$

For the RQA, we decided to set its vale to $RR = 0.01$ and ϵ was equal to 0.055, 0.14, 0.175, 0.09 for $f_0 = 0.25, 0.50, 0.75$ and 1.00 , respectively. The aim was to keep the RR constant providing the same number of black points in each of RPs.

Some indications of the system dynamics can be obtained by visualizing the recurrence matrix \mathbf{R} plotted in Fig. 9.6a–d. The presents of long diagonal lines with a constant distance between these lines (Fig. 9.6a, d) implies a periodic behaviour, while shorter lines (Fig. 9.6b, c) stand for a locally correlated signal nonperiodic behavior—chaotic motion. For better clarity, we used different RR vales for various cases.

More insight into the system dynamics can be provided through recurrence variables for diagonal and vertical lines. They are presented in Tables 9.1 and 9.2. Similarly to Fig. 9.6, in the cases of the diagonal lines statistics, we observe for

Table 9.2 Recurrence variable based on vertical lines for various f_0

f_0	ϵ	LAM	$VMAX$	TT	VE
0.25	0.055	0	5	0	0
0.50	0.14	0.1876	9	5.595	1.0375
0.75	0.175	0.2161	11	6.034	1.3760
1.0	0.09	0.1064	6	5.021	0.1045

For each case $RR = 0.01$

$f_0 = 0.25$ and 1.00 fairly long maximal lengths $LMAX$, higher predictability in terms of average length L and finally a lower entropy LE . Simultaneously, the cutting coefficient values $f_0 = 0.50$ and 0.75 lead to the opposite conclusions indicating a considerable difference. On the other hand, the DET is on the same level for all the cases.

For the vertical lines, we observe similar differences showed by the vertical lines entropy variable VE . Interestingly, the LAM was also different—higher for $f_0 = 0.50$ and 0.75 with respect to $f_0 = 0.25$ and 1.00 , indicating that the system transits from the “laminar” to “turbulent” behavior. Finally, the TT and $VMAX$ appeared to be larger for chaotic dynamics (see Tables 9.1 and 9.2). They are caused by variations of recurrence times in cases of a nonperiodic evolution of the system [33].

9.4 Conclusions

We have examined the periodic and chaotic responses of the cutting process system. The application of the RP and RQA appeared to be very useful and comprehensive to system response identification. The periodic and chaotic vibrations mapped on the corresponding RPs differ significantly. In the first class of regular solutions, one may notice that the RP consists of long diagonal lines, with many isolated points. These were present account for the multiperiodic solution. In the second case, one can find a fairly wide distribution of diagonal lines in the corresponding RPs and some evidences of appearance of short vertical structure. Note, these qualitative changes are visible in a quantitative way through the RQA variables. In particular, the appearance of a nonperiodic response was shown by the significant differences of Shannon entropies LE and VE and the laminarity variable LAM .

The obtained results have been also confirmed by such standard techniques as Poincaré maps, bifurcation diagrams and the phase portraits. The identification procedure may be useful for developing reliable methods of controlling and minimizing detrimental vibrations such as chatter. It is worth to add that the similar approach (RP and RQA) has been also successfully applied to classify experimental data of the turning process [36] while looking for characteristic chatter vibrations. The disadvantages of recurrence techniques relate to the need for multiple RQA

parameter selections as well as the complex dependence of the behavior of the dynamical system on the threshold, ϵ .

Acknowledgements Financial support the Structural Funds under the Operational Programme—Innovative Economy (IE OP) financed from the European Regional Development Fund—Project Modern material technologies in aerospace industry, No. POIG.01.01.02-00-015/08-00, is gratefully acknowledged.

References

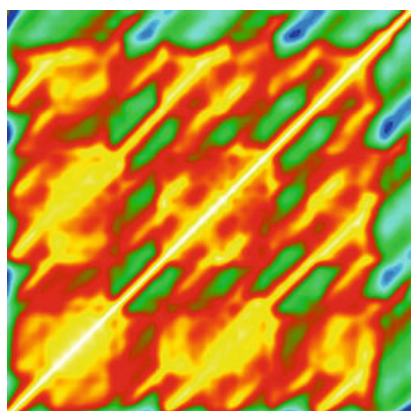
1. I. Grabec, Chaos generated by the cutting process. *Phys. Lett. A* **8**, 384–386 (1986)
2. I. Grabec, Chaotic dynamics of the cutting process. *Int. J. Mach. Tools Manufact.* **28**, 19–32 (1988)
3. M. Wiercigroch, Chaotic vibration of a simple model of the machine tool-cutting process system. *J. Vib. Acoust.* **119**, 468–475 (1997)
4. M. Wiercigroch, A. Cheng, Chaotic and stochastic dynamics of orthogonal metal cutting. *Chaos Solitons Fractals* **8**, 715–726 (1997)
5. M. Wiercigroch, A.M. Krivtsov, Frictional chatter in orthogonal metal cutting. *Philos Trans. R. Soc. A* **359**, 713–738 (2001)
6. J. Warminski, G. Litak, M.P. Cartmell, R. Khanin, M. Wiercigroch, Approximate analytical solutions for primary chatter in the nonlinear metal cutting model. *J. Sound Vib.* **259**, 917–933 (2003)
7. S. Chatterjee and T.K. Singha, Controlling chaotic instability of cutting process by high-frequency excitation: a numerical investigation. *J. Sound Vib.* **267**, 1184–1192 (2003)
8. J. Gradisek, E. Govekar, I. Grabec, A chaotic cutting process and determining optimal cutting parameter using neural networks. *Int. J. Mach. Tools Manufact.* **36**, 1161–1172 (1996)
9. A.K. Sen, G. Litak, A. Syta, Cutting process dynamics by nonlinear time series and wavelet analysis. *Chaos* **17**, 023133 (2007)
10. G. Stepan, T. Kalmar-Nagy, Nonlinear regenerative machine tool vibrations, in *Proceedings of 16th ASME Biennial Conference on Mechanical Vibrations and Noise, ASME Design and Technical Conference*, Sacramento, 14–17 September 1997, pp. 1–11
11. G. Stepan, Delay-differential equation models for machine tool chatter, in *Dynamics and Chaos in Manufacturing Processes*, ed. by F.C. Moon (Wiley, New York, 1998), pp. 165–192
12. G. Stepan, Modelling nonlinear regenerative effects in metal cutting. *Phil. Trans. R. Soc. Lond. A* **359**, 739–757 (2001)
13. M.S. Fofana, Delay dynamical systems and applications to nonlinear machine-tool chatter. *Chaos Solitons Fractals* **17**, 731–747 (2003)
14. J.R. Pratt, A.H. Nayfeh, Design and modeling for chatter control. *Nonlinear Dyn.* **19**, 49–69 (1999)
15. A.H. Nayfeh, C.M. Chin, J.R. Pratt, Applications of perturbation methods to tool chatter dynamics, in *Nonlinear Dynamics of Material Processing and Manufacturing*, ed. by F.C. Moon (Wiley, New York, 1997), pp. 123–193
16. G. Litak, Chaotic vibrations in a regenerative cutting process. *Chaos Solitons Fractals* **13**, 1531–1535 (2002)
17. X.S. Wang, J. Hu, J.B. Gao, Nonlinear dynamics of regenerative cutting processes—Comparison of two models. *Chaos Solitons Fractals* **29**, 1219–1228 (2006)
18. G. Litak, S. Schubert, G. Radons, Nonlinear dynamics of a regenerative cutting process. *Nonlinear Dyn.* **69**, 1255–1262 (2012)
19. I.N. Tansel, C. Erkal, T. Keramidas, The chaotic characteristics of three-dimensional cutting. *Int. J. Mach. Tools Manufact.* **32**, 811–827 (1992)

20. D.B. Marghitu, B.O. Ciocirlan, N. Craciunoiu, Dynamics in orthogonal turning process. *Chaos Solitons Fractals* **12**, 2343–2352 (2001)
21. G. Litak, R. Rusinek, A. Teter, Nonlinear analysis of experimental time series of a straight turning process. *Meccanica* **39**, 105–112 (2004)
22. J. Gradisek, I. Grabec, S. Sigert, R. Friedrich, Stochastic dynamics of metal cutting: bifurcation phenomena in turning. *Mech. Syst. Signal Process.* **16**, 831–840 (2002)
23. J. Gradisek, E. Govekar, I. Grabec, Chatter onset in non-regenerative cutting: a numerical study. *J. Sound Vib.* **242**, 829–838 (2001)
24. J. Gradisek, E. Govekar, I. Grabec, Time series analysis in metal cutting: chatter versus chatter-free cutting. *Mech. Syst. Signal Process.* **12**, 839–854 (1998)
25. J. Gradisek, E. Govekar, I. Grabec, Using coarse-grained entropy rate to detect chatter in cutting. *J. Sound Vib.* **214**, 941–952 (1998)
26. G. Litak, A. Syta, M. Wiercigroch, Identification of chaos in a cutting process by the 0-1 test, *Chaos Solitons Fractals* **40**, 2095–2101 (2009)
27. W.F. Hastings, P. Mathew, P.L.B. Oxley, A machining theory for predicting chip geometry, cutting forces etc. from material properties and cutting conditions. *Proc. R. Soc. Lond. A* **371**, 569–587 (1980)
28. A.M. Fraser, H.L. Swinney, Independent coordinates for strange attractors from mutual information, *Phys. Rev. A* **33**, 1134–1140 (1986)
29. H. Kantz, T. Schreiber, *Non-linear Time Series Analysis* (Cambridge University Press, Cambridge, 1997)
30. R. Hegger, H. Kantz, T. Schreiber, Practical implementation of nonlinear time series methods: the TISEAN package. *Chaos* **9**, 413–435 (1999)
31. J.-P. Eckmann, S. O. Kamphorst, D. Ruelle, Recurrence plots of dynamical systems. *Europhys. Lett.* **5**, 973–977 (1987)
32. C.L. Webber, Jr., J.P. Zbilut, Dynamical assessment of physiological systems and states using recurrence plot strategies. *J. Appl. Physiol.* **76**, 965–973 (1994)
33. N. Marwan, J. Kurths, Line structures in recurrence plots. *Phys. Lett. A* **336**, 349–357 (2005)
34. N. Marwan, M. C. Romano, M. Thiel, and J. Kurths, Recurrence plots for the analysis of complex systems. *Phys. Rep.* **438**, 237–329 (2007)
35. N. Marwan, A historical review of recurrence plots. *Eur. Phys. J. Spec. Top.* **164**, 3–12 (2008)
36. G. Litak, R. Rusinek, Dynamics of a stainless steel turning process by statistical and recurrence analyses. *Meccanica* **47**, 1517–1526 (2012)

Chapter 10

Dynamical Patterns in Seismology

Tamaz Chelidze and Teimuraz Matcharashvili



Abstract Earthquakes are important components of complex dynamical Earth systems known as geocomplexity. One of the main contributors to geocomplexity is the seismic process which notoriously displays nonlinear behaviors including the self-organization of many interacting components (tectonic plates, faults). These processes result in seismic (macro-scale) events of collective behaviors in the temporal, spatial and energy domains. In the chapter the results of both active and passive experiments on triggering/synchronization are presented. The dynamic patterns of seismicity are revealed by the application of nonlinear dynamics tools to time series from: “laboratory” earthquakes (acoustic emission during natural stick-slip and stick-slip under weak periodic forcing), regional seismicity of the Caucasus Mountains, local seismicity in the area of large reservoir during its construction and regular exploitation, as well as from analysis of variations in local seismicity in a Central Asia test area due to application of strong electric pulses. The review of recent results on dynamic triggering of local seismicity by remote earthquakes is also presented. It is shown that relatively weak external forcings can induce clear changes in modeled and real earthquake dynamics. Recurrence Quantification Analysis (RQA) proved to be an efficient method for finding hidden nonlinear structures in seismic time series. Quoting Webber and Zbilut (2005): “... *whatever the case, whether it be forecasting dynamical events in the medical field, geophysics, or meteorology, the future of recurrence analysis looks bright and promising.*”

T. Chelidze (✉) • T. Matcharashvili
Institute of Geophysics, Tbilisi, Georgia
e-mail: tamaz.chelidze@gmail.com

10.1 Introduction

Earthquakes (EQs) are important components of complex dynamical Earth systems known as geocomplexity. One of the main contributors to geocomplexity is the seismic process which notoriously displays nonlinear behaviors including the self-organization of many interacting components (tectonic plates, faults). These processes result in seismic (macro-scale) events of collective behaviors in the temporal, spatial and energy domains. The attributes of complexity include critical phenomena, scaling properties, formation of time-space patterns, intermittency, and extreme sensitivity to initial conditions and weak forcings. This last chaotic property implies that the manifestation of seismic processes such as EQs can be triggered by relatively weak external forcings.

The Earth is dynamically subjected to oscillating force fields of differing origins (foci) over an extremely wide range of frequencies (periods of seconds to months to years). Since these oscillating forces are not isolated from the Earth's plates, they produce rhythmic stress/strain fields within the Earth's crust expressed as triggering and synchronization phenomena. This sensitivity to weak external forcings implies that the affected system is close to its critical state, where the sensitivity of the system to, weak impacts is very high.

Early observations of triggering/synchronization (T/S) were not validated by trustworthy field observations or strict mathematical procedures. Thus they should be considered as mainly qualitative if not unreliable. However, the situation has changed dramatically over the last decade with the implementation of dense seismic networks equipped by high sensitivity broadband devices. Now it is possible to identify reliable triggering events as spatially and temporally correlated local events following strong EQs. In the important domain of earthquake seismology where strong remote EQs trigger local seismicity patterns, experimental measurements trump mathematical theory.

Active laboratory experiments on the electromagnetic (EM) and mechanical triggering and synchronization of mechanical instability (slip) in the spring-slider system are herein considered in detail. In this case the main variables of the stick-slip (which is a small-scale model of seismic process) can be controlled. The forcing of local seismicity by periodic reservoir loading and strong electromagnetic pulses, wave trains from remote earthquakes can be considered as semi-active experiment where only some of process-determining variables are known. Lastly, analysis of seismic catalogs for revealing hidden periodicities represent passive experiments where most of key variables are unknown.

The robust theoretical analysis of T/S phenomena was also strongly facilitated by recent developments in the field of modern nonlinear dynamics which allows for quantitative interpretation of many of the mentioned unconditional effects. For relaxation (integrate and fire type) processes, when the signal is composed of well-defined marked events (slips, EQs), the generalized phase difference can be used for computing and comparing phases between events. Other possible measures for quantification of synchronization include Recurrence Quantification

Analysis (RQA), mutual information (MI), Shannon and Tsallis entropies, conditional probability of phases, flatness of stripes of synchrograms (stroboscopic approach), coefficients of phase diffusion, fractal dimension, Lempel–Ziv complexity, and unstable periodic orbits. For revealing hidden periodicities in short and noisy time series, which can be related to synchronization effects, the methods of Singular Spectral Analysis (SSA) and Detrended Fluctuation Analysis (DFA) have been suggested [1, 2]. Lastly, in the case of very restricted statistics (tens of events), less demanding methods such as Shuster test can be used.

From the famous “butterfly effect” of Edward Lorenz we ask the question: Can weak external forcings trigger or synchronize seismic processes which are driven by enormous tectonic forces? It is accepted [3] that the Earths’ lithosphere has active seismic regions close to their critical state which means that even weak impacts can trigger massive EQs. Many experimental observations confirm the possibility that significant events can be elicited by a wide variety of small impact forces: coulomb stress changes induced by previous strong EQs [4]; mining activity and water reservoir loadings [5, 6]; nuclear explosions [7]; snow melts [8]; strong electrical pulses [9, 10]; tides [11, 12]; wave trains of strong EQs [12–15]; solar activity; atmospheric pressure and magnetic field variations.

Nonlinear dynamics provides a solid theoretical basis for explaining observed unconventional results [16–22]. New tools for quantification of the strength of T/S between the phases of synchronized process and synchronizing impact have been developed over the last years making the corresponding assessments much more reliable [23].

Analysis of the susceptibility of the region’s seismic process to a small impact may serve as a new method of EQ prediction. This idea is founded on the hypothesis that closer is the system to the critical state, the stronger is its sensitivity to small forcings [24]. Application of nonlinear dynamics to time series related to various catastrophic phenomena reveals striking universality in the precursory patterns of such different events as seismic shocks, financial crises, cardiac attacks and epileptic seizures [25, 26].

The physical modeling of seismic T/S phenomenon is not well developed. In particular there are problems related to driving/forcing stress amplitudes and periods, the role of fluids, inhomogeneities, etc. that favor/suppress triggering.

10.2 Elements of Seismology

10.2.1 *Physics of Earthquakes*

Earthquakes (EQ) constitute important components of complex dynamical Earth systems. They are caused by the slow drift (1–8 cm/year) of Earth’s lithosphere plates driven by thermal convection in the underlying molten asthenosphere. EQs represent one of the main expressions of the release of stress which accumulates

due to this motion [3, 27]. Non-EQ manifestation of stress build-up is slow creep (several mm/year). As a rule EQs occur on active faults delineating the borders between lithosphere plates and smaller blocks. Main variables characterizing EQs include: magnitude (logarithmic measure of EQ size or energy); location; depth and timing. Bak et al. [28] considered seismicity as a self-organized process.

10.2.2 Stick-Slip as a Physical Model of the Seismic Process

Seismicity is considered as a nonlinear frictional process of the “integrate and fire” type. Here shear stresses build up (integrate) on the fault plane until the friction force is exceeded. This leads to a sudden drop in stress (fire) after which a new epoch of build-up begins again. Hence the process is non-periodic cyclic [27]. This recurrent stick-slip process can be modeled both theoretically and experimentally by single or multiple spring-slider systems.

10.2.3 Seismic Time Series

Seismic time series represent continuous recordings of Earth’s surface weak vibrations or microseisms with relatively rare strong events (EQs). Seismic catalogs are data tables containing information of the main characteristics of EQs (date, location and magnitude) and organized according to the date of events. Declustered from foreshocks and aftershocks, strong events in EQ catalogs are mainly Poisson distributed. Several methods of synthetic seismic catalog compilation have been suggested. One good example is the Epidemic-type aftershock sequence model [29] which produces results mimicking those found in natural EQ catalogs.

10.2.4 Scaling Relations

In 1954 the observation was made that scaling relationship existed between EQ frequency and EQ magnitude. This was known as the Gutenberg–Richter law (G–R law), one of the first fractal laws. The G–R law states that: $\log N = a - bM_s$, where M_s is the surface wave magnitude, N is the annual frequency of EQs, and a and b are regional constants. In general it has been shown that power law distribution pointing to self-similarity or scale invariance holds in EQ magnitude, temporal and spatial patterns. It should be noted that as a rule the strongest EQs follow special distributions (e.g. Gumbel probability density function) for extreme events.

10.2.5 Patterns of Seismicity

Different models have been developed to describe the statistical properties of EQ distributions in time, space and magnitude domains [30]. According to one model, strong EQ events on a given fault line occur quasi-periodically. A second model assumes that EQs occur independently and follow the Poisson distribution (i.e. seismic events are randomly distributed). Recent results indicate that some complex seismic processes have patterns of nonrandom dynamical structuring [31–33]. For example, Matcharashvili et al. [32] showed that the qualitative and quantitative analysis of sequences of EQ magnitude and waiting times in the Caucasus Mountains showed evidence of being low-dimensional and nonlinear. This result was similar to the findings of Goltz [31] for the temporal distribution of Japanese EQs (1997). Then Sobolev [34] reported existence of deterministic chaos in (smoothed) seismic time series. In any case the disagreements are substantial enough to preclude concluding that the real seismic possess possesses dynamical patterns that are strictly deterministic or completely random. Many intermediary possibilities exist.

10.2.6 Earthquake Prediction

Opinions vary on the possibility of predicting EQs from the mainly positive arguments [35–37] to the strongly negative arguments [38]. EQ prediction is a major challenge to Earth sciences and implies assessment of three main characteristics of impending EQ: location; magnitude and time of occurrence. The assessment of prediction success is an object of complex statistical evaluations [39, 40]. The problem seems intractable since even if the first two characteristics (location and magnitude) are more or less well forecasted using geological, historical and instrumental data, prediction of the timing of strong events still evades prediction. Application of modern methods from nonlinear dynamics allow for the revealing new fine structures in geophysical time series which might be useful for finding predictors of strong events [41]. One positive observation is that the waiting times of EQs possess the property of low-dimensionality/memory. On the negative side, prediction of EQ magnitudes should be much more difficult due to the high dimensionality of their distributions [32].

10.2.7 Earthquake “Control”

Experiments show that relatively weak forces, such as reservoir exploitation, mining, water injection in wells, etc., can affect dynamical features of seismic processes. Such effects are sometimes understood as T/S processes or seismic

activity controls [5, 6, 42, 43]. Investigations are still underway, but so far the control of seismicity by weak forcings resides more in the domain of science fiction than science fact. This is not to say that positive outcomes may not be realized in the future [44]. The phenomenon of triggering and synchronization of local seismicity by the wave trains generated from remote earthquakes can be mentioned as an interesting related effect [13, 45]. Such views on possible EQ control are supported by recent discoveries on behavior of complex systems under external influences. Phase and coherence synchronization phenomena are examples of such effects [23, 46, 47]. Indeed as in any complex system, depending on the conditions (parameter settings) weak forcings may exert a control effect as happens with chaotic oscillators driven by a series of weak impacts. In this last case a chaotic attractor containing an infinite number of unstable periodic orbits can be stabilized by a tiny perturbation of system many orders of magnitude less than the main driving force [48, 49]. The control can even be achieved by non-persistent (single pulse) perturbations [50].

10.3 Methods of Revealing Dynamic Patterns in Acoustic/Seismic Time Series

It seems informative to consider the relationship of T/S phenomena to such concepts as ordering, control, self-organization, coupling of systems, randomness and determinism. The phenomenon of triggering is related to the coupling of two dynamical systems in which one is driven by a relatively strong force and the second one by a relatively weak force. Still this small impact introduces significant nonrandom (isolated) action in the former system (e.g. triggered discharge). Synchronization, on the other hand is related to a fundamental property of dynamical systems, specifically their recurrence. Synchronization can be considered as a specific kind of ordering in a series of data, namely, regular recurrence of some characteristic events caused by coupling of two or more systems in master and slave manner. Coupling of even decentralized dynamical systems can lead to self-organization phenomenon (like coherent flashing of firefly swarms), which can be also considered as self-synchronization of oscillators possessing different properties.

There are several kinds of synchronization between oscillating autonomous systems each with a natural frequency (ω_0) and forcing frequency (ω): partial; strong and weak; imperfect (non-locking); generalized (strong coupling with tight functional correlation); complete (nearly identical oscillations in two systems); and phase synchronization (PS). If the amplitudes are irregular and uncorrelated, but the forcing frequency ω and resulting frequency of forced system Ω are adjusted, there is a regular phase shift between ω and Ω . An interesting effect is a high-order or ($n:m$) synchronization, when observed Ω and forcing frequencies ω satisfy equation [23]: $n\omega = m\Omega$, where n and m are integer numbers. The ratio n/m is called the winding or rotation number ρ and is defined as: $\rho = \Omega/\omega$ or in terms of the oscillators' phases: $|n\phi - m\phi_{obs}| < const$, where ϕ is the phase of the forcing and ϕ_{obs} is that of the kicked oscillator.

There are several main approaches to quantify the complexity/synchronization of processes by analysis of measured time series [23, 33, 51]. Some approaches are rooted in dynamical systems theory and fractals and include Lyapunov exponents and fractal dimensions (i.e. correlation dimensions). These methods are based on reconstruction and testing of objects having phase spaces equivalent to the unknown dynamics. The possible calculated measures for quantification of synchronization are: generalized phase difference; mutual information [52]; first Fourier mode of distribution of relative phases; conditional probability of phases [53]; flatness of stripes of synchrograms or stroboscopic technique [54]; and coefficients of phase diffusion [55]. We suppose that analysis of unstable periodic orbits can also be used to reveal synchronization and ordering in coupled time series [44]. The other methods stem from the information theory including Shannon or Tsallis entropy [56], algorithmic complexity [57], etc. It should be stressed that both laboratory stick-slip and EQ source evolution can be considered as the relaxation (integrate and fire type) processes, where the acoustic/seismic signals contain well-defined marked events corresponding to slips/EQs. To deal with relatively short time series new tests have been proposed such as recurrence plots (RP) and recurrence quantitative analysis (RQA). In the case of very restricted statistics (tens of events) less demanding methods, such as Shuster test [58] can be used.

Here we provide a cursory overview of methods. More detailed descriptions can be found in many other sources (e.g. [59–68]).

The term recurrence plot was introduced by Eckmann et al. [60]. The RP construction method was inspired by the fact that nonrandom dynamical systems are characterized by returns to specific state-space locations. Thus recurrence is a fundamental property of any deterministic dynamical system. Recurrence properties also hold for systems which are not exactly deterministic but have nonrandom dynamical structures. Generally, recurrence properties of such systems are stronger if their behavior in the phase space is close to a so-called deterministic dynamical structure (attractor). RPs graphically visualize moments when two trajectories visit the same region in phase space. It is at these points when the system displays recurrences.

According to Eckmann et al. [60] the RP can be mathematically expressed as:

$$R_{i,j}(\varepsilon) = \Theta\left(\varepsilon - \left\|\vec{x}_i - \vec{x}_j\right\|\right) \quad i, j = 1, \dots, N \quad (10.1)$$

where N is the number of measured x_i points. Phase space vectors \vec{x}_i are higher-dimensional reconstructions from measured one-dimensional observations by applying Takens' time-delay method. Each vector of the reconstructed phase-space trajectory is tested for proximity to another vector of the trajectory by testing whether the distance between them is less than a specified threshold ε (radius). Finding of reliable threshold value is not an easy task as has been much discussed in various publications (e.g. [61, 64, 69, 70]). Generally it is reasonable to recognize that the selection of the recurrence threshold is neither straightforward nor optimal. Indeed, radius choice really depends on the particular problem and data set at

hand [70]. It is not necessary to repeat these well-known views about threshold calculation, but for the purposes of this chapter a threshold was typically set to 20 % of the mean of all distances in the distance matrix and l_{min} was set to 2–4 as recommended by others [66, 71].

According to Eq. (10.1), the recurrence matrix contains elements of 0 values (non-recurrent vectors) or 1 values (recurrent vectors). The square recurrence matrix is easily visualized by plotting elements i vs. j as black points if $R_{i,j} = 1$ or white points if $R_{i,j} = 0$. Since i and j are both functions of time, time travels up the ubiquitous line of identity where $R_{i,j} = R_{j,i} = 1$ (always recurrent by self-identity). The main advantage of the RP approach is that it enables the two-dimensional visualization of the recurrence matrix projected from higher-dimensional phase-space trajectories.

As beautiful as RPs appear, the evaluation of the RPs remains a visualization task or qualitative art. For this reason RQA was developed in order to quantify differently appearing RPs based on the small-scale structures contained therein. As a quantitative extension of the RP methodology, RQA can be used as a tool for studying the temporal dynamics of a complex time series. RQA was introduced by Zbilut and Webber [61], Webber and Zbilut [72]. This technique quantified visual features in an $N \times N$ distance matrix recurrence plot and defined five measures of complexity based on diagonal line structures in the recurrence plot: recurrence rate (%RR), determinism (%DET), maximum diagonal line length (DMAX), entropy (ENT) and trend (TND). Three of these variables were used in the present research (%RR, %DET, ENT).

By definition %RR is the ratio of number of all recurrent states (recurrence points) to the number of all possible states and is therefore the recurrence probability of a certain state. There exists a scaling relationship between %RR and the radius which mimicks the correlation dimension [63]. That is why it is crucial to keep the radius parameter low so as not to saturate the RP with recurrence points which are in fact distant from one another. %DET is the ratio of the number of recurrence points forming diagonal structures to the total number of recurrence points in the RP. Specifically, %DET is the proportion of recurrence points forming diagonal structures of a length l_{min} or greater. Periodic systems (e.g. sine waves) inscribe long diagonals and are highly deterministic. Conversely, chaotic signals produce very short diagonals, yet these systems are fully deterministic. Stochastic signals or wildly fluctuating (transient) data produce no diagonals or only short diagonals (only per chance). ENT is the Shannon entropy computed from the histogram distribution of diagonal line lengths of length l_{min} or longer. ENT measures the complexity of the dynamical system under scrutiny.

Over a decade later Marwan extended the RQA approach by introducing three additional measures including, laminarity (%LAM), maximum vertical line length (VMAX) and trapping time (TT). These variables describe characteristics of the vertical line structures in RPs [66]. %LAM is defined as the fraction of recurrent points forming vertical lines with a length of l_0 or greater. This measure is related to the density of laminar states in the system because a vertical/horizontal line marks a length of period in which a state does not change or changes very slowly. Then

TT is defined as the average length of the vertical lines as contrasted with VMAX which registers just the longest vertical line structure. Two of these variables were used in the present research (%LAM, TT).

For interpretation purposes, it is assumed that the quantification of diagonal structures by RQA is able to reveal transitions between ordered and disordered states of a system. Alternatively, RQA measures based on vertical (horizontal) structures are more suitable to identify transitions between different disordered conditions.

Beside the difficulty of setting the radius, at least two additional parameters must be selected before RQA can commence. The delay time is best estimated by submitting the test signal to mutual information (MI) analysis. The embedding dimension is best estimated by the false nearest neighbors (FNN) approach.

To conclude this short overview we note that RQA does not require that the input data (time series) be of any particular distribution, be stationary, be noise free, or be particularly long. In fact, input data can contain as few as 30 points. These characteristics of RQA make it a powerful tool when dealing with real-world data sets [63, 66].

10.4 Triggering and Synchronization of Stick-Slip: Laboratory Model of Seismic Process

In this review the results of both active and passive experiments on T/S are presented. In the case of active (laboratory) experiments on the electromagnetic (EM) and mechanical triggering and synchronization of mechanical instability (slip) in the spring-slider system, the main variables of the stick-slip (which is a small-scale model of seismic process) can be controlled. The forcing of local seismicity by periodic reservoir loading, strong electromagnetic pulses and wave trains from remote earthquakes can be considered as semi-active experiments where only some of process-determining variables are known. Lastly, the analysis of seismic catalogs for revealing hidden periodicities is a passive approach in which key characteristics of the process are mostly unknown.

Many publications over the last decades have addressed EQ triggering (especially dynamic triggering), but the problem is far from being resolved due to the complexity of natural seismic processes [3, 11, 13, 19, 27, 73]. We believe that a reasonable understanding of the T/S phenomena can best be obtained from controlled experiments carried out on model systems [24, 74–78]. It is generally accepted that the most relevant model of the seismic process is unstable frictional movement or stick-slip [79–81]. This is a special case of the relaxation, integrate and fire model. In our stick-slip experiments, acoustic emissions (AE) signaling slip events and electromechanical (EM) or mechanical forcings were recorded on two separate channels of a sound card (Figs. 10.1, 10.2, 10.3, 10.8a, 10.9a, and 10.10).

Laboratory experiments were carried out on the slider-spring system with superimposed pulse or periodic electromagnetic (EM) or mechanical forcing both of which are weak compared with the main dragging force of the spring. The

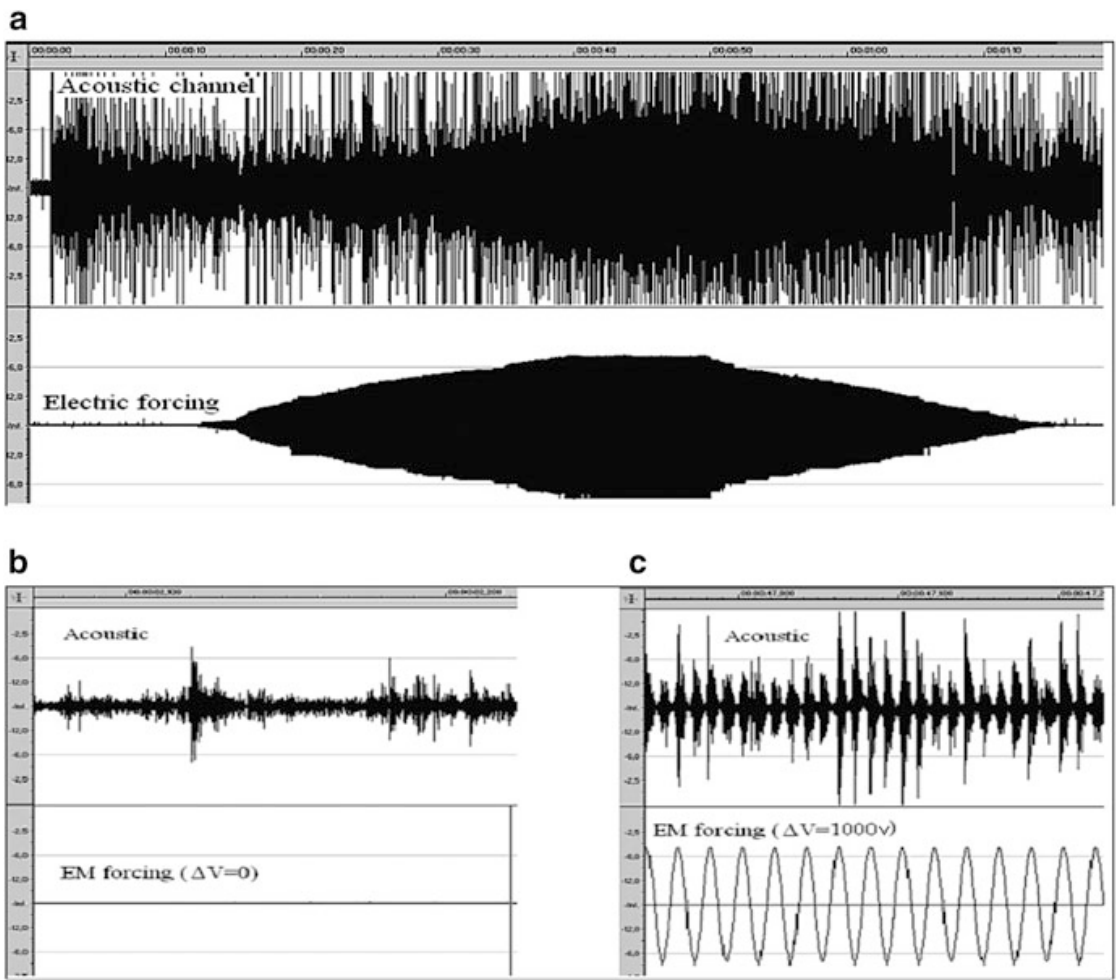


Fig. 10.1 (a) Acoustic emission during slip under variable from zero to 1,000 V external periodical voltage. (b) The extended part of record with zero EM forcing. (c) The extended middle part of record under maximal EM forcing with complete phase synchronization (PS)

details of experimental setups can be found elsewhere [83, 84]. Relative forcing (RF) or coupling strength was calculated as a ratio of the forcing intensity to the main driving force. In our laboratory experiments RF was of the order of 10^{-2} with EM forcings and of the order of 10^{-3} – 10^{-5} with mechanical forcings. For selecting AE signals from the noisy records we use an approach modified from Akaike Information Criterion (AIC), which gives good coincidence with manual selections.

10.4.1 Electromagnetic Triggering of Slip

The first set of triggering experiments was prompted by experiments carried out from 1983 to 1988 by the Institute of High Temperatures (Russia) at the Bishkek test area. After performing a series of MHD soundings as well as “cold” discharges,

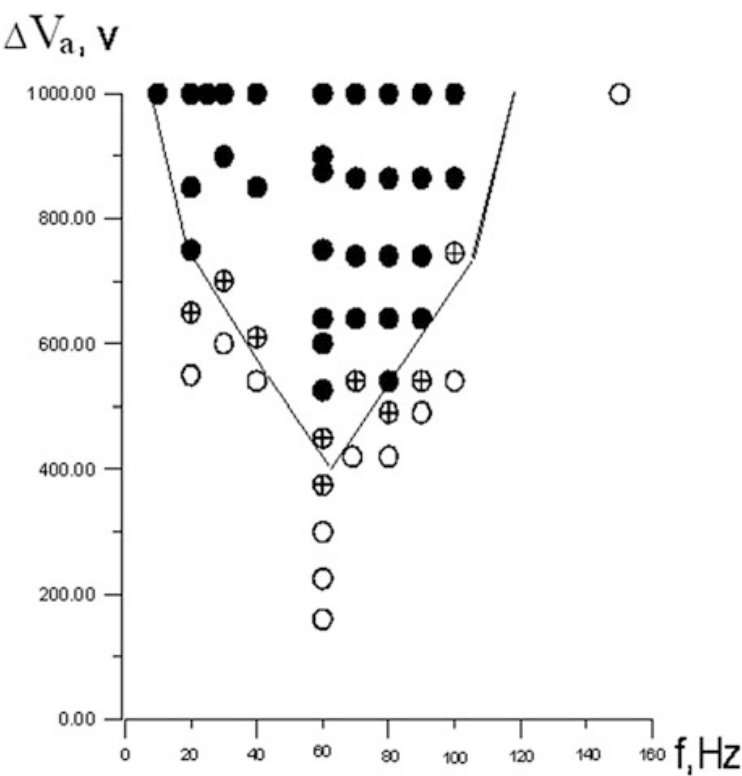


Fig. 10.2 Stick-slip synchronization area (Arnold's tongue) for various intensities (V_a) and frequencies (f) of the external EM forcing. *Filled circles*: perfect synchronization; *circles with crosses*: intermittent synchronization; and *empty circles*: absence of synchronization [82]

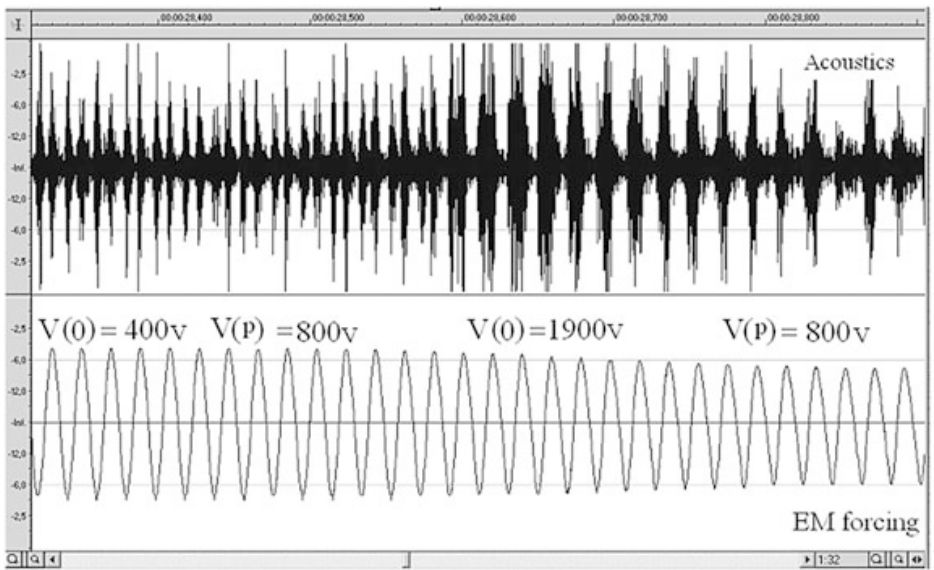


Fig. 10.3 Transition (bifurcation) in stick-slip from 1:2 synchronization (period doubling) to 1:1 synchronization at simultaneous action of direct $V(0)$ and periodic $V(p)$ voltages; transition occur at $V(0) > V(p)$ [82]

an unexpected effect of micro-seismicity activation by strong EM pulses was discovered [9, 10]. We modeled this effect using a rock block placed on the inclined fixed block at the angle less than the critical repose angle. At this angle the system was stable (non-moving) for days. Application of strong electromagnetic (EM) pulses initiated (with some probability) mechanical instability (slip) provided the electrostriction force parallel to the slip surface. On the contrary, slip was hampered if the EM field was applied normal to the surface. The EM field contribution F_{pi} to the friction force F_f can be written as:

$$F_f = \mu (F_n + F_{pi}), \quad (10.2)$$

where μ is friction coefficient, F_n is normal force including pore pressure contribution and F_{pi} is electrostriction force acting in the gap between blocks' contacting surfaces. Using known expression for electrostriction force [85] we obtain:

$$\vec{F}_{pi} = \pm \frac{\varepsilon_{eff} S}{8\pi} \left(\frac{\Delta V_{eff}}{d_{eff}} \right)^2 \vec{n}, \quad (10.3)$$

where $\Delta V_{eff}/d_{eff}$ is the effective voltage gradient in the gap, ε_{eff} is the effective dielectric constant of the gap and S is the area of the slip surface. The positive sign (+) is applied when the EM field \vec{E} is parallel to the external normal \vec{n} to the surface and the negative sign (−) when the field \vec{E} is normal to \vec{n} [77].

It is remarkable that the expression (10.2) is similar to Byerlee's friction law which takes into account fluid pore pressure contribution to the friction force [86]. A significant difference is that additional EM force can be positive or negative.

10.4.2 Electromagnetic Stick-Slip Synchronization

Synchronization experiments with electromagnetic (EM) forcings were carried out on the system consisting of two (horizontal) plates of roughly finished basalt. A constant pulling force F_p of the order of 10 N was applied to the upper (sliding) plate. In addition, the same plate was subjected to periodic electric forcing of variable frequency and amplitude (from 0 to 1,000 V) which was mechanically much weaker (of order of 1 N) compared to the pulling force. The electric field was normal to the sliding plane. Acoustic bursts accompanying slip events were registered by the PC sound card. Details of the setup and technique are previously published [84].

Acoustic emissions during slip with superimposed periodic low-frequency EM field ($f \approx 60$ Hz) of varying intensity, oriented normally to the slip surface are presented in Fig. 10.1 [82]. The significant synchronization at this frequency occurs at $V_a \geq 500$ V. At such conditions excitation the AE events (micro slips) is synchronized with EM field and occur twice per period (Fig. 10.1C).

Synchronization was observed only at some definite values of set of parameters (spring stiffness K_s , frequency f and intensity V_a of forcing). The “phase diagram” for variables f and V_a or so-called Arnold’s tongue [23] is presented in the Fig. 10.2.

It should be noted that the phenomenon of synchronization was observed only with the EM field directed normally to the slip surface. When the EM field was applied roughly parallel to the slip plane, synchronization was absent. We conclude that the synchronization is related to “electromagnetic braking” of slip at large enough values of (normal) sinusoidal EM forcing (large positive EM field contribution F_{pi} in Eq. 10.2) and a sudden slip after accumulation of critical stress provided by spring pull (as was noted in the previous section, the effect of EM “braking” was discovered in slip experiments on the inclined slope).

Also observed is a transition (bifurcation) in stick-slip from 1:2 or period doubling synchronization when two slip events occur per period of EM forcing to 1:1 synchronization when one slip event occur per period of EM forcing (Fig. 10.3). At simultaneous action of direct $V(0)$ and periodic $V(p)$ voltages the transition occur at $V(0) > V(p)$. A simple explanation of the period doubling effect follows from the principle of electrostriction in which direct and periodic voltages are simultaneously applied to a dielectric material [77].

10.4.3 Measuring EM Synchronization Strength of Stick-Slip

Several quantitative analysis tools were applied to the recordings of stick-slip processes where the superimposed periodic EM field intensity was raised monotonously from zero to 1,000 V and then decreased back to zero (Fig. 10.1a). In order to assess synchronization in a quantitative manner various tools of nonlinear dynamics (synchronization) theory described in [18, 23, 64, 87] were applied to experimental data (Fig. 10.1a) obtained under variable intensity of forcing. Specifically, we calculated: phase difference $\Delta\Phi$ between acoustic emission bursts and external sinusoidal forcings; phase diffusion coefficient D of phase differences $\Delta\Phi$; mutual information (MI) between phase of forcings and AE bursts and recurrence plots (RP) and RQA %DET measures of stick-slip generated AE time series. The results are shown in Figs. 10.4, 10.5, 10.6, and 10.7. All these methods confirm strong synchronization of forcing and AE at maximal EM forcing.

10.4.3.1 High Order EM Synchronization of Stick-Slip

High-order or $(n:m)$ synchronization means that observed T_{obs} and forcing T_f periods of oscillators satisfy equation [23]: $nT_{obs} = mT_f$, where n and m are some integer numbers. The ratio n/m is called the winding or rotation number ρ and is defined as:

$$\rho = T_f / T_{obs}. \quad (10.4)$$

Fig. 10.4 Phase difference $\Delta\Phi$ between the sequence of maximums of acoustic emissions' bursts (AE catalogue) and external sinusoidal EM forcing versus time for a whole record, Fig. 10.1a. Note that $\Delta\Phi$ remains practically constant at forcings stronger than the threshold value of V

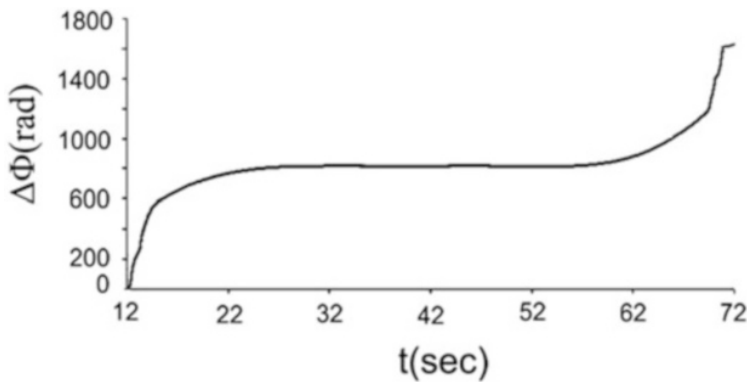


Fig. 10.5 Variation of phase diffusion coefficient D of phase differences $\Delta\Phi$ versus time, calculated for consecutive sliding windows, containing 500 events for data presented in Fig. 10.1a—for a whole record. Note that D remains rather constant near zero at the forcings stronger than the threshold value of V

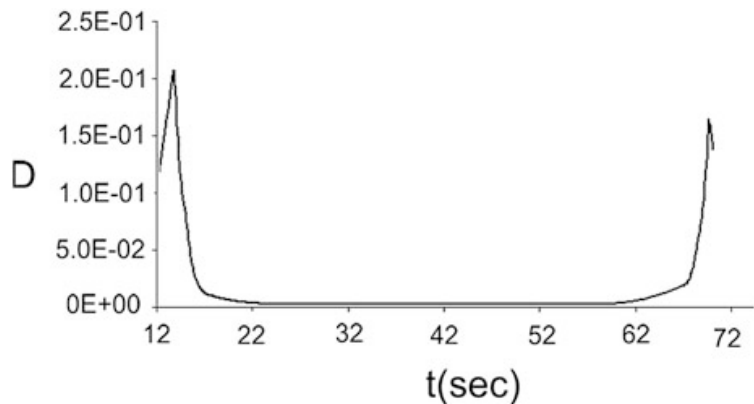
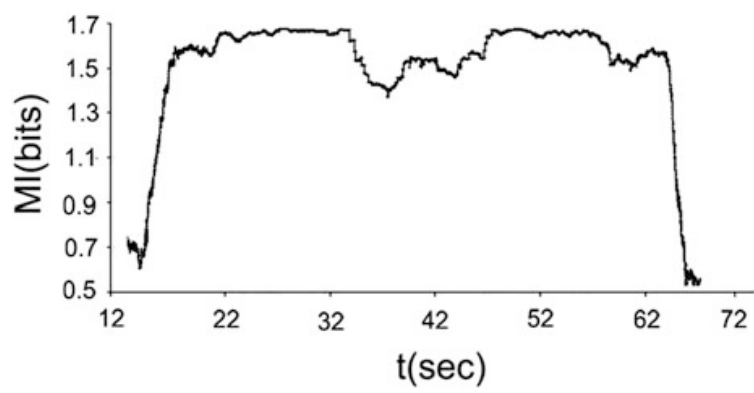


Fig. 10.6 Variation of mutual information between the phase of forcing and AE bursts versus time calculated for consecutive sliding windows, containing 500 events (for a whole record presented in Fig. 10.1a). *Note:* large values of MI in the same range of t , where $\Delta\Phi$ is constant and D is very small



This condition of high-order synchronization can also be presented as a ratio $n:m$; in this case the winding number is: $\rho = n/m$.

The experiments [77, 89] show that the addition of a strong enough component of the constant electric field to the high frequency EM forcing signal (40 Hz) invokes transition from 1:2 to 1:1 synchronization. Figure 10.8 represents an example of high order ($n/m = 1:2$) synchronization as well as the high-order synchronization (HOS) of stick-slip at EM forcing by alternating pulses of different duration.

Both the onsets of AE swarms and those of the individual events within swarm sequences turn out to be very well organized which is evident from the synchrograms (Fig. 10.8b, c). The synchrogram (stroboscopic plot) implies a fixing of the phase of forcing EM signal at the moments of spiking of the acoustic pulses [23]. If the moments of n th spiking in the swarm ($n = 1, 2, 3$, etc.) always occur at

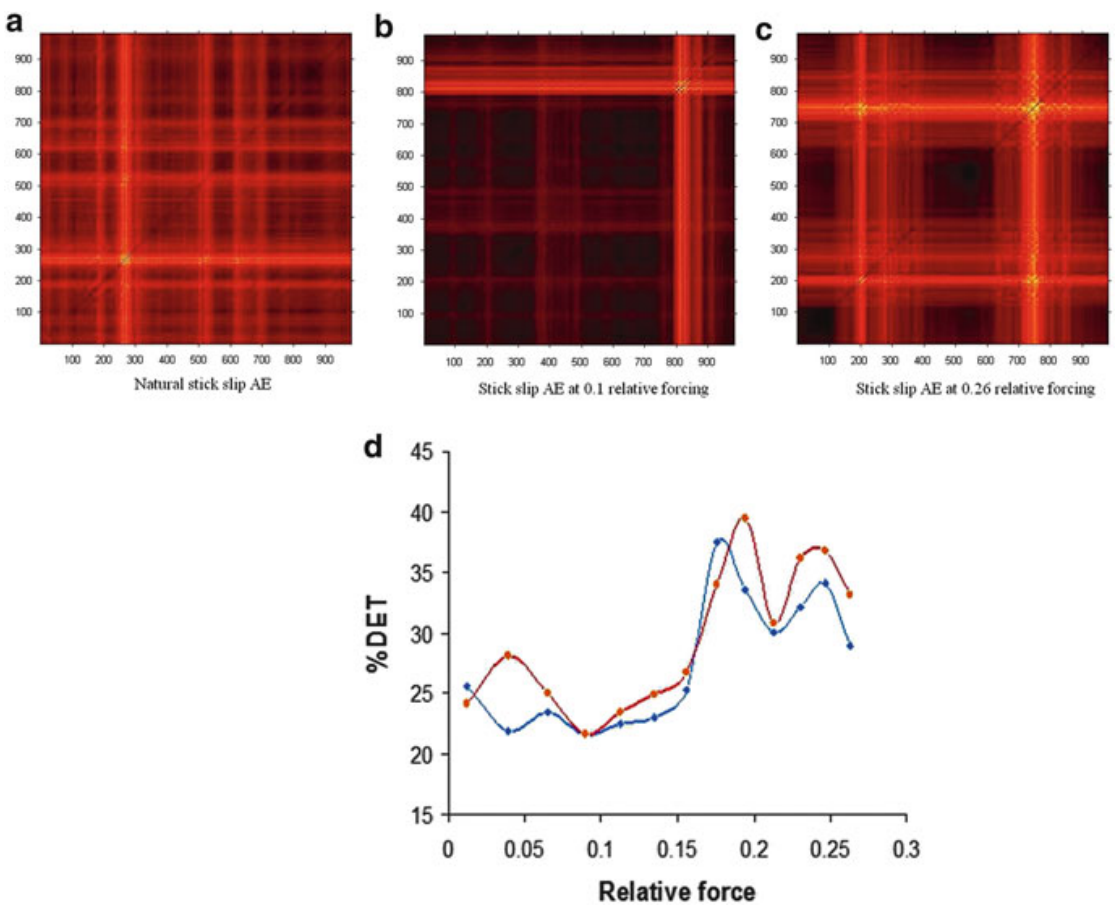


Fig. 10.7 Recurrence plots compiled using VRA program [88]; (a) RP of natural stick-slip. (b) RP of stick-slip at the relative forcing $RF = 0.1$. (c) RP of stick-slip at the relative forcing $RF = 0.26$; (d) %DET versus relative force (RF), calculated for onsets (red) and maximums (blue) of AE bursts for the first half of the record presented in Fig. 10.1a

the definite phase of the forcing, then both oscillators are synchronized and the plot of delays of n th spike relative to the onset of the forcing signal versus time (here the number of cycle for a periodic forcing) reveals n stripes parallel to the time axis. In Fig. 10.8b the synchrogram for the onsets of swarms of AE bursts and sequential AE signals within swarms generated during forcing by short EM pulses corresponds to the winding number $\rho = n:m = 1/3$. In the case of longer forcing pulses, $\rho = 1/4$ or $1/5$.

Figure 10.9a presents even more populated AE swarms generated by longer EM pulses. In this case the swarms contain up to 40 AE events, so $\rho = n:m \approx 1/40$. Again, the AE bursts within the swarm are surprisingly well organized. The corresponding synchrogram (Fig. 10.9b) shows that the first ten AE bursts show almost constant phase shift relative to the EM forcing onsets. The following bursts demonstrate regular small increase of the phase delay in sequential swarms [82]. It is striking how Fig. 10.9a, b are similar to synchronization pattern of neuron spiking under LF external electrical forcing ([23], Fig. 6.7). This constitutes a vivid illustration of universality of synchronization phenomenon.

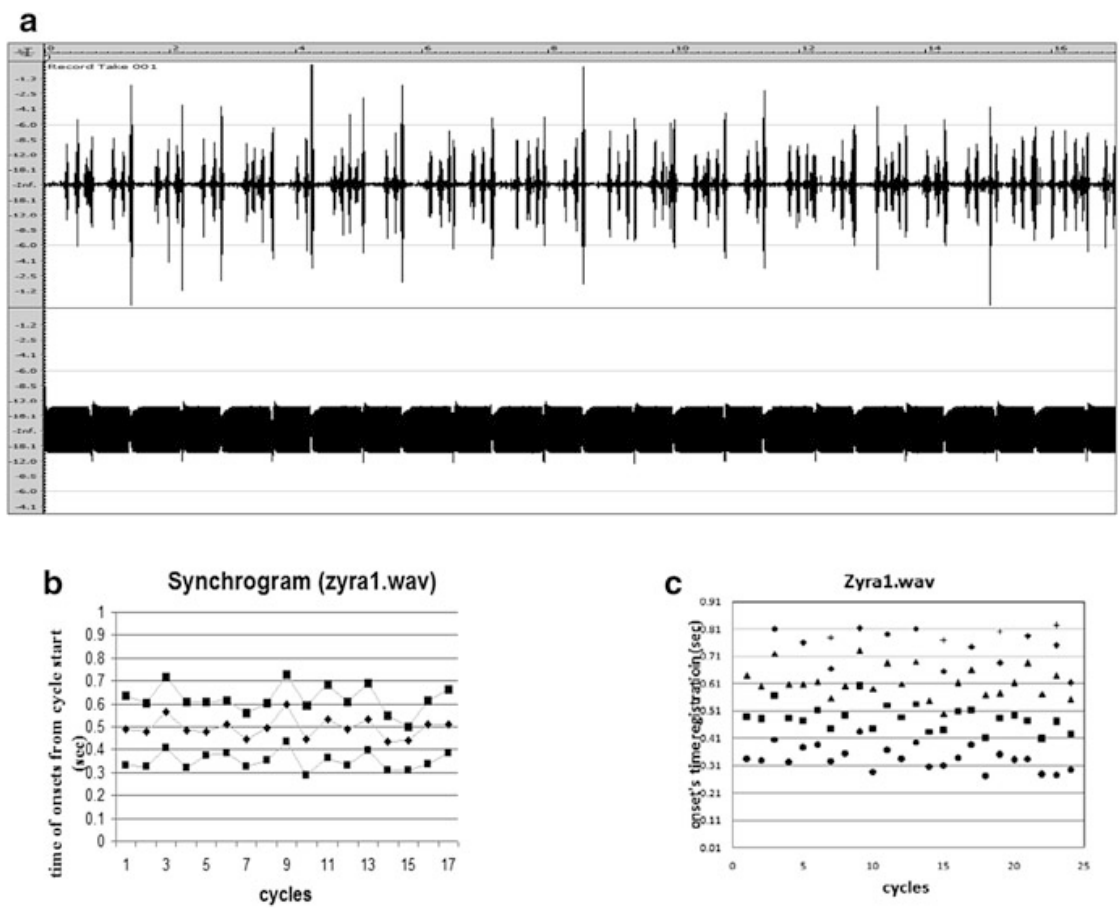


Fig. 10.8 (a) High order synchronization: $n < m$ (exp. Zyra1). Note that the onsets of slip swarms and these of the individual events within sequences (swarms) turn out to be very well organized (see text). (b) Synchrogram for the onsets of swarms of AE bursts and AE signals within swarms generated during forcing by short EM pulses versus number of forcing pulses (cycles), here $n:m = 1/3$. (c) A similar synchrogram for longer forcing pulses where $n:m = 1/4$ or $1/5$. *Note:* in this figure and Fig. 10.9 the LF forcing pulses are visualized by high-frequency (HF) modulation as the sound card does not respond to the direct current; HF is applied only to the computer channel (not to the sample)

10.4.4 Dynamic Mechanical Synchronization of Stick-Slip

Synchronization of the spring-slider system under weak periodic mechanical forcing has been investigated [77, 78, 90]. These experiments can be considered as laboratory modeling of the recently discovered effect of dynamical triggering/synchronization of local seismicity by the wave trains of strong remote EQs (Sect. 10.5.2). Two modes of mechanical forcing experiments were tested: (1) the forcing normal to the slip surface; and (2) the forcing parallel to the slip surface. For brevity we will refer to them as normal and tangential forcing accordingly. We calculated the maximum value of mechanical forcing which corresponds to the maximum measured voltage applied to the mechanical vibrator (i.e. when the voltage applied to the vibrator equals 6.5 V). The mass of the oscillating

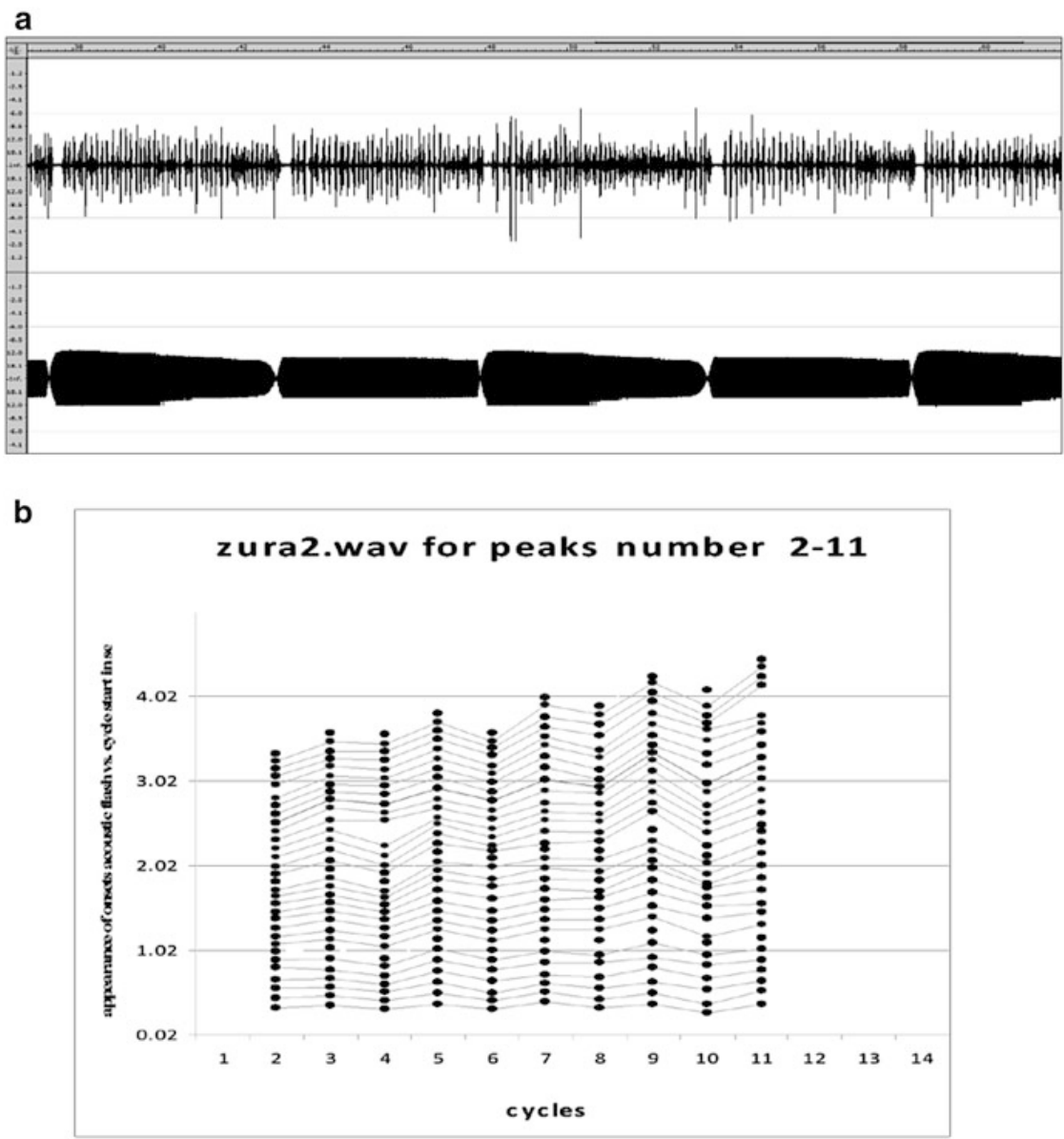


Fig. 10.9 High order synchronization: $n:m \approx 1/40$ (exp. Zura2). **(a)** Swarms of AE pulses (slips) generated during forcing by very long EM pulses. **(b)** Stroboscopic diagram (synchrogram) for the first 31 AE bursts in the swarms generated by 11 forcing EM pulses; note the stripe structure of synchrogram, which shows that the phase shift between onsets of electromagnetic (EM) forcing and acoustic emission (AE) pulses is almost constant for the first ten slips in the swarm [82]

element of the vibrator m is ≈ 20 g, so we obtain for the natural frequency f of the vibrator's oscillating element: $f = \sqrt{k/m} = 5$ Hz where k is the stiffness of the vibrator spring. Substituting into expression for maximal force $F_{max} = k x_{max}$ values of $k = 25$ and $m = 0.5$ N/m and that of the maximum deflection x_{max} of the oscillating element at the applied voltage 6.5 V (here $x_{max} \approx 4 \times 10^{-3}$ m), we obtain for the corresponding (maximal) intensity of mechanical forcing F_{max} : $F_{max} = k x_{max} \approx 2 \times 10^{-3}$ N.

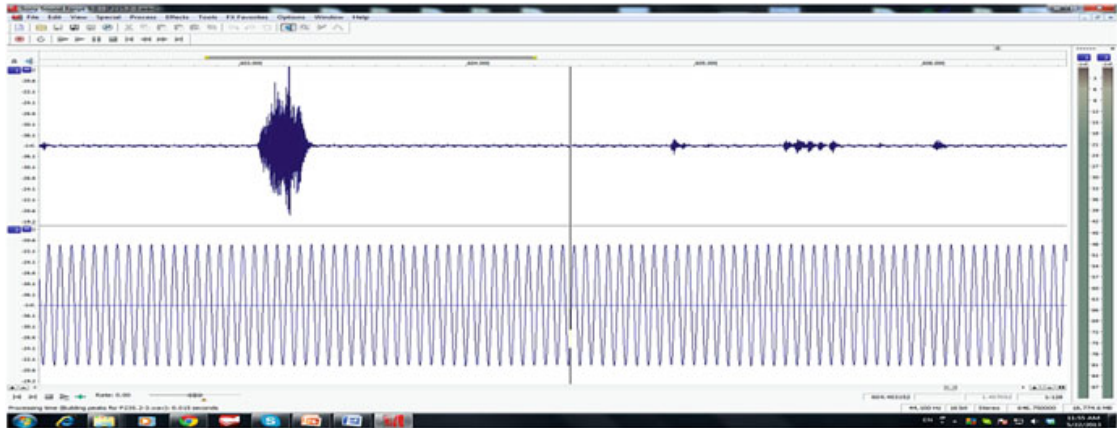


Fig. 10.10 Typical record of AE (*upper channel*) and simultaneous tangential mechanical forcing (*lower channel*). This example corresponds to vibration intensity generated by application of 5 V voltage to the vibrator. The *arrow head* demark the onset of slip (AE pulse). In the right part of record AE signals arise spontaneously without impulsive onsets and are analogs of non-volcanic tremors (see Sect. 10.5.2)

At lesser voltages forcing is much weaker: our assessment for 1 V intensity of vibrator input is approximately 5×10^{-4} N. Thus the forcing was always much less than the driving force (spring) $F = 4$ N.

The typical recording of AE with mechanical forcing is shown in Fig. 10.10 and distributions of acoustic emission onsets relative to phases of forcing period T_f at various conditions is shown in Fig. 10.11a, b. Note that at mechanical forcing we have the high order synchronization (HOS) with winding number < 1 (0.01–0.005).

Voltage (and correspondingly, the intensity of mechanical forcing) increase the density of the AE bursts population in certain parts of the forcing period, namely in the first and the last decimals of forcing phase. In between lies a gap or forbidden zone. We suppose that increasing the voltage applied to the vibrator promotes synchronization of AE offsets with external forcing. In case of normal forcing AE activity is maximal at force phases corresponding to minimal oscillating normal stress (phases $0-72^\circ$ and $288-360^\circ$). Slips are absent in the area of maximal normal stress σ_n (phases $102-252^\circ$). A similar pattern is observed for the tangential forcing (Fig. 10.11b, left columns). Such distributions can be expected from the general expression of friction law: $\tau = c + \mu\sigma_n$ in the case of normal forcing as the increase of the normal stress suppresses slip. The slips distribution in the case of tangential forcing reveals the same pattern as at normal forcing. This seems to contradict to the above friction law. The only explanation of this phenomenon is that the real contacting (rubbing) surfaces are not flat. Protuberances of one surface are often sunk into hollows of the opposite surface. That means that both kinds of forcing increase resistance to slip at certain phases of forcing: (1) in the case of strong normal forcing the tip of protuberance is pressed onto bottom of the trough which prevents slip; (2) in the case of strong tangential forcing the edge of the cusp is pressed against the edge of the through and this again prevents slip. Consequently, the slip can be suppressed at the same phases of both kinds of intensive forcing and the optimal condition for slipping are realized in complementary phases of forcing.

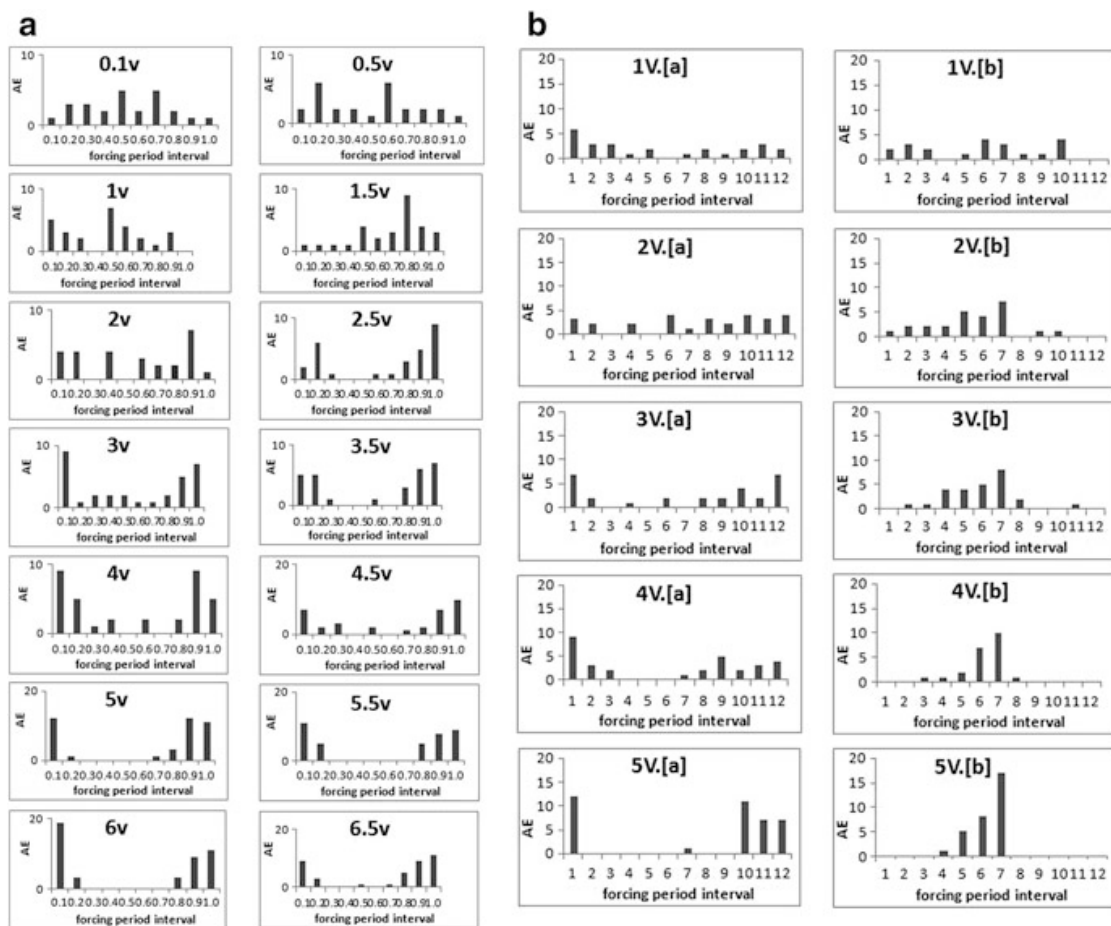


Fig. 10.11 (a) Distribution of acoustic emission onsets relative to phases of forcing period T_f (in decimals of T_f) for different intensities of normal forcing (forcing frequency $f = 20$ Hz). (b) Distribution of AE onsets (the *left column*) and terminations (the *right column*) relative to the (mechanical) forcing phase (in twelfths of the forcing period) for different intensities of tangential forcing (forcing frequency of 80 Hz). Clearly, AE events can be synchronized both by the onsets and offsets of the mechanical forcings

Additional arguments on similarity of stick-slip patterns for both kinds of forcing can be deduced from the recent results on the physical mechanism of the stick-slip. According to [91], the slip begins only after rising of the upper (sliding) block over a fixed one to a some height. This means that resistance to both normal and tangential forcings decreases as the area of contacting cusps and troughs decreases. We presume that in the case of periodical forcing this optimal configuration of real rubbing surfaces are observed at certain phases of forcing which are the same for both normal and tangential impact. The similar arguments can be applied to the slip termination distribution (Fig. 10.11b, right columns). Here the phases of optimal termination forcing fill the gap of forbidden phases for onsets (Fig. 10.11b, left columns). In other words termination is promoted at phases forbidden for onsets which seems logical as conditions for terminations and onsets should be mutually exclusive.

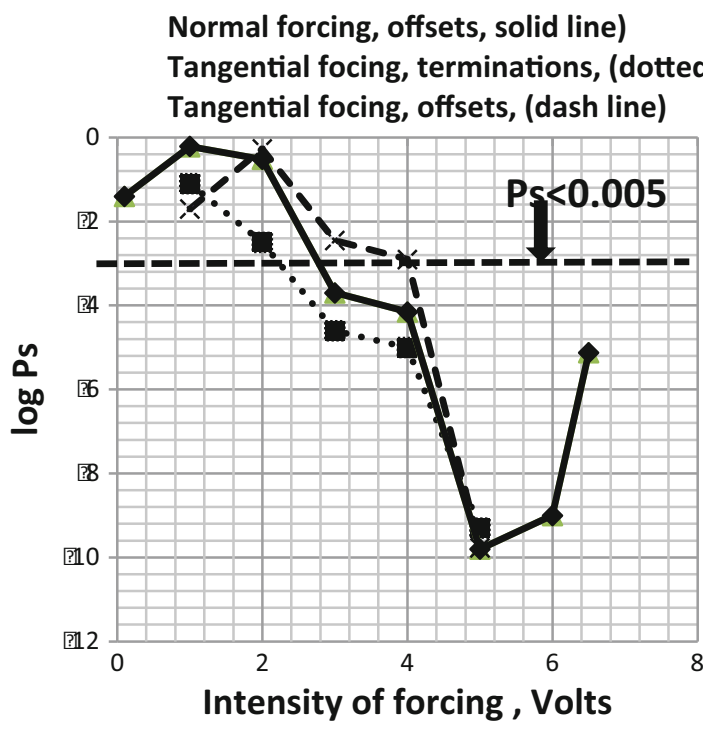


Fig. 10.12 Shuster probability P_s versus mechanical forcing intensity for experimental data (*solid line*—normal forcing offsets, $f = 20$ Hz; *dash line*—tangential forcing offsets, $f = 80$ Hz; *dotted line*—tangential forcing terminations, $f = 80$ Hz). The threshold probability P_s , for which the slips and the forcing are strongly correlated ($P_s < 0.005$) is marked by *horizontal line with arrow*; this means that the probability of random distribution of slips in the area below the line is less than 0.005. Note that at forcing $F \approx 2 \times 10^{-3} N$ (corresponding to the voltage > 4 V) the probability of random distribution of slips is of the order of 10^{-10}

A modified Shuster test [58, 75] was applied to experimental data (namely, to the data for normal forcing offsets, $f = 20$ Hz, tangential forcing offsets, $f = 80$ Hz and tangential forcing terminations, $f = 80$ Hz). Significant “Shuster” probability P_s values indicate that the slips do not occur randomly with respect to the higher forcing intensities (Fig. 10.12). Generally $P_s < 0.05$ is accepted as a reliable threshold, though for confidence we accepted the significance level $P_s < 0.005$; this threshold is shown in Fig. 10.12. Note that at forcing of the order of $10^{-3} N$ (voltage > 4 V) the probability of random distribution of slips is of the order of 10^{-10} . The remarkable detail is also the close similarity of synchronization trends for different forcing modes as well as for both offsets and terminations.

It was a surprise to discover that not only AE onsets can be synchronized by a weak mechanical forcing, but so too can AE offsets at the same coupling strength. This is shown in Fig. 10.11 (right columns) and Fig. 10.12. It is striking that the AE bursts are so well synchronized with mechanical forcing though the waiting interval of AE bursts varies between 100 and 200 periods of forcing. That is, to initiate a single phase-synchronized slip event the forcing oscillator needs to generate enough energy delivered in of hundreds of oscillations to the slider-spring system.

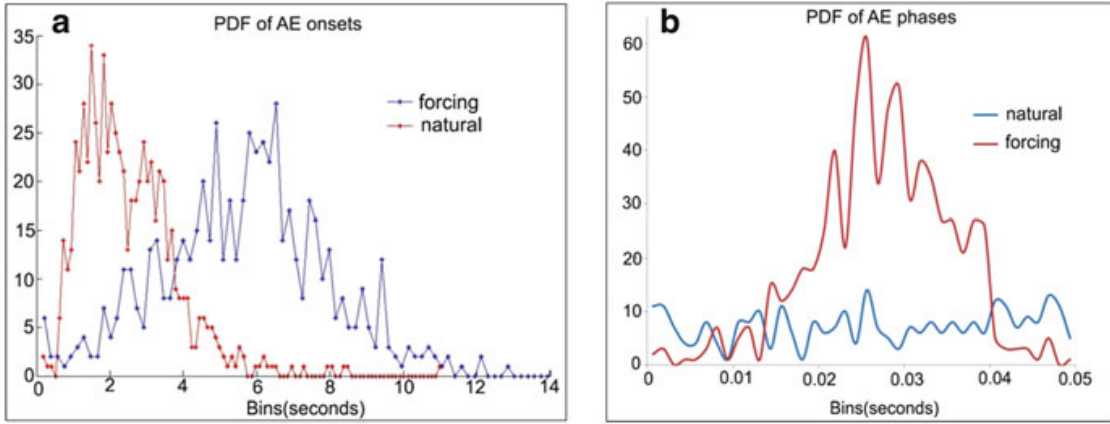


Fig. 10.13 (a) AE waiting time distributions for natural (red line) and forced (blue line) stick-slip in seconds. (b) Distribution of AE phase shifts $\Delta\phi$ relative to the forcing sinusoid phase for natural (blue line) and forced (red line) stick-slip in seconds

AE waiting time (ΔT) distributions for spring stiffness of 235 N and velocity of 0.7 mm/s is shown in Fig. 10.13. The distributions are similar for natural and forced stick-slip except that the mean ΔT is 3.03 and 2.4 s respectively (Fig. 10.13a). This means that forcing shortens waiting times, which is expected from earlier work [90]. In contrast, distributions of phase shifts $\Delta\phi$ between AE and maxima of periodic forcing are quite different for natural and stick-slip forced at 20 Hz (Fig. 10.13b). Of course for natural stick-slip the onsets of AE bursts were compared to maxima of virtual sinusoid with the frequency 20 Hz. There is no preferred $\Delta\phi$ value for the natural stick-slip case, however there is a strong synchronization peak in the $\Delta\phi$ distribution of the forced stick-slip case.

Next we present RPs of series of time intervals between consecutive AE bursts (Fig. 10.14). All RPs in this chapter are plotted using Visual Recurrence Analysis software [88]. These waiting time sequences can be considered as values indicating accumulation of strain energy which is finally released as AEs. Thus within the context of our experimental model, these time interval series can be regarded as proxies to the sequences of “seismic” energy released in consecutive events (seismic rate). In this case we applied a data smoothing procedure similar to what was done for the real seismic processes [34]. Smoothing acts as a low-pass filter and reduces the variability of inter-event times.

We now address RP analysis of AE waiting times and phase shifts $\Delta\phi$ between AEs for natural and forced stick-slip processes as described above. The appropriate delay time and phase space dimension were chosen using MI and FNN approaches respectively. The forcing function was a sinusoidal wave (20 Hz) with a vibration intensity of 4 V. The RPs of AE are shown before (Fig. 10.14a, b) and after (Fig. 10.14c and d) the application of a Savitzky–Golay filter to the AE waiting time (ΔT) data. Although visual inspection hints that the forcing may alter the RP structure (Fig. 10.14a, b), RQA of the short time series showed no difference between the natural and forced stick-slip processes. Smoothing of the original ΔT and $\Delta\phi$ data reveals stronger differences in the two conditions (Fig. 10.14c and d).

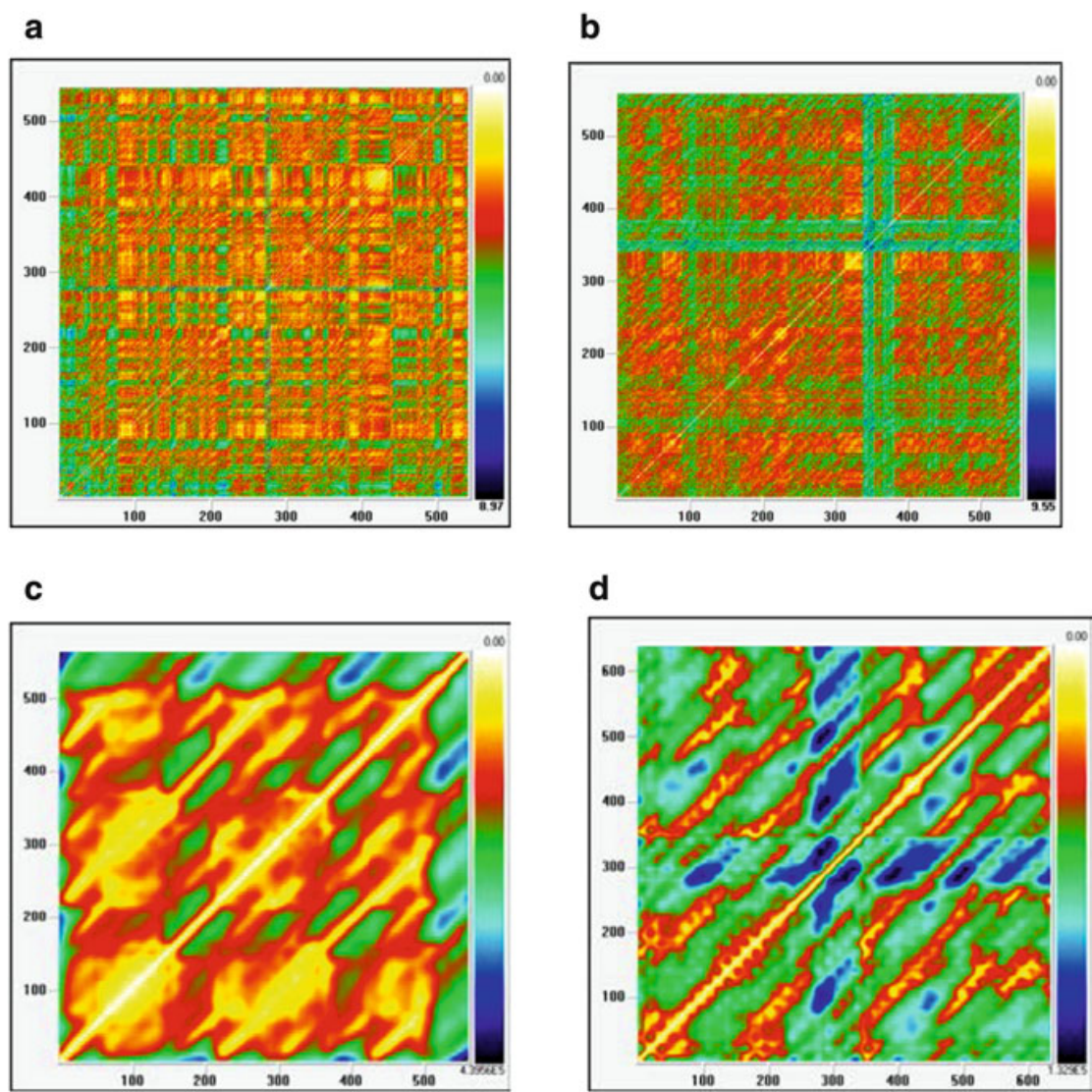


Fig. 10.14 (a) Visual recurrence analysis (VRA) plots of waiting times sequences of acoustic emissions (AE) generated during natural stick-slip (laboratory model of natural seismic process) from original, non-smoothed data. (b) VRA plots of waiting times sequences of AE generated during stick-slip with forcing (laboratory model of forced seismic process). (c) VRA plots for waiting times sequences of AE generated during natural stick-slip smoothed by Savitzky–Golay filter. (d) VRA waiting times sequences of AE generated during stick-slip with forcing smoothed by Savitzky–Golay filter

In any case, it should be noted that in these forced experiments there was high order synchronization with a winding number $n:m \approx 0.2$.

Regarding the phase shifts, RPs of $\Delta\varphi$ are shown before (Fig. 10.15a, b) and after (Fig. 10.15c and d) the application of a Savitzky–Golay filter to the phase shift $\Delta\varphi$ data. The phase delay RPs appear to be more ordered (phase synchronized) under external forcing and this was verified by RQA. For example, compared with the natural stick-slip state, forcings caused increases in both %DET (9 % natural to 28 % forced) and %LAM (12 % natural to 40 % forced) values. Smoothing the data

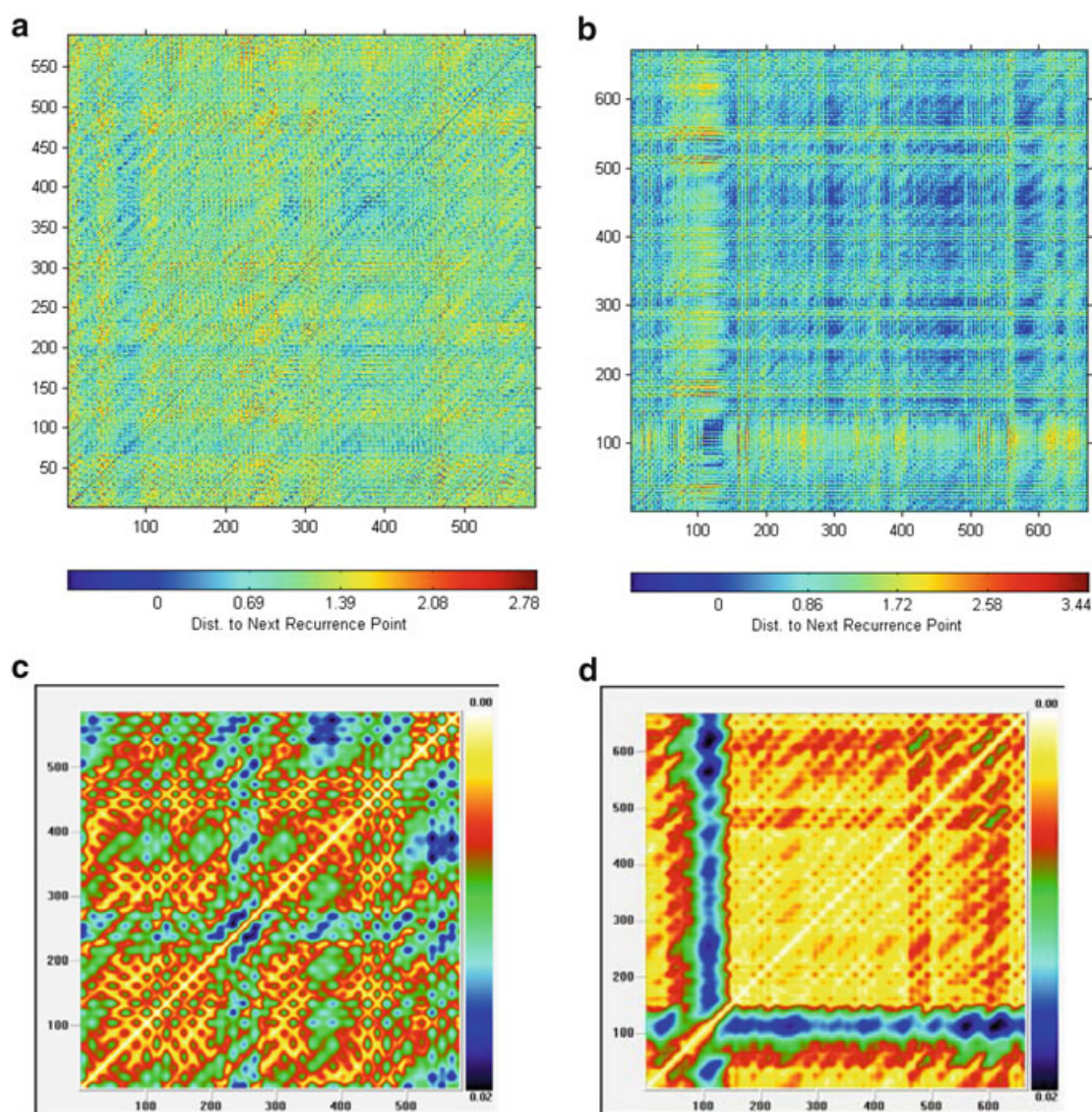


Fig. 10.15 Recurrence plots of time series of phase shift values $\Delta\phi$. **(a)** RP of phase shifts between onsets of acoustic emission (AE) wave trains generated during natural stick-slip process and maximums of virtual 20 Hz sinusoid. There are no clear recurrent structures. **(b)** RP of phase shifts between onsets of acoustic emission (AE) wave trains generated during forced stick-slip process and maximums of imposed external periodic forcing (20 Hz). Recurrent structures are more obvious. **(c)** RP of smoothed by Savitzky–Golay filter time series of phase shifts between onsets of AE bursts during natural stick-slip process and maximums of virtual sinusoid. **(d)** RP of smoothed by Savitzky–Golay filter time series of phase shifts between onsets of AE bursts and maximums of external periodic forcing (20 Hz) during forced stick-slip process

gave similar results with increases in %DET (67 % natural to 84 % forced), but not %LAM (60 % natural to 62 % forced).

RP plots for original non-smoothed data (Fig. 10.14a, b) contain horizontal and vertical lines/clusters, which means that during both natural and driven stick-slip some states are “laminar.” That is they change slowly or do not change at all during some (integrate, trapping) time interval. These laminar states of AE waiting times

can be compared to the periods of time the upper plate was stuck until the strain from the spring became equal to the friction force.

One important conclusion arises from the HOS data above (Sect. 10.4.4) as pertains to the much discussed interaction between tidal deformation and EQs [11, 74, 75]. That is, in order to uncover tidal effects we should look not only for direct 1:1 correlations between periods and forcings, i.e. for increases/decreases of seismicity exactly at the tidal periods (e.g. 12 h etc.). For EQs, high order phase synchronization can occur at multiples of tidal periods as was shown in laboratory forced stick-slip experiments. It is important to note that phase synchronization of AE or seismic events can be observed irrelative to the duration of event nucleation. Of course, the optimal condition of synchronization (minimal forcing) should correspond to the forcing period which is close to the natural event nucleation time. Additional complication arises from the phenomenon of delay. The phase of response can be shifted quite significantly from the phase of forcing if the last one is weak [82]. Thus the question of tidal forcing of EQs should be reconsidered taking into account these new experimental and theoretical evidences.

10.5 Dynamical Patterns in Seismicity

Analysis of the temporal features of EQ occurrences represents a focus of intensive interdisciplinary research. In this section we study the features of temporal variations of seismicity in the Caucasus Mountains, in general the T/S effects of remote strong earthquakes. We also examine the temporal aspects of seismicity in the areas where natural seismic processes are forced by man-made impacts. The test sites are the high Enguri dam (Georgia) and the magnetohydrodynamic laboratories in Central Asia.

10.5.1 *Recurrence Patterns in Seismic Catalogs*

Matcharashvili et al. [32] presented results of nonlinear analysis of magnitude and waiting time sequences for EQs in the Caucasian region using correlation dimension and algorithmic complexity analyses. It was shown that EQ waiting time series are much more ordered than magnitude sequences. In the present research we used RQA to investigate the complexity of catalogued EQs. Based on the Caucasian EQ catalogue spanning from 1960 to 2011, we investigated the variations in temporal distribution of earthquakes. Daily and monthly occurrences of EQ data sets have been derived both from the original catalogues as well as from declustered catalogues according to the Reasenber algorithm [92, 93] applying a magnitude threshold $M \geq 3.0$. For example, daily and monthly EQ occurrence frequencies are presented for the declustered Caucasian catalogue in Fig. 10.16. From this catalogue we calculated the number of EQ occurrences over consecutive days and months and

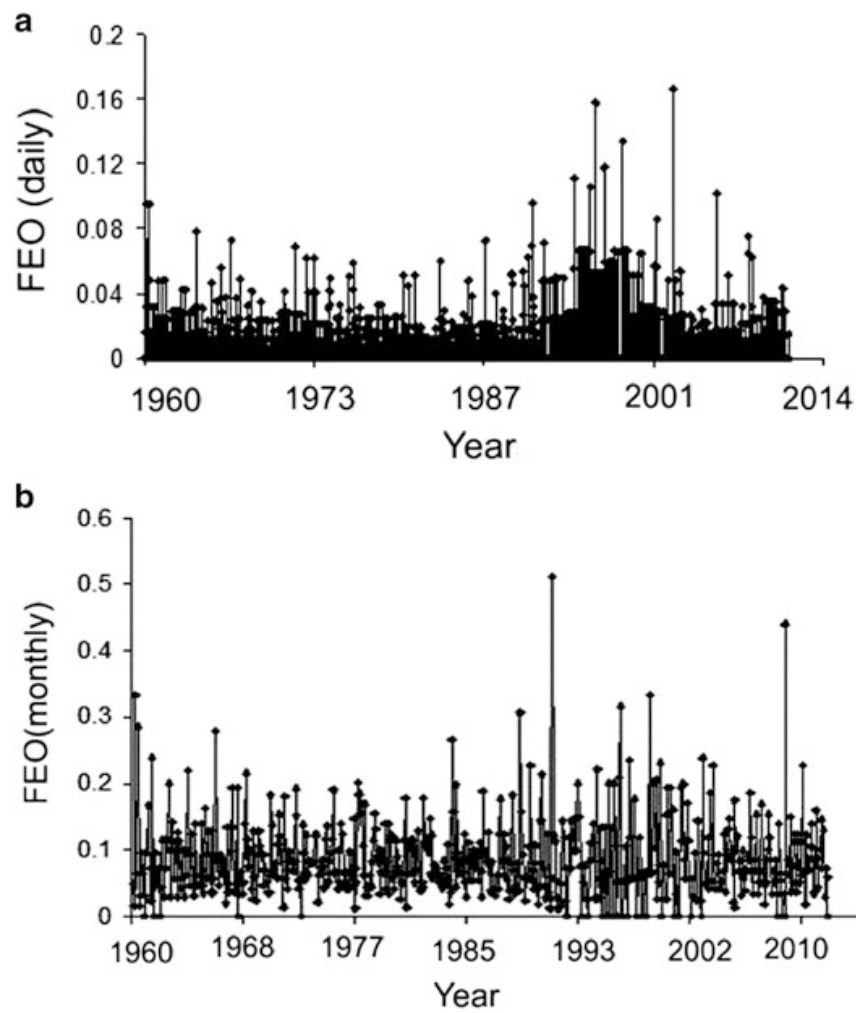


Fig. 10.16 (a) Daily and (b) monthly frequency of EQ occurrences (FEO) vs. years as compiled from the declustered Caucasian EQ catalogue (1960–2011, $M \geq 3.0$)

divided both by the total number of yearly occurrences. These data are identified as frequency of earthquake occurrences (FEO). The daily and monthly FEO series were then normalized to a zero mean and unit variance.

All eight RQA variables were calculated for the daily FEO time series, but here we only report the results for %DET and %LAM calculations. In Fig. 10.17 the results of sliding window RQA calculations are shown for daily FEO values within 1-year windows moved forward in steps of 1-year. There are wide fluctuations in the non-filtered FEO data, but applying the Savitzky–Golay filter smooths the resultant %DET and %LAM values. This filter approximates the data with an n th degree polynomial preserving up to n th moments of the data. Thus, Savitzky–Golay filtering has the advantage over, for instance, a moving average filter since the magnitude of the variations in the data (local extremes) are to a large extent preserved [94].

In order to decrease possible variations caused by the short windows length of 365 days, the same RQA calculations were repeated using a longer window length

Fig. 10.17 (a) Recurrence plots of daily FEO. (b) %DET (black circles) and %LAM (black triangles) of daily frequency of earthquake occurrence time series for 365 days length sliding window and 365 days step. %DET (gray circles) and %LAM (gray triangles) after Savitzky–Golay smoothing

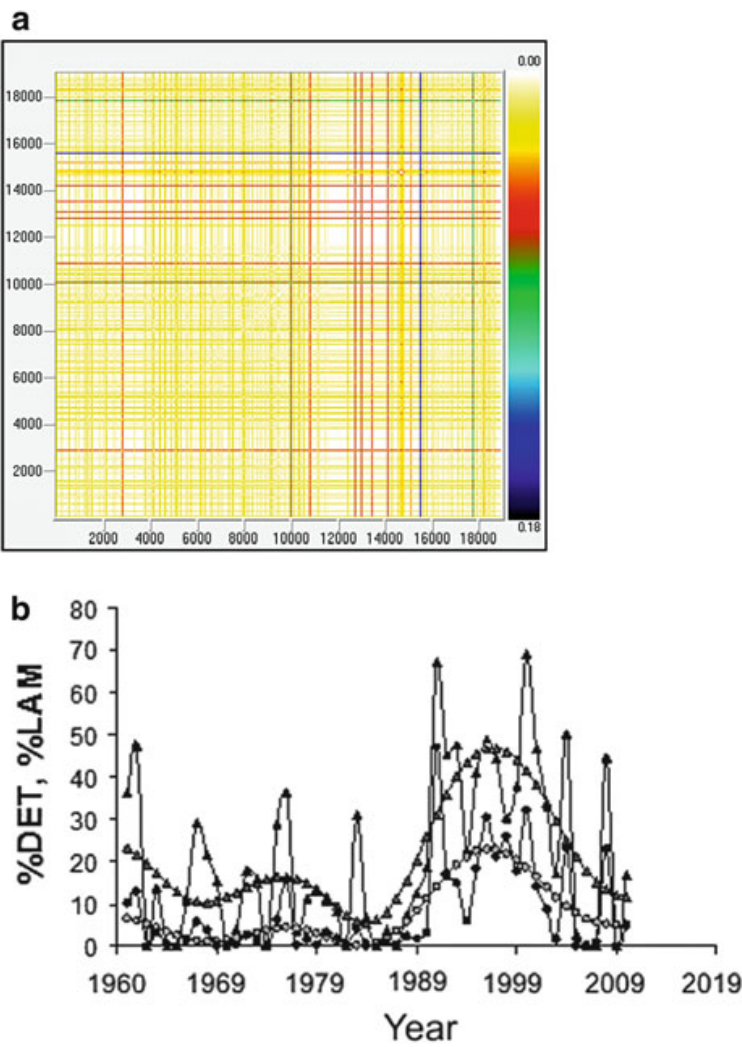
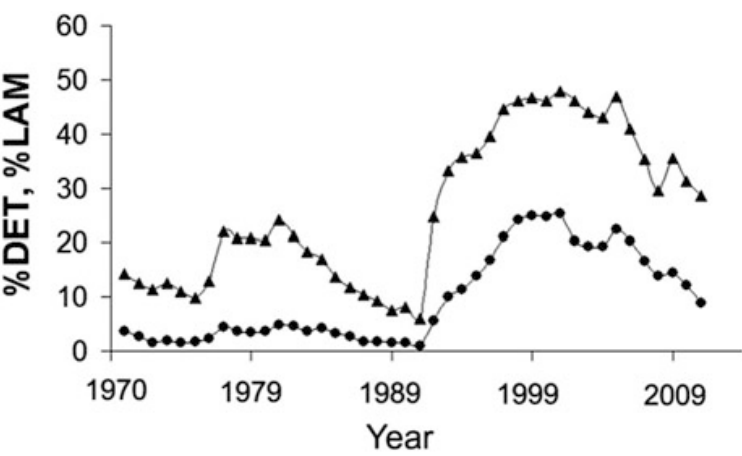


Fig. 10.18 %DET (circles) and %LAM (triangles) of daily frequency of EQ occurrence time series for 3,650 days length sliding window and 365 days step



of 3,650 days on the same 10 year daily FEO data. In these cases increases in %DET and %LAM variables are seen in windows corresponding to the period from 1990 to 2005 (Fig. 10.18).

Results obtained for the monthly FEO data using 10 year windows are shown in Fig. 10.19. These results are similar to those obtained for daily data and indicate that

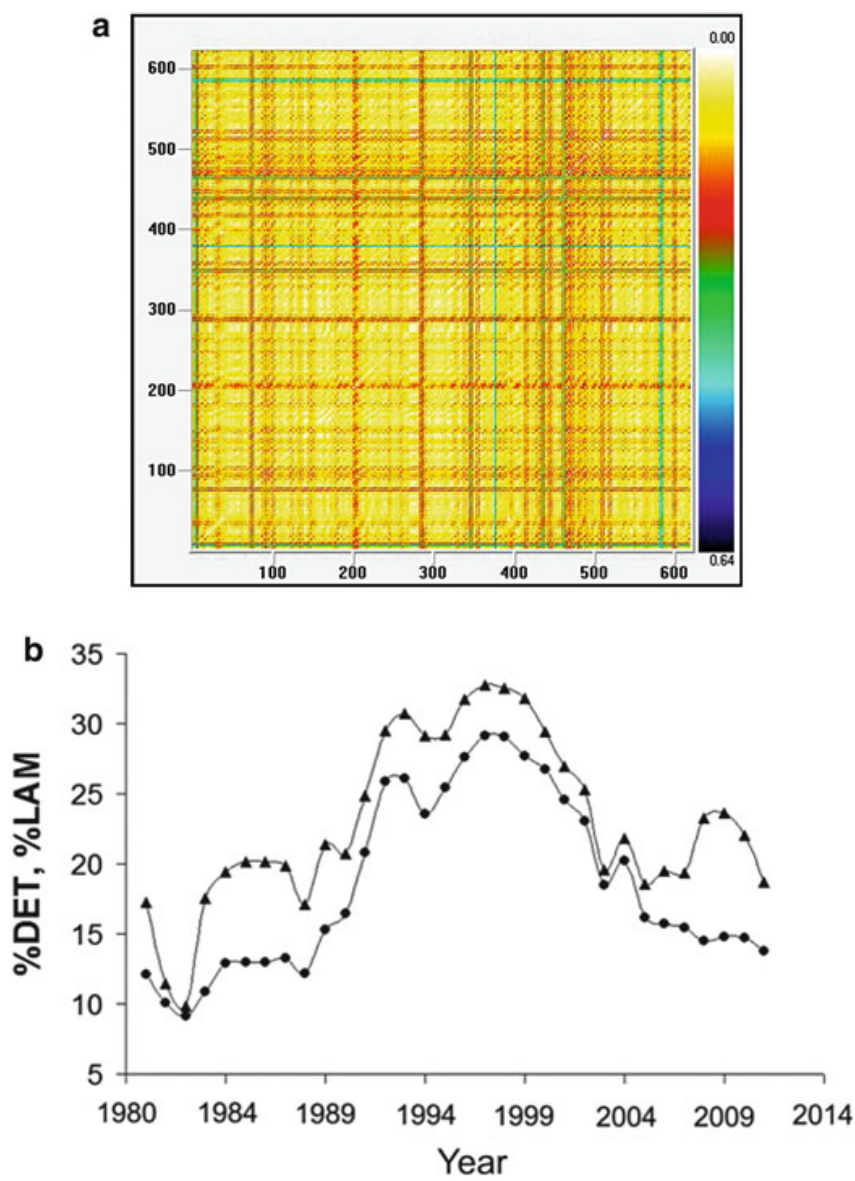
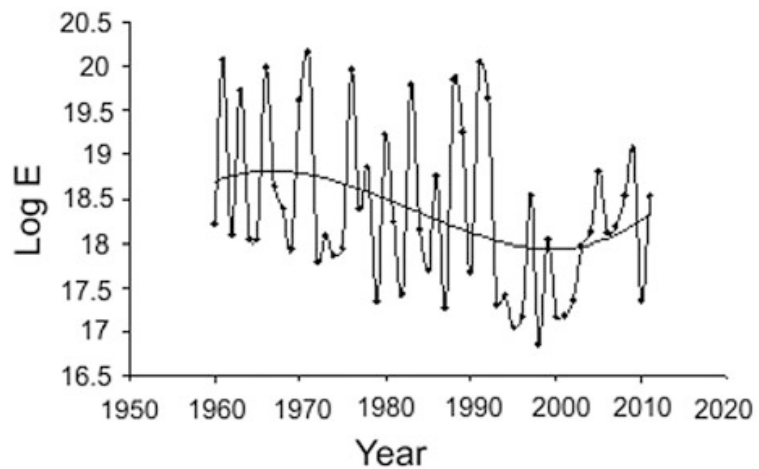


Fig. 10.19 (a) Recurrence plots of FEO. (b) %DET (*circles*) and %LAM (*triangles*) of monthly frequency of EQ occurrences for 240 month length sliding window and step of 12 months

the most important changes in temporal structure of considered process (frequency of EQ occurrence) also took place in the time period from approximately 1990 to 2005.

Although in most of the presented RPs above clear regular structures are not discernable qualitatively, it is important to focus on the behaviors of the %DET and %LAM variables which report on quantitative features of EQ occurrences over the years 1990–2005 (Figs. 10.17, 10.18, and 10.19). It should be emphasized that the demonstrated increases in %DET and %LAM measures data coincide with the period of decreased release of seismic energy in the Caucasus Mountains. This is shown in Fig. 10.20.

Fig. 10.20 Yearly release of seismic energy in the Caucasus Mountains (for $M > 3.0$) from 1960 to 2011 as calculated from the declustered catalogue. *Solid curve* is a third order polynomial fit



We hypothesize that the increase in seismic determinism is connected with formation of the “stress shadow” after strong EQs earthquakes of magnitude $M = 6.9$ (Spitak EQ, 1988; Racha EQ, 1991) in the region of Caucasus Mountains [95, 96]. After strong events, tectonic stresses relax (stress shadow appears) meaning that the share of relatively weak, quasi-periodic forces (e.g. tides, seasonal loadings, etc.) increase relative to the main driving tectonic forces. This makes the coupling of these weak forcings with tectonic stresses stronger and renders the time series more ordered (more deterministic).

10.5.2 *Dynamic Triggering and Synchronization of Seismicity*

Finding reliable proofs of seismic T/S dynamics has become possible over last decades due to the accumulation of numerous seismic observations. One kind of dynamic T/S is connected with volcanoes which produce tremors because, unlike EQs, they do not have impulsive onsets. Rather volcanos present one wave package of long duration without clear arrival times for subsequent waves. Stich et al. [97] reported on swarms of volcanic tremors with periodic occurrences at Deception Island in Antarctica. The characteristic inter-event periods for individual swarms ranged from approximately 10–20 s and were close to integer multiples of the dominant periods of the oceanic microseism (5 s). This evinces synchronization of volcanic tremor activity with the phase of oceanic noise which generates strains in the order of 10^{-7} . It appears that this case is an example of high order synchronization (HOS) of tremor activity with oceanic microseisms (see also Sect. 10.4.4). The 2 s periodicity inside the swarm also looks very similar to our laboratory modeling results (Figs. 10.8 and 10.9).

Another kind of T/S event was discovered over the last several years [12–15, 45, 73, 98, 99]. These so called non-volcanic or tectonic tremors (TT) are connected with activation of local seismic activity by wave trains (mostly Love and Rayleigh waves) from remote yet strong EQs. The TTs are singled out temporal and spatial

correlations of anomalous increases of local seismicity following the remote strong EQ [13]. The peak dynamic values of stress T_p due to the wave trains forcings is very low (of the order of 7–8 kPa) compared to tectonic stresses at the tremor source depth (several MPa). One approach for detecting TTs is to look for any excess seismic activity over the background level in the area in which wave trains arrive from a remote strong EQ. This method should be corrected for artifacts connected with the arrival of strong aftershock wave trains from remote sources. It should be noted that even if wave train stress exceeds the mean triggering threshold TTs are not generated everywhere. This is due to the impact of another important factor, namely, the local (site) strength of Earth material, which is highly heterogeneous. Our intuition is that what matters is not the absolute value of T_p , but the difference between local stress and local strength (failure strength or friction resistance). This is why in some areas with high T_p local seismicity is not triggered and, on the contrary, in some areas TTs are observed even at a low peak stresses. These dynamic differences are due to the competition between stress and strength in the different areas. One of the main factors reducing local strength is the pore pressure of fluids which is within the scope of relatively new field of hydroseismology [100, 101]. Reduced strength can explain high frequency of TTs in hydrothermal areas [45] as high pore pressures reduce the resistance to slip.

Another method of revealing TT seems more reliable. In this case TT events are singled out from the remote strong EQ broadband recording by applying a HF bandpass filter (2–8 Hz). One example of TT is shown in Fig. 10.21. Note that by visual inspection TTs and Love/Rayleigh wave peaks seem to be synchronized. Strict quantitative analysis of synchronization between TT events and wavetrains from remote EQs is very difficult to perform due to: (1) small statistics of tremors (as a rule, several tens of events); and (2) simultaneous forcing by Love (L) and Rayleigh (R) waves. The accepted approach is to compute correlation coefficients between tremor envelopes and L/R waves separately. R/L correlation coefficient ratios can also be computed. Visual analysis of Fig. 10.21 shows that there is not a one-to-one correspondence between amplitudes of surface waves and tremors. That is, strong tremors occur even in periods of weak forcing and vice versa. It seems that the R and L waves just control (stabilize) one of the unstable periodic orbits of chaotic system (seen here as active fault activation under forcing) resulting quasi-periodic tremors [50].

10.5.3 Phase Synchronization of Seismic Activity Induced by Water Level Variations in Reservoir

It has been documented by many authors that large water reservoirs can cause reservoir-induced seismicity (RIS) [5, 102, 103] or according to a new terminology, reservoir-triggered seismicity (RTS) [104]. The analysis of RTS by nonlinear dynamics methods is of particular interest as the water level change cause

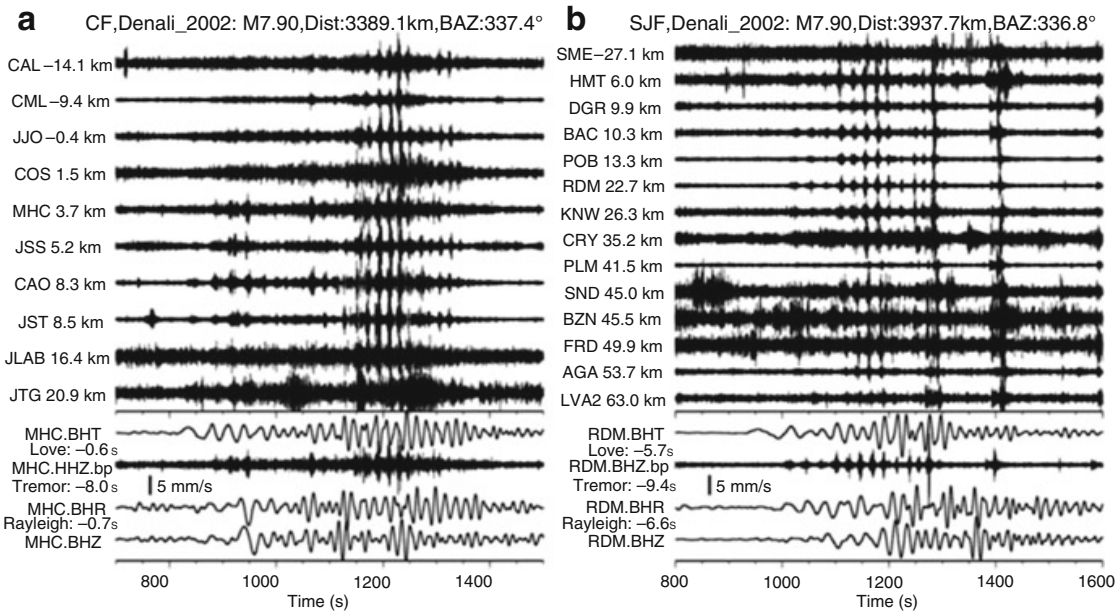


Fig. 10.21 (a) Top: 2–8 Hz band-pass-filtered vertical seismograms showing the move-out of tremor along the Calaveras fault (CF) in northern California (NC) triggered by the 2002 Mw 7.9 Denali fault EQ. The along-strike distance to the tremor source and the station names are marked in the seismograms. The event name and the occurrence year, its magnitude (M), and the epicenter distance (Dist) and back azimuth (BAZ) relative to the broadband station are all shown above the seismograms. *Bottom*: A comparison between the instrument-corrected transverse (BHT), radial (BHR), and vertical (BHZ) velocity seismograms and the 2–8-Hz band-pass-filtered seismogram recorded at the broadband station MHC. The zero time corresponds to the origin time of the main shock. The velocity seismograms have been time-shifted back to the tremor sources. The adjusted times of Love waves, Rayleigh waves and tremor are marked below the station names. The *thick vertical bar* marks the amplitude scale of surface waves. (b) Top: 2–8 Hz band-pass filtered vertical seismograms showing the move-out of tremor triggered by the Denali fault Earthquake along the San Jacinto fault (SJF) in the Anza network in southern California (SC). *Bottom*: A comparison between the velocity and the 2–8 Hz band-pass-filtered seismograms recorded at the broadband station RDM [14]

stress-regime variations in the Earth crust which can be correlated with local seismicity. The 271 m high Enguri Arch dam, still the highest arch dam in operation in the world, was built in the canyon of the Enguri river (West Georgia) in the 1970s. We investigated the change in regional seismic activity around the Enguri reservoir in different periods related to different stress regimes from the natural state (before dam construction) to periodic loading-unloading during the regular exploitation regime [105–107]. Three distinct periods were defined for individual analysis: (1) before impoundment; (2) during flooding and reservoir filling; and (3) during periodic changes of water level.

Figure 10.22 shows the daily record of the water level in the Enguri dam reservoir from 1978 to 1995. Values of the emitted daily seismic energy, as a measure of regional seismic activity, essentially increased at all magnitude scales during the period of territory flooding and reservoir filling. We found that after 8–9 years, the daily seismic energy release fell below those measured before reservoir filling. This

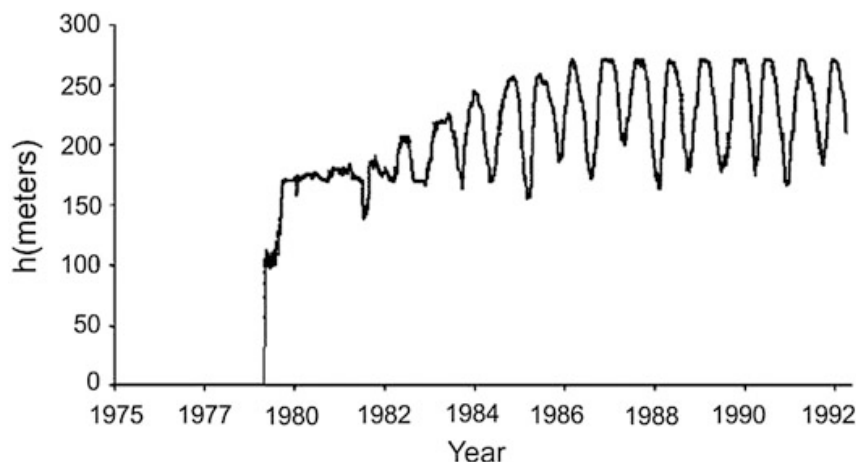


Fig. 10.22 Record of the daily water level height (h) in the lake behind the Enguri Dam versus time from 1975 to 1993

energy release decrease occurred most significantly in the range of large events. The effect of reservoir dynamics is rather small compared to the tectonic stress field. The level of water in the lake changes by 100 m, which means that the pressure at the bottom changes by 10 bar. At a depth of 10 km where the most of hypocentres of EQs are located, the pressure would be much less since the reservoir is of finite volume. For crude assessment of the order of load decrease we used a well-known expression for the stress field increment. Here $\Delta\sigma$ is due to slot-like defects of the size a at the distance r from it: $\Delta\sigma/\sigma \approx (a/r)^{1/2}$ where σ is applied stress, and a is the size of defect. Substituting $\sigma = 10$ bar, $a = 100$ m (here a is the depth of lake) and $r = 10,000$ m (the average depth of hypocenters) we obtain the value of order of 1 bar for $\Delta\sigma$. This is much less than the tectonic stress on this depth, which is of the order of several kilobars.

The period of reservoir filling ended in 1980 and after that, the water level variation became quasi-periodic. We investigated the interevent time (in minutes) of EQ sequences for the whole period, 1960–2012 as shown in Fig. 10.23. Of interest was how variations in reservoir water levels affected temporal features of local seismic activity. We considered both the original and the declustered [92] catalogues applying threshold magnitude of 2.0.

The results are reported as a qualitative RP (Fig. 10.24a) and quantitative, sliding-window RQA (Fig. 10.24b) on the EQ catalogue of the Enguri area for events occurring within 90 km from the location of the dam (1960–2012). It follows from results of EQ magnitude sequences (windows of 1 year repeated in steps of 1 year) that %DET and %LAM both increase after the start of high dam construction and first reservoir impounding (about 1977, Fig. 10.24b). As a caveat it should be emphasized that the first strong increases in the RQA variables may have been caused by engineering explosions during dam construction. These cannot be excluded since the threshold value applied ($M > 2.0$) was relatively low (high sensitivity). Higher threshold values lead to very short data sets disallowing reliable RQA calculations (low sensitivity). At the same time it seems quite logical that

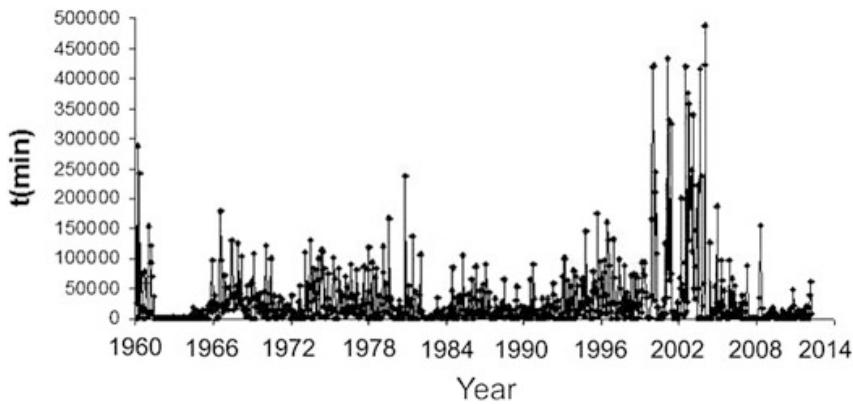


Fig. 10.23 Inter-event time intervals from the original seismic catalogue of Enguri Dam area from 1960 to 2012

increases in %DET and %LAM occurring in the right part of Fig. 10.24b, 10.25 could be explained by the influence of periodic variations of the reservoir water level commencing around 1982.

RQA patterns of earthquake waiting-time distributions also manifest essential temporal changes. As show in Fig. 10.25 the regularity in waiting time series strongly decreases during dam construction and irregular water impoundment and increases during quasiperiodic loading-unloadings of the reservoir. It is interesting to compare %DET of local seismicity waiting times (Fig. 10.25) with our earlier results [107, 108] of %DET of Earth crust tilts time series recorded in the foundation of Enguri Dam for the following periods: (1) long before the reservoir filling; (2) immediately before the start of filling; (3) just after the start of filling, (4) after the second stage of filling; (5) after the third stage of filling; (6) after the fourth stage of filling; and (7) long after the completion of the reservoir filling (Fig. 10.26). There is strong correspondence between %DET tilt patterns and EQ waiting times. We attribute both of these patterns to man-made effects (artifacts). That is, before construction of the dam the high determinism of natural Earth crust dynamics can be due to some regular forcings like tides or seasonal factors. This natural regime was disrupted by construction of dam, but the high determinism was re-established after start of regular load–unload cycles of the reservoir.

The physical mechanism of increased determinism in both EQ waiting times and tilts time series should be due to the periodic load–unload cycles of the reservoir. Arguing in favor of this interpretation are the findings from Singular Spectral Analysis which provides the resultant power spectra of the first four reconstructed components of both the monthly number of earthquakes and the mean water level (see [2]). Thus it was shown that the 1-year periodicity of seismicity in the spectrum correlates well with 1-year water level cycle. Also the weak 4-month period in seismicity can be considered as HOS with the main forcing period of 1-year.

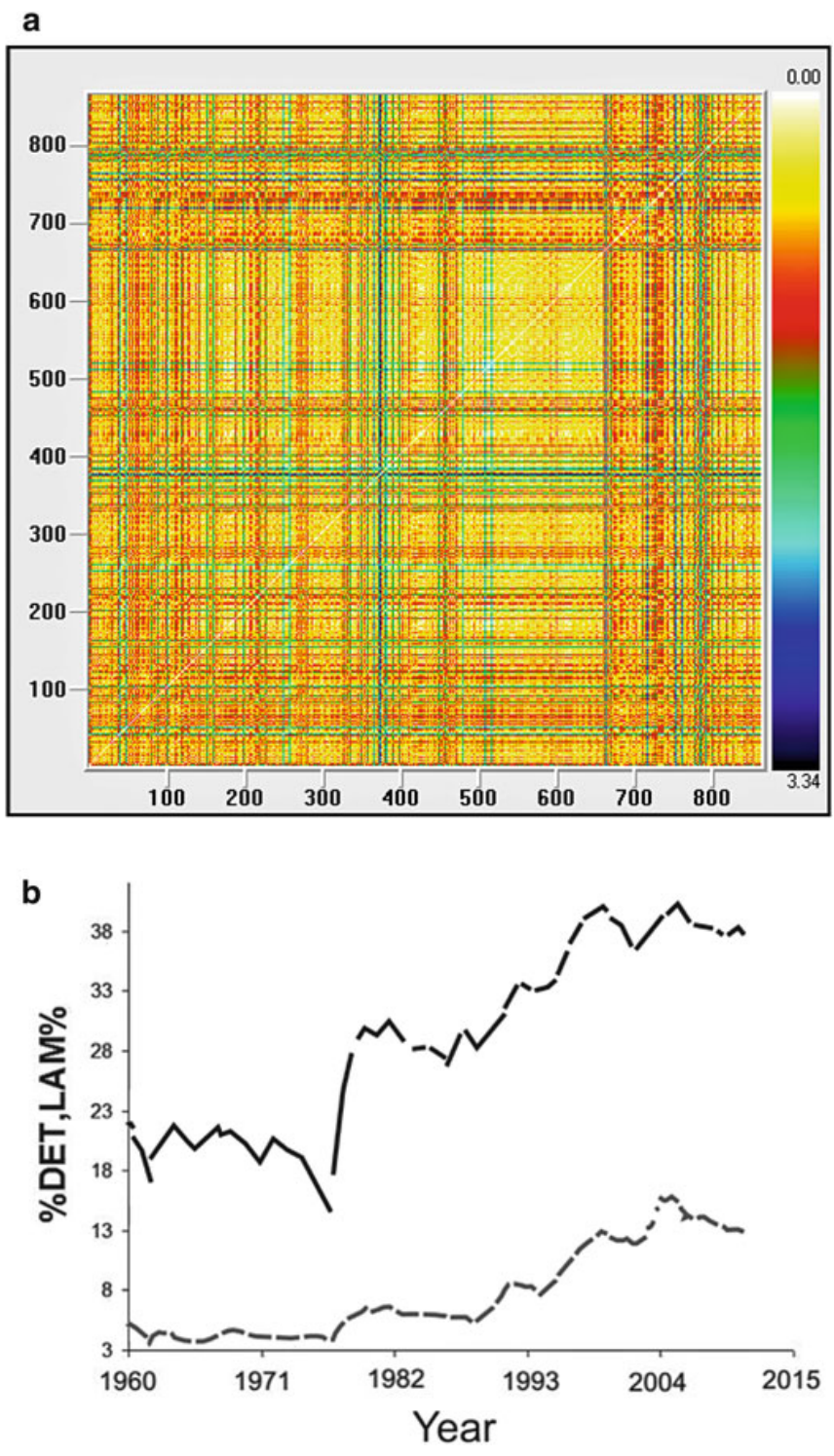


Fig. 10.24 (a) Recurrence plots; (b) %DET (*black*) and %LAM (*grey*) of sequence of EQ magnitudes from the Enguri EQ catalogue (1960–2011) at a representative threshold of $M > 2.0$. Window length is 365 days (1 year) with sliding step size of 1 day

Fig. 10.25 %DET (black) and %LAM (gray) of inter-event time sequences from the Enguri EQ catalogue (1960–2011) at a representative threshold of $M > 2.0$. Window length is 365 days (1 year) with sliding step size of 1 day

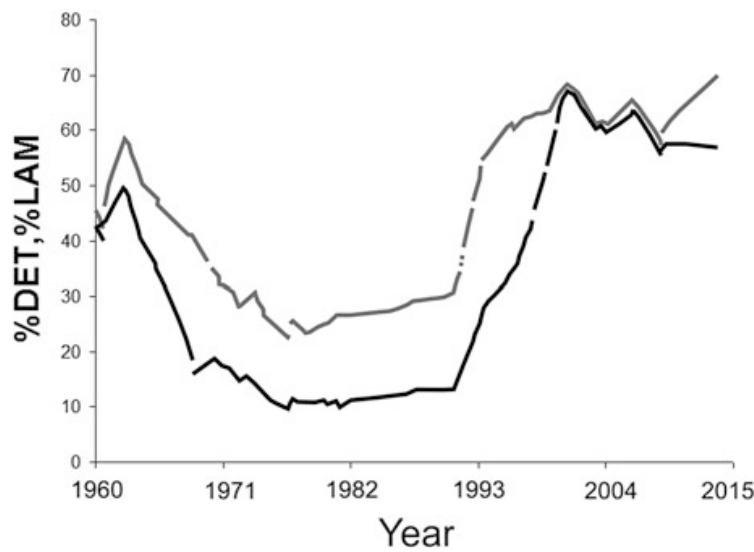
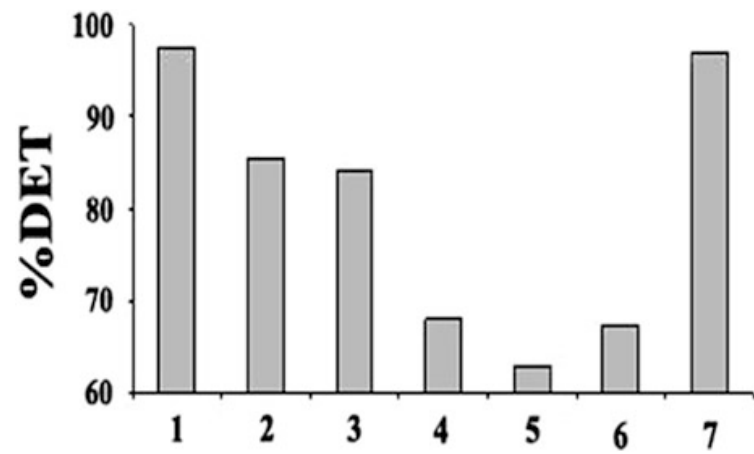


Fig. 10.26 %DET calculated for Earth crust tilt data series for different stages of observation in the foundation of Enguri dam. Numbers on abscissa correspond to periods of observation (see text)



10.5.4 Action of High Energy Electromagnetic Pulses on Local Seismicity

Unique field experiments were carried out by Institute of High Temperatures of Russian Academy of Sciences [9, 10, 109] at Bishkek test area (Central Asia). These studies showed that the actions of high energy electromagnetic pulses radiated by MHD (magnetohydrodynamic) generators or those of sets of batteries cause substantial changes in the number of comparatively small seismic events (energy $K \approx 7$) in the surrounding area with delays of 2–3 days. Laboratory modeling related to these results were considered previously (Sect. 10.4).

The Bishkek area is seismically active as shown in Fig. 10.27. Here are depicted some 14,100 seismic events extracted from the EQ catalogue over two decades. At the test site from January 8, 1983 to March 28, 1990, multiple series of strong current pulses (up to 2.5 kA) were released as 4.5 km long electrical (grounded) dipoles. The energy of separate EM pulses were in the order of $1\text{--}2 \cdot 10^7$ J. This is close to energy level of seismic events ($K \approx 7$) which were preferably excited by

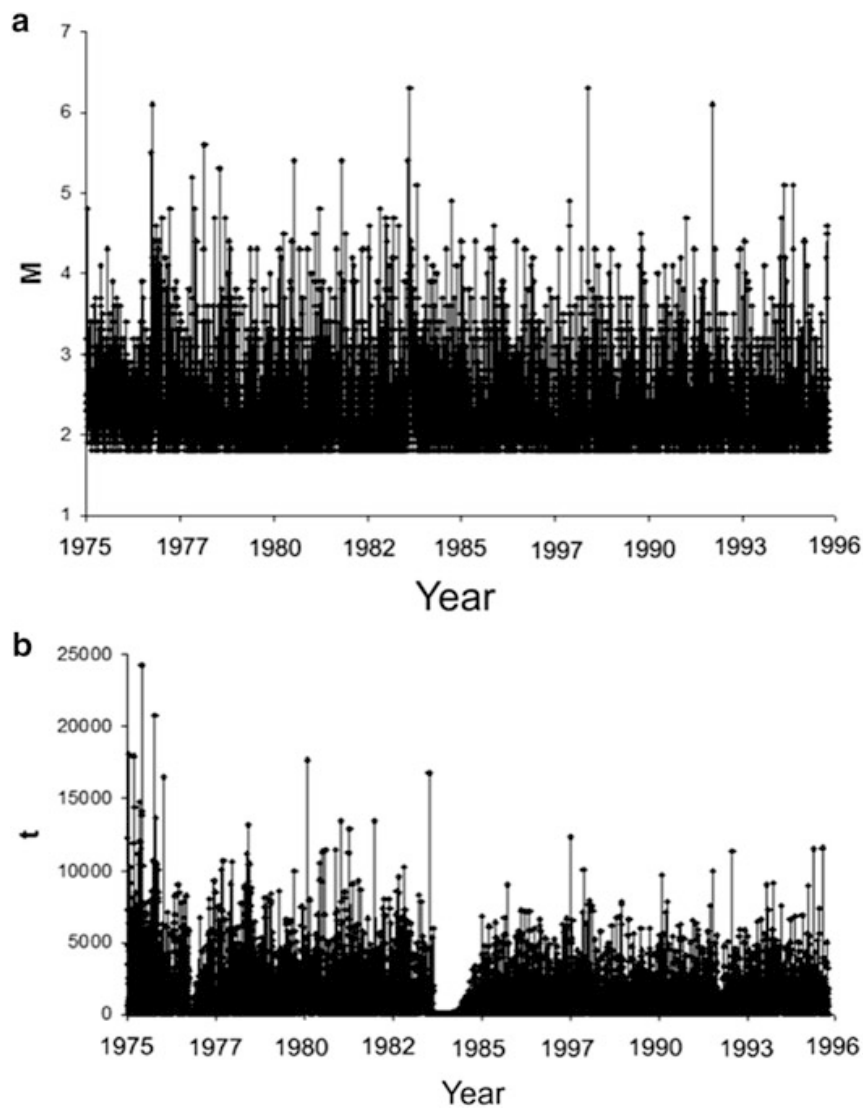


Fig. 10.27 (a) Sequence of EQ magnitudes for threshold $M \geq 1.8$. (b) Inter-event time intervals from the EQ catalogue of Central Asia over the period 1975–1996

injections of EM pulses. Quantitative nonlinear analysis of these data are reported elsewhere [106, 110].

RQA was performed on sequences of EQ waiting times (in minutes) from the Central Asia seismic catalogue which was divided into three parts: (1) 1975–1983 prior to EM experiments; (2) 1983–1990 during EM experiments; and (3) 1988–1996 after termination of EM experiments. The catalogue was tested for completeness [110]. Time intervals between EQs with threshold magnitudes $M > 1.8$ and $M > 2.5$ were investigated.

In Fig. 10.28 %DET and %LAM variables are shown for the waiting time series. Other RQA variables were computed, but are not shown. Using windows with 600 points shifted by 100 points, there are increases in both %DET and %LAM from 1983 to 1986. This time period corresponds to when the EM experiments were carried out and on which the %DET and %LAM calculations were also made.

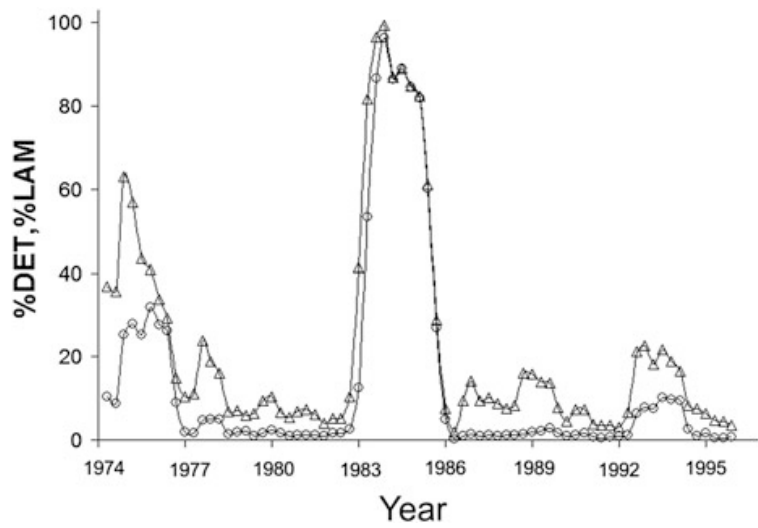


Fig. 10.28 %DET (*circles*) and %LAM (*triangles*) of waiting times sequences from the Central Asia EQ catalogue (1975–1996) at a representative threshold $M > 1.8$

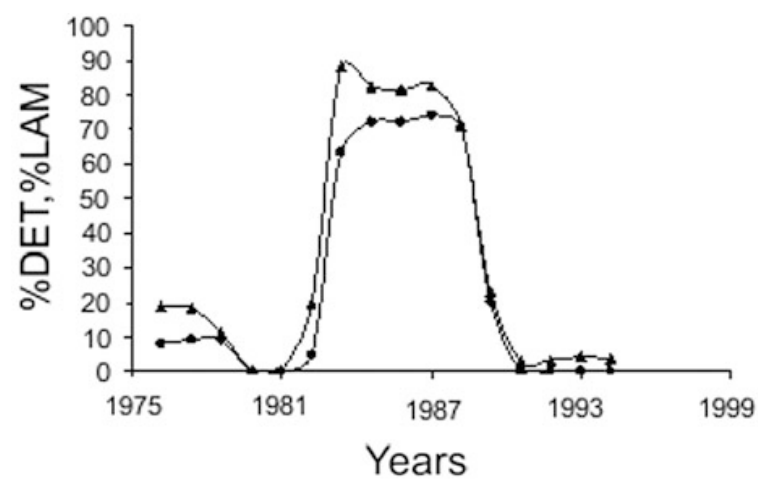


Fig. 10.29 %DET (*circles*) and %LAM (*triangles*) of sequence of inter-event times from Central Asia EQ catalogue (1975–1996) at a representative threshold $M > 2.5$. Window length is 500 days with sliding step size of 100 days. *Note:* strong increases in %DET and %LAM during active EM experiments (1983–1988)

%DET and %LAM values of the randomized waiting times sequences were both flat. These results indicate that the original time series had deterministic structuring in both the diagonal and vertical directions.

In order to avoid mistakes due to problems associated with small earthquake registrations at $M > 1.8$, the representative threshold was increased to $M > 2.5$. Results of this analysis are presented in Fig. 10.29. Again observe clear increases in %DET and %LAM between 1983 and 1989 corresponding to the period of active EM experiments. As before, all structures in the original time series were destroyed by randomization of the waiting-time sequence.

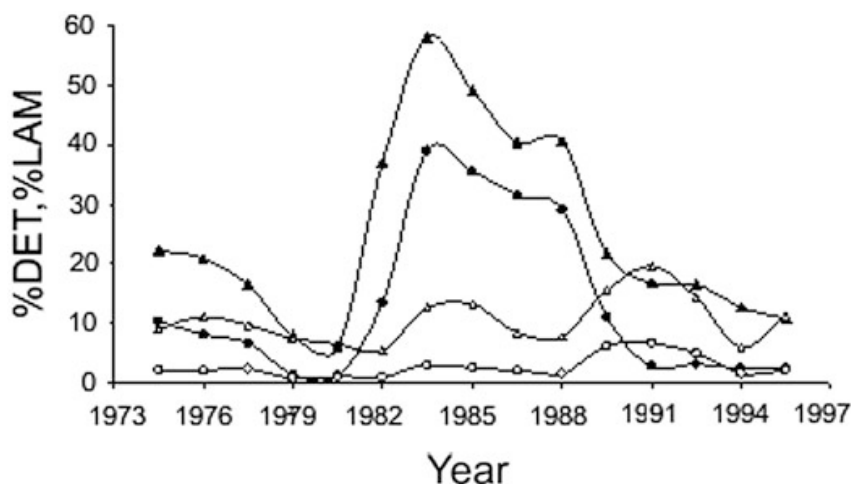


Fig. 10.30 %DET (black circles) and %LAM (black triangles) of waiting time sequences from the declustered Central Asia EQ catalogue (1975–1996) at a representative threshold of $M > 2.5$. Window length is 500 days with sliding step size of 100 days. Open symbols correspond to randomized data sets showing attenuation of deterministic structures

Similar increases in %DET and %LAM are also observed for the declustered EQ catalogue as shown in Fig. 10.30. This particular analysis ensures that obtained changes in RQA measures are not caused by aftershocks of strong earthquakes ($M \approx 6.1$ – 6.3) that occurred in the periods before, during and after EM soundings.

Figure 10.31 shows recurrence plots of waiting times sequences at Bishkek test area for a complete EQ catalogue (threshold $M \geq 2.5$): (1) before (1975–1983), (2) during (1983–1988), and (3) after performing the EM experiments (1988–1992). Note the compact structuring of RPs during EM experiments.

According to the present (new) RQA results and previous analysis [24, 110] we conclude that strong EM discharges lead to an increased regularity of the temporal distribution of EQs. After cessation of EM experiments the temporal distribution of EQs become more random than before the experiments were conducted. The source of the increased determinism is not quite clear. One explanation can be the presence of several almost regular patterns in the distribution of EM discharge waiting-times.

10.5.5 Geocomplexity Analysis for Earthquake Prediction

The wide variety of new and modern tools for the quantitative analysis of complexity in time series has greatly increased the ability to discover fine variations in seismic and other geophysical time series. Such subtleties are caused by the nonstationarity of underlying processes including anomalies connected with “silent” dynamics prior to strong EQ events [111]. For example, the idea that synchronization phenomena can be correlated with (linked to) closeness of the system to its critical state (EQ manifestation) has been suggested Chelidze et al. [24]. Likewise, systematic exploration of variations of different synchronization characteristics of low-frequency

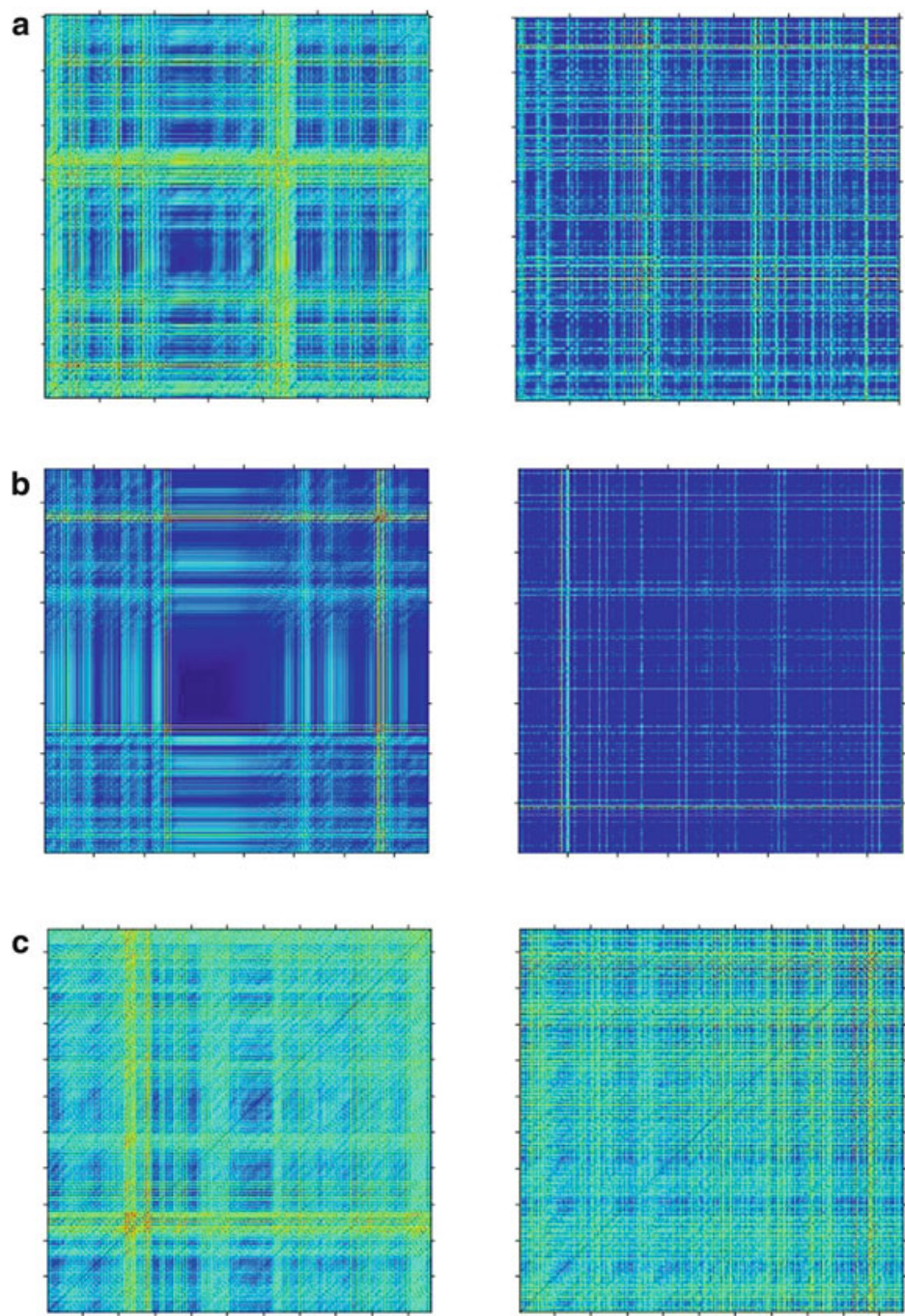


Fig. 10.31 *Left column:* recurrence plots analysis of EQs waiting times sequences at Bishkek test area (complete catalogue, $M \geq 2.5$). **(a)** Before EM experiments (1975–1983), **(b)** during EM experiments (1983–1988), and **(c)** after accomplishing of experiments (1988–1992). *Note:* compact structure in RP during EM experiments. *Right column:* shuffled EQ catalogues for the same three periods

microseismic noise fields obtained during tens of years on the seismic network of Japan have been studied [41]. As a result, several months before great EQ in Japan (Tohoku) on March 11, 2011 ($M = 9$) official predictions on the possibility of strong earthquake in Japan starting from June 2011.

10.6 Conclusions

The dynamic patterns of seismicity for analysis of triggering/synchronization phenomenon were revealed by application of tools from nonlinear dynamics including RQA to various time series: (1) “laboratory” earthquakes measured as acoustic emissions (AEs) during natural stick-slip and stick-slip under weak periodic forcings; (2) regional seismicity of Caucasus Mountains; (3) local seismicity in the area of a large reservoir during its construction and regular exploitation; and (4) local seismicity in Central Asia test area before, during and after the application of strong electric pulses.

For quantitative measuring of synchronization strength several modern tools of nonlinear dynamics were used such as mean effective phase diffusion coefficient, Shannon entropy based characteristic phase synchronization measure, Mutual Information, Recurrence Plots and Recurrence Quantitative Analysis (e.g. percent determinism, %DET and percent laminarity, %LAM, etc.).

The phase synchronization of stick-slip processes induced by a weak electromagnetic or mechanical periodic forcings were investigated on small-scale laboratory experiments. Application of varying frequencies and intensities of forcing allowed for generation of Arnold’s tongue diagrams for EM forcings. It was discovered that not only are the onsets/maxima of AE signals synchronized with forcings, but so too are AE wave train terminations.

The effect of high order synchronization of stick-slip events by weak electromagnetic or mechanical periodic forcings was demonstrated. Two kinds of high order synchronization were discovered: (1) occurrence of one or more AE bursts during a single forcing period; and (2) occurrence of a single AE burst during many forcing periods.

Reviews of recent results of dynamic triggering of local seismicity by remote earthquakes is presented, as well as other examples of triggering/synchronization by reservoir exploitation and strong electromagnetic pulses.

In all these studies from laboratory scale to actual earthquakes, RQA proved to be a very efficient method for revealing deterministic structuring and synchronization in related time series. These results point to the possibility of revealing new fine details in the stick-slip process which can be very useful for refining the understanding of the physical mechanism of frictional motion in general. These findings may also be useful in discovering new but hidden regularities in seismic time series.

Acknowledgments The authors express their gratitude to the Georgian (Rustaveli) National Science Foundation, INTAS foundation and Open Partial Agreement on Major Disasters at Council of Europe (EUR-OPA) for financial support. Authors are grateful for the kind permission of World Scientific and Engineering Academy and Society (WSEAS), Bulletin of Seismological Society of America, Journal of Georgian Geophysical Society and Bulletin of Georgian National Academy of Sciences to reuse Figures from published in these editions papers in the present review. The help of Prof. O. Lursmanashvili, Dr. N. Varamashvili and researchers N. Zhukova, E. Mepharidze and D. Tephnadze is acknowledged.

References

1. D. Broomhead, G. King, On the qualitative analysis of experimental dynamical systems, in *Nonlinear Phenomena and Chaos*, ed. by S. Sarkar (Adam Hilger, Bristol, 1986), pp. 113–144
2. L. Telesca, T. Matcharashvili, T. Chelidze, N. Zhukova, Relationship between seismicity and water level in the Enguri high dam area (Georgia) using the singular spectrum analysis. *Nat. Hazards Earth Syst. Sci.* **12**, 2479–2485 (2012)
3. C.H. Scholz, *Mechanics of Earthquakes and Faulting* (Cambridge University Press, Cambridge, 2003)
4. G. King, Fault interaction, earthquake stress changes and the evolution of seismicity, in *Earthquake Seismology*, volume ed. by H. Kanamori (Elsevier, Amsterdam, 2009), pp. 225–255
5. H.K. Gupta, *Reservoir-Induced Earthquakes* (Elsevier, New York, 1992). 264 pp
6. J.-R. Grasso, D. Sornette, Testing self-organized criticality by induced seismicity. *J. Geophys. Res.* **103**, 29965–29987 (1998)
7. A.V. Nikolaev (ed.), *Induced Seismicity* (“Nauka”, Moscow, 1994) 220 p (in Russian)
8. K. Heki, Snow load and seasonal variation of earthquake occurrence in Japan. *Earth Planet. Sci. Lett.* **207**, 159–164 (2003)
9. N.T. Tarasov, Crustal seismicity variation under electric action. *Trans. (Doklady) Russ. Acad. Sci.* **353A**(3), 445–448 (1997)
10. N. Tarasov, H. Tarasova, A. Avagimov, V. Zeigarnik, The effect of high-power electromagnetic pulses on the seismicity of Central Asia and Kazakhstan. *Volcanol. Seismol. (Moscow)* **N4–5**, 152–160 (1999) (in Russian)
11. C.H. Scholz, Earthquakes: good tidings. *Nature* **425**, 670–671 (2003). doi:[10.1038/425670a](https://doi.org/10.1038/425670a)
12. T. Iwata, Earthquake triggering caused by the external oscillation of stress/strain changes. *Community Online Resource for Statistical Seismicity Analysis* (2012). doi:[10.5078/corssa-65828518](https://doi.org/10.5078/corssa-65828518). <http://www.corssa.org>
13. D. Hill, S. Prejean, Dynamic triggering, in *Earthquake Seismology*, volume ed. by H. Kanamori (Elsevier, Amsterdam, 2009), pp. 257–293
14. K. Chao, Z. Peng, W. Ch, C.-C. Tang, C.-H. Lin, Remote triggering of non-volcanic tremor around Taiwan. *Geophys. J. Int.* **188**, 301–324 (2012)
15. K. Chao, Z. Peng, A. Fabian, L. Ojha, Comparisons of triggered tremor in California. *Bull. Seismol. Soc. Am.* **102**, 900–908 (2012)
16. R. Meyers (ed.), *Encyclopedia of Complexity and Systems Science* (Springer, New York, 2009)
17. H. Abarbanel, L.S. Tsimring, The analysis of observed chaotic data in physical systems. *Rev. Mod. Phys.* **65**, 1331–1392 (1993)
18. H. Kantz, T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, 1997)
19. J. Rundle, D. Turcotte, W. Klein (eds.), *Geocomplexity and Physics of Earthquakes* (AGU, Washington, DC, 2009)
20. D. Sornette, *Critical Phenomena in Natural Sciences* (Springer, Heidelberg, 2000)
21. J. Sprott, *Chaos and Time-Series Analysis* (Oxford University Press, Oxford, 2003)
22. S. Strogatz, *Nonlinear Dynamics and Chaos* (Perseus Books Group, Westview, 2000)
23. A. Pikovsky, M.G. Rosenblum, J. Kurths, *Synchronization: Universal Concept in Nonlinear Science* (Cambridge University Press, Cambridge, 2003)
24. T. Chelidze, O. Lursmanashvili, T. Matcharashvili, M. Devidze, Triggering and synchronization of stick slip: waiting times and frequency–energy distribution. *Tectonophysics* **424**, 139–155 (2006)
25. D. Sornette, *Why Stock Markets Crash* (Princeton University Press, Princeton, 2003)
26. A. Bunde, J. Kropp, H. Schellnhuber (eds.), *The Science of Disasters* (Springer, Berlin, 2002)
27. H. Kanamori, E.E. Brodsky, The physics of earthquakes. *Rep. Prog. Phys.* **67**, 1429–1496 (2004)

28. P. Bak, C. Tang, K. Wiesenfeld, Self-organized criticality. *Phys. Rev.* **A38**, 364–374 (1988)
29. Y. Ogata, Statistical models for earthquake occurrences and residual analysis for point processes. *J. Am. Stat. Assoc.* **83**, 9–27 (1988)
30. D. Sornette, M. Werner, Statistical physics approaches to seismicity, in *Encyclopedia of Complexity and Systems Science* (Springer, New York, 2009), pp. 7872–7891
31. C. Goltz, *Fractal and Chaotic Properties of Earthquakes* (Springer, Berlin, 1997)
32. T. Matcharashvili, T. Chelidze, Z. Javakhishvili, Nonlinear analysis of magnitude and interevent time interval sequences for earthquakes of the Caucasian region. *Nonlinear Processes Geophys.* **7**, 9–19 (2000)
33. T. Chelidze, T. Matcharashvili, Complexity of seismic process, measuring and applications – a review. *Tectonophysics* **431**, 49–61 (2007)
34. G. Sobolev, Seismicity dynamics and earthquake predictability. *Nat. Hazards Earth Syst. Sci.* **11**, 445–458 (2011). doi:[10.5194/nhess-11-445-2011](https://doi.org/10.5194/nhess-11-445-2011)
35. T. Rikitake, *Earthquake Forecasting and Warning* (Center for Academic Publications, Tokyo, 1982)
36. K. Mogi, *Earthquake Prediction* (AP, Tokyo, 1985)
37. G.A. Sobolev, A.V. Ponomarev, *Physics of Earthquakes and Precursors* (“Nauka”, Moscow, 2003) (in Russian)
38. R.J. Geller, Earthquake prediction: a critical review. *Geophys. J. Int.* **131**, 425–450 (1997)
39. M. Wyss, D. Booth, The IASPEI procedure for the evaluation of earthquake precursors. *Geophys. J. Int.* **131**, 42–424 (1997)
40. Y. Ogata, K. Katsura, G. Falcone, K. Nanjo, J. Zhuang, Comprehensive and topical evaluations of earthquake forecasts in terms of number, time, space, and magnitude. *Bull. Seismol. Soc. Am.* **103**, 1692–1708 (2013). doi:[10.1785/0120120063](https://doi.org/10.1785/0120120063)
41. A. Lyubushin, Prognostic properties of low-frequency seismic noise. *Nat. Sci.* **4**, 659–666 (2012). doi:[10.4236/ns.2012.428087](https://doi.org/10.4236/ns.2012.428087)
42. D.M. Evans, Denver area earthquakes and the rocky mountain arsenal disposal well. *Mt. Geol.* **3**, 23–26 (1966)
43. S. Hainzl, T. Kraft, J. Wassermann, H. Igel, E. Schmedes, Evidence for rainfall-triggered earthquake activity. *Geophys. Res. Lett.* **33**, L19303 (2006). doi:[10.1029/2006GL027642](https://doi.org/10.1029/2006GL027642)
44. T. Chelidze, T. Matcharashvili, Electromagnetic control of earthquake dynamics? *Comput. Geosci.* **29**, 587–593 (2003)
45. S. Prejean, D. Hill, Dynamic triggering of earthquakes, in *Encyclopedia of Complexity and Systems Science*, ed. by R.A. Meyers (Springer, New York, 2009), pp. 2600–2621
46. S. Boccaletti, C. Grebogi, Y.-C. Lai, H. Mancini, D. Maza, The control of chaos: theory and applications. *Phys. Rep.* **329**, 103–119 (2000)
47. A. Koronovskii, O. Moskalenko, A. Hramov, Mechanisms behind the generalized synchronization conditions. *Tech. Phys.* **51**(2), 143–150 (2006)
48. E. Ott, M. Spano, Controlling chaos. *Phys. Today* **5**, 34–40 (1995)
49. W.L. Ditto, S.N. Rausero, M.L. Spano, Experimental control of chaos. *Phys. Rev. Lett.* **65**, 3211–3214 (1990)
50. J. Starrett, Control of chaos by occasional bang–bang. *Phys. Rev. E* **67**, 036203, 1–4 (2003)
51. I.I. Blekhman, *Synchronization in Science and Technology* (ASME Press, New York, 1988)
52. M. Palus, Detecting phase synchronization in noisy systems. *Phys. Lett. A* **227**, 301–308 (1997)
53. M. Rosenblum et al., Phase synchronization: from theory to data analysis, in *Handbook of Biological Physics*, ed. by F. Moss, S. Gielen. *Neuro-Informatics*, vol 4 (Elsevier, Amsterdam, 2001), pp. 279–321
54. E. Toledo, M. Rosenblum, J. Kurths, S. Akselrod, Cardiorespiratory synchronization: is it real phenomenon? in *Computers in Cardiology*, ed. by A. Murrey, S. Swiryn (IEEE, Hannover, 1999), pp. 237–240
55. V. Anishchenko, *Dynamical Chaos. Models and Experiments* (World Scientific, Singapore, 1995)

56. C. Tsallis, Possible generalization of Boltzmann–Gibbs statistics. *J. Stat. Phys.* **52**, 479–487 (1988)
57. A. Lempel, J. Ziv, On the complexity of finite sequences. *IEEE Trans. Inf. Theory* **IT-22**, 75–81 (1976)
58. A. Schuster, On lunar and solar periodicities of earthquakes. *Proc. R. Soc. Lond.* **61**, 455–465 (1897)
59. F. Takens, Detecting strange attractors in turbulence, in *Dynamical Systems and Turbulence*, ed. by D.A. Rand, L.S. Young. Springer Lecture Notes in Mathematics, vol. 898 (Springer, Berliner, 1981), pp. 366–381
60. J.-P. Eckmann, S. Kamphorst, D. Ruelle, Recurrence plots of dynamical systems. *Europhys. Lett.* **4**, 973–977 (1987)
61. J.P. Zbilut, C.L. Webber Jr., Embeddings and delays as derived from quantification of recurrence plots. *Phys. Lett. A* **171**, 199–203 (1992)
62. J.P. Zbilut, C. Webber Jr., Detecting deterministic signals in exceptionally noisy environments using cross-recurrence quantization. *Phys. Lett. A* **246**, 122–128 (1998)
63. C.L. Webber Jr., J.P. Zbilut, Recurrence quantification analysis of nonlinear dynamical systems, in *Tutorials in Contemporary Nonlinear Methods for the Behavioral Sciences*, ed. by M.A. Riley, G.C. Van Orden (National Science Foundation, Arlington, VA, 2005), pp. 26–94
64. M. Marwan, Encounters with neighborhood, Ph.D. Thesis. Potsdam, 2003
65. N. Marwan, M. Thiel, N. Nowaczyk, Cross recurrence plot based synchronization of time series. *Nonlinear Processes Geophys.* **9**, 325–331 (2002)
66. N. Marwan, M.C. Romano, M. Thiel, J. Kurths, Recurrence plots for the analysis of complex systems. *Phys. Rep.* **438**, 237–329 (2007)
67. C.L. Webber Jr., N. Marwan, A. Facchini, A. Giuliani, Simpler methods do it better: success of recurrence quantification analysis as a general purpose data analysis tool. *Phys. Lett. A* **373**, 3753–3756 (2009)
68. T. Matcharashvili, T. Chelidze, Nonlinear dynamics as a tool for revealing synchronization and ordering in geophysical time series: application to Caucasus seismicity, in *Synchronization and Triggering: From Fracture to Earthquake Processes*, ed. by V. de Rubeis, Z. Czechowski, R. Teisseyre (Springer, Berlin, 2010), pp. 3–21
69. M. Thiel, M.C. Romano, J. Kurths, R. Meucci, E. Allaria, F.T. Arecchi, Influence of observational noise on the recurrence quantification analysis. *Physica D* **171**, 138–152 (2002)
70. N. Marwan, How to avoid potential pitfalls in recurrence plot based data analysis. *Int. J. Bifurcation Chaos* **21**, 1003–1017 (2011)
71. J.P. Zbilut, J.M. Zaldivar-Comenges, F. Strozzi, Recurrence quantification based Lyapunov exponents for monitoring divergence in experimental data. *Phys. Lett. A* **297**, 173–181 (2002)
72. C.L. Webber Jr., J.P. Zbilut, Dynamical assessment of physiological systems and states using recurrence plot strategies. *J. Appl. Physiol.* **76**, 965–973 (1994)
73. D. Hill, Surface wave potential for triggering tectonic (nonvolcanic) tremor. *Bull. Seismol. Soc. Am.* **100**, 1859–1878 (2010). doi:[10.1785/0120090362](https://doi.org/10.1785/0120090362)
74. N.M. Beeler, D.A. Lockner, Why earthquakes correlate weakly with the solid Earth tides: effects of periodic stress on the rate and probability of earthquake occurrence. *J. Geophys. Res.* **B108**, 2391–2405 (2003)
75. N. Bartlow, D. Lockner, N. Beeler, Laboratory triggering of stick-slip events by oscillatory loading in the presence of pore fluid with implications for physics of tectonic tremor. *J. Geophys. Res.* **117**, B11411 (2012). doi:[10.1029/2012JB009452](https://doi.org/10.1029/2012JB009452)
76. T. Chelidze, T. Matcharashvili, J. Gogiashvili, O. Lursmanashvili, M. Devidze, Phase synchronization of slip in laboratory slider system. *Nonlinear Processes Geophys.* **12** (2005)
77. T. Chelidze, O. Lursmanashvili, T. Matcharashvili, N. Varamashvili, N. Zhukova, E. Mepharidze, Triggering and synchronization of stick-slip: experiments on spring-slider system, in *Synchronization and Triggering: From Fracture to Earthquake Processes*, ed. by V. de Rubeis, Z. Czechowski, R. Teisseyre (Springer, Berlin, 2010), pp. 123–164
78. T. Chelidze, T. Matcharashvili, Triggering and synchronization of seismicity: laboratory and field data – a review, in *Earthquakes – Triggers, Environmental Impact and Potential Hazards*, ed. by K. Konstantinou (Nova Science Pub, New York, 2013), pp. 165–231

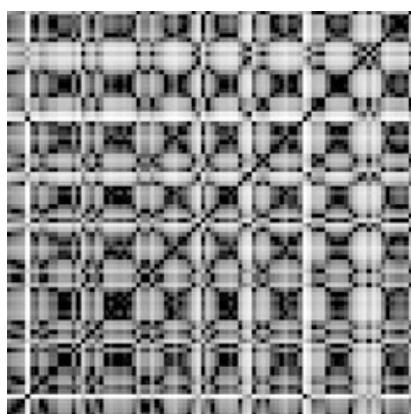
79. W.E. Brace, I.D. Byerlee, Stick slip as a mechanism for earthquakes. *Science* **153**, 990–992 (1966)
80. J.H. Dieterich, Modeling of rock friction 1. Experimental results and constitutive equations. *J. Geophys. Res.* **84B**, 2161–2168 (1979)
81. A. Ruina, Slip instability and state variable friction laws. *J. Geophys. Res.* **88B**, 10359–10370 (1983)
82. T. Chelidze, T. Matcharashvili, O. Lursmanashvili, N. Varamashvili, Acoustics of stick-slip deformation under external forcing: the model of seismic process synchronization, in *2nd IASME/WSEAS International Conference on GEOLOGY and SEISMOLOGY (GES '08)* (Cambridge, 2008)
83. T. Chelidze, N. Varamashvili, M. Devidze, Z. Chelidze, V. Chikhladze, T. Matcharashvili, Laboratory study of electromagnetic initiation of slip. *Ann. Geophys.* **45**, 587–599 (2002)
84. T. Chelidze, O. Lursmanashvili, Electromagnetic and mechanical control of slip: laboratory experiments with slider system. *Nonlinear Processes Geophys.* **20**, 1–8 (2003)
85. I. Tamm, *Fundamentals of Theory of Electricity*. 6th Edition (1956) Moscow, State Publishing House of Technical-Theoretical Literature (in Russian)
86. R. Sibson, Crustal stress, faulting and fluid flow, in *Deformation and Fluid Flow Geological Society*, London, Special Publications, vol 78 (1994), pp. 69–84
87. M.G. Rosenblum, A. Pikovsky, J. Kurths, Phase synchronization of chaotic oscillators. *Phys. Rev. Lett.* **76**, 1804–1808 (1996)
88. E. Kononov, Visual Recurrence Analysis (2006), www2.netcom.com/~eugenek/page6.html
89. T. Chelidze, O. Lursmanashvili, T. Matcharashvili, N. Varamashvili, N. Zhukova, E. Mepharidze, High order synchronization of stick-slip process: experiments on spring-slider system. *Nonlinear Dyn.* (2009). doi:[10.1007/s11071-009-9536-6](https://doi.org/10.1007/s11071-009-9536-6)
90. N. Varamashvili, T. Chelidze, O. Lursmanashvili, Phase synchronization of slips by periodical (tangential and normal) mechanical forcing in the spring-slider model. *Acta Geophys.* **56**, 357–371 (2008)
91. O. Ben-David, S. Rubinstein, J. Fineberg, Stick-slip and the evolution of frictional strength. *Nature* **463**, 76–79 (2010). doi:[10.1038/nature08676](https://doi.org/10.1038/nature08676)
92. P. Reasenberg, Second-order moment of central California seismicity, 1969–1982. *J. Geophys. Res.* **90**, 5479–5495 (1985)
93. M. Matthews, P. Reasenberg, Statistical methods for investigating quiescence and other temporal seismicity patterns. *Pure Appl. Geophys.* **126**, 357–372 (1988)
94. W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, *Numerical Recipes. The Art of Scientific Computing* (Cambridge University Press, New York, 1997)
95. R.N. Vakarchuk, R.E. Tatevossian, Zh. Ya. Aptekman, V.V. Bykova, The 1991 Racha earthquake, Caucasus. *Izvestiya, Phys. Solid Earth* **49**, 653–659 (2013)
96. G. Sobolev et al., The optical disk “Spitak earthquake of 1988”. *Earthquake Hazard Risk* **6**, 231–238 (1996)
97. D. Stich, J. Almendros, V. Jiménez, F. Mancilla, E. Carmona, Ocean noise triggering of rhythmic long period events at Deception Island volcano. *Geophys. Res. Lett.* **38**, L22307 (2011). doi:[10.1029/2011GL049671](https://doi.org/10.1029/2011GL049671)
98. K. Obara, Nonvolcanic deep tremors associated with subduction in Southwest Japan. *Science* **296**, 1679–1681 (2002)
99. Z. Peng, J.E. Vidale, A.G. Wech, R.M. Nadeau, K.C. Creager, Remote triggering of tremor along the San Andreas fault in central California. *J. Geophys. Res.* **114**, B00A06 (2009). doi:[10.1029/2008JB006049](https://doi.org/10.1029/2008JB006049)
100. C.-J. Wang, M. Manga, Earthquakes and water, in *Lecture Notes in Earth Sciences* (Springer, 2010)
101. J. Costain, J. Bollinger, Review: research results in hydroseismicity from 1987 to 2009. *Bull. Seismol. Soc. Am.* **100**, 1841–1858 (2010). doi:[10.1785/0120090288](https://doi.org/10.1785/0120090288)
102. D. Simpson, Triggered earthquakes. *Ann. Rev. Earth Planetary Sci.* **14**, 21–42 (1986)
103. P. Talwani, On nature of reservoir-induced seismicity. *Pure Appl. Geophys.* **150**, 473–492 (1997)

104. Reservoirs and Seismicity: State of knowledge. Bulletin of International Commission on Large Dams, #137, (Paris, 2011)
105. J. Peinke, T. Matcharashvili, T. Chelidze, J. Gogiashvili, A. Nawroth, O. Lursmanashvili, Z. Javakhishvili, Influence of periodic variations in water level on regional seismic activity around a large reservoir: field data and laboratory model. *Phys. Earth Plan. Interiors* **156**, 130–142 (2006)
106. T. Chelidze, V. de Rubeis, T. Matcharashvili, P. Tosi, Dynamical changes induced by strong electromagnetic discharges in earthquakes' waiting time distribution at the Bishkek Test Area (Central Asia), in *Synchronization and Triggering: From Fracture to Earthquake Processes*, ed. by V. de Rubeis, Z. Czechowski, R. Teisseyre (Springer, Berlin, 2010), pp. 339–360
107. T. Matcharashvili, T. Chelidze, V. Abashidze, N. Zhukova, E. Meparidze, Changes in dynamics of seismic processes around Enguri high dam reservoir induced by periodic variation of water level, in *Synchronization and Triggering: From Fracture to Earthquake Processes*, ed. by V. de Rubeis, Z. Czechowski, R. Teisseyre (Springer, Berlin, 2010), pp. 273–286
108. T. Matcharashvili, T. Chelidze, V. Abashidze, N. Zhukova, E. Meparidze, Evidence for changes in the dynamics of Earth crust tilts caused by the large dam construction and reservoir filling at the Enguri dam international test area (Georgia). *Nonlinear Dyn.* (2011). doi:[10.1007/s11071-010-9930-0](https://doi.org/10.1007/s11071-010-9930-0)
109. N. Jones, The quake machine. *New Scientists*. June 30 (2001), 2297, pp. 34–37
110. T. Chelidze, V. De Rubeis, T. Matcharashvili, P. Tosi, Influence of electro magnetic strong discharges on the dynamics of earthquakes time distribution at the Bishkek test area. *Ann. Geophys.* **49**, 989–1000 (2006)
111. V. Keilis-Boirok, A. Gabrielov, A. Soloviev, Geocomplexity and earthquake prediction, in *Encyclopedia of Complexity and System Science*, ed. by R. Meyers (Springer, New York, 2009), pp. 4178–4194

Chapter 11

Long Time-Scale Recurrences in Ecology: Detecting Relationships Between Climate Dynamics and Biodiversity Along a Latitudinal Gradient

Raphaël Proulx, Lael Parrott, Lenore Fahrig, and David J. Currie



Abstract Climate is an important driver of ecological dynamics. However, many quantitative methods still ignore the fact that both ecological and climatic dynamics are inherently non-linear. While temporal variability is commonly measured as

R. Proulx (✉)

Canada Research Chair in Ecological Integrity, Centre de recherche sur les interactions bassins versants-écosystèmes aquatiques, Université du Québec à Trois-Rivières, 3351 Boul. Des Forges, Trois-Rivières, QC, Canada G9A 5H7
e-mail: raphael.proulx@uqtr.ca

L. Parrott

Earth and Environmental Sciences, Geography and Biology, Irving K Barber School of Arts and Sciences, University of British Columbia, Okanagan Campus, 3333 University Way, Kelowna, BC, Canada V1V 1V7
e-mail: lael.parrott@ubc.ca

L. Fahrig

Geomatics and Landscape Ecology Laboratory, Department of Biology, Carleton University, 1125 Colonel By Drive, Ottawa, ON, Canada K1S 5B6
e-mail: lenore_fahrig@carleton.ca

D.J. Currie

Department of Biology, University of Ottawa, 30 Marie Curie street, Ottawa, ON, Canada K1N 6N5
e-mail: dcurrie@uottawa.ca

the standard deviation of the records in a time-series, temporal determinism and predictability can be measured using the approach of Recurrence Plots–Recurrence Quantification Analysis (RP–RQA). In this study, we explore the relationship between climate dynamics and biodiversity of two taxonomic groups (mammal species and tree family richness) over the North- and South-American landmasses. We found that higher biodiversity levels in both taxonomic groups are associated to lower values of climate predictability, determinism and variability in monthly temperature data. Our results also revealed a multiplicity of climate–biodiversity relationships, suggesting that the mechanisms underlying large-scale geographic variations in biodiversity may be more complex than originally envisioned.

11.1 Background

Ecology is the study of how organisms relate to one another and to their surroundings. Because organisms interact in many ways with both their biotic and abiotic environment, ecological dynamics may be highly non-linear and have inspired expressions such as: “on the edge of chaos” [1], “dynamically transient” [2], and “subject to critical transition” [3]. Yet, when confronted with such an array of dynamical behaviors, many ecologists do not know how or what to analyze in a time-series. In fact, most time-series in ecology are still characterized using mean and standard deviation measures, thus discarding information that may (or may not) be relevant to the understanding of ecological processes.

Climate is an important driver of ecological dynamics. Independently of biotic factors such as competition for resources, reproductive success, or predator–prey interactions, populations tend to be more abundant if climatic conditions are favorable to them and to decline if conditions are too harsh. Community ecology in particular has a long history of relating the global geographic distribution of species richness and diversity to contemporary climatic conditions [4]. Early successes of correlating taxonomic biodiversity to climatic variables, such as the long-term mean of annual temperature and precipitation, have bolstered the search for a mechanistic underpinning of these relationships (reviewed in [5]).

Other studies have emphasized that time-series of climatic variables can be described and interpreted in contrasting ways. Using long time-series of air temperatures at a spatial resolution of 0.5° over the globe, von Bloh et al. [6] showed that even though monthly temperatures at low latitudes do not fluctuate much, seasonal and inter-annual temperature dynamics are harder to forecast near the equator than towards the poles. In another study conducted in the Qinghai-Tibet plateau region, Zhao et al. [7] showed that the temporal determinism and predictability of temperature dynamics has decreased over recent decades, more prominently so in croplands and urbanized areas. The above studies have in common that they both used the method of recurrence plots to describe the determinism and predictability of time-series.

Recurrence plots are an excellent method for detecting dynamical patterns in time series [8], and, coupled with recurrence quantification analysis, permit the characterisation and detection of deterministic structures that may not be captured using classical statistical methods. Recurrence analysis should be of particular interest to ecologists as it is well-adapted to the analysis of short time series and, in its multivariate form, provides a valuable tool for detecting and describing nonlinear relationships between dynamical processes. In this chapter, we demonstrate the use of recurrence quantification analysis to explore the relationship between climate and biodiversity over large geographic scales.

11.2 Recurrence Plots and Recurrence Quantification Analysis

Recurrence plots (RP) were developed in statistical physics by Eckmann et al. [9] to visualize the system states in a phase space. System states are reconstructed by the method of time-delay embedding (see [8]). To illustrate how embedding works, let us create a time-series of plant species abundances surveyed over six consecutive years: 15, 25, 5, 7, 10, and 28 individuals per square meter. We define the dimension (the number of independent variables needed to capture a particular dynamic such as plant abundances e.g., $m = 3$) and time-delay (the time lag which minimizes autocorrelation, e.g., $d = 2$) of the embedding. Using these two parameters we reconstruct the system states from the above time-series, yielding the two successive states of: [15, 5, 10] and [25, 7, 28]. If n is the time-series length, one can only reconstruct $n - d(m - 1)$ system states. For example, embedding the above time-series with parameters of $m = 4$ and $d = 1$, three reconstructed states are obtained: [15, 25, 5, 7], [25, 5, 7, 10] and [5, 7, 10, 28]. The RP is then a square matrix, with time on both axes, of pairwise Euclidean distances between the reconstructed system states to which a distance threshold (*thresh*) is applied. A black dot in the plot indicates the presence of a “recurrence”; i.e., two system states that differ by less than the threshold (Fig. 11.1). If the threshold is too large, most states are considered similar and the plot is mainly black. Conversely, if the threshold is too small, most states are considered different and the plot is mainly white. RPs can be visually interpreted to detect non-stationary dynamics with either smooth or abrupt transitions, as well as identifying the presence of periodic and non-periodic processes (e.g., [10]). Several algorithms have been proposed to assist the analyst with the choice of m , d and *thresh* parameters [8].

Recurrence quantification analysis (RQA) forms a toolbox of mathematical measures for characterizing RPs [11]. In particular, RQA is capable of detecting the signature of chaotic dynamics in long time-series [12, 13]. In the ecological sciences, however, the time-series are usually too short or stochastic for detecting chaos or assessing the dimensionality of the system (i.e., the minimum number of variables required for capturing the system’s dynamics). Nevertheless, RQA allows

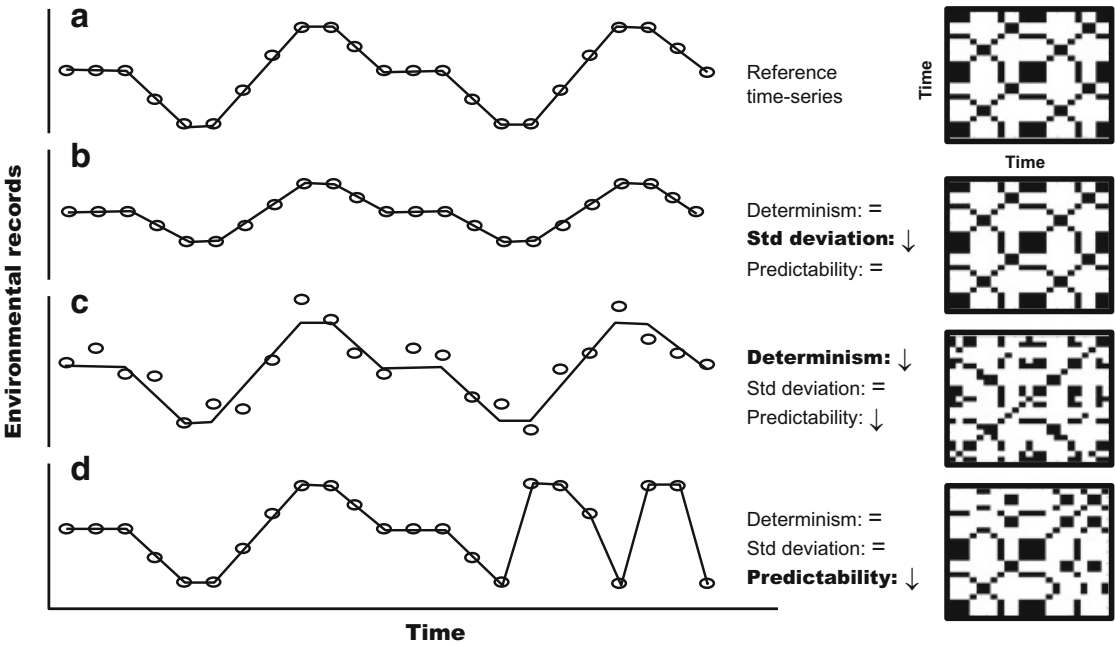


Fig. 11.1 Representation of a reference time-series of environmental (ecological or climatic) records (a) to which modifications were applied by: reducing variance (b), increasing stochasticity (c), introducing regime shifts (d). Next to the time-series are reported the effects of these modifications (↓ or =) on measures of temporal variability (std. deviation), determinism and predictability. Recurrence plots (with $m = 1$, $d = 1$, and 10 % recurrence rate) are displayed on the right-hand side

Table 11.1 Three fundamental measures of ecological and climatic dynamics

Measure	Definition	Measure	Comment
Temporal variability	Fluctuation of environmental records in a time-series	Standard deviation (<i>SD</i>) of the environmental records	Temporal variability depends on the scaling of the records and has no upper bound
Temporal determinism	Recurrence of two environmental states over time (see text)	Percentage of recurrences that are part of diagonal lines with 2 or more recurrences (<i>DET</i>)	Temporal determinism is a reciprocal measure of temporal stochasticity
Temporal predictability	Time period over which environmental states can be predicted into the future	The average length of diagonal lines in the RP (<i>length</i>)	Temporal predictability can be low even if determinism is high

us to calculate dynamical measures (Table 11.1) that are *a priori* unrelated to those of central tendency and variance and which may provide some indication of the degree of determinism in a system, even for short time series.

11.3 Ecological and Climatic Dynamics

Ecological and climatic dynamics can be characterized along three major axes (Table 11.1): (1) temporal variability, (2) temporal determinism, and (3) temporal predictability. Temporal variability is most typically measured as the standard deviation of the environmental records in a time-series. From a mathematical standpoint, temporal variability is not independent of the time-series' mean and does not account for temporal periodicities. The two additional axes, temporal determinism and temporal predictability, can be represented by dynamical measures calculated using the approach of Recurrence Plots–Recurrence Quantification Analysis (RP–RQA). Determinism is the reciprocal of stochasticity. Temporal determinism represents the percentage of all recurrences in the RP that are part of diagonal lines. A diagonal line is one that contains two or more recurrences forming an uninterrupted sequence running parallel to the line of identity; i.e., the longest diagonal line splitting the RP in two halves. Determinism will be zero if all recurrences are single dots. The second measure, temporal predictability, represents the average forecasting horizon of the system states and is expressed in time units. Temporal predictability is calculated as the mean of the frequency distribution of diagonal lines in the RP, excluding the line of identity.

Other RP–RQA measures, such as the tendency of records in time-series to drift away from the long term mean (*Trend*; [8]), may also be useful for characterizing transiency in ecological and climatic dynamics (see Chap. 12). In general, the recurrence plot approach provides alternative ways of describing dynamical systems but is not without its own shortcomings. RP–RQA measures cannot be interpreted beyond sampling limitations in terms of missing data, time-series length and temporal resolution. These limitations are discussed in detail in Marwan [14].

11.4 A Case Study of Climate–Biodiversity Relationships

11.4.1 Context

The relationship between climate and biodiversity is one of the most pervasive patterns of community ecology. Several studies, covering a range of ecosystems, have shown that large-scale geographic variations in taxonomic biodiversity correlate to energy- and water-related climatic variables (reviewed in [15, 16]). Over large geographic distances, the variation in plant and mammal biodiversity has been associated to the long-term mean or temporal variability of temperature and precipitation (e.g., [17–20]).

Non-spatial competition models in community ecology suggest that, unless environmental conditions vary unpredictably, even small functional differences among species may eventually lead to competitive exclusion (e.g., [4, 21]). Thus, one could hypothesize that lower levels of temporal determinism and predictability

in climatic dynamics may slow the rate of competitive exclusion, allowing more species to coexist. Lower climatic predictability would thus translate into greater environmental heterogeneity in time, creating more opportunities for species coexistence.

In this case study, we explore the validity of this premise through the use of RP–RQA measures to quantify temporal determinism and predictability in climatic time series. We then explore the relationships between biodiversity and climate dynamics at different locations along a north–south gradient. For the first time, these relationships will be contrasted against those obtained using conventional measures of climatic dynamics (i.e., long-term mean and standard deviation).

11.4.2 *Methods*

We used as biodiversity variables the number of mammal species [22] and angiosperm tree families [18] over the North- and South-American landmasses at a resolution of 1° and 2°, respectively. We extracted monthly climatic time-series from the global dataset compiled by the Climate Research Unit on grid-cells of 0.5°, for the period 1920–2002 (CRU TS 2.1; [23]). We then calculated the mean and standard deviation of monthly temperature (°C) and precipitation (mm) time-series. Finally, we built two recurrence plots (RPs) for each grid-cell, one for monthly temperature and the other for precipitation. Following von Bloh et al. [6], we chose an embedding dimension of $m = 3$, a time-delay of $d = 1$, as well as adjustable threshold values (*thresh*) allowing us to fix the recurrence rate to 10 % in the RPs. We used the CRP Matlab toolbox [8] to calculate temporal determinism and predictability measures.

For each climatic variable (temperature and precipitation), we spatially averaged the measures of determinism, predictability, mean and standard deviation at the grid-cell resolution of each biodiversity dataset. We graphed bivariate climate–biodiversity relationships and calculated their coefficient of determination (R^2) using recursive trees. A recursive tree is comparable to a regression model without the statistical assumptions of linearity and normality of the predictors. Recursive trees were fitted using the R package “rpart” version 2.15.3 [24].

11.5 Results and Discussion

The three dynamical measures—temporal variability, determinism and predictability measures—were weakly correlated with each other for the precipitation data (Fig. 11.2a), suggesting that determinism and predictability contain important dynamical information not captured by the standard deviation. The highest association was between determinism and predictability (Pearson’s $r^2 = 0.46$) and the lowest between determinism and variability (Pearson’s $r^2 = 0.13$).

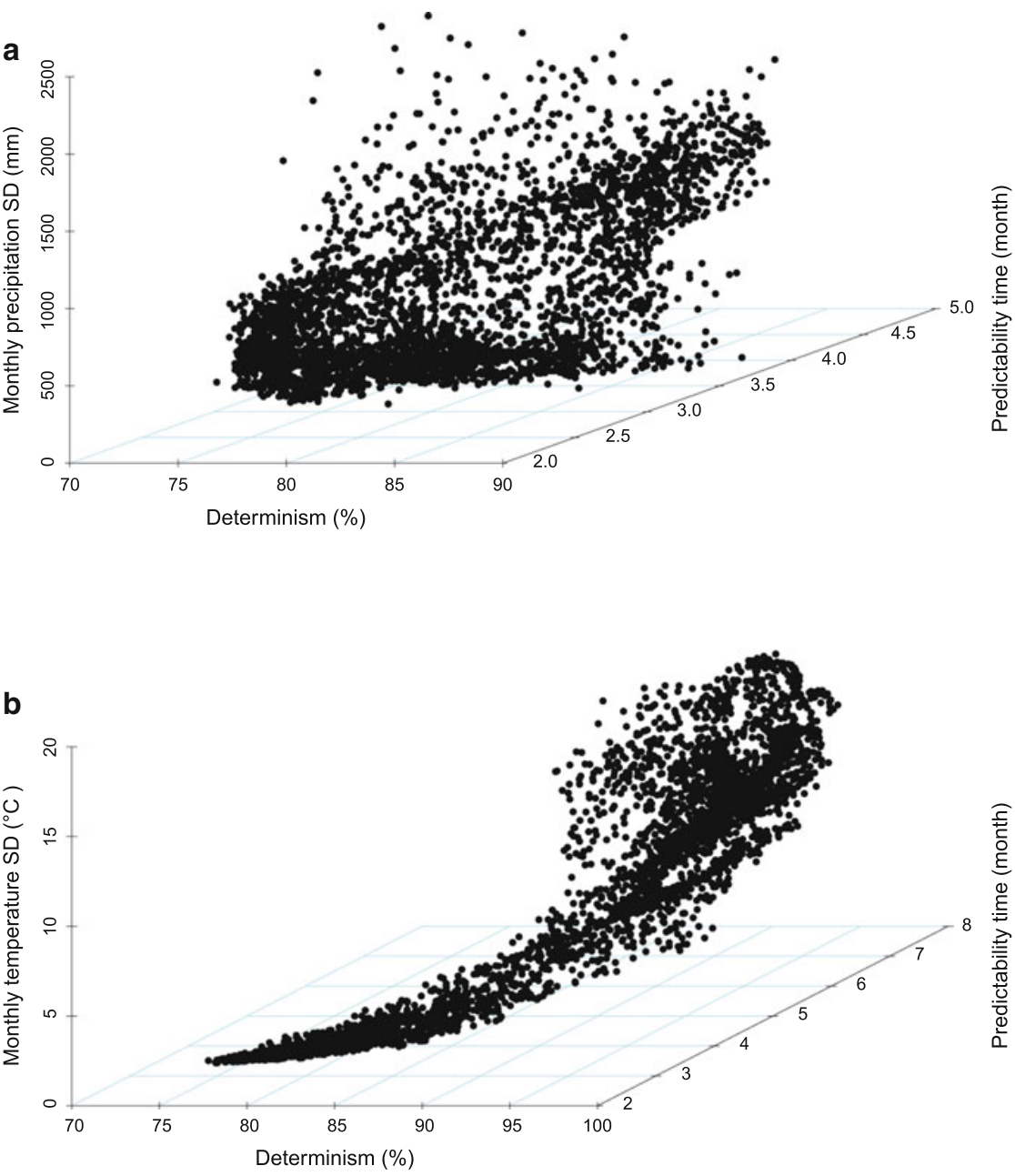


Fig. 11.2 Relationships among measures of temporal variability (standard deviation; SD), determinism and predictability for (a) precipitation and (b) temperature data. Monthly climatic time-series covered the period 1920–2002 at a spatial resolution of 0.5° over North- and South-America. Each point represents one location

In comparison, the three measures were rather strongly correlated for the temperature data (Fig. 11.2b), with the highest association between determinism and predictability (Pearson’s $r^2 = 0.85$) and the lowest between predictability and variability (Pearson’s $r^2 = 0.66$).

Bivariate biodiversity–climate relationships for the two datasets revealed contrasting patterns (Fig. 11.3), extending from none (e.g., mammal species richness vs precipitation determinism.) to strongly linear (e.g., mammal species richness

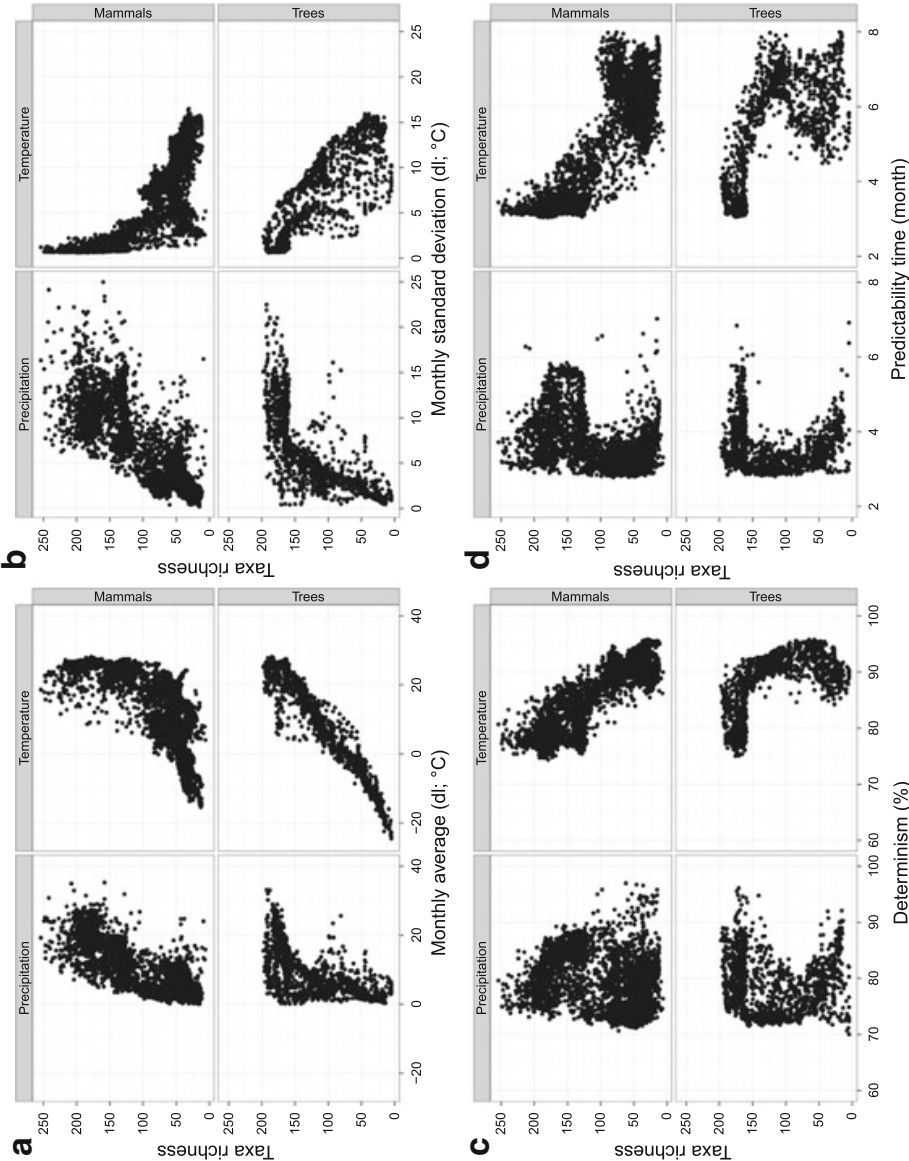


Fig. 11.3 Relationship between dynamical measures of climatic time-series (temperatures, precipitations) and taxonomic richness (mammal species, tree families) for North- and South-America. Figure panels show the relationships with: (a) long-term mean (monthly average), (b) temporal variability (monthly standard deviation), (c) temporal predictability (average forecasting length in months). Monthly mean and standard deviation measures of precipitation data are expressed in decilitres (dl). In each panel, *left* and *right* columns show relationships to precipitation and temperature data, respectively, whereas *top* and *bottom* show relationships to mammal species richness and tree family richness, respectively

vs temperature mean.), and including asymptotic (e.g., tree family richness vs. precipitation standard deviation) as well as more complex functions (e.g., tree family richness vs. temperature predictability). Both the number of mammal species and tree families were positively related to long-term averages of monthly temperature and precipitation time-series (Fig. 11.3; panel A). No such generality existed for the other variables, although some patterns can be observed (Fig. 11.3; panels B–D). At higher biodiversity levels, taxonomic richness decreased with increasing temporal variability, determinism and predictability of monthly temperatures. In comparison, richness related positively to the temporal variability of monthly precipitation, especially at lower biodiversity levels (Fig. 11.3; panel B).

How then might the global geographic distribution of biodiversity relate to the determinism and predictability of climatic variables? Our preliminary findings suggest that biodiversity increases with increasing temperature stochasticity and decreasing predictability, at least in the upper range of biodiversity values. Hubbell [25] suggested that species can coexist because they are essentially indistinguishable from one another in their realized demographic traits; that is the long-term persistence of species is independent of per capita birth–death rates. Subsequent tests have detected differences, although they are probably small (e.g., [26]). Alternatively, Clark et al. [27] proposed that species coexistence is possible because individuals within species differ and interact with their environment in many different ways. Maintenance of biodiversity in the above theories requires that, independently of the specific mechanisms involved, within-species variability in demographic traits is large in comparison to between-species variability (i.e., that ecological niches are high-dimensional). Both theories thus suggest a tight coupling between the dimensionality of ecological niches and environmental dynamics.

In atmospheric sciences, high-dimensional climatic systems are typified by dynamical measures of low temporal determinism and predictability [28]. Species may have higher-dimensional ecological niches if their resources respond to varying (e.g., less deterministic and predictable) climatic conditions, providing more opportunities for the niches of species to be distinguished [21]. Consequently, species may show different responses to their common varying climate and experience strongest intraspecific competition when favored by the environment, thus buffering population growth. This equalizing mechanism of species coexistence has been termed the “storage effect” (e.g., [21]). The storage effect has received support in a recent study by Usinowicz et al. [29] who showed that coexistence in tropical forests is facilitated through asynchronous variation in seed production among tree species. Runkle [30] had previously argued that the storage effect could explain the latitudinal gradient in forest diversity because in his own words: “Winter decreases effective environmental variation and imposes a synchronization in seasonal phenology upon species of trees in temperate zones. Therefore the probability that a rare species may encounter the unusual favorable period which will enable it to persist in the community is reduced”. It remains to be examined if latitudinal differences exist in the strength of the storage effect.

The multiplicity of climate–biodiversity relationships seems to call for a multiplicity of hypotheses, but this may only confound the search for a general

Table 11.2 Coefficients of determination (R^2) calculated from the recursive trees fitted to each bivariate climate–biodiversity relationship

Measure	Precipitation		Temperature	
	Mammal species	Tree family	Mammal species	Tree family
Long-term average (Mean)	0.617	0.461	0.723	0.912
Temporal variability (SD)	0.701	0.687	0.831	0.739
Temporal determinism (DET)	0.123	0.054	0.802	0.438
Temporal predictability (L)	0.219	0.049	0.809	0.528

mechanism. When considering only the bivariate relationship strengths, one might conclude that mean temperature controls the number of tree families in the Americas, but that the geographic distribution of mammal species richness is determined by temporal dynamics in temperature (Table 11.2). In fact, 12 of the 16 measures of climatic dynamics were associated to nearly, or more than, 50 % of the geographic variation in biodiversity (Table 11.2). When the eight climatic variables were included in a model of recursive partitioning, the coefficient of determination attained values of 0.882 (88 %) and 0.922 (92 %) for mammals and trees, respectively. In models considering only the variables of mean precipitation and mean temperature, the coefficient of determination was distinctly lower for mammals ($R^2 = 0.785$), but only slightly lower for trees ($R^2 = 0.912$).

Alternatively, the multiplicity of relationships may indicate that biodiversity is constrained by more fundamental climatic processes, such as the latitudinal gradient in solar radiation dynamics. At the global scale, yearly solar irradiation at the top of the atmosphere is strongly negatively correlated to absolute latitude. Solar irradiation on Earth defines the maximum amount of energy that can, ultimately, be transformed into biomass and shared among species (the species-energy hypothesis; e.g., [15, 16, 31]). Moreover, latitudinal differences in solar heating drive the vertical and horizontal motion of air masses, which in turn entrain the broad-scale temporal dynamics in temperature and precipitation. Thus, over large geographical regions, different climatic variables may all relate to the same underlying variable, solar irradiation. In partial support of this idea, if we enter only the latitude coordinates of grid-cells in a recursive tree model, the coefficients of determination we obtained were as high as $R^2 = 0.857$ and $R^2 = 0.934$ for mammals and trees, respectively.

In summary, we found evidence that higher biodiversity levels of two taxonomic groups are associated with lower values of predictability, determinism, variability and higher long-term means in monthly temperature data. Our results also revealed a multiplicity of climate–biodiversity relationships, suggesting that identifying the specific mechanisms of large-scale geographic variations in biodiversity may be more complex than originally envisioned. In this context, the study of Nevo [32] is instructive as he compared the biodiversity of a large number of plant and animal taxonomic clades on two opposite south- and north-facing slopes at the Evolution Canyon in Mount Carmel, Israel. He found that plants (angiosperms) and animals (many insect and arthropod groups, reptiles and birds) were more diverse on the south-facing slope than on the north-facing slope. Genetic diversity was also higher

on the south-facing slope. Nevo described the south-facing (high biodiversity) slope in the following words: "... warmer and drier (receiving up to 300 % more solar radiation), micro-climatically less predictable than the north-facing slope". He added: "The south-facing slope represents a 'spatiotemporally broader niche' ...". [32]. We propose that representation of a spatiotemporally broader niche may be extended to include large-scale variations in climate dynamics.

11.6 Perspectives

The RP-RQA approach can contribute to the investigation of other research hypotheses in community ecology. For example, the mismatch hypothesis (e.g., [33]) states that the survival of populations and communities is impaired if ecological processes, such as plant greening and flowering, animal migration, or breeding events, are desynchronized over time. Loss of seasonal or inter-annual predictability in the timing of climatic events is known to be detrimental to the functioning of plant and animal species [34, 35]. In this context, the vulnerability of communities to temporal mismatch may depend on the predictability of climatic and ecological processes. Temporal mismatch between ecological and climatic processes could be readily assessed with the approach of cross-recurrence plots (CRP; [8]).

Another hypothesis posits that species-rich communities host a variety of life strategies that respond differently to environmental conditions and contribute to ecological functioning in different ways, thus increasing the temporal stability of ecosystems (e.g., [36, 39]). Considering that the concept of stability is necessarily multifaceted, different measures of temporal variability, determinism and predictability are needed to uncover the mechanisms of biodiversity–stability relationships. For example, using a multi-species competition model with environmental forcing, Proulx et al. [10] found a positive biodiversity–stability relationship when community stability was defined using measures of temporal determinism and predictability. No such relationship was observed when measures of temporal variability were used to define community stability [10]. Multivariate recurrence plots (MRP; [37]) can be used to calculate measures of temporal determinism and predictability of dynamical systems with many interacting variables.

In conclusion, recurrence plots and recurrence quantification analysis hold promise for describing and detecting dynamical patterns in ecological time series. They provide a simple alternative to other methods of non-linear analysis (e.g., attractor reconstruction, Lyapunov exponents, correlation dimension analysis) that require long time series and have rarely provided convincing results for ecological data (e.g., [38]). In contrast, the RP-RQA approach provides a more subtle description of determinism and predictability in a time series, and thus may prove more useful than other approaches for detecting deterministic structure in ecological data. Ecological systems have a dynamic that is inherently non-linear; however, many quantitative methods in ecology ignore this fundamental aspect of the time series to which they are applied. RP-RQA thus provides a good complement to the ecologist's quantitative toolbox.

References

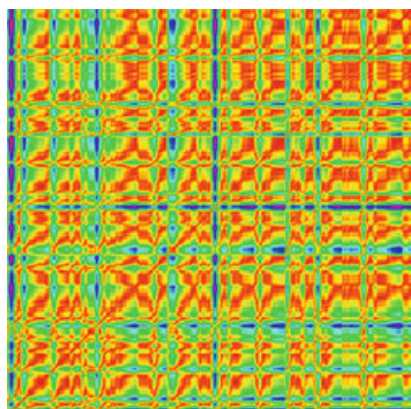
1. P. Turchin, S.P. Ellner, Living on the edge of chaos: population dynamics of Fennoscandian voles. *Ecology* **81**(11), 3099–3116 (2000)
2. A. Hastings, Transients: the key to long-term ecological understanding? *Trends Ecol. Evol.* **19**(1), 39–45 (2004)
3. M. Scheffer, J. Bascompte, W.A. Brock, V. Brovkin, S.R. Carpenter, V. Dakos, G. Sugihara, Early-warning signals for critical transitions. *Nature* **461**(7260), 53–59 (2009)
4. M.A. Huston, M.A. Huston, *Biological Diversity: The Coexistence of Species* (Cambridge University Press, Cambridge, 1994)
5. D.J. Currie, G.G. Mittelbach, H.V. Cornell, R. Field, J.F. Guégan, B.A. Hawkins, J.R.G. Turner, Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. *Ecol. Lett.* **7**(12), 1121–1134 (2004)
6. W. Von Bloh, M.C. Romano, M. Thiel, Long-term predictability of mean daily temperature data. *Nonlinear Processes Geophys.* **12**(4), 471–479 (2005)
7. Z.Q. Zhao, S.C. Li, J.B. Gao, Y.L. Wang, Identifying spatial patterns and dynamics of climate change using recurrence quantification analysis: a case study of Qinghai-Tibet plateau. *Int. J. Bifurcation Chaos* **21**(04), 1127–1139 (2011)
8. N. Marwan, M. Carmen Romano, M. Thiel, J. Kurths, Recurrence plots for the analysis of complex systems. *Phys. Rep.* **438**(5), 237–329 (2007)
9. J.P. Eckmann, S.O. Kamphorst, D. Ruelle, Recurrence plots of dynamical systems. *EPL (Europhys. Lett.)* **4**(9), 973 (1987)
10. R. Proulx, P. Côté, L. Parrott, Use of recurrence analysis to measure the dynamical stability of a multi-species community model. *Eur. Phys. J. Spec. Top.* **164**(1), 117–126 (2008)
11. J.P. Zbilut, C.L. Webber, Embeddings and delays as derived from quantification of recurrence plots. *Phys. Lett. A* **171**(3), 199–203 (1992)
12. L.L. Trulla, A. Giuliani, J.P. Zbilut, C.L. Webber, Recurrence quantification analysis of the logistic equation with transients. *Phys. Lett. A* **223**(4), 255–260 (1996)
13. C. Mocenni, A. Facchini, A. Vicino, Comparison of recurrence quantification methods for the analysis of temporal and spatial chaos. *Math. Comput. Model.* **53**(7), 1535–1545 (2011)
14. N. Marwan, How to avoid potential pitfalls in recurrence plot based data analysis. *Int. J. Bifurcation Chaos* **21**(04), 1003–1017 (2011)
15. B.A. Hawkins, R. Field, H.V. Cornell, D.J. Currie, J.F. Guégan, D.M. Kaufman, J.R. Turner, Energy, water, and broad-scale geographic patterns of species richness. *Ecology* **84**(12), 3105–3117 (2003)
16. A. Clarke, K.J. Gaston, Climate, energy and diversity. *Proc. Roy. Soc. B Biol. Sci.* **273**(1599), 2257–2266 (2006)
17. P. Andrews, E.M. O'Brien, Climate, vegetation, and predictable gradients in mammal species richness in southern Africa. *J. Zool.* **251**(2), 205–231 (2000)
18. A.P. Francis, D.J. Currie, A globally consistent richness–climate relationship for angiosperms. *Am. Nat.* **161**(4), 523–536 (2003)
19. H. Kreft, W. Jetz, Global patterns and determinants of vascular plant diversity. *Proc. Natl. Acad. Sci.* **104**(14), 5925–5930 (2007)
20. R. Carrara, D.P. Vázquez, The species–energy theory: a role for energy variability. *Ecography* **33**(5), 942–948 (2010)
21. P. Chesson, Mechanisms of maintenance of species diversity. *Annu. Rev. Ecol. Syst.* **31**(1), 343–366 (2000)
22. B.D. Patterson, G. Ceballos, W. Sechrest, M.F. Tognelli, T. Brooks, L. Luna, P. Ortega, I. Salazar, B.E. Young, *Digital Distribution Maps of the Mammals of the Western Hemisphere*, version 2.0. NatureServe, Arlington, 2005
23. T.D. Mitchell, P.D. Jones, An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *Int. J. Climatol.* **25**(6), 693–712 (2005)

24. T.M. Therneau, B. Atkinson, B. Ripley, *rpart: Recursive Partitioning*. R package version, 3, (2010) pp. 1–46
25. S.P. Hubbell, Tree dispersion, abundance, and diversity in a tropical dry forest. *Science* **203**(4387), 1299–1309 (1979)
26. J. Chave, Neutral theory and community ecology. *Ecol. Lett.* **7**(3), 241–253 (2004)
27. J.S. Clark, M. Dietze, S. Chakraborty, P.K. Agarwal, I. Ibanez, S. LaDeau, M. Wolosin, Resolving the biodiversity paradox. *Ecol. Lett.* **10**(8), 647–659 (2007)
28. K. Fraedrich, Estimating the dimensions of weather and climate attractors. *J. Atmos. Sci.* **43**(5), 419–432 (1986)
29. J. Usinowicz, S.J. Wright, A.R. Ives, Coexistence in tropical forests through asynchronous variation in annual seed production. *Ecology* **93**(9), 2073–2084 (2012)
30. J.R. Runkle, Synchrony of regeneration, gaps, and latitudinal differences in tree species diversity. *Ecology* **70**(3), 546–547 (1989)
31. D.H. Wright, Species–energy theory: an extension of species–area theory. *Oikos* **41**(3), 496–506 (1983)
32. E. Nevo, Asian, African and European biota meet at ‘Evolution Canyon’ Israel: local tests of global biodiversity and genetic diversity patterns. *Proc. Roy. Soc. Lond. Ser. B Biol. Sci.* **262**(1364), 149–155 (1995)
33. J.M. Durant, D.O. Hjermann, G. Ottersen, N.C. Stenseth, Climate and the match or mismatch between predator requirements and resource availability. *Climate Res.* **33**(3), 271 (2007)
34. M.E. Visser, C. Both, Shifts in phenology due to global climate change: the need for a yardstick. *Proc. Roy. Soc. B Biol. Sci.* **272**(1581), 2561–2569 (2005)
35. G.R. Walther, E. Post, P. Convey, A. Menzel, C. Parmesan, T.J. Beebee, F. Bairlein, Ecological responses to recent climate change. *Nature* **416**(6879), 389–395 (2002)
36. A.R. Ives, S.R. Carpenter, Stability and diversity of ecosystems. *Science* **317**, 58–62 (2007)
37. M.C. Romano, M. Thiel, J. Kurths, W. von Bloh, Multivariate recurrence plots. *Phys. Lett. A* **330**(3), 214–223 (2004)
38. A. Hastings, C.L. Hom, S. Ellner, P. Turchin, H.C.J. Godfray, Chaos in ecology: is mother nature a strange attractor? *Annu. Rev. Ecol. Syst.* **24**(1), 1–33 (1993)
39. R. Proulx, C. Wirth, W. Voigt, A. Weigelt, C. Roscher, S. Attinger, J. Baade et al. Diversity promotes temporal stability across levels of ecosystem organization in experimental grasslands. *PLoS one.* **5**(10), e13382 (2010)

Chapter 12

Recurrence Quantification and Recurrence Network Analysis of Global Photosynthetic Activity

Holger Lange and Sven Boese



Abstract In this chapter, the potential of Recurrence Analysis (RA) for applications in the biogeosciences is demonstrated. We investigate the fraction of absorbed photosynthetically active radiation (FAPAR), an index based on multispectral reflectance properties of land surfaces which relates to the carbon uptake by plants. FAPAR is available with global coverage from satellites. We combine observations from two sensors, SeaWiifs on board SeaStar and MERIS on board Envisat, to produce time series with 10 days resolution for a period of 14 years (1998–2011) at a spatial resolution of 0.5° latitude \times 0.5° longitude. After careful quality checking and gap-filling, more than 30,000 individual time series are obtained covering all terrestrial ecosystems and climates apart from Antarctica and major deserts. To characterize the different dynamical behaviour as a function of spatial location, we employ Recurrence Quantification Analysis (RQA) and Recurrence Network Analysis (RNA). They deliver detailed information on the nonlinear dynamics in phase space through embedding. RQA and network measures are calculated for individual time series using identical recurrence parameters, and results are

H. Lange (✉)

Norwegian Forest and Landscape Institute, P.O. Box 115, 1431 Ås, Norway

e-mail: holger.lange@skogoglandskap.no

S. Boese

Max-Planck-Institute for Biogeochemistry, P.O. Box 10 01 64, 07701 Jena, Germany

e-mail: sboese@bgc-jena.mpg.de

visualized on world maps. Taken together, the recurrence analysis leads to a partitioning of the terrestrial biosphere into regions with distinct dynamical patterns of photosynthetic activity.

12.1 Introduction

Satellite-based remote sensing provides time series of geophysical and biophysical variables which reflect the status of the biosphere. The observation and understanding of global biosphere dynamics is crucial for many purposes, including an understanding of ecological processes acting on different spatial and temporal scales, practical applications such as forest fire detection, deforestation including illegal logging, hazard monitoring, or on longer time scales also climate change effects.

Understanding and quantification of carbon dynamics is of utmost importance in this context. Uptake and release of carbon from vegetation represent two large fluxes of opposite direction. The sign of their difference determines whether an ecosystem, or the biosphere as a whole, acts as carbon sink or carbon source to the atmosphere, and thus the type of feedback to Earth's climate. Unfortunately, the differences of two large quantities of comparable magnitude obtained with reasonable uncertainties is plagued with a rather large (relative) inaccuracy. Every effort to reduce the imprecision of incoming and outgoing carbon fluxes is more than welcome. The problem is intriguing for at least two reasons: on one hand, processes are involved which act on rather disparate time scales, from turbulent exchange of CO₂ molecules at vegetation surfaces, quantified through Eddy Covariance measurements at a rate of 20 measurements per second, to changes in carbon stocks in forest soils which are significant at decadal scales only. On the other hand, feedback mechanisms exist which couple the sink strength of a given compartment (soils, the ocean) to the actual level of carbon input, with saturation effects which are partially unknown.

In this chapter, we use a data-oriented approach to investigate the spatiotemporal dynamics of vegetation activity. We use the *fraction of absorbed photosynthetically active radiation* (FAPAR) as a proxy for primary productivity, i.e. biomass increase rate. FAPAR has been utilized in multiple geoecological studies [1, 2] and was defined as *essential climate variable* for the *global climate observing system* [3].

In a broader context, these historical biomass changes can be related to land use changes, large wildfires, or changing local climate. Using FAPAR is one out of several possibilities. Another choice is the Normalized Difference Vegetation Index (NDVI), which is available from satellite measurements since the late 1970s. It is however, affected by a saturation effect at high biomass densities, rendering it problematic e.g. in tropical rainforests. Nevertheless, one attempt to characterize NDVI dynamics using Recurrence Quantification Analysis has been made previously [4].

From a modelling perspective, datasets of the global biosphere are challenging as they comprise periodic fluctuations, chaotic components and noise [5]. *Time series analysis* is therefore needed to systematically study their dynamics. A notorious problem with biosphere data is that the lengths of the considered time series are often too short for most methods of analysis. Instationarity and missing observations are further problems that are found in geoecological time series. Variables directly related to vegetation activity are expected to show nonlinear dynamics, such as threshold behaviour, since they are observables extracted from nonlinear dynamical systems. Thus, appropriate methods have to be chosen. *Recurrence Quantification Analysis* (RQA) is a particularly promising candidate for this task, as it delivers robust results on comparatively short time series and was specifically created for the analysis of nonlinear systems. RQA can track a variety of the system's dynamical properties such as recurring regular patterns, stationarities and dynamical complexity.

In this chapter, we combine two available FAPAR datasets from two sensors, namely SeaWiifs (1998–2005) and MERIS (2002–2011), i.e. we produce a 14-years record of global vegetation activity at a temporal resolution of 10 days (around 500 individual observations for each location). To stitch together two time series from different sensors is not without problems, as we discuss below. This investigation represents the very first analysis of the tailored data set. We use both RQA and *Recurrence Network Analysis* (RNA) on all individual time series, with a special emphasis on the qualitative differences between the insights obtainable from both approaches. RQA and RNA are jointly referred to as *Recurrence Analysis* (RA) in the remainder of this text. This chapter is organized as follows. In the next section, we present the two data sets and the peculiarities of gap-filling and concatenation. Then we provide a short overview of the methods utilized in this analysis, in particular RQA and RNA, where we present the quantities that are derived from the two methods and used in this study. Results are presented and discussed for some basic analysis methods, RQA and RNA applied to FAPAR. The final section concludes on the insights gained and provides an outlook on further investigations using the methods for globally covering remote sensing products in general.

12.2 Data Sets and Preprocessing

This section describes the two data sets for FAPAR which were combined to produce a continuous record of 14 years (1998–2011). Gap-filling and exclusion of impaired time series had to be performed in a first step. As two different sensors were used for the construction of the full-length FAPAR data set, the respective parts of the time-series were corrected for systematic and time-dependent deviations. As FAPAR quantifies the fraction of incoming radiation not returned to the atmosphere, it is always a real number bounded strictly by 0 and 1. However, negative values originated from the gap-filling procedure. These values were replaced by random numbers within the range of the assumed noise floor. Time series with a mean value

lower than the noise floor were excluded from the analysis. As a final step prior to the recurrence analysis, the time series were normalized to zero mean and unit standard deviation (z scores).

12.2.1 Data Basis

Many properties of plants like leaf area index and chlorophyll content can be obtained from field experiments, through direct inspection and in some cases in a destructive manner. Measurements from satellites in contrast necessarily rely on certain surrogates that are attainable via emission or reflection of electromagnetic radiation. Photosynthesis is a process where the absorption of incoming shortwave radiation takes place mainly in wavelengths of 400–500 nm and 600–700 nm. The fraction that is absorbed by the surface of the earth at a given place is dependent on plant density, plant type, season (phenology), health status of the vegetation, and water availability, among other factors. Photosynthesis is the process which leads to primary production of chemical energy by synthesizing sugar from water and carbon dioxide. Its quantification is of key interest for ecology.

The spectral signatures of photosynthesis are utilized by the vegetation index FAPAR (fraction of absorbed photosynthetically active radiation). Through the algorithm described in [6], the uptake of radiative energy by the biosphere is inferred. The product is provided by the *Joint Research Center* of the European Union in Ispra, Italy. The algorithms and corrections associated with the data-collection and correction were described in depth for SeaWifs [7] and for MERIS [8].

In essence, the computation of FAPAR is based on closing the energy balance of the 600–700 nm range in the canopy. Preprocessing of the index includes orbital and sun-angle related geometric corrections, and corrections for scattering and absorption in the atmosphere. In this study, data from two different satellite missions were used. The mission of the *Sea-viewing Wide Field-of-view Sensor* (SeaWifs) was launched onboard the *SeaStar* spacecraft in 1997, originally to monitor the development and status of algae in oceans on a global scale.¹ The FAPAR data-set of SeaWifs spans the period from 1998 to 2005. The mission ceased in 2006.

The *Medium Resolution Imaging Spectrometer* (MERIS) was a sensor on board the European *EnviSat* mission. It was launched in 2002, holding a wide range of environmental monitoring systems. The mission stopped with a malfunction in 2012. FAPAR data from MERIS are available from 2003 to 2011.

The FAPAR products have different spatial resolutions of 1.2 km (SeaWifs) and 1.5 km (MERIS) respectively, but were then aggregated to a pixel-size of $0.5^\circ \times 0.5^\circ$. Both sensors provided global coverage with a repeat cycle of at most 3 days (less at high latitudes); however, mainly due to cloud cover, the datasets are heavily compromised with gaps at this resolution. The final product used in this study was an

¹<http://oceancolor.gsfc.nasa.gov/SeaWiFS/BACKGROUND/>.

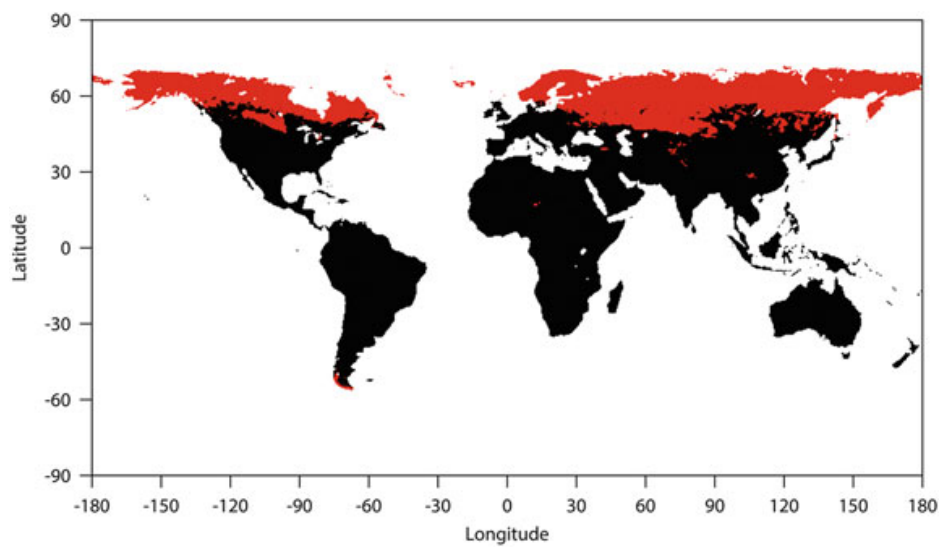


Fig. 12.1 Geographical distribution of pixels with more than 30 % gaps (in red). The northern latitudes are heavily affected by missing values due to persisting snow cover during the winter months

aggregation to three values per month. When merged, the resulting data set consists of 504 time steps.

Nevertheless, the resulting datasets still contain gaps at this stage. This affects mainly the northern latitudes with a persisting snow cover during winter. Figure 12.1 shows the spatial coverage of time series that contained more than 30 % gaps. Thus, a procedure to fill these gaps was required.

12.2.2 Gap-Filling

As with any remote sensing method based on optical information obtained from passive sensors, the FAPAR data-sets are plagued with gaps (no value retrieved) basically for two reasons: intense cloud cover or snow cover. They lead to a substantial fraction of missing values. Since even highly sophisticated gap-filling algorithms deteriorate at very high gap fractions, we decided to exclude all time series where the number of gaps in the combined series (SeaWifs and MERIS) exceeded 30 %. For the remainder, *Singular Spectrum Analysis* (SSA) was used. SSA is a method widely used for gap-filling [9, 10]. The method performs a spectral decomposition of the time series—the resulting components can then be used to estimate the signal for occurring gaps. As this invokes the lag-covariance matrix and is also based on embedding, the analyst has to define the embedding dimension or window length, which is a compromise between the conflicting goals of detectability of long periods (demanding a high window length) and statistical reliability through many repetitions (small window length). In this work, a moving window of length 200, corresponding to 5.6 years, has been used. It is also necessary to restrict the

number of components (eigenfunctions) in the gap-filling to avoid spurious high frequency oscillations. This has been done in a data-adaptive way, keeping control on the accuracy achieved already in gap-free parts of the time series, and in many cases, just a few components were sufficient to explain 99 % of the total variance; but for any time series, a maximum of 20 components was used. Any value actually present in the records was left unchanged, i.e. there was no substitution with the SSA interpolated value. No protection against negative values was carried out; thus, when the FAPAR values are rather low, SSA occasionally produces slightly negative values in the gaps. These are unphysical and were replaced by uniform noise with values between 0 and the noise floor assumed (0.05).

12.2.3 Combining the Two Datasets

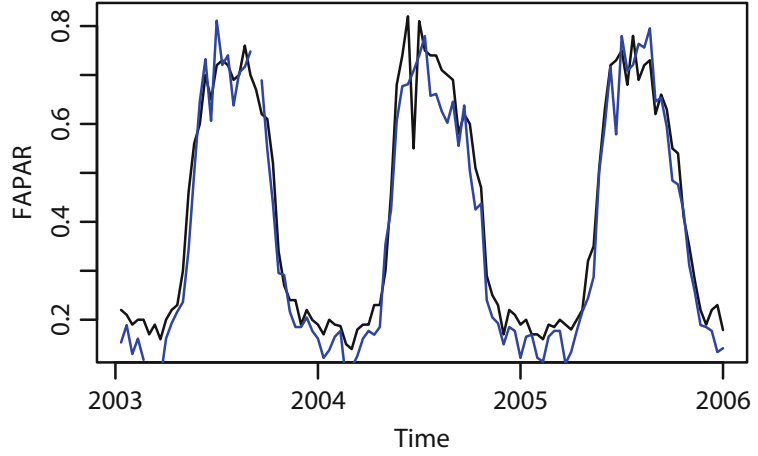
At the given spatial resolution, the land surface of the globe excluding Antarctica is covered by 57,500 pixels. Of all time series associated to each of these pixels, 16,385 contained more than 30 % gaps, mostly at high Northern latitudes. They were excluded from the analysis. There were 16,403 (39.9 %) complete time series which required no gap-filling. The remaining 24,712 (60.1 %) series were sent to the SSA algorithm to fill the gaps with interpolated values. As the time series to be investigated are from two different sensors, we have to check whether the two parts are homogeneous enough, or whether the breakpoint between them is a major issue. The noise level of the second sensor (MERIS, 2003–2011) appears to be roughly half that of the older one, or 0.05 in absolute terms as a rule of thumb. In addition, the algorithm producing the FAPAR variable has an uncertainty of around 0.02 [6]. Differences between FAPAR values smaller than that are of no significance, and time series with values <0.1 (say) cannot be trusted as proper signals. These occur typically at desert locations where vegetation cover is so sparse that they might simply be ignored at the relatively coarse resolution available. All FAPAR time series with a mean value <0.02 were excluded.

A tiny fraction (98 of 16,403, or 0.6 %) of the gapless time series exhibits spurious half-year cycles, stronger than the yearly cycle, only from the MERIS sensor. As a consequence, some of the RNA measures were not even significantly correlated in the overlap period (2003–2005). These few series were excluded as well.

As a result, the main analysis was done on a set of 30,253 time series. Since the noise level was lower for the MERIS data, which was obvious in the overlap region (Fig. 12.2), we decided to use them from 2003 onwards.

To further homogenize the data set, all time series were transformed to zero mean and standard deviation one for all subsequent analysis. This has to be undone if their spatial relatedness is investigated. In this chapter, each time series is analyzed at its own, independent from the others (univariate).

Fig. 12.2 Time series from the two sensors during the overlap period for a randomly selected pixel. SeaWifs(*black*) and MERIS (*blue*)



12.3 Nonlinear Methods Used for FAPAR Analysis

Within the theory of dynamical systems, the most salient feature of a nonlinear system is its phase space. A conventional way to characterize properties of the phase space of unknown dimension, given just a time series of a single observable (univariate), is *embedding*. Embedding is the method to infer the multi-dimensional phase space from a one-dimensional time series. Motivated by the embedding theorem [11], the time series can be embedded into an m -dimensional space with a certain delay τ for this purpose. The new variable v obtained by the embedding of the time series is given by

$$v_i = (x_i, x_{i-\tau}, \dots, x_{i-\tau \cdot (m-1)}) , \quad i = 1 + \tau \cdot (m-1), \dots, n . \quad (12.1)$$

To determine an appropriate choice for m , we estimate the number of *false nearest neighbors*. For small values of m , FNN is typically high due to coalescing projections of points onto the low dimensional space. The embedding is chosen appropriately when the FNN is small and often at a minimum; an excessively high m will cause them to increase again [12]. Implementation in the programming environment R [13] was provided by the package `tseriesChaos`.

12.3.1 Average Mutual Information

A widespread technique for selecting a suitable τ is the *average mutual information* (AMI). Consider the variable $\epsilon \in E$ at the time step i and the variable $\eta \in H$ at time step $i + \tau$, where E and H are two random processes. The AMI is defined as

$$\text{AMI}(\tau) = \sum_{\epsilon, \eta} p_{\epsilon, \eta}(\tau) \cdot \ln \frac{p_{\epsilon, \eta}(\tau)}{p_{\epsilon} \cdot p_{\eta}} , \quad (12.2)$$

where $p_{\epsilon,\eta}$ is the joint probability density function of E and H while p_ϵ and p_η are the marginal distributions of the two processes. $\text{AMI}(\tau)$ can be seen as a measure for information redundancy between states that are separated by the lag τ . Hence, the aim is to find the τ of the first local minimum of $\text{AMI}(\tau)$. This method was also implemented in `tseriesChaos`.

12.3.2 Correlation Length

As a simple conventional measure to characterize the dynamics of time series, we use the *length of autocorrelations* (LAC). LAC is the time (τ) a signal needs until its sample autocorrelation function falls below the level of a significant correlation for the first time. In a *white-noise* process, the autocorrelations after $\tau = 0$ can be assumed to be approximately normally distributed. Based on the distribution for white noise, a significance test can be conducted. The given LAC was calculated with $\alpha = 5\%$.

12.3.3 Recurrence Plots

The analysis was performed with the *python* package *pynetwork*.² The methods applied to univariate time series were RQA and RNA. Both are based on the sufficiently close recurrence of initial states in future trajectories of the system. As our time series are short to medium length only with 504 values each, and in addition noisy and nonstationary, RA is a particularly appropriate choice within the analysis methods for nonlinear systems. Since their invention more than 25 years ago [14], their application domains have widened all the time. The bibliography maintained by N. Marwan³ contains more than 1,000 references, and there are over 100 new papers per year during the last few years on the subject. An overview over ecological and hydrological applications is provided by [15]. As usual, we start with the recurrence matrix

$$\mathbf{R}_{i,j}(\varepsilon) = \Theta(\varepsilon - \|x_i - x_j\|) \quad , \quad i, j = 1, \dots, N \quad , \quad (12.3)$$

where ε is the recurrence threshold, Θ is the Heaviside-function and N is the number of data points. Distances between them can be calculated using any suitable norm.

Two possibilities to obtain the recurrent states were considered. Either a fixed threshold ε (FTHR) or a fixed recurrence rate (FRR, cf. the next section) can be chosen. In the latter case, the phase space distances are calculated beforehand and ε

²Written by Jonathan F. Donges, Jobst Heitzig, Hanna Schultz and Jakob Runge.

³<http://www.recurrence-plot.tk/bibliography.php>.

is then modified to get the selected recurrence rate. Visualizing the matrix \mathbf{R} , e.g. as a black-white pixel image, then gives the Recurrence Plot (RP).

To quantify the visual impressions one gets from looking at RPs, one may obtain quantities extracted from either the recurrence matrix directly (RQA) or from the adjacency matrix derived from the recurrence matrix but considered as that of a network (RNA).

12.3.4 Recurrence Quantification Analysis

Recurrence Quantification Analysis (RQA) intends to quantify the structures that are visible in the RP, such as diagonal lines, vertical lines and vertical white-space. First employed by [16, 17], the associated methods were successively widely augmented [18]. The subsequent definitions are taken from the latter reference, and are summarized here, as far as they are relevant for this chapter, for the sake of convenience.

12.3.4.1 Recurrence Rate

The ratio between points of recurrence and the extent of the recurrence matrix is called recurrence rate (RR). It is calculated by summing over \mathbf{R} and dividing by N^2 .

$$\text{RR} = \frac{1}{N^2} \sum_{i,j=1}^N \mathbf{R}_{i,j}(\varepsilon) . \quad (12.4)$$

The measure depends strongly on the chosen threshold ε , or may be set to a fixed value (possibly in windows gliding over the time series), which then requires the calculation of the corresponding ε .

12.3.4.2 Determinism

The ratio between the amount of recurrent points forming diagonal lines of at least length l_{\min} and the total number of recurrent points is termed *determinism* (DET). As for the following measures that are based on the characteristics of diagonal lines, the distribution of line lengths occurring has to be established first.

$$P(\varepsilon, l) = \sum_{i,j=1}^N (1 - \mathbf{R}_{i-1,j-1}(\varepsilon)) (1 - \mathbf{R}_{i+l,j+l}(\varepsilon)) \prod_{k=0}^{l-1} \mathbf{R}_{i+k,j+k}(\varepsilon) . \quad (12.5)$$

The determinism is then calculated as

$$\text{DET} = \frac{\sum_{l=l_{\min}}^N lP(\varepsilon, l)}{\sum_{l=1}^N lP(\varepsilon, l)} . \quad (12.6)$$

Similar to DET, one can also consider the fraction of recurrent points which form *vertical lines* (longer than a given threshold length), which is called *laminarity* (LAM). For artificial chaotic maps, laminarity peaks at chaos-chaos transitions to laminar states [18]. One may also compute the average line length (ALL) in either diagonal or vertical direction.

12.3.4.3 Entropy

Based on the distribution of diagonal lines given in Eq. (12.5), an *entropy* (ENT) can be computed. The measure is based on the works of Shannon [19], where it quantifies the information content of random processes. It is defined as

$$\text{ENT} = - \sum_{l=l_{\min}}^N p(\varepsilon, l) \ln p(\varepsilon, l) , \quad (12.7)$$

where $p(\varepsilon, l) = P(\varepsilon, l)/N(l)$. For a time series where all line lengths appear equally often, ENT will have a high value, whereas it is low for data which are dominated by one line length. ENT is generally associated with the complexity of an RP. The RP of a sine-wave would show a narrow distribution of relatively long diagonal lines, which would result in a low ENT. The ENT of a complex biological signal could be expected to be high.

12.3.4.4 Trend

The instationarity of a time series can be quantified by the measure *trend* (TRE). It represents the thinning of recurrence points towards the upper left and lower right corners of the respective RP and is calculated as a least squares regression:

$$\text{TRE} = \frac{\sum_{\tau=1}^{\tilde{N}} (\tau - N/2)(\text{RR}_{\tau} - \langle \text{RR}_{\tau} \rangle)}{\sum_{\tau=1}^{\tilde{N}} (\tau - N/2)^2} , \quad (12.8)$$

where τ denotes the number of the diagonal line excluding the main diagonal and \tilde{N} the maximum diagonal line number covered in the analysis, intending to minimize uncertainties due to the short diagonal lines towards the corners of the RP. \tilde{N} is a parameter of the method and is thus chosen by the analyst. $\langle \text{RR}_{\tau} \rangle$ is the expected RR on the diagonal τ . It is calculated as the mean RR of the \tilde{N} diagonals.

12.3.5 Recurrence Network Analysis

To consider time series as representing complex networks is a relatively new approach in time series analysis [20, 21]. Network construction is based on interpreting the recurrence plot as the *adjacency matrix* of a complex network [22],

$$A_{i,j}(\varepsilon) = R_{i,j}(\varepsilon) - \delta_{i,j} , \quad (12.9)$$

where δ is the Kronecker delta.

Thus, each moment of observation in a time series is considered as a node in a network. Each node might be connected to any other node or not through an undirected edge. The *edge density* $\rho(\varepsilon)$, i.e. the number of realized edges divided by the number of all possible edges, is equal to the *recurrence rate* RR. Once established, the network can be analyzed by means of graph theory. Similar to the RQA in the last section, we list here a number of measures which have been proposed to quantify network properties, and which turned out to be useful for our purposes. They are mostly taken from [22].

12.3.5.1 Average Path Length

Taking the shortest connections between all pairs of nodes in the network and averaging the lengths delivers the *average path length* (\mathcal{L}). It is calculated as

$$\mathcal{L} = \langle l_{i,j} \rangle . \quad (12.10)$$

The average path lengths of time series vary substantially when considering dynamical transitions [23].

12.3.5.2 Transitivity

In a complex network, the *transitivity* (\mathcal{T}) gives the probability that two nodes connected to a third node are also directly connected:

$$\mathcal{T} = \frac{\sum_{i,j,k} A_{i,j} A_{j,k} A_{k,i}}{\sum_{i,j,k} A_{k,i} A_{k,j}} . \quad (12.11)$$

Regular dynamics like periodic orbits are generally associated with a higher transitivity, while chaotic systems show comparatively low values of transitivity [22].

12.3.5.3 Global Clustering Coefficient

The average interconnectivity of neighbours of a certain node is also quantified by the *global clustering coefficient* (C). It is calculated as the mean of the *local clustering coefficients* which are defined as

$$C_n = \frac{2}{k_n(k_n - 1)} N_n^\Delta, \quad (12.12)$$

where n denotes a single node, N_n^Δ all possible closed triangles for this node and k_n the respective degree (number of edges) of the node. The mean is then calculated as

$$C = \frac{1}{N} \sum_{n=1}^N C_n. \quad (12.13)$$

12.3.5.4 Assortativity

The inclination of the nodes of a complex network to connect to nodes with a comparable number of neighbours is termed *assortativity* (\mathcal{A}). This means that a network in which nodes with three neighbours are linked to nodes with three neighbours while nodes with two neighbours tend to be connected to nodes with two neighbours is classified as *assortative*. The assortativity is calculated as a Pearson product-moment correlation coefficient among all pairs of degrees k_i and k_j :

$$\mathcal{A} = \frac{\frac{1}{L} \sum_{j>i} k_i k_j A_{i,j} - \left\langle \frac{1}{2}(k_i + k_j) \right\rangle_{i,j}^2}{\frac{1}{L} \sum_{j>i} \frac{1}{2}(k_i^2 + k_j^2) A_{i,j} - \left\langle \frac{1}{2}(k_i + k_j) \right\rangle_{i,j}^2}. \quad (12.14)$$

where $L = \sum_{j>i} A_{ij}$ is the number of edges in the network.

The recurrence networks obtained from our FAPAR data were visualized by the *Fruchtermann–Reingold* force-directed algorithm [24] implemented in the software *Gephi*.⁴

12.3.6 Fixing the RQA Parameters

First, we investigated the effect of merging the two datasets on the RQA and RNA measures. The overall differences between RQA measures of the first or second half and the total period were comparatively small. Network measures, on the other hand, were generally weaker correlated. Splitting one of the sensor time series in

⁴<http://gephi.org>.

Table 12.1 Correlations of RA-measures for different embedding dimensions

m		DET	LAM	\mathcal{L}	\mathcal{T}	\mathcal{C}	\mathcal{A}
1–2	ρ	0.940	0.932	0.476	0.695	0.040	0.487
	p -value	< 0.001	< 0.001	< 0.001	< 0.001	0.674	< 0.001
2–3	ρ	0.990	0.990	0.925	0.962	0.947	0.757
	p -value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
1–3	ρ	0.917	0.894	0.404	0.582	-0.031	0.295
	p -value	< 0.001	< 0.001	< 0.001	< 0.001	0.743	0.001

Recurrence measures were calculated for 100 time series in each case. The minimal line length (affecting DET and LAM) was set to 3, and the recurrence rate was $FRR = 0.1$

halves revealed that there is also a within-series variability, indicating that time series length is important as expected. Some of the RNA measures were not even significantly correlated between the two halves. As a rule of thumb, RQA quantities turned out to be more robust than RNA ones against changes in time series length or the time window used for calculation.

12.3.6.1 Effects of the Embedding Dimension

False Nearest Neighbors calculations did not lead to conclusive results for the choice of the embedding dimension that would fit all time series. This is common for time series containing noise [15]. Different embedding dimensions were tested with respect to the changes of resulting recurrence measures. Table 12.1 contains the three combinations of the embedding dimensions 1, 2 and 3 against each other. The RQA-measures DET and LAM were significantly and strongly correlated across all combinations of embedding dimensions. By contrast, in case of the RNA-measures, the same quality of correlation could only be observed for the comparison of 2-d versus 3-d embeddings. While false nearest neighbours might influence the measures with $d = 2$ quite strongly, $d = 3$ should be sufficiently high while accounting for the limited time series length which prohibits the use of even higher values. Thus, $d = 3$ was chosen in all cases.

12.3.6.2 Effect of the Delay

The embedding lag τ was expected to have a significant impact on the recurrence measures. To test the robustness of the results with regard to τ , three different delays – $\tau = 3$, $\tau = 6$ and $\tau = 9$ – were chosen.

The two RQA-measures DET and LAM showed strong and significant correlations for all comparisons (Table 12.2). Among the RNA-measures \mathcal{T} and \mathcal{C} showed weaker ($\rho \approx 0.7 \dots 0.85$) yet significant correlations. \mathcal{L} was weakly ($\rho \approx 0.2 \dots 0.3$) but again significantly correlated. The correlation was weakest for

Table 12.2 Correlations of RA-measures for different embedding delays

τ		DET	LAM	\mathcal{L}	\mathcal{T}	\mathcal{C}	\mathcal{A}
3–6	ρ	0.951	0.926	0.296	0.722	0.763	0.352
	p -value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
6–9	ρ	0.982	0.961	0.288	0.780	0.832	−0.033
	p -value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
3–9	ρ	0.920	0.901	0.228	0.702	0.744	0.051
	p -value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001

The embedding dimension was 3 for all calculations. RQA measures were calculated with $FRR = 0.1$, RNA-measures were calculated with $\varepsilon = 0.9$

\mathcal{A} ($\rho \approx 0 \dots 0.35$) yet also significant. Thus, the variability between different time series prevented to find a unique value for τ . However, when we use the location of the first minimum of the average mutual information as optimal embedding delay, a delay of $\tau = 9$ emerges as a suitable choice for many time series. This is not a surprising value since for series with a (strong) annual cycle, the first minimum of the AMI (and at the same time the first zero crossing of the autocorrelation function) would be expected at a lag of $1/4$ year, corresponding to 9 values for our temporal resolution.

Influence of Time Series Length

A further test was conducted to evaluate the robustness of the results for varying lengths of the time series. Five randomly selected FAPAR time series were therefore gradually shortened by complete years. Resulting lengths were 14, 13, 12, 11 and 10 years.

DET could be considered as stable, as the *ranking* of the time series did not change with the number of included years (Fig. 12.4). There was a certain but not very pronounced tendency of the determinism values to decrease with increasing length. By contrast, TRE exhibited considerable changes in magnitude and even in relative ranking. One should add that the value obtained for TRE also depends on the number of basic periods (annual cycles) included—which we deliberately chose to be integers to suppress this effect. This means that a careful interpretation of the measure is required. The RNA-measure \mathcal{T} was found to show a consistent ranking, with the exception of one point-pair at 10–11 included years, and overall not very variable for individual time series. The same holds for the assortativity \mathcal{A} , for which only one point-pair at 13–14 included years was inconsistent.

Main Analysis

A fixed threshold (FTHR) ε led to a substantial variance of RR (Fig. 12.3). To minimize the influence of the varying RR on the results, we opted for a FRR that

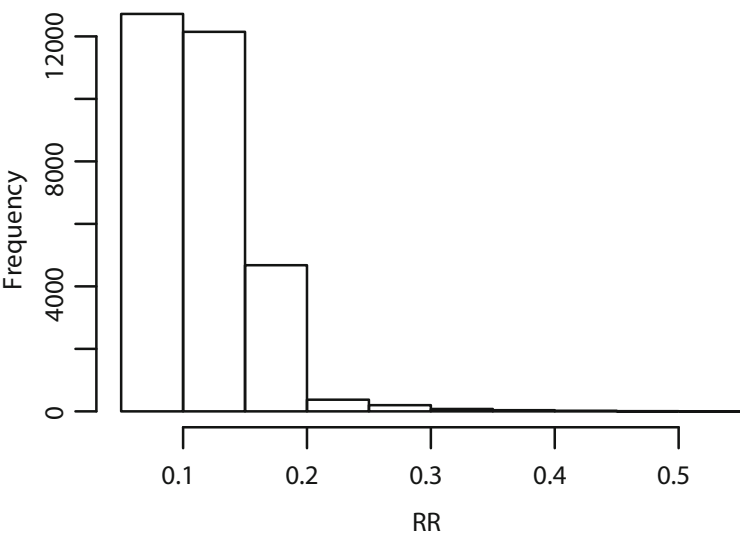


Fig. 12.3 Histogram of RR for FTNR

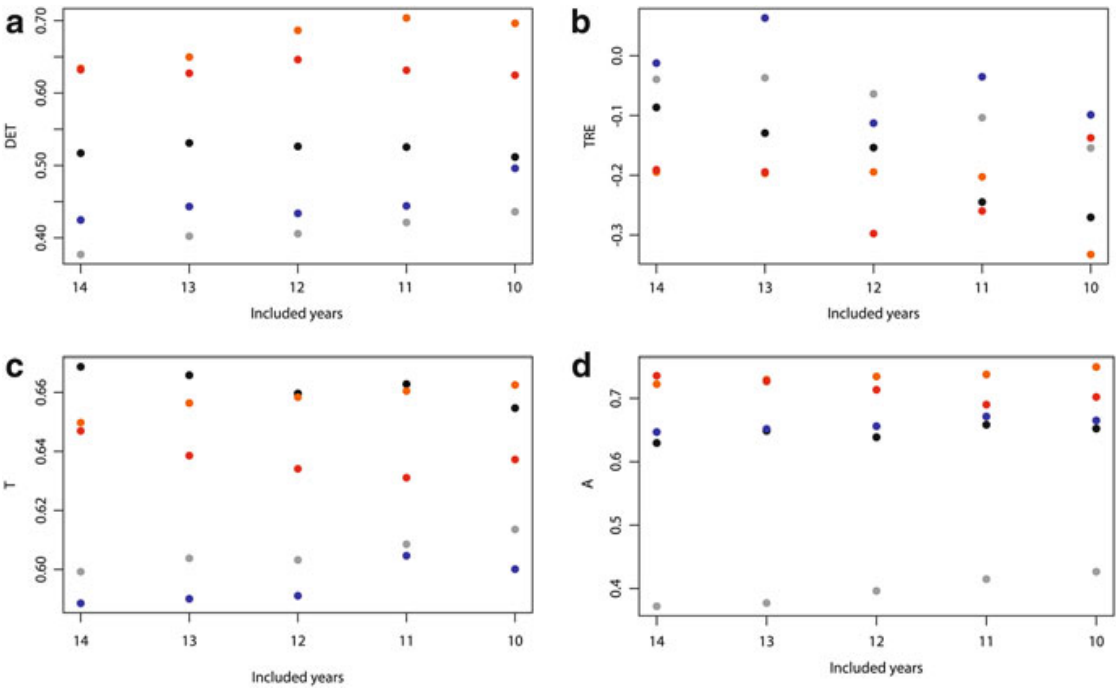


Fig. 12.4 Variation of RA-measures with time series length. Colors denote the different time series. (a) DET. (b) TRE. (c) \mathcal{T} . (d) \mathcal{A}

was both suitable for the RQA and yet close to the maximum of the distribution seen in Fig. 12.3. The main RQA analysis was thus performed with $FRR = 0.1$. By contrast, the network measures of the RNA were expected and observed to be highly dependent on the chosen ε , which would be different for each time series when using FRR. A FTNR of 0.9 (in units of standard deviations, i.e. referring to z scores) was therefore chosen for them. The embedding was done with $m = 3$ and $\tau = 9$. A minimum length of $l_{\min} = 3$ was set for subsequent points to be counted

as either diagonal or vertical line. Distances in phase space were calculated with the Euclidean norm.

12.4 Results and Discussion for the Recurrence Analysis of FAPAR

We present results for each individual time series separately, i.e. there is no explicit spatial analysis involved. Each recurrence measure is a scalar quantifier and can, using the corresponding pixel coordinates, be put on a world map. No interpolation or smoothing has been performed. Thus, any consistent spatial pattern observed is amenable to interpretation, e.g. in terms of known properties of the landscape elements.

12.4.1 Mean Values and Autocorrelation Lengths

FAPAR reflects primary productivity in a landscape element. In Fig. 12.6a, we show overall mean values for each FAPAR time series. In this map, black pixels indicate no available data as discussed above. The inner tropics showed the expected high mean values for FAPAR, while regions with low values were located near deserts. Semiarid regions of the western US also had low mean values around 0.1–0.2, and there is a rather sharp boundary to higher values east of around 100° W. Western Europe, especially Ireland, showed high FAPAR means. The American tropics are variable at a high level, whereas the African tropics seem to be more homogeneous with a lower overall mean.

The length of the autocorrelation (LAC) was calculated as outlined above. Several areas with LAC of more than half a year were found at southern latitudes such as southern Argentina and Australia, where an interesting dependence from the distance to the coast can be observed (Fig. 12.6b). While large parts of the globe revealed a spatially heterogeneous distribution of LAC, it was zonally organized in tropical Africa.

The patterns of mean FAPAR values were consistent with the expectation. As tropical rain forests have no seasons and no dormancy, they exhibit the highest values on average. The mild oceanic climate caused by the Gulf Stream can be seen as the reason for high values for FAPAR in Ireland. Water availability is among the most crucial limiting factors for biological activity in general. This is reflected in low mean values in arid regions. The short vegetation periods in mountainous regions are another limiting factor for plant growth. The length of autocorrelations indicated some areas with short-term fluctuations. This included tropical as well as subtropical and boreal areas. High LAC (≥ 6 months) was seen in arid areas in central Australia and southern Argentina. It is not clear at present whether this could result from the interannual variability of the *El-Niño-Southern-Oscillation* (ENSO).

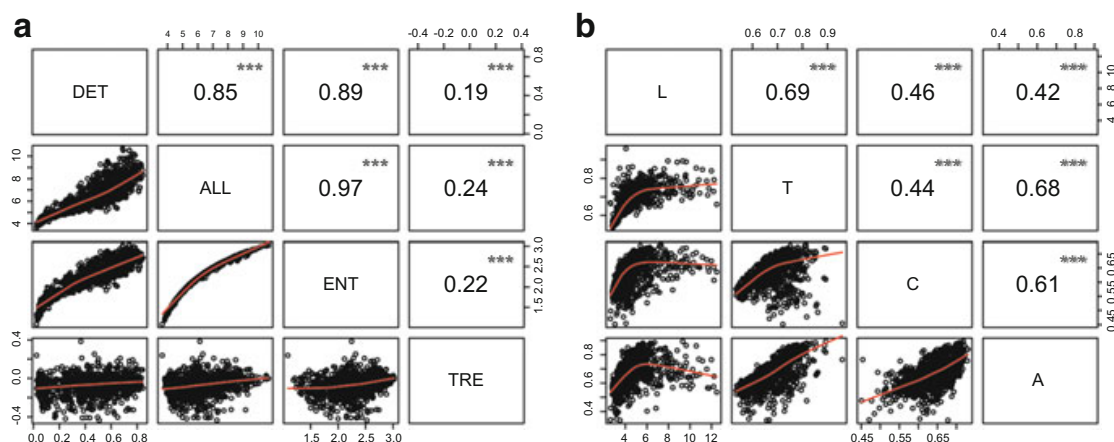


Fig. 12.5 Correlations between the recurrence measures for FAPAR. Pearson correlation coefficients and scatter-plots based on 1,000 sampled time series. Asterisks denote level of significance (* * *: = 1 %). (a) Correlation matrix for RQA-measures. (b) Correlation matrix for RNA-measures

12.4.2 RQA Measures

For the FAPAR time series, the RQA variables DET, ALL and ENT were highly correlated (Fig. 12.5a) with each other, contrary to TRE. We present therefore results for DET and TRE only.

The measure DET shown in Fig. 12.6c reflects the occurrence of regular patterns (predictability) in a time series. Low values of about 0.1 could be found in the inner tropics of South America, Central Africa and South East Asia. High values above 0.8 prevailed in the outer tropics of Africa and in central areas of Asia. In Europe, one could find gradients with low values in coastal areas and the British Isles and higher values in South Eastern Europe and Southern Spain. In North America, low values dominate in the Western parts and higher values in the East, but there is also a hotspot of DET in the Dakotas, Nebraska and Kansas.

Thus, the measure DET was strongly associated with regular vegetation patterns. In the temperate zone, strong seasonalities in the eastern US were contrasted with lower seasonal variation in western Europe. Low DET (≈ 0.05) in the inner tropics also showed that the vegetation is not dominated by regular dynamics in these areas. By contrast, the outer tropical areas of Africa showed high DET which can be explained by the strong seasonality induced by vegetation driving rain periods, and the monsoon in India may cause the likewise high DET there.

TRE indicates instationarities and changes in the dynamical behaviour of a system over the length of the time series. Large parts of the global land surface showed values around zero (Fig. 12.6d). Negative values were located in parts of eastern and western Australia. Other regions with comparatively low TRE appeared disseminated in North and South America, Asia and Africa. Several regions with slightly positive values of TRE stood out. In Africa, positive values were found in most parts except the areas bordering the Namib desert and the Horn of Africa.

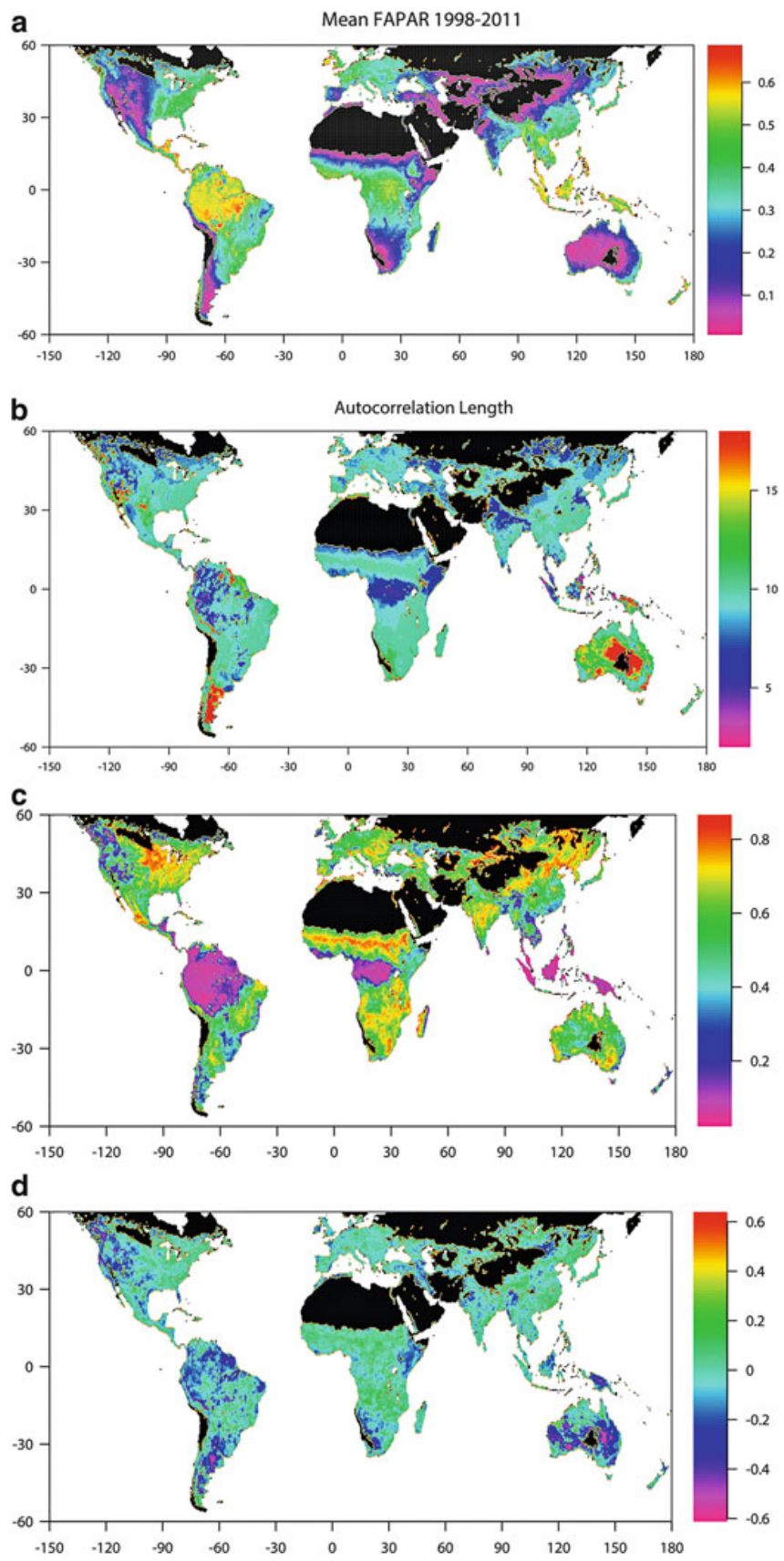


Fig. 12.6 World maps of basic analysis (a, b) and RQA (c, d). (a) Mean FAPAR. (b) LAC. (c) DET of FAPAR. (d) TRE of FAPAR

A gradient could be seen in North America, where the northwestern parts showed predominantly negative values while positive values were found in the south-east.

The complicated pattern of TRE values indicate that this measure is rather sensitive against small random changes in the time series. Nevertheless, some larger areas with consistently negative values are found. They could be related to increased vegetation activity (*greening*) as also observed in [25]. However, there are also connected areas with positive TRE values which are difficult to explain. This could point to the limited length of the considered time series. On the other hand, a significance test for the TRE values has not been performed, they might have appeared just as random fluctuations.

12.4.3 Recurrence Network Analysis

In this section, we focus on the average path length \mathcal{L} , the network transitivity \mathcal{T} , the global clustering coefficient \mathcal{C} and the assortativity \mathcal{A} extracted from RNA of each FAPAR time series. These are not completely independent from each other, as Fig. 12.5b shows, but the correlations are generally weaker compared to those between the RQA measures. Figure 12.7 shows the spatial distribution of the calculated RNA-measures. The average path length (Fig. 12.7a) had a globally rather homogeneous distribution. Noticeably higher values were found in distinct regions in North America and eastern Asia. These regions appeared as highly structured and heterogeneous. In North America, the eastern parts of the US and Canada had high values, especially around the Great Lakes. Several patches with values ≤ 2.5 were seen in southeastern US and northern Mexico. South America showed consistently homogeneous values of $3 \leq \mathcal{L} < 3.5$ with only slightly elevated values in western Brasil. In Europe, only eastern parts showed slightly raised values, compared to the rest of the continent where intermediate values dominated. Apart from a confined area of low values in eastern Africa, the rest of the continent also showed the intermediate values that prevailed on the rest of the globe—again with marginally elevated values at the outer tropics. The most pronounced pattern was found in Northeast Asia, where high values ($\mathcal{L} \geq 3.5$) could be seen. It is possible that a linear scale is not best suited for the visualization of \mathcal{L} , but a percentile-based rather.

In general, a low \mathcal{L} is associated with a network of highly connected nodes. The two areas that have markedly low average path lengths are located in the northernmost part of the nemoral zone. Their placement in the east of the respective continents leads to a strong seasonal amplitude of vegetation. Although the time series were transformed to unity standard deviation prior to the analysis, the lower relative level of noise could effect an almost continuous seasonal increase and decrease and thereby a low connectivity in the recurrence network.

Figure 12.7b depicts the network transitivity on a world map. North America is separated into west and east by a sharp line at 100° W, coinciding with that found for the FAPAR mean value, with lower values in the West. Central America shows

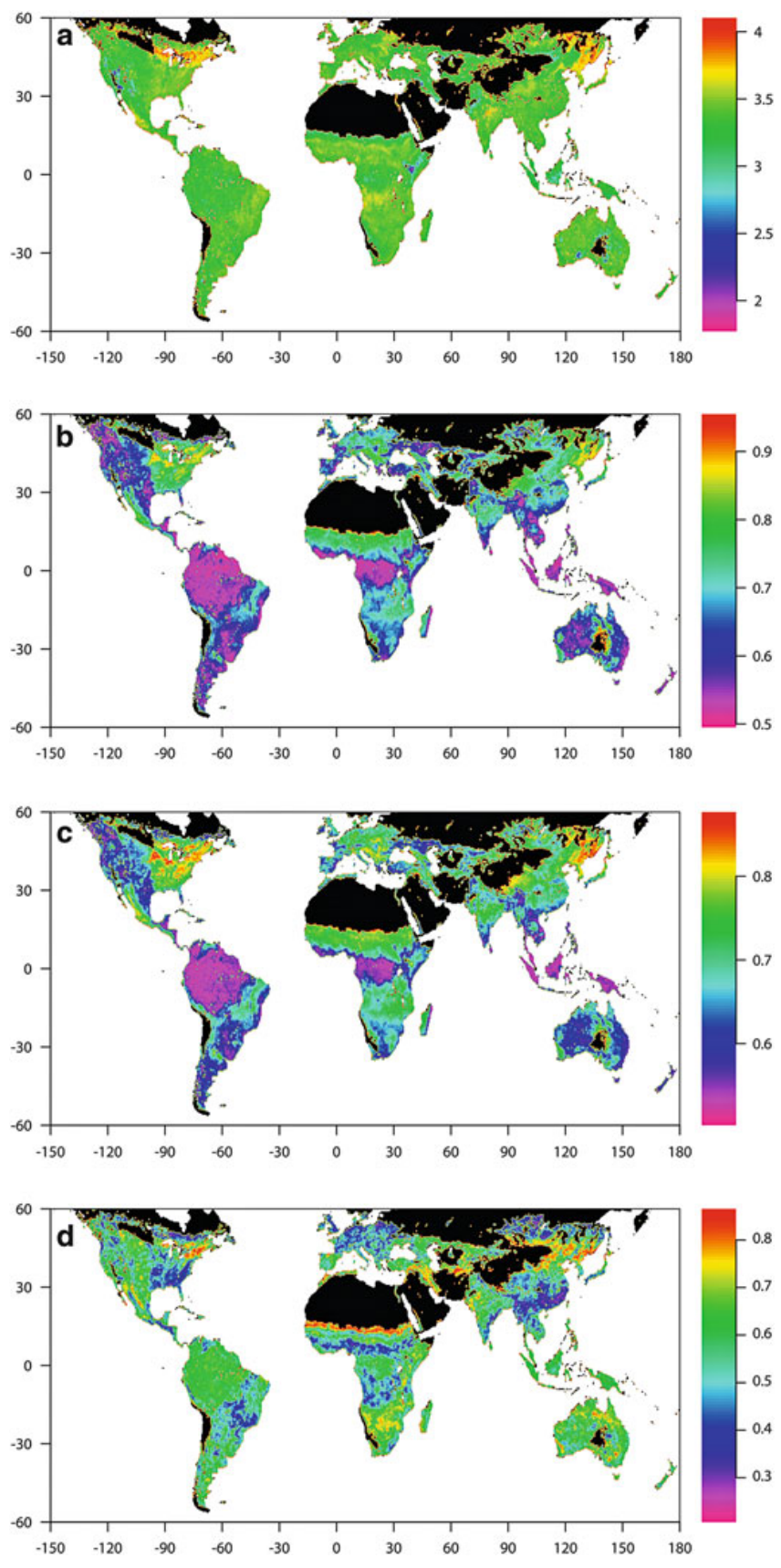


Fig. 12.7 World maps of four RNA measures for FAPAR. (a) \mathcal{L} of FAPAR. (b) \mathcal{T} of FAPAR. (c) \mathcal{C} of FAPAR. (d) \mathcal{A} of FAPAR

a gradient from the Atlantic to the Pacific. In South America, low \mathcal{T} (≤ 0.6) is found in most parts of the continent, most notably in the Amazon basin. Slightly higher values were seen in western Brasil, Argentina and the areas flanking the Atacama desert. Europe shows a heterogeneous distribution of \mathcal{T} : western Europe has lower values than central and eastern Europe. In Africa, the highest values for \mathcal{T} (≥ 0.9) were found directly south of the Sahara. The Kongo basin by contrast had low values (≈ 0.5). The spatial distribution of \mathcal{T} was comparatively complex in Asia. The western and central parts showed various areas with both high and low \mathcal{T} . Regions bordering the mountainous and arid central areas tended towards higher values. By contrast, Southeast Asia showed notably lower \mathcal{T} (≤ 0.6). Northeastern China then again was dominated by higher values ($\mathcal{T} \geq 0.85$). In Australia, regions surrounding the central area of excluded values showed very high \mathcal{T} (≥ 0.9).

\mathcal{T} highlights areas comparable to DET. This is intuitive, insofar as \mathcal{T} quantifies the regularity of the respective dynamical system. Nevertheless, there are differences to be seen, for example in the Sahel. \mathcal{T} emphasized the pixels which were directly bordering the zone of excluded pixels, whereas the zone of maximal DET was located further south.

The spatial distribution of the global clustering coefficient is given in Fig. 12.7c. In general, similar regions were highlighted by \mathcal{C} as for \mathcal{T} . In particular, the sharp boundary in Northern America appears again, although the distribution of values in both the western and the eastern part is more complex. Overall, \mathcal{C} adds little extra information once \mathcal{T} already has been calculated.

That \mathcal{C} shows a distribution similar to \mathcal{T} is comprehensible, as both measures quantify similar network characteristics (see Sect. 12.3.5). The similarity pertained also to structures on a smaller scale, such as the pattern of elevated values to the southwest of the Great Lakes.

This is very different from the assortativity \mathcal{A} (Fig. 12.7d). In North America, low \mathcal{A} (≤ 0.45) was found in the southeastern US, intermediate \mathcal{A} ($0.5 < \mathcal{A} \leq 0.7$) in the western US/Canada and high values ($\mathcal{A} \geq 0.75$) to the east of the Great Lakes. In South America, most parts had intermediate values ≥ 0.55 , only southern Brasil had a coherent area of lower \mathcal{A} . Europe is comparatively heterogeneous at relatively low values, Spain—with higher values—being the exception. Several zones could be distinguished in Africa. Very high values (≥ 0.8) prevailed directly to the south of the Sahara. This was followed by intermediate and then low values in the marginal tropics. The Kongo basin and the southern coast of West Africa was found to have intermediate values again. Another zone of low values was then succeeded by higher values (≥ 0.7) that dominated southern Africa. In Asia, high values (≥ 0.7) prevailed in the central parts of the continent. By contrast, lower values were seen in northern Asia, southern China and eastern India. Australia displayed high values with the exception of various patches in the central and southern parts.

RNA-Assortativity quantifies the homogeneity of the recurrence network. The highest values were located in areas that show strong seasonal variation (nemoral forests in East Asia), but not necessarily a high FAPAR-amplitude (Sahel). By contrast, low seasonal variation, as seen in the inner tropics, did not automatically

lead to low \mathcal{A} . In these areas, all states of the associated RN could be assumed to have a similar degree, not allowing for a low \mathcal{A} .

The distribution of \mathcal{A} deviated considerably from the other measures. This confirms the assertion of [22] that its quantification of network continuity and attractor fragmentation is unparalleled by other RA-measures.

In order to investigate the clear zonal structure of \mathcal{A} in Africa, we take a transect at longitude 20.25° E, extending northwards from the equator to the southern Sahara border at latitude around 15° N. We demonstrate the diversity of network structures obtained from the different time series. Values of $\mathcal{A} \approx 0.6$ were found in the Kongo basin, with a highly connected recurrence network (Fig. 12.8a). The values dropped below 0.5 northwards, and the networks had smaller amounts of edges (Fig. 12.8b, c). The time series had notable seasonal periodicity. The FAPAR minima appeared as sharp incisions. Further to the north (Fig. 12.8d), more extended seasonal minima were found, and the network structure gets more circular. This was accompanied with an elevated \mathcal{A} . The next time series showed a significantly lower peak width, while the associated recurrence network (Fig. 12.8e) was dominated again by a circular structure, however exhibiting several deviations mostly of more recent states. In this case, \mathcal{A} was low again at ≈ 0.45 . \mathcal{A} rose again towards the Sahara, and the corresponding recurrence network (Fig. 12.8f) is the sparsest one in this selection.

12.5 Conclusions

It could be demonstrated that recurrence analysis is able to describe and quantify various dynamical properties of global photosynthetic activity. The results for the recurrence determinism showed that seasonal phenology is a crucial aspect of ecosystem dynamics. Determinism was highest in regions with a pronounced annual or semi-annual periodicity of vegetation activity, and lowest for inner tropical rainforests with high but non-seasonal photosynthetic activity. A further step to investigate the regularity of FAPAR time series in more detail would be to calculate DET for deseasonalized time series. A significant DET in these time series also would indicate inter-annual variability of vegetation dynamics. However, the total length of the observation period sets limits to this approach. The recurrence trend identified regions with growing dynamical dissimilarity, including linear trends, instationarities and changes in the type or strength of the seasonal periodicity. The most pronounced TRE values (negative, but highest in absolute value) were found in both South America and Australia. This could indicate a connection to ENSO, although longer observations would probably be required to confirm this. We demonstrated that RNA-measures like the assortativity complemented established RQA-measures, being non-redundant to the latter—a finding which was anticipated from the underlying theory. However, the interpretation in terms of apparent characteristics of the respective time series proved to be more complicated for RNA-measures.

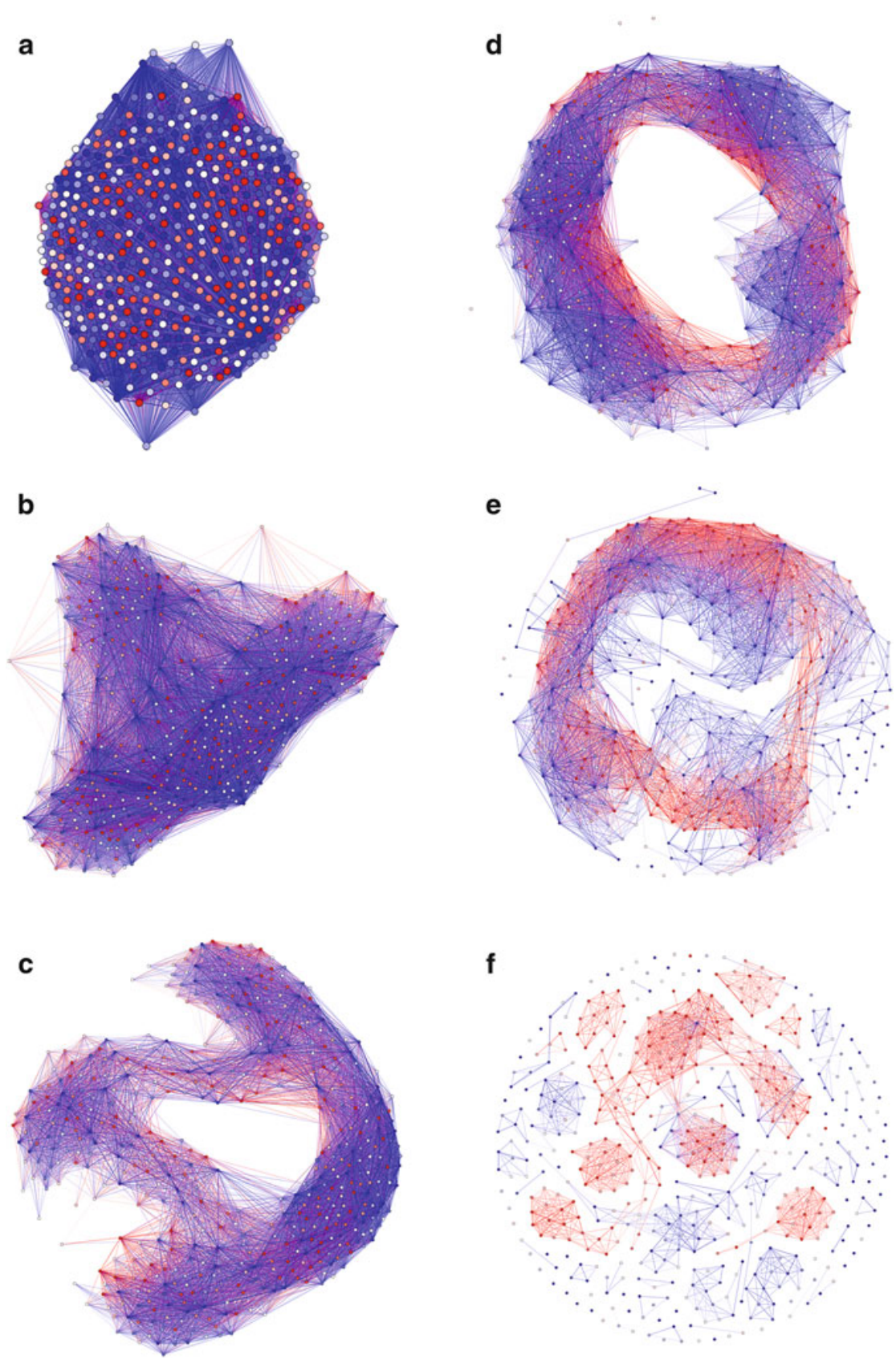


Fig. 12.8 Recurrence Networks of time series from a North–South transect at 20.25° longitude. Colours denote the time of the depicted states (1998: *red*—2011: *blue*). Visualization with the Fruchtermann–Reingold force-directed algorithm. (a) 1.25° N. (b) 5.25° N. (c) 7.75° N. (d) 9.75° N. (e) 11.25° N. (f) 13.25° N

The assortativity could be seen as an outstanding example for a finding that also concerned other measures: high sensitivity to time series properties that would hardly be visible by looking at time-series charts, and yet are forming remarkably coherent regions—as exemplified in Subsaharan Africa. This illustrates that RNA is sensitive to tiny details, which is an asset for well-defined measurements but potentially problematic when a large noise level is present.

However, the conducted tests regarding the influence of noise were inconclusive. As significance tests have yet to be developed for the given methods, uncertainty is not explicitly accounted for in this chapter. However, in light of the spatial coherence of the patterns observed, this drawback is moderated to a certain extent. The recurrence trend emerged from the tests as a particularly problematic measure.

The length of the time series investigated was 14 years or 504 time steps only. However, most measures proved to be considerably stable with respect to time series length. Yet in the case of TRE, 14 years turned out to be insufficient as the measure exhibited substantial variation when the time series were shortened by full years. It is therefore disappointing that the end of the *EnviSat* satellite mission precludes a seamless extension of the data set.

For further studies applying RA to a broad range of different time series, the choice of embedding parameters and the recurrence threshold remains a challenge. If one attempts to obtain maximum comparability, both a fixed threshold and a fixed RR can be compromising. To exclude spurious results introduced by either of the variants, one could combine the two. If the time series does exhibit high regularities—and thus high DET—this should hold for both FTTH and FRR.

The choice also affects the RNA-measures heavily, as they depend critically on the edge density. This influences also the network visualizations, where a fixed edge density can impede the comparison of fundamentally different time series. As demonstrated in this chapter, the visualization of the corresponding recurrence networks is an important step and offers a new perspective on ecosystem dynamics. While the interpretation of network measures might be difficult, the representation with connected states can facilitate the examination of temporally embedded time series.

Further research could proceed in either of two directions. Investigating the relations between the patterns found with recurrence analysis and fundamental processes is one option. A number of candidates for driving variables controlling in part the vegetation activity are available at the same spatiotemporal resolution. From a basic understanding of the processes involved, particularly promising in this context are air temperature, precipitation, and shortwave radiation. The RA could also be performed for these; in addition, they open up for the pixel-based *Joint Recurrence Analysis* [26] which allows the quantification of synchronization processes, or to conclude on relevant lags in vegetation responses. This could lead (e.g.) to a delineation of areas which are mainly temperature-driven and those which are precipitation-driven. A problem with this extension could be the coarse spatiotemporal resolution; for the spatial grid available, a single pixel might contain rather different vegetation types, and the relevant time scales of lagged responses might be below the temporal resolution of 10 days.

The other, complementary route to follow consists in a new *classification* of the land surface, obtained through homogeneous patches of one or several of the RA measures. The description of the visual impressions from the world map could be lifted up to a quantitative level through cluster analysis. The new classification obtained through RA and the subsequent clustering is completely ignorant of land use types, biomes or other qualitative descriptors, and also of climate. It would be interesting to see how it compares to existing classifications based on these descriptors. We expect the new classification based on nonlinear dynamics to be superior (more accurate, more detailed) to the “static” classifications based exclusively on climate classes and vegetation types. A first attempt into that direction is represented by [27].

Finally, it would also be possible although computationally challenging to extend the RNA presented here to a multivariate “network of networks” approach, where each node in the higher-order network is a complete time series, either for FAPAR at different locations, or for different driving variables. The most relevant issue here would be the hitherto unknown correspondence between network topology and ecological processes.

Acknowledgements We would like to thank C.S. Webber for the invitation to contribute a chapter to this ebook on Recurrence Plots. We are indebted to N. Gobron for providing us the two different FAPAR datasets. Two anonymous reviewers provided valuable hints for improvement.

References

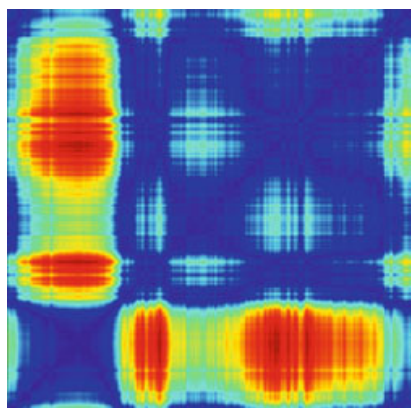
1. T. Kaminski, W. Knorr, M. Scholze, N. Gobron, B. Pinty, R. Giering, P.-P. Mathieu, Consistent assimilation of MERIS FAPAR and atmospheric CO₂ into a terrestrial vegetation model and interactive mission benefit analysis. *Biogeosci. Discuss.* **8**, 10761–10795 (2011)
2. M.D. Mahecha, L.M. Fürst, N. Gobron, H. Lange, Identifying multiple spatiotemporal patterns: A refined view on terrestrial photosynthetic activity. *Pattern Recognit. Lett.* **31**, 2309–2317 (2010)
3. GCOS, Implementation plan for the global observing system for climate in support of the UNFCCC. *World Meteorological Organization* (2010)
4. L.C. Li, Z.Q. Zhao, F.Y. Liu, Identifying spatial pattern of NDVI series dynamics using recurrence quantification analysis. *Eur. Phys. J. Spec. Top.* **164**, 127–139 (2008)
5. P.C. Stoy, A.D. Richardson, D.D. Baldocchi, G.G. Katul, J. Stanovick, M.D. Mahecha, M. Reichstein, M. Detto, B.E. Law, G. Wohlfahrt, N. Arriga, J. Campos, J.H. McCaughey, L. Montagnani, U.K.T. Paw, S. Sevanto, M. Williams, Biosphere-atmosphere exchange of CO₂ in relation to climate: a cross-biome analysis across multiple time scales. *Biogeosciences* **6**, 2297–2312 (2009)
6. N. Gobron, M. Taberner, *Landsat 7 Enhanced Thematic Mapper JRC-FAPAR – Algorithm Theoretical Basis Document*. EUR Report 23554 EN. Institute for Environment and Sustainability, Joint Research Center (2008)
7. N. Gobron, B. Pinty, F. Mélin, M. Taberner, M.M. Verstraete, *Sea Wide Field-of-View Sensor (SeaWifs) – An optimized FAPAR algorithm – Theoretical basis document*. EUR Report 20148 EN. Institute for Environment and Sustainability, Joint Research Center (2002)
8. N. Gobron, *Envisat’s Medium Resolution Imaging Spectrometer (MERIS) – Algorithm Theoretical Basis Document: FAPAR and Rectified Channels over Terrestrial Surfaces*. EUR Report 24844 EN. Institute for Environment and Sustainability, Joint Research Center (2011)

9. D. Kondrashov, M. Ghil, Spatio-temporal filling of missing points in geophysical data sets. *Nonlinear Process. Geophys.* **13**(2), 151–159 (2006)
10. N. Golyandina, E. Osipov, The “caterpillar”-SSA method for analysis of time series with missing values. *J. Stat. Plan. Infer.* **137**(8), 2642–2653 (2007)
11. F. Takens, Detecting strange attractors in turbulence, in *Dynamical Systems and Turbulence*, ed. by D.A. Rand, L.-S. Young. Lecture Notes in Mathematics, vol. 898 (Springer, Heidelberg, 1981), pp. 366–381
12. M.B. Kennel, R. Brown, H.D.I. Abarbanel, Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys. Rev. A* **45**(6), 3403–3411 (1992)
13. R Core Team, R: A Language and Environment for Statistical Computing (R Core Team, Vienna, 2012). <http://www.R-project.org>
14. J.-P. Eckmann, S.O. Kamphorst, D. Ruelle, Recurrence plots of dynamical systems. *Europhys. Lett.* **4**(9), 973 (1987). <http://stacks.iop.org/0295-5075/4/i=9/a=004>
15. H. Lange, Recurrence quantification analysis in watershed ecosystem research. *Int. J. Bifurcat. Chaos* **21**(4), 1113–1125 (2011)
16. J.P. Zbilut, C.L. Webber, Embeddings and delays as derived from quantification of recurrence plots. *Phys. Lett. A* **171**(3–4), 199–203 (1992). ISSN: 0375-9601. doi:[10.1016/0375-9601\(92\)90426-M](https://doi.org/10.1016/0375-9601(92)90426-M). <http://www.sciencedirect.com/science/article/pii/037596019290426M>
17. C.L. Webber, J.P. Zbilut, Dynamical assessment of physiological systems and states using recurrence plot strategies. *J. Appl. Physiol.* **76**(2), 965–973 (1994)
18. N. Marwan, M.C. Romano, M. Thiel, J. Kurths, Recurrence plots for the analysis of complex systems. *Phys. Rep.* **438**(5–6), 237–329 (2007). ISSN: 0370-1573. doi:[10.1016/j.physrep.2006.11.001](https://doi.org/10.1016/j.physrep.2006.11.001). <http://www.sciencedirect.com/science/article/pii/S0370157306004066>
19. C.E. Shannon, A mathematical theory of information. *Bell Syst. Tech. J.* **27**(3), 379–423 (1948)
20. Z. Gao, K. Li, Evolution of traffic flow with scale-free topology. *Chin. Phys. Lett.* **22**, 2711–2714 (2005)
21. J. Zhang, M. Small, Complex network from pseudoperiodic time series: Topology versus dynamics. *Phys. Rev. Lett.* **96**, 238701 (2006)
22. R.V. Donner, Y. Zou, J.F. Donges, N. Marwan, J. Kurths, Recurrence networks—a novel paradigm for nonlinear time series analysis. *New J. Phys.* **12**(3), 033025 (2010)
23. N. Marwan, J.F. Donges, Y. Zou, R.V. Donner, J. Kurths, Complex network approach for recurrence analysis of time series. *Phys. Lett. A* **373**, 4246–4254 (2009)
24. T.M.J. Fruchterman, E.M. Reingold, Graph drawing by force-directed placement. *Softw. Pract. Exp.* **21**(11), 1129–1164 (1991)
25. N. Gobron, A. Belward, B. Pinty, W. Knorr, Monitoring biosphere vegetation 1998–2009. *Geophys. Res. Lett.* **37**, L15402 (2010)
26. M.C. Romano, M. Thiel, J. Kurths, W. von Bloh, Multivariate recurrence plots. *Phys. Lett. A* **330**, 214–223 (2004)
27. J. Zscheischler, M.D. Mahecha, S. Harmeling, Climate classifications: the value of unsupervised clustering. *Procedia Comput. Sci.* **9**, 897–906 (2012)

Chapter 13

Recurrence Analysis Applications to Short-Term Macroscopic and Microscopic Road Traffic

Eleni I. Vlahogianni, Matthew G. Karlaftis, and John C. Golias



Abstract The applicability of the theory of Recurrence Plots and Recurrence Quantification Analysis to short-term traffic flow is demonstrated through three distinct road traffic case studies. The first focuses on short-term traffic patterns of volume and speed in urban freeway sections. The second case study examines urban traffic flow dynamics under different traffic conditions and associates them to specific short-term statistical characteristics. The third case study discusses the use of the Recurrence Analysis for modeling the dynamics of the microscopic car following behavior on freeways. The applicability is discussed at a conceptual level and each case study is then presented. Finally, the modeling implications of the results on traffic flow prediction are discussed.

13.1 Introduction

Traffic flow instabilities, including rapid variations in speed and traffic density and observed stop-and-go driving conditions, form a major problem for traffic management and affect the overall traffic operation in road networks. Instabilities may lead to a significant reduction of the total traffic throughput on highways and to

E.I. Vlahogianni (✉) • M.G. Karlaftis • J.C. Golias

Department of Transportation Planning and Engineering, School of Civil Engineering, National Technical University of Athens, 5, Iroon Polytechniou Str., Athens 15773, Greece

e-mail: elenivl@central.ntua.gr

an increase accident risk. They may also induce higher fuel consumption and impose significant delays to drivers. Instabilities are reflected to the time series of traffic variables as irregularities and highly non-stationary spatio-temporal characteristics. Instabilities may be observed from different angles. From a stationary observer, instabilities are usually understood as cyclic variations in average travel speed, along with inverse variations in density. From a still video caption, instabilities are seen as waves of density that propagate “upstream” of the location of an initial disturbance. From the view of a single vehicle, instabilities may be observed as quasi-periodic variations in vehicle velocity.

There are several different ways of explaining the emergence of instabilities, as well as their implications to the efficiency of the road system [1–13]. A variety of traffic flow models have been developed that may be broadly categorized into four classes: microscopic, mesoscopic, fluid-dynamical and truly macroscopic models (for a review see [14–19]). The goal of traffic modeling is to understand the fundamental macroscopic or microscopic dynamics that may occur in space and time.

Although such understanding is invaluable, traffic complexity is still unmet especially in metastable traffic conditions. The recent emerging information technologies for monitoring traffic have enabled the accurate retrospective study of traffic that aims to retrieve useful knowledge on the traffic flow spatio-temporal evolution and construct the underlying behavior of traffic flow through the use of abundant data and powerful computational approaches. This knowledge focuses on the identification of instabilities, regimes and transitions and may lead to improving the understanding of the frequently observed boundary conditions induced by congestion and non-recurrent incidents and, thus, enhance traffic flow modeling.

Recurrence Plots introduced by Eckmann et al. [20], quantified by Zbilut and Webber [21], Webber and Zbilut [22] and extended by Marwan et al. [23] is based on the concepts of nonlinear dynamics and may be used to reveal and characterize traffic instabilities and quantify the observed complexities. Given extended series of traffic data, the methodology applied provides the ability to study traffic’s short-term dynamics and its statistical characteristics without prior assumptions on data’s non-stationary and nonlinear features. The specific methodological approach may be extended to account for multivariate relationships between different traffic variables.

We aim to demonstrate the manner recurrence plots and their subsequent quantification may be used to analyze short-term traffic flow in both univariate and multivariate settings. The applicability of recurrence analysis will be discussed in a conceptual level and will be further evaluated through three distinct case studies. The first refers to macroscopic traffic and focuses on the short-term traffic patterns of volume and speed in urban freeway sections. The second case study is based on urban arterial short-term traffic flow and aims to identify the traffic dynamics in the different traffic conditions and associate them to specific short-term statistical characteristics. The third case study centers to microscopic car following behavior and attempts to study the short-term evolution of a subject vehicle’s speed and spacing, as well as its interaction with the lead and following vehicles.

13.2 Analysis of Recurrences of Traffic Flow

The analysis of recurrences is based on two fundamental concepts of nonlinear dynamics [24]: (1) Some states of a system may occur over and over again and (2) Similar patterns may most likely evolve in a similar manner. Recurrence Plots (RPs) and Recurrence Quantification Analysis (RQA) provide a methodological framework to study the short-term traffic dynamics. The term dynamics is used to describe the temporal evolution of traffic flow patterns, but may be also extended to consider spatial patterns.

A traffic flow pattern may be defined in both a univariate and a multivariate setting. In the univariate setting, a traffic pattern can be considered as a sequence of states of a traffic variable; for example in the case of a traffic volume, the vector $\mathbf{V}_t = \{V(t), V(t - \tau), \dots, V(t - (m - 1)\tau)\}$ describes the pattern of traffic volume. The task of transforming a scalar variable to a vector is known as embedding in the phase-space. Parameters τ and m are the time delay and dimension respectively and define the depth of information a traffic pattern conveys. Choosing lower value of time delay τ than needed, will result to vectors with redundant information and will lead to poor predictions as there will be no difference between the elements of the delay vectors [25], whereas larger τ values than needed, will result to vector's coordinates that are almost uncorrelated and the information becomes unrelated [26]. Moreover, choosing a wrong dimension m will result to a wrongful embedding and an unfeasible representation of traffic's temporal characteristics. There exist various methodologies for the estimation of the optimum values of the τ and m . In this work the mutual information and the false nearest neighbors will be utilized. In brief, mutual information measures the information flow between sequential time delays [27], while the false nearest neighbors examines the behavior of near neighbors under changes in the dimension from m to $m + 1$ [28].

A RP is a graphical tool to visualize the characteristics of the temporal evolution of traffic. Using the embedded time series in the Phase-Space, a symmetric matrix of Euclidean distances may be constructed by computing the distances between all pairs of embedded vectors. The recurrence of a state x the time i and j is given by [29]:

$$R_{i,j}^{m,\varepsilon_i} = \Theta \left(\varepsilon_i - \left\| \vec{x}_i - \vec{x}_j \right\| \right), \quad \vec{x}_i \in \mathbb{R}^m, \quad i, j = 1 \dots N \quad (13.1)$$

where, N is the number of states x_i in the time window of study, ε_i is the distance threshold (Euclidean distance) for the distances $r_{i,j}$, $\|\cdot\|$ a norm (Euclidean norm), and $\theta()$ the Heaviside function: $\theta(y < 0) = 0$, $\theta(y \geq 0) = 1$. Given a constant value of threshold ε_i , the RP is symmetric. Moreover, because $R_{i,i} = 1$ ($i = 1, \dots, N$), RP consists of a diagonal line with an angle of $\pi/4$ called Line of Identity (LOI). The purpose of RPs is to visually distinguish several patterns that indicate the statistical properties of the time series such as non-stationarity, nonlinearity, drifts in data and so on. The different large and small scale structural patterns that may emerge along with their statistical interpretation are reviewed in Gao and Cai [30] and Marwan et al. [24].

RPs can be extended to incorporate multivariate relationships through Cross-Recurrence Plots (CRPs) and Joint Recurrence Plots (JRPs). CRP is a bivariate extension of RP and may be used to analyze two traffic variables (e.g. volume and speed) by comparing their states; it may be seen as a nonlinear cross-correlation function [24]:

$$\mathbf{CR}_{i,j}^{\vec{x},\vec{y}}(\varepsilon) = \Theta\left(\varepsilon - \left\|\vec{x}_i - \vec{y}_j\right\|\right), \quad i = 1, \dots, N, j = 1, \dots, M \quad (13.2)$$

where x_i, y_j the two time series, ε_i is the threshold (Euclidean distance) for the distances $r_{i,j}$, $\|\cdot\|$ a norm (Euclidean norm), and $\Theta(\cdot)$ the Heaviside function. The JRP considers the recurrences of the trajectories of the two time series separately and attempts to find times where these two time series recur [24]:

$$\begin{aligned} \mathbf{JR}_{i,j}^{\vec{x},\vec{y}}\left(\varepsilon^{\vec{x}}, \varepsilon^{\vec{y}}\right) \\ = \Theta\left(\varepsilon - \left\|\vec{x}_i - \vec{x}_j\right\|\right) \Theta\left(\varepsilon - \left\|\vec{y}_i - \vec{y}_j\right\|\right), \quad i, j = 1, \dots, N \end{aligned} \quad (13.3)$$

where $\varepsilon^{\vec{x}}, \varepsilon^{\vec{y}}$ the two different thresholds of distances of the two systems.

CRPs are more appropriate to investigate relationships between the parts of the same system, for example speed and volume at a specific location in a road system, with the condition that their ranges need to be mutually adjusted, whereas JRPs are more appropriate for the investigation of two interacting systems which influence each other, for example volume for two consecutive locations in a road [24].

Based on the density of the recurrent points in the RP it is possible to further statistically evaluate the observed patterns by introducing several recurrence measures [22]: the (%) recurrence rate (RR) (the density of the recurrence points in the PR), the (%) determinism (DET) (the percentage of recurrent points that form lines parallel to the main diagonal), the entropy (ENT) (the Shannon entropy of line segment distribution), the L_{\max} (reciprocal of the longest diagonal line segment) and the trend (TREND) (measure of the positioning of recurrent points away from the central diagonal). The above statistics form the Recurrence Quantification Analysis (RQA) that may be further extended with two other measures, the LAM and TT that describe the laminarity of traffic states and the chaos to chaos transitions [23].

The RR measure is an indication of correlation and may correspond to the definition of the correlation sum [24]. DET is related to deterministic structure that may have high or low complexity, a characteristic that is quantified by the ENT, a measure analogous to Shannon Entropy. Moreover, L_{\max} is inversely related to the largest positive Lyapunov Exponent [30]; low values indicate strong nonlinearity. TREND is a least square regression from the diagonal to the plot's corner and provides evidence of changing dynamics in the system [29]. LAM measures the laminarity in traffic patterns, whereas TT the trapping time; both measures allow for investigating intermittency even in short non-stationary time series [23]. It is to note that the same approach for quantifying the statistical characteristics of

traffic patterns based on the recurrent point densities may be extended to CRPs. For a review of the above, as well as some supplementary measures that may be introduced see [24].

The PRs and RQA, as well as their multivariate extensions, are an efficient constraint-free way to visualize and identify complex statistical features of traffic. These approaches may be used to uncover time correlations between data that are not based on linear or nonlinear assumptions and cannot be distinguished through the direct study of one-dimensional series of volume. The dynamics of traffic patterns, implying the manner in which traffic volume states propagate through time, can be revealed by the study of the evolution of the above RQA measures in sliding windows of time W_i :

$$W_i = (V_{t_1+(i-1)T}, V_{t_2+(i-1)T}, \dots, V_{t_N+(i-1)T}) \quad (13.4)$$

W_i satisfies $N + (i - 1) T \leq n$

where N is the number of measurements in W_i and T is the time interval.

13.3 RQA Applications to Road Traffic

13.3.1 *Macroscopic Freeway Traffic Dynamics Using Recurrences*

We utilize the RPs and the RQA approach in order to visualize and statistically quantify the properties of lane volume and travel speed time series in urban freeways. As shown in Fig. 13.1, data were obtained from 16 loop detectors and cameras located at major traffic interchanges that connect the center of Athens with the southern, western and eastern regions of the city. Lane volume and travel speed are collected every 90 s. For the specific application, data between July 2006 and June 2007 are utilized.

Figure 13.2 shows the time-series of travel speed (km/h) for selected weekdays and for Saturday during a typical week of the year. As can be observed a variable temporal evolution in both speed and volume can be identified. The onset of morning peak period is marked by the significant increase in lane volume during the morning hours (6:00–10:00). The observed speed drops may signify the formation of queues in a downstream section; such behavior is detected during the morning peak 6:00–10:00 and the afternoon peak 16:00–20:00—but with a variable duration and magnitude.

This variability in traffic can be also observed in the speed–volume relationship. As seen in Fig. 13.3 two regions are distinguished: (a) the free-flow (unqueued) region characterized by high speeds and densely located traffic states with volume values ranging from zero to effective capacity and (b) the congested (queued) region (area below the dashed line) marked by the wide scattering of speed–volume data.

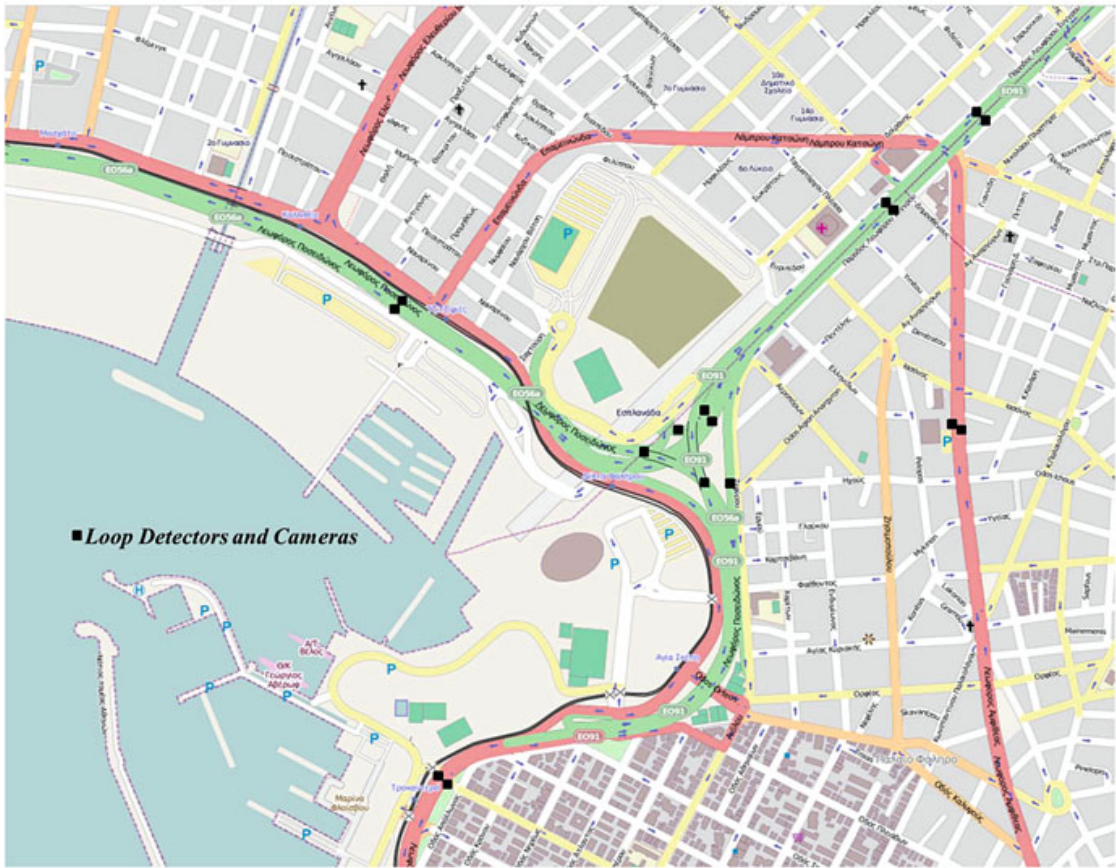


Fig. 13.1 Graphical representation of the freeway sections under study and the controlled locations. Points signify the locations of loop detectors and cameras

Figures 13.4 and 13.5 demonstrate the RPs for three different weekdays and Saturday of a typical week where no adverse conditions were observed (e.g. accidents, extreme weather conditions, strikes, rallies etc.). As can be observed a square-like structure is apparent in all RPs indicating a cyclic evolution of both volume and speed in 90 s intervals, whereas vertical and horizontal lines/clusters are indication of laminar states [24].

The stochasticity of free flow during the early morning period is depicted as a band of warm colors in the volume RPs, where the abrupt onset of congestion (stop-and-go conditions) during the evening observed as very low speed values are depicted by warmer colored bands in the RPs of speed. The temporal evolution of both variables is quite distinct and differs between weekdays and weekends, or even between the working days.

The RQA approach is conducted in a sliding time window of 1 h updated every 1 interval (90 s). Conducting an RQA analysis within a sliding window provides the opportunity to incorporate the sudden changes of traffic flow and their relevant statistical characteristics. Figure 13.6 shows the times series of recurrence (RR) of volume and speed along with the time series of volume and occupancy for a typical Friday. Two distinct periods are depicted: a. the onset of congestion from 6:00 to 10:00 (241–400 count), and b. the afternoon peak period from 16:00 to 20:00 (641–800 count). Recurrences are largely affected by the onset of morning

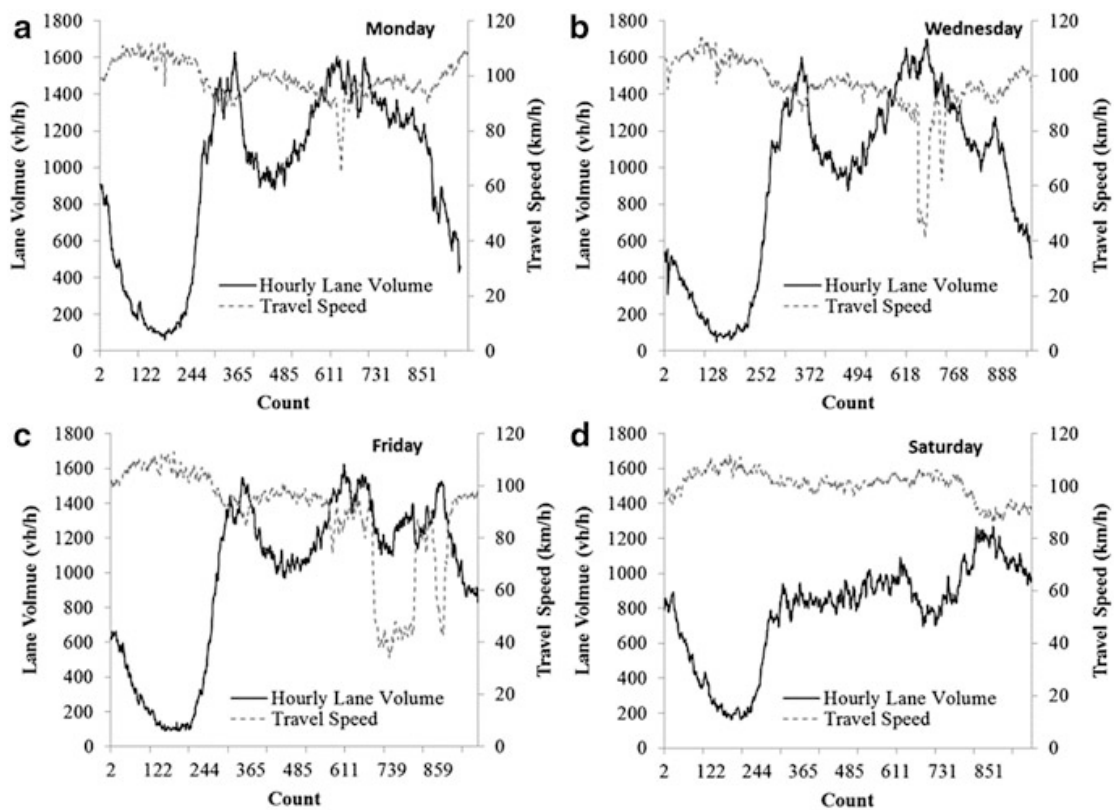


Fig. 13.2 Speed and volume time-series for the busiest weekdays and weekend in the study area.

congestion, as well as the afternoon peak period. The fluctuations in the volume's and travel speed's RR statistics are indicative of a piece-wise evolution of traffic flow, an issue that should be considered in the model selection process of short-term traffic forecasting. The peaks observed in RR time series may be considered as transitions that take place in traffic and, if associated to specific traffic conditions (joint consideration of volume and speed) may indicate the time interval where a different modeling approach should be applied. This may have significant implications to the process of short-term traffic prediction, as models will be asked to provide prediction under stable traffic conditions.

The subsequent analysis of the time series of DET, L_{\max} , ENT, LAM and TT may reveal more specific statistical characteristics on the evolution of short-term traffic flow (Fig. 13.7).

Strong determinism in the temporal evolution of volume may be traced during the formation of congested traffic flow conditions in the morning. A sharp transition from periodic to chaotic temporal evolution is observed at the beginning of the morning peak period, as seen in the L_{\max} series of traffic volume.

The volume increase at the morning is marked by a sudden increase in the entropy (ENT), laminarity (LAM) and trapping time (TT). After the morning traffic volume increase, the temporal evolution is associated to a constant periodic-to-chaotic (peaks in L_{\max}) and chaotic-to-chaotic (peaks in LAM) transitions.

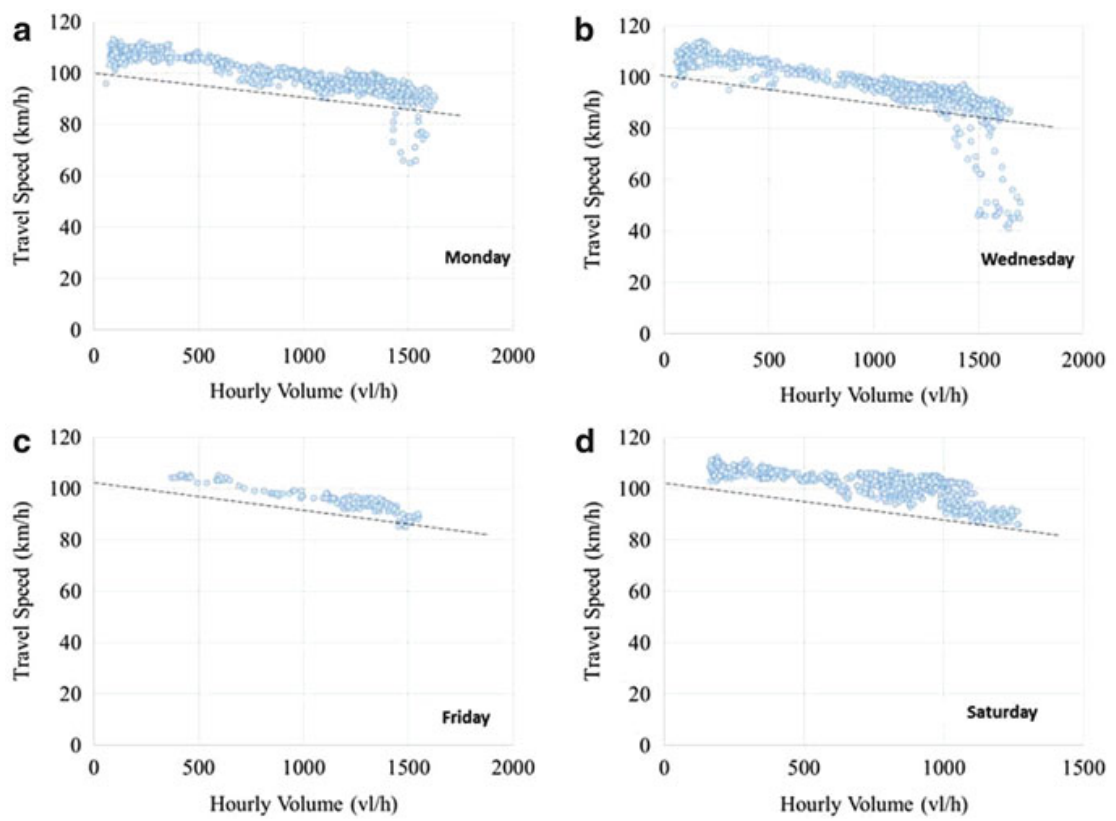


Fig. 13.3 Speed–volume relationship for the busiest weekdays and weekend of the year. The area below the *dashed* line is the congested area

Speed’s temporal evolution follows a different behavior. The interest is on the evening peak period where a sudden speed drop is observed (around count 650 as seen in Fig. 13.6); before the occurrence of the high speed drop, there is a short period where an abrupt drop in all RQA statistics is observed. This may signify that the onset of the congested conditions may be marked by a sudden shift of traffic to stochastic temporal behavior, where the long memory properties of traffic fade; this outcome is in line with recent findings using fractionally integrated time series models [31].

Table 13.1 summarizes the results of a nonparametric test on the comparison of distributions of RQA variables between weekdays and weekend, peak and off-peak periods, as well as morning and evening peak. The test has the null hypothesis that the two samples tested are from populations with the same distribution function. As can be observed, the RQA variables have distributions that vary according to the type of day, the period within a day, or even between two different peak periods of a weekday.

The above provide some evidence that urban freeway traffic has an irregular temporal statistical evolution. Irregularities are observed in both volume and speed evolution with different frequency and magnitude. These irregularities may provide a guidance on which prediction strategy to use (linear vs. non-linear,

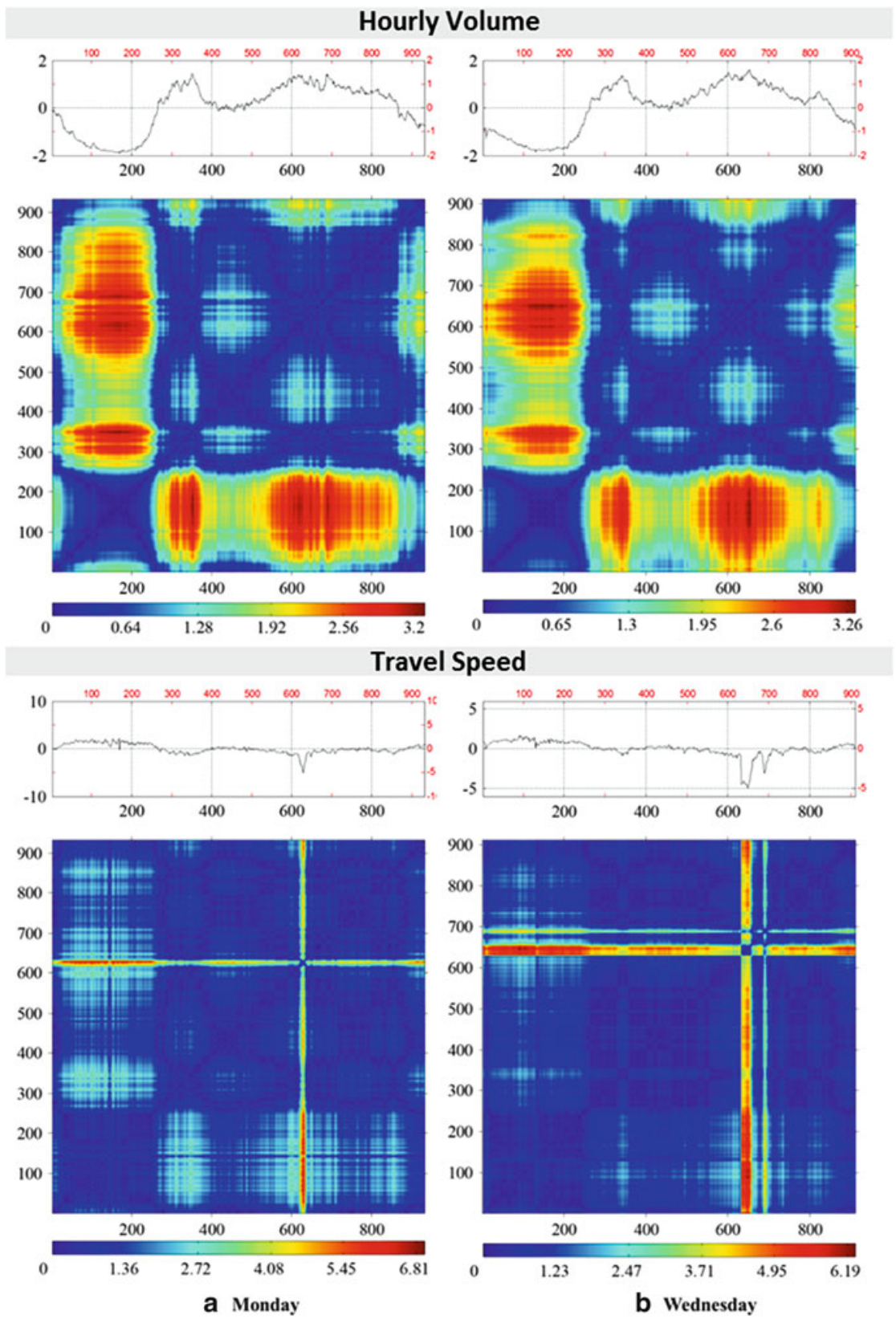


Fig. 13.4 Recurrence plots for Monday and Wednesday traffic volume and speed during a typical week ($\tau = 1$, $m = 3$, $\varepsilon = 0.15 \sigma$, *color band* indicates distance to next recurrence point)

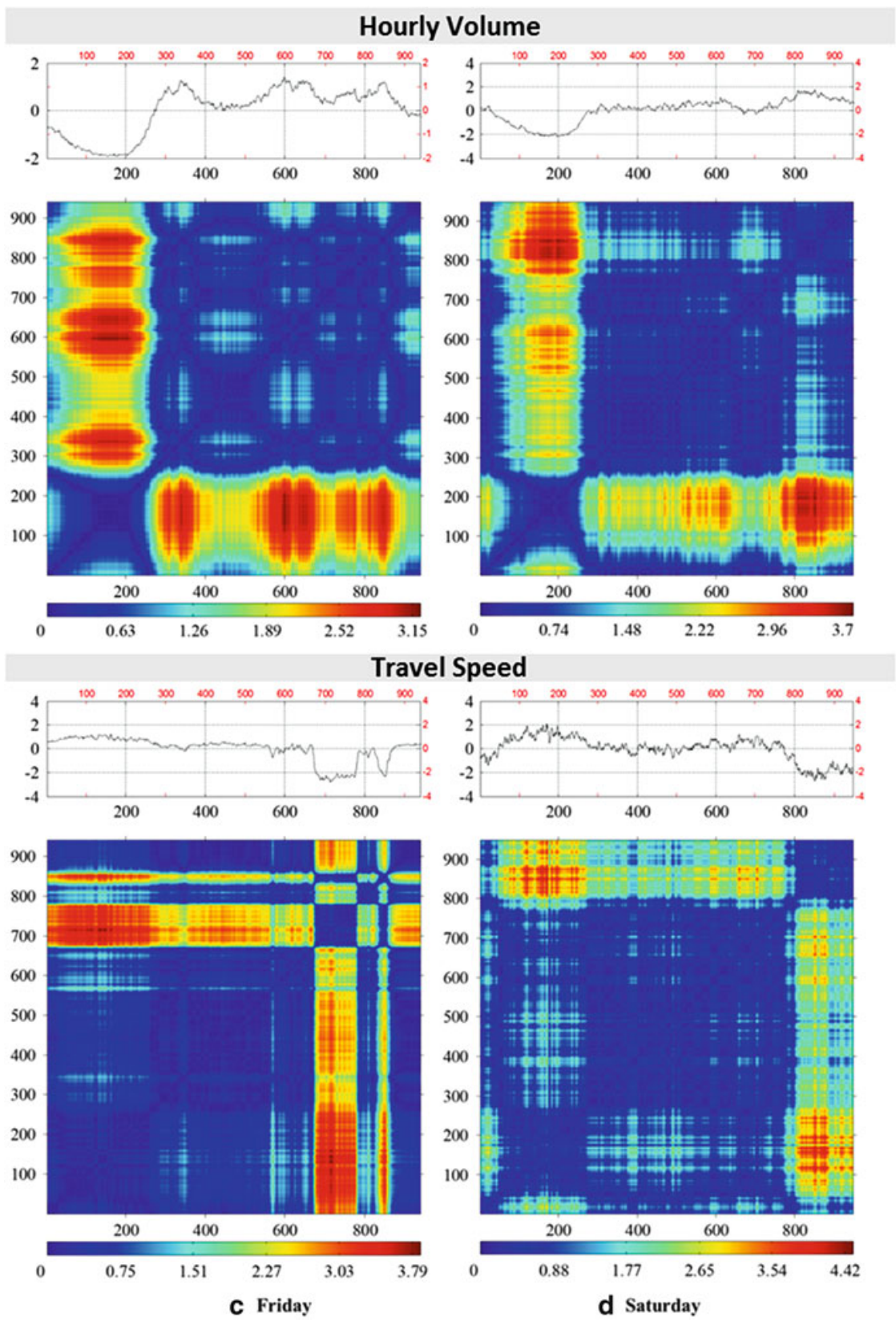


Fig. 13.5 Recurrence plots for Friday and Saturday traffic volume and speed during a typical week ($\tau = 1$, $m = 3$, $\varepsilon = 0.15 \sigma$, *color band* indicates distance to next recurrence point)

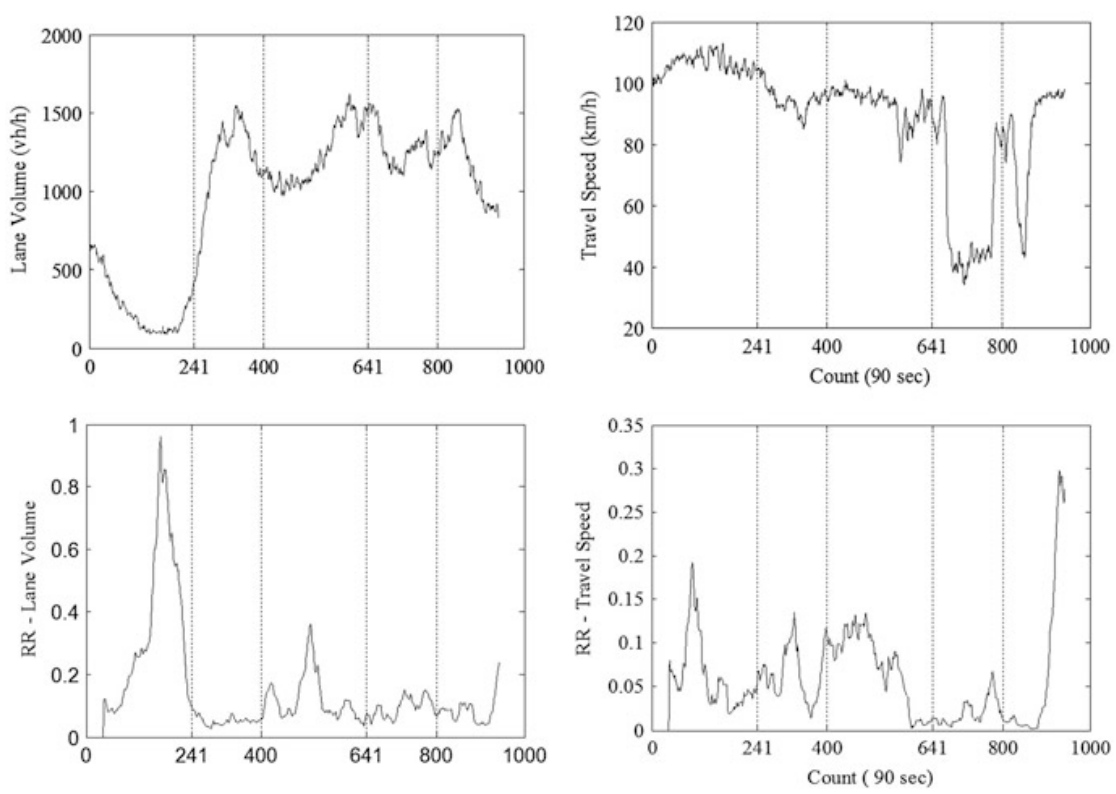


Fig. 13.6 Recurrence evolution of volume and speed during Friday ($m = 3$, $\tau = 1$, $\varepsilon = 0.06 \sigma$, $W = 40$ updated every 90 s)

non-stationarity treatment, treatment in the mean or variance and so on) and when to shift to a different prediction strategy to ensure a constant acceptable level of prediction accuracy.

13.3.2 Macroscopic Urban Traffic Flow Analysis and Congestion Detection

In the specific application the aim is to describe the statistical characteristics of traffic flow jointly considering the series of volume and occupancy. These characteristics will be associated to specific traffic conditions (free-flow, critical flow and jammed conditions). Transitional behavior will be also analyzed. The proposed approach for identifying traffic flow’s pattern-based evolution relies solely on data and implicitly considers the effect of signalization as a reflection on short-term traffic patterns.

Volume and occupancy measurements from arterial links in the center of Athens (Greece) are used; the arterial links are controlled by loop detectors (midblock) located 90 m from the stop-line. Figure 13.7 (first two graphs) depicts the volume and occupancy time-series for a week in a specific arterial location. As can be observed, the time-series under study exhibit a significant temporal variability and periods of strong short-term oscillations.

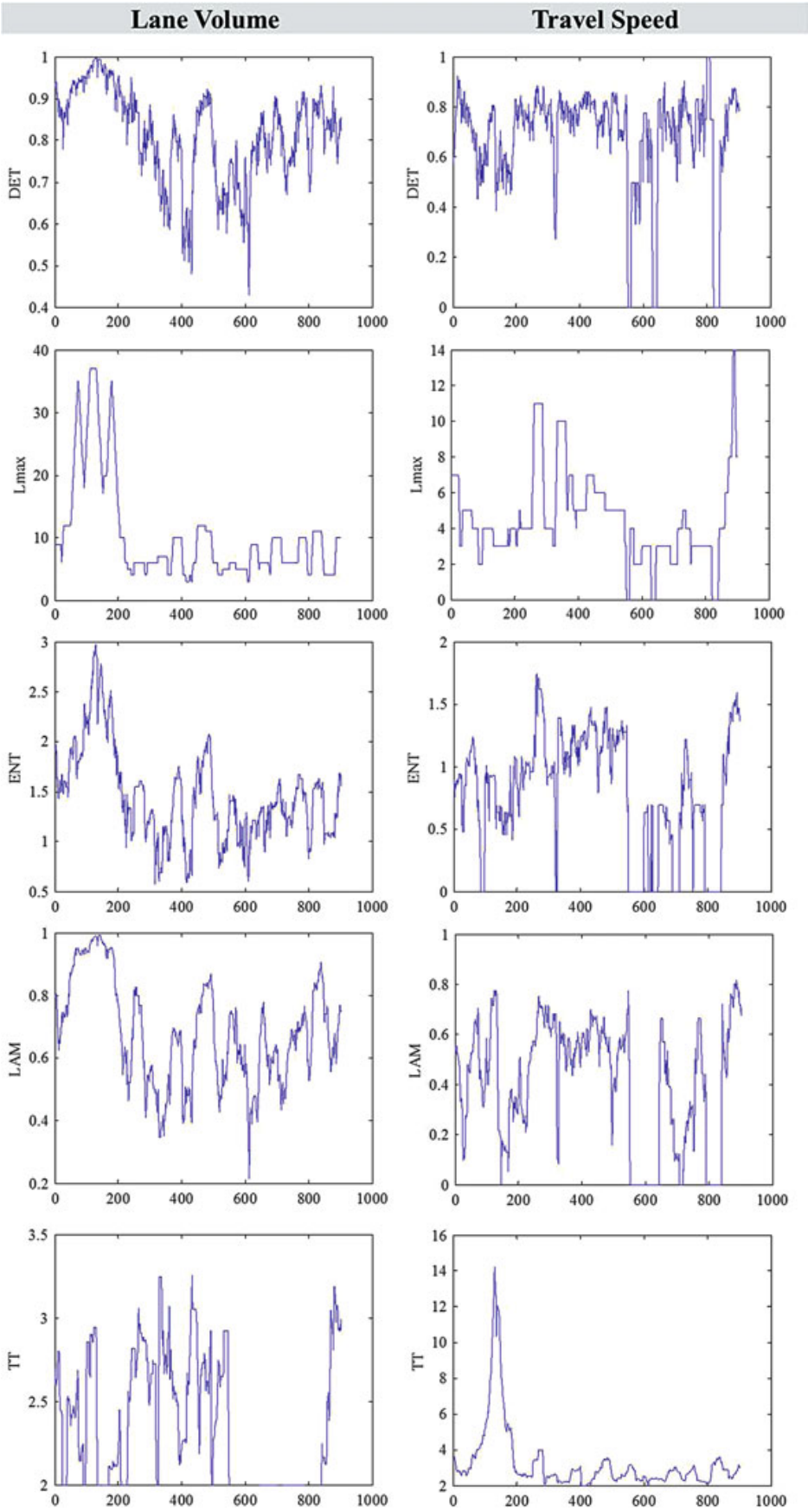


Fig. 13.7 RQA variables for volume and speed for a typical Friday ($m = 3$, $\tau = 1$, $\varepsilon = 0.06 \sigma$, $W = 40$ updated every 90 s)

Table 13.1 Mann–Whitney test results (t values) for RQA variables between weekdays and weekends, weekday morning peak and off-peak and afternoon peak and off peak periods

Distributions	Weekday vs weekend	Weekday morning peak and off-peak	Weekday afternoon peak and off-peak
RR	−56.55*	−18.46*	−9.49*
DET	−52.33	−9.15*	−8.59*
L	−56.64*	−12.276*	−12.90*
ENT	−47.47*	−13.79*	−12.25*
LAM	−48.26*	−10.24*	−12.54*
TT	−48.01*	−10.78*	−13.11*

*Significance level is 0.05

Volume and occupancy series are embedded in the phase-space by applying the mutual information and the false nearest neighborhood algorithms. Embedding is, first, applied to various 1-day time series of volume and occupancy that gradually shrinks to time windows of 1-h. This procedure yield $\tau = 1$ and $m = 5$ as the embedding parameters for both volume and occupancy, suggesting that the dynamics of traffic at time t should be studied in patterns of the following form:

$$\mathbf{V}\{t\} = \{V(t), V(t - 1), \dots, V(t - 4)\}, \quad \mathbf{O}\{t\} = \{O(t), O(t - 1), \dots, O(t - 4)\}.$$

Figure 13.8 shows the time series of the CRQA variables computed in a sliding time window of study of 1 h length that is updated every 1 time interval (90 s) and provide resourceful information on the complexity in arterial traffic flow. As expected, the daily traffic has a systematic cyclicity that is also observed in the CRQA statistics time series. Nevertheless, the rich traffic dynamics within the day are very difficult to quantify as both the traffic and CRQA statistics show a variable temporal evolution. Peaks in traffic variables may demonstrate extreme traffic flow conditions, whereas peaks in the CRQA variables may hint periodic-to-chaotic or chaotic-to-chaotic transitions that could be critical to the predictability of traffic flow time series. The selection of the proper methodology to predict traffic is significantly influenced by the statistical characteristics of the temporal evolution of traffic variables especially regarding non-stationarity, long-memory and nonlinearity.

Recently, a two-level clustering approach based on artificial intelligence and classical clustering algorithms has been implemented in order to associate traffic conditions to specific statistical characteristics of the evolution of traffic [32]. The first level incorporated a Kohonen Self-organizing Map to produce a set of prototypes of the form $\mathbf{w}_j = [meanV_j, meanO_j, \%DET_j, L_{maxj}]$, where $meanV_j$ and $meanO_j$ is the mean values of volume and occupancy in a sliding time window of 1 h, and DET and L_{max} the associated determinism and non-linearity if their joint temporal evolution. A subsequent k-means algorithm was also implemented to cluster the prototypes and produce the prevailing regions of pattern-based short term traffic evolution.

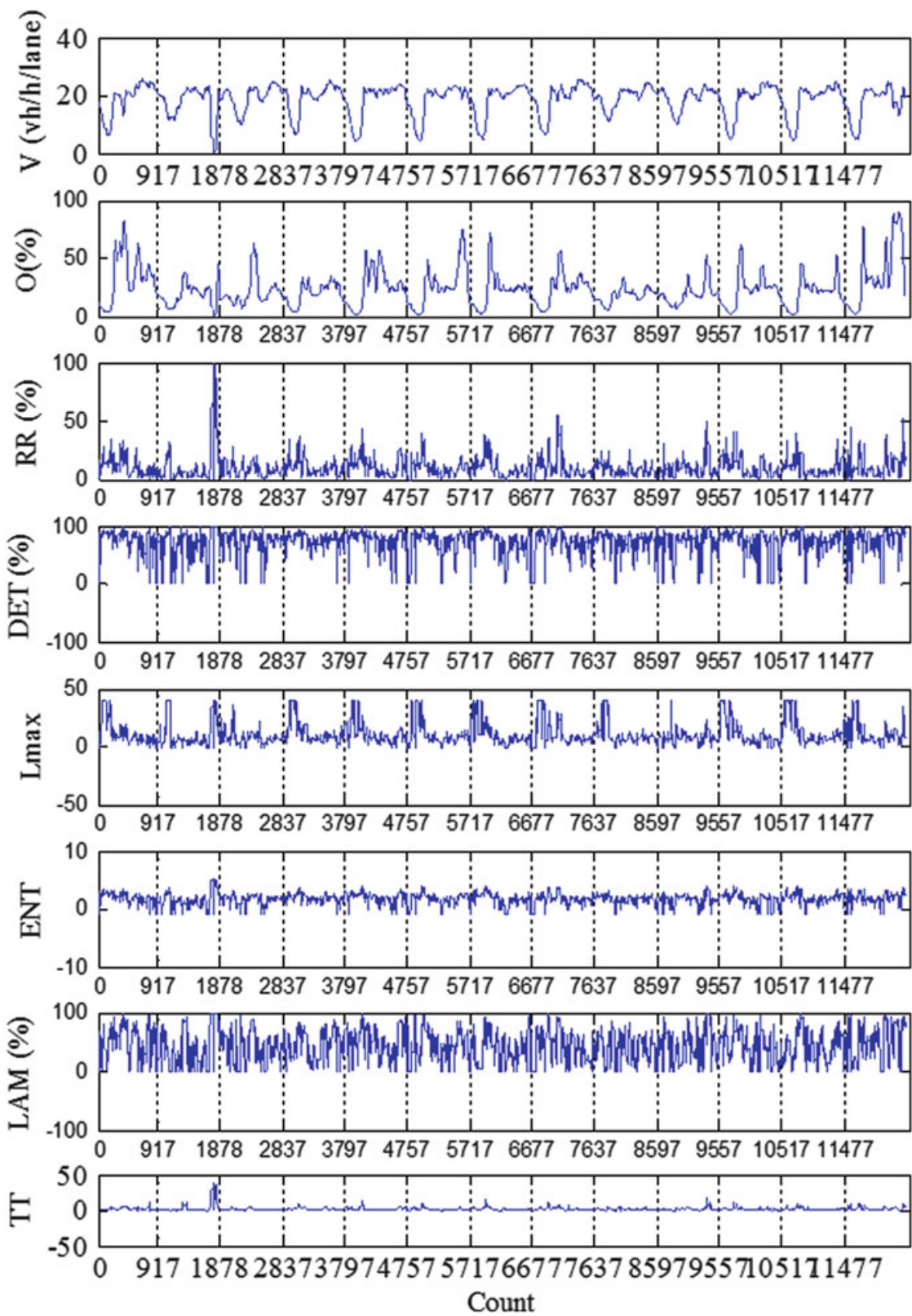


Fig. 13.8 CRQA variables of urban traffic flow for 13 sequential days ($\tau = 1, m = 5, \varepsilon = 0.25 \sigma$)

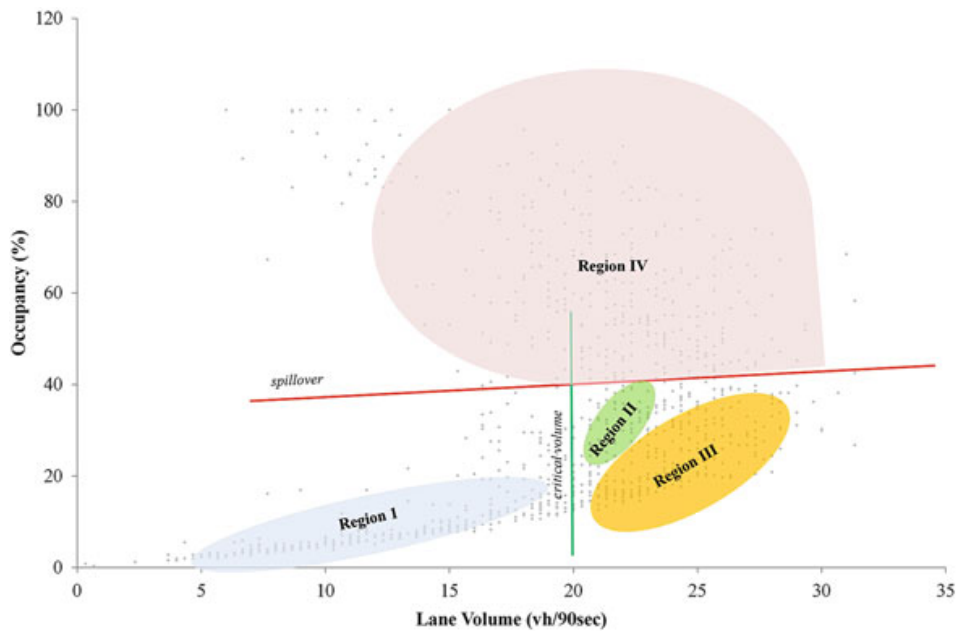


Fig. 13.9 Regions of pattern-based traffic evolution

Overall, four distinct regions were identified; results are seen in Fig. 13.9. Region I observed in average volume 28 ± 9 (veh/90 s) and average occupancy 6.18 ± 3.25 (%) is described by a stable weakly deterministic pattern-based short-term evolution (Medium values of DET and L_{\max}). Region IV (mean volume 64 ± 6 (veh/90 s) and mean occupancy 54.4 ± 14.93 (%)) reflects a pattern based short-term traffic evolution that is stable deterministic (high values of DET, medium values of L_{\max}). Interestingly, the Regions II and III near congested conditions for a critical area of flow are indicative of an unstable evolution of short-term traffic flow. In Region II (mean volume 62 ± 9 (veh/90 s), mean occupancy 9.01 ± 5.95 (%)), increased stochasticity is observed (low values of DET, low values of L_{\max}), whereas in Region III (mean volume 63 ± 6 (veh/90 s), mean occupancy 19.01 ± 5.95 (%)) the traffic flow evolution may be characterized as unstable deterministic (high values of DET, low values of L_{\max}). These features are indicative of the oscillating nature observed by the series of flow and occupancy (a finding also reported by [9, 33]); instability and chaotic-like behavior are observed near capacity.

Although the reasons why such traffic behavior occurs cannot be readily identified, the frequently documented in literature instability near capacity, may be distinguished into two different statistical evolutions; the deterministic structure that dominates the area of synchronized flow and a stochastic short-term evolution in synchronized flow, probably encompassing non-recurrent incidents. The observed regimes of statistical evolution are indicative of traffic's macroscopic complexity. Based on the popular Edie's hypothesis on the discontinuity observed in near to capacity traffic conditions, when traffic reaches a critical occupancy point, it may break down to an unstable state (flow decreases and density increases) or reach a bi-stable state of higher density and flow values than critical density point, until a discontinuity leads traffic to congestion [1]. Kerner [34] describes in part this

Table 13.2 Kruskal–Wallis test results for CRQA variables between the different regions

Region <i>i</i> – Region <i>j</i>	RR	DET	L _{max}	ENT	LAM	TT
I–II	10.356*	8.454*	12.185*	9.326*	0.328	6.106*
I–III	11.431*	8.950*	14.198*	10.162*	−1.432	5.991*
II–III	1.872	0.896	3.443*	1.466	−2.912*	−0.137
II–IV	−5.991*	−5.897*	−3.884*	−6.168*	−8.573*	−6.913*
III–IV	−7.011*	−6.371*	−5.792*	−6.960*	−6.904*	−6.804*
I–IV	2.888*	1.550	5.940*	1.980	−6.641*	−1.011

H₀: the two samples’ distributions are the same
*Two-sided test, rejection at significance level 0.05

behavior by introducing the notion of synchronized flow in order to describe the tendency to equilibrate the speed across lanes in dense enough traffic. Kerner and Rehborn [8] have provided empirical evidence that traffic in synchronized flow possesses different non-linear properties in comparison both with free and jammed traffic leading to describe congestion as a two phase process encompassing synchronized and wide moving jam flow.

Table 13.2 summarizes the results of a non-parametric statistical test with the null hypothesis that the extended CRQA variables distributions are equal across two different regions. As can be observed, all regions encompass unique distributional CRQA variables characteristics. Interestingly, Regions II and II differ in the L_{max} and LAM distributional characteristics.

The above regions may be related to the formation and dissipation of queues in arterials due to signalization. The boundary between free-flow (Region I) and near to capacity (Regions II and II) was found to be in the maximum volume value for which queue lengths do not extend beyond the detector is given by [35]:

$$V_i^* = \frac{k_{jam}}{\frac{k_{jam}}{C} + \frac{r}{L_d}}$$

(13.5)

where k_{jam} is the jam density and L_d is the detector length from the stop line, C is the cycle length and r is the duration of the red phase. Moreover, the onset of jammed conditions (Region IV) may be approximated by kinematic wave theory that identifies spillovers on the basis of the volume occupancy relationship and a critical value of occupancy O_{sp} [35]:

$$O_{sp} = \frac{L_{eff} \cdot V}{u_f} + \frac{r}{C}$$

(13.6)

where C is the cycle length, V (vehicles/time interval) is the average volume data (measured by the detectors), r is duration of the red phase, L_{eff} the typical vehicle length and u_f the free-flow speed. If the observed occupancy is larger than the critical value, then a queue spillover exists.

As for transitions, findings indicate that congestion is most likely to be reached through a nonlinear deterministic behavior (Region III), but may also occur in a highly stochastic manner (from Region II). Traffic flow departs free-flow conditions either through a nonlinear deterministic manner (and moves to Region III) or a sudden stochastic shift (and moves to Region II). The further study of transitions and traffic propagation may result to useful insights on the duration of various traffic phenomena, e.g. duration of the onset of congested conditions, duration of congestion, and duration of instabilities in critical traffic conditions.

13.3.3 Dynamics of Car Following Patterns Using Cross and Joint Recurrence Analysis

Since Chandler et al. [36] has published on the car following dynamics, there has been extensive research on the manner drivers tend to follow each other on a highway. The recent technological advances have provided an abundance of data on the microscopic kinematic characteristics of traffic and have enabled the thorough monitoring of the car following behavior in freeways and the instabilities in a microscopic level. Instabilities may emerge when a driver that has the tendency to keep a safe distance, while sustain a desired speed, will have to react to an external stimulus or a sudden change in the kinematic characteristics of the lead vehicle.

The Cross Recurrent and Joint Recurrent Analyses are implemented to study the short-term dynamics of the car following behavior on a single freeway lane. Two different cases are investigated: in the first, the subject vehicle is solely considered and the spot speed and spacing from the lead vehicle are studied using CRQA. In the second case, a system of two vehicles—either the subject with the lead vehicle or subject with the following vehicle—is considered and the dynamics of the interactions between those vehicles are further studied using the Joint Recurrence Analysis of the corresponding speeds.

Data from the I-80 freeway in San Francisco, California are utilized in order to study the short-term microscopic patterns of car following behavior in freeway traffic. I-80 freeway is located in an urban area and consists of 6 traffic lanes including the high occupancy vehicle lane. The data to be further analyzed is from April 13, 2005 for two time periods, 4:00–4:15 and 5:00–5:15. Two different datasets are available and will be jointly considered: the first is vehicle trajectory data transcribed from the video using the software developed for NGSIM (<http://ngsim-community.org/>). The second refers to extracted microscopic variables such as vehicle speed, acceleration, and spacing. The vehicles, whose trajectories are analyzed travel to the northbound direction of the freeway and are collected using video cameras. Data are available in 10 Hz resolution.

First, we focus on the traffic dynamics of a subject vehicle following a lead vehicle within the same lane. Figure 13.10 shows the speed spacing relationship and the corresponding time series of speed and spacing for a platoon of two vehicles.

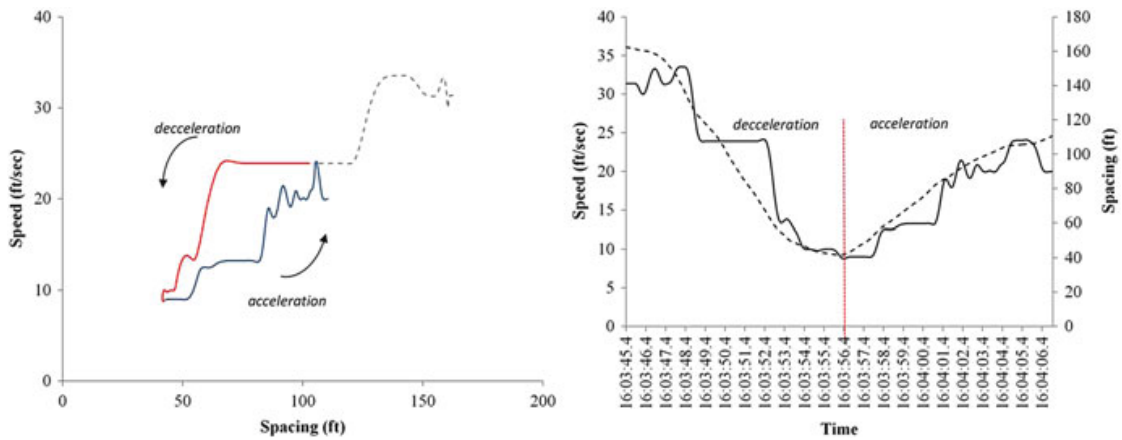


Fig. 13.10 The speed spacing relationship and the corresponding time series of speed and spacing for a platoon of two vehicles

As seen in Fig. 13.10, the subject vehicle decelerates as a response to the changes observed in the front vehicle and the upper branch of the speed–spacing relationship is formed leading from high speed and spacing values to lower values of the same traffic flow variables (Fig. 13.10), until some point in time where the subject vehicle starts to accelerate and the lower branch in the speed–spacing relationship emerges. The loop seen in Fig. 13.10 may correspond to a hysteresis phenomenon and is known as the Newell’s hypothesis of the asymmetric behavior of drivers between accelerating and decelerating conditions [3]; the accelerating and decelerating conditions in a speed–spacing graph are distinguished by shifts in the value of spacing. The above described behavior is reflected in the series of speed and spacing as a drop to lower values. Interestingly, by observing the CRQA variables produced by the cross-analysis of speed and spacing time series for the same case of a system of two vehicles of Fig. 13.11 produces some interesting results.

The deceleration phase is associated to low recurrent behavior, low deterministic structure, as well as low laminarity and indication of strong stochastic behavior (count 0–50 in Fig. 13.11). The acceleration phase (51–267) is associated with a variable statistical behavior. Patterns of high nonlinear deterministic structures and laminarity are replaced by low RR values, when the subject vehicle stops to accelerate and keeps stable speed for a short period.

Second, the dynamics of the interaction of the subject with the lead and following vehicle, as a system of two vehicles, are studied through the joint recurrence analysis of the corresponding speeds. Figure 13.12 demonstrates the time series of the subject, lead and following vehicle speeds, the corresponding covariances with the lead and following vehicles, as well as the joint RR with the lead and following vehicle (Counts refer to 10 Hz intervals).

The time series of speeds in Fig. 13.12 reflect two consecutive decelerations and accelerations of the lead vehicle and the manner the subject and following vehicle adapt their own speed. The subject vehicle first react to the first deceleration of the lead vehicle, reduces its speed, and then gradually increases its speed until

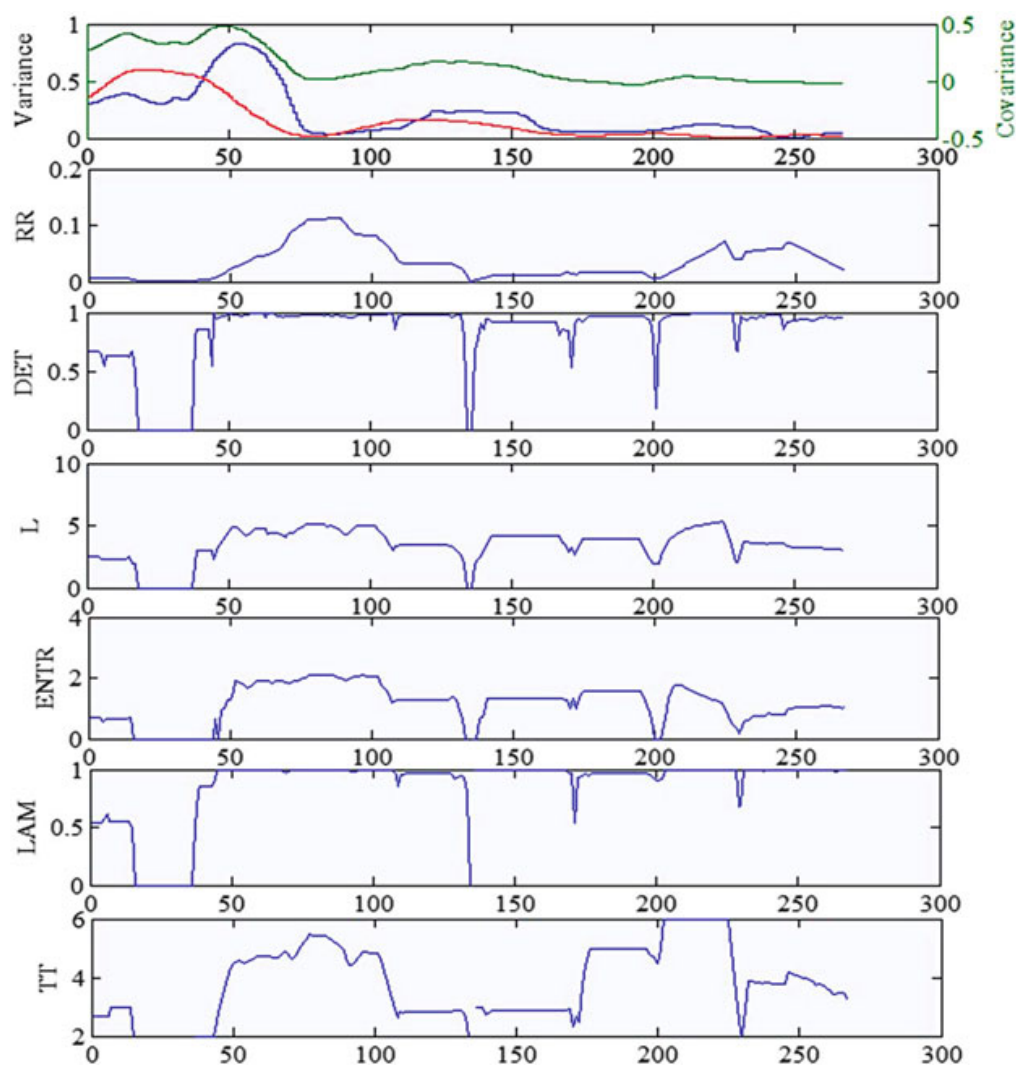


Fig. 13.11 CRQA variables of a system of two vehicles ($\tau = 1$, $m = 5$, $\varepsilon = 0.1 \sigma$, $W = 50$)

the second deceleration of the lead vehicle. Then, it seems that the subject vehicle cautiously follows the changes of the kinematic characteristics of the lead vehicle; the change in speeds observed in the second deceleration and acceleration of the lead vehicle are smooth.

As can be observed in Fig. 13.12, a different temporal dependence between the subject vehicle and the lead and following vehicles is identified. The joint recurrence of the subject with the lead vehicle is low at the first time the lead vehicle starts to decelerate and remains low until the second time of deceleration. The recurrent behavior of the two vehicles seems to be synchronized at the second time the lead vehicle decelerates leading to assume that the driver of the subject vehicle has adapted to the first change of the kinematic characteristics of the lead vehicle and drives in a recurrent manner that coincides to the recurrent driving of the lead vehicle.

The interaction with the following vehicles is quite the opposite. The driving of the following vehicle is synchronized with the subject vehicle until the acceleration

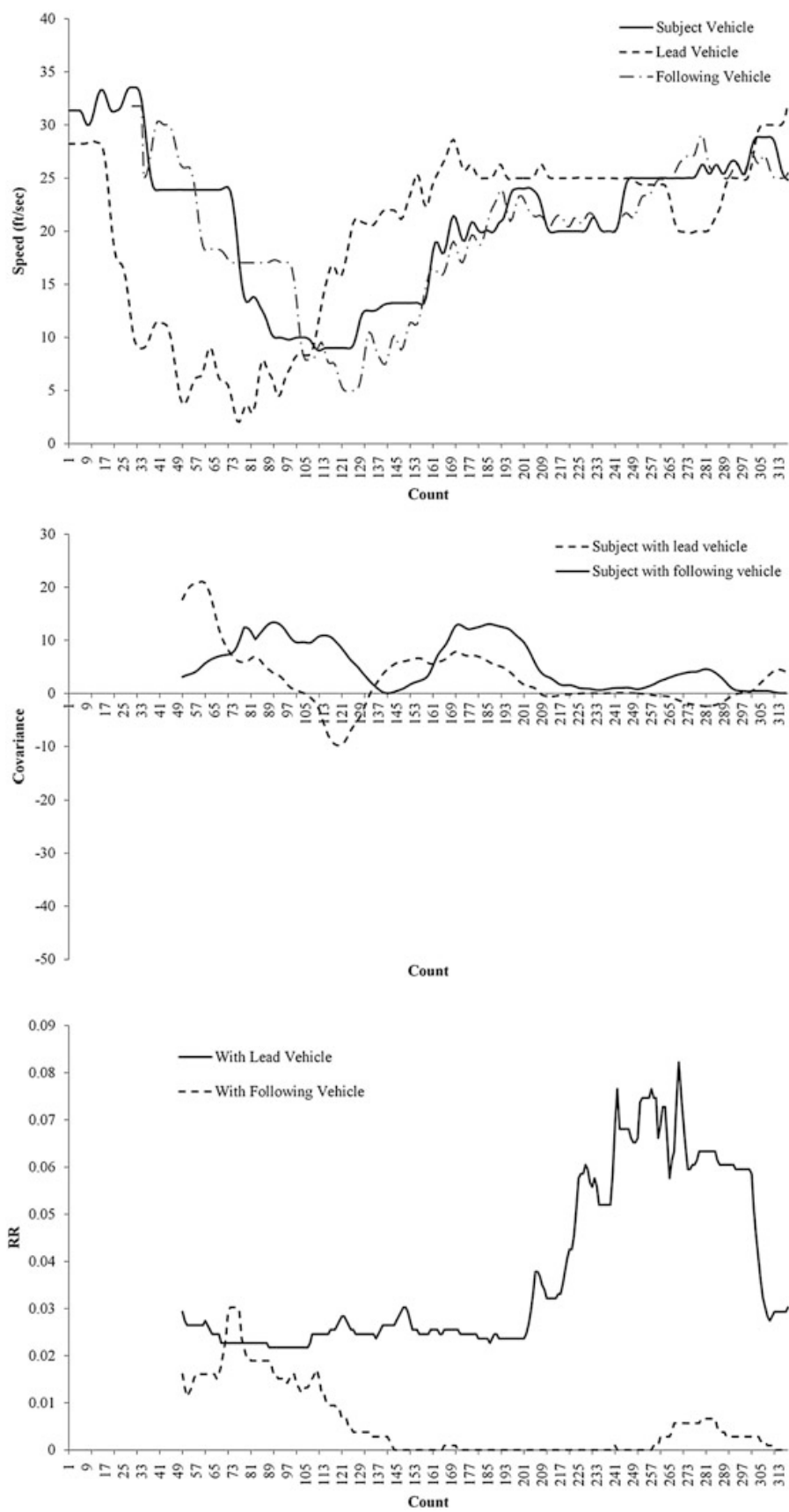


Fig. 13.12 The time series of the subject, lead and following vehicles, the covariance with the lead and following vehicle and the joint RR with the lead and following vehicles ($\tau = 1$, $m = 5$, $\varepsilon = 0.1 \sigma$, $W = 50$)

of the subject vehicle where the system loses its synchronization. This may be attributed to an anticipated lag between the reaction times of the subject and the following vehicles to changes of the lead vehicle. This lag and the loss of synchronization may be further associated to the manner a disturbance travel upstream its first occurrence (wave speed) and demonstrates that driving during car following conditions is a task done on the verge of instability. Although no generalization can be done based on the above observations, JRPs may be used to identify the synchronization between vehicles and define the information transfer between consecutive vehicles on road networks. This information transfer may be used for control purposes in order to suppress instabilities and stabilize traffic in a desired value of throughput.

13.4 Discussion on Modeling Implications and Conclusions

The applicability of RPs and RQA on both univariate and multivariate short-term traffic flow has been demonstrated through three distinct case studies. The first application showed that traffic flow has a variable statistical evolution and this should be taken into consideration when attempting to provide short-term predictions. The conceptual problem of short-term traffic flow prediction seems not to be a simple time series problem that may be conducted using either a linear or a nonlinear approach; findings indicated that traffic flow exhibits discontinuities in its temporal evolution and the adopted prediction method should be robust in treating patterns, rather than be based on data stream continuity.

The second application showed that if the statistical characteristics of traffic flow are associated with the prevailing traffic flow conditions we may come up with some interesting results on the manner traffic flow evolve with time, as well as which are the critical characteristics for shifting between the different traffic flow conditions. In literature, it is systematically supported that in each traffic flow condition (free-flow, synchronized flow or congestion) each variable's temporal evolution has a different effect in the transitional behavior of traffic flow [34]. Therefore, it should be more useful to develop multiple simple prediction models to account for the complex temporal traffic evolution in relation to anticipated traffic conditions and transitional phenomena and assure a stable prediction performance, rather than strangling to come up with one complex algorithmic structure to account for the prediction of more than one variable.

The third application focused on the car following characteristics of microscopic traffic. The variable statistical characteristics of the microscopic short-term traffic patterns identified, as well as the ability to detect synchronization between the subject and the lead and following vehicle may have significant implications to the investigation of the manner a disturbance may propagate upstream to a platoon of vehicles. In the above several critical phenomena related to the reaction times of drivers and their tendency to shift from aggressive to timid driving and form hysteretic phenomena in the macroscopic level are reflected. The knowledge of

the probability that vehicles move in a recurrent manner or are synchronized with respect to the kinematic characteristics of the lead or following vehicle may be used to develop advanced vehicle control strategies that may target the overall suppression of microscopic traffic instabilities in order to achieve the maximization of the throughput of road networks.

The applications presented showed that RPs and RQA, as well as their multivariate extensions are a robust methodological platform for analyzing short-term traffic flow in either macroscopic or microscopic level. By applying these approaches resourceful information on the statistical characteristics of traffic time series may be identified especially related to non-stationarity and nonlinearity. Evidently, this information cannot be used to explain why certain phenomena—e.g. spontaneous breakdowns, stop-and-go waves, hysteresis and transitions—may occur that is usually the case with traffic flow modeling, but rather improve the accuracy of predictions by clarifying the statistical characteristics of short-term traffic evolution and improve the process of model choice.

Acknowledgments Recurrence Quantification Analysis was implemented using RQA v14 (<http://homepages.luc.edu/~cwebber/>) and CRP Toolbox 5.5 (<http://tocsy.pik-potsdam.de/CRPtoolbox/>).

References

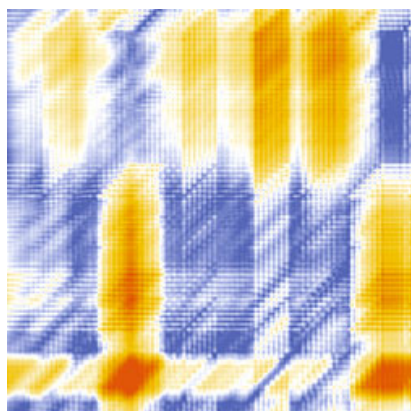
1. L. Edie, Car-following and steady-state theory for non-congested traffic. *Oper. Res.* **9**(1), 66–75 (1961)
2. T.W. Forbes, Human factor consideration in traffic flow theory. *Highway Res. Rec.* **15**, 60–66 (1963)
3. G.F. Newell, Instability in dense highway traffic: a review, in *Proceedings of the Second International Symposium on Transportation and Traffic Theory*. (London, 1965) pp. 73–83
4. I. Prigogine, R. Herman, *Kinetic Theory of Vehicular Traffic* (Elsevier, New York, 1971)
5. J. Treiterer, J. Myers, Hysteresis Phenomenon in Traffic Flow, in *Proceedings of the Sixth International Symposium on Transportation and Traffic Theory*, ed. by D. Buckley (Elsevier, New York, 1974), pp. 13–38
6. K. Nagel, M. Schreckenberg, A cellular automaton model for freeway traffic. *J. Phys. I (France)* **2**, 2221 (1992)
7. G. Newell, A simplified theory of kinematic waves in highway traffic, i. general theory, ii. queuing at freeway bottlenecks, iii. Multi destination flows. *Transport. Res.* **27B**, 281–313 (1993)
8. B. Kerner, H. Rehborn, Experimental properties of phase transitions in traffic flow. *Phys. Rev. Lett.* **79**, 4030 (1997)
9. C. Daganzo, M. Cassidy, R. Bertini, Possible explanations of phase transitions in highway traffic. *Transport. Res. Part A* **33**, 365 (1999)
10. D. Helbing, M. Schreckenberg, Cellular automata simulating experimental properties of traffic flow. *Phys. Rev. E* **59**, R2505 (1999)
11. H.M. Zhang, A mathematical theory of traffic hysteresis. *Transport. Res. Part B: Methodol.* **33**(1), 1–23 (1999)
12. E. Tomer, L. Safonov, S. Havlin, Presence of many stable nonhomogeneous states in an inertial car-following model. *Phys. Rev. Lett.* **84**, 382–385 (2000)

13. J.A. Laval, L. Leclercq, A mechanism to describe the formation and propagation of stop-and-go waves in congested freeway traffic. *Philos. Trans. R. Soc. A* **368**(1928), 4519–4541 (2010)
14. D. Chowdhury, L. Santen, A. Schadschneider, Statistical physics of vehicular traffic and some related systems. *Phys. Rep.* **329**, 199 (2000)
15. A. Schadschneider, Statistical physics of traffic flow. *Physica A* **285**, 101–120 (2000)
16. D. Helbing, Traffic and related self-driven many-particle systems. *Rev. Mod. Phys.* **73**, 1067–1141 (2001)
17. K. Nagel, P. Wagner, R. Woesler, Still flowing: approaches to traffic flow and traffic jam modeling. *Oper. Res.* **51**, 681–710 (2003)
18. S. Maerivoet, B. De Moor, Cellular automata models of road traffic. *Phys. Rep.* **419**(1), 1–64 (2005)
19. G. Orosz, R.E. Wilson, G. Stépán, Traffic jams: dynamics and control. *Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci.* **368**(1928), 4455–4479 (2010)
20. J.P. Eckmann, S.O. Kamphorst, D. Ruelle, Recurrence plots of dynamical systems. *Europhys. Lett.* **5**(9), 973–977 (1987)
21. J.P. Zbilut, C.L. Webber Jr., Embeddings and delays as derived from quantification of recurrence plots. *Phys. Lett. A* **171**(3–4), 199–203 (1992)
22. C.L. Webber Jr., J.P. Zbilut, Dynamical assessment of physiological systems and states using recurrence plot strategies. *J. Appl. Physiol.* **76**, 965 (1994)
23. N. Marwan, N. Wessel, U. Meyerfeldt, A. Schirdewan, J. Kurths, Recurrence plot based measures of complexity and its application to heart rate variability data. *Phys. Rev. E* **66**(2), 026702 (2002)
24. N. Marwan, M.C. Romano, M. Thiel, J. Kurths, Recurrence plots for the analysis of complex systems. *Phys. Rep.* **438**(5–6), 237–329 (2007)
25. M.C. Casdagli, Recurrence plots revisited. *Physica D* **108**, 12–44 (1997)
26. H. Kantz, T. Schreiber, *Non-Linear Time Series Analysis. Cambridge Non-linear Science: Series 7* (Cambridge University Press, Cambridge, 1997)
27. A.M. Fraser, H.L. Swinney, Independent coordinates for strange attractors from mutual information. *Phys. Rev. A* **33**(2), 1134–1140 (1986)
28. M.B. Kennel, R. Brown, H.D.I. Abarbanel, Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys. Rev. A* **45**, 3403 (1992)
29. J.P. Zbilut, A. Giuliani, C.L. Webber Jr., Recurrence quantification analysis and principal components in the detection of short complex signals. *Phys. Lett. A* **237**, 131 (1998)
30. J. Gao, H. Cai, On the structures and quantification of recurrence plots. *Phys. Lett. A* **270**, 75–87 (2000)
31. M.G. Karlaftis, E.I. Vlahogianni, Memory properties and fractional integration in transportation time-series. *Transport. Res. Part C: Emerg. Technol.* **17**(4), 444–453 (2009)
32. E.I. Vlahogianni, M.G. Karlaftis, J.C. Golias, Temporal evolution of short-term urban traffic flow: a non-linear dynamics approach. *Comput.-Aided Civil Infrastruct. Eng.* **22**(5), 317–325 (2008)
33. T. Nagatani, Effect of irregularity on vehicular traffic through a sequence of traffic lights. *Phys. A: Stat. Mech. Appl.* **387**(7), 1637 (2008)
34. B. Kerner, The physics of traffic: empirical freeway pattern features, engineering applications, and theory, in *Understanding Complex Systems Series* ed. by J.A. Scott Kelso (Springer, 2004) ISBN: 3-540-20716-3
35. E.I. Vlahogianni, C.L. Webber Jr., N. Geroliminis, A. Skabardonis, Statistical characteristics of transitional queue conditions in signalized arterials. *Trans. Res. Part C* **15**(6), 345–404 (2007)
36. R.E. Chandler, R. Herman, E.W. Montroll, Traffic dynamics: studies in car following. *Oper. Res.* **6**(2), 165–184 (1958)

Chapter 14

Interpersonal Couplings in Human Interactions

Kevin Shockley and Michael A. Riley



Abstract As inherently social beings people routinely interact with others. Interpersonal activities such as dancing, conversation, or team sports require people to coordinate at several different levels, ranging from the coordination of physical movements and physiological states of the body to the coordination of mental states and cognitive or linguistic activity. One of the challenges confronted by researchers in this interdisciplinary field has been to find ways to objectively quantify interpersonal coupling on the basis of brief, noisy, nonstationary, and complex time series of human behavioral sequences. Given their robustness to these challenges, recurrence-based strategies have played a very important role in the development of this field of research. This chapter provides a review of current behavioral, cognitive, and physiological research that has used recurrence methods to quantify interpersonal coupling.

14.1 Introduction

As inherently social beings people routinely interact with others, such as when navigating busy sidewalks, engaging in a friendly debate with a colleague, playing a game of pick-up basketball, or helping a friend move a heavy piece of furniture. Such interpersonal activities require people to coordinate at several different levels,

K. Shockley (✉) • M.A. Riley

Department of Psychology, Center for Cognition, Action, & Perception, University of Cincinnati, Cincinnati, OH, 45219-0376 USA

e-mail: kevin.shockley@uc.edu

ranging from the coordination of physical movements and physiological states of the body to the coordination of mental states and cognitive or linguistic activity. Behavioral coordination is ubiquitous and is highly likely to occur—even when people do not intend for it to occur—whenever there is present some medium of interpersonal coupling that serves to link people together, whether that medium is a physical connection between people, visual information about another person's movements, or linguistic information that is exchanged during a conversation.

Interpersonal or social coordination is very important in everyday life as well as in many practical settings such as surgical teams or military operations. There are potentially adverse consequences when breakdowns in interpersonal coupling occur. For example, if communication links among members of a firefighting team are interrupted and the firefighters' behaviors become uncoordinated, the results can be potentially very dangerous and even life-threatening. Social coordination and interpersonal coupling may also be compromised as a result of neurological deficits such as autism spectrum disorder. For these and many other reasons, interpersonal coupling is a rapidly growing field of interdisciplinary research, spanning experimental psychology, cognitive science, ergonomics, neuroscience, movement science, and sport science.

Recurrence methods (introduced conceptually by Eckmann and Kamphorst [1]) have played a very important role in the development of this field of research. One of the challenges confronted by researchers has been to find a means of objectively quantifying interpersonal coupling. Advances in motion-capture and eye-tracking technology have made it relatively easy to collect highly precise data and have rendered obsolete time-consuming and subjective methods such as qualitative hand-coding of video sequences of interpersonal interactions. But the advent of these technologies and the wealth of data they permitted introduced the new challenge of identifying metrics to quantify patterns of coordination in the data. Time series of human behavioral sequences, such as movements of the limbs during gesturing or movements of the eyes, can be noisy, nonstationary, and complex, and often the time series—particularly during real-world activity—are relatively brief. Recurrence-based methods are ideally suited to meet those challenges [2–4], and thus recurrence-based methods have advantages over many other methods. For example, simple cross-correlation in the time domain, or its equivalent in the frequency domain, spectral coherence, are both linear measures and both assume stationarity of the time series being analyzed. If the assumptions of linear interactions and stationarity are violated, the methods may give incomplete, or sometimes even misleading, results.

Interpersonal coupling has been observed and studied in a wide variety of behavioral domains. These include the coupling of movements of limbs, of overall body posture, of whole-body positions of athletes, of musical performance, of eye movements, of speech, of cognitive states such as attention, and of linguistic information. Recurrence quantification analysis (RQA; [2–4]), cross-recurrence quantification analysis (CRQA; e.g., [5–7]), and novel variations of these methods have all been utilized to study interpersonal coupling. Many important insights into

interpersonal coupling have been gained from this research, yet many questions remain unanswered and many new questions have emerged. This chapter provides a review of current behavioral, cognitive, and physiological research that has used recurrence methods to quantify interpersonal coupling.

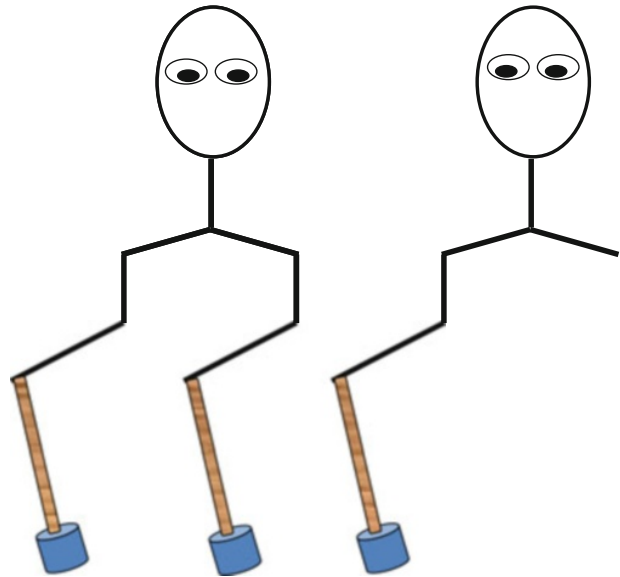
14.2 Interpersonal Rhythmic Motor Coordination

One of the most basic forms of interpersonal coupling occurs when two or more people synchronize or coordinate rhythmic movements of their bodies with each other. The coordination could be of limb segments or limbs, the head, the torso, the whole body, or even the eyes. This coordination can be intentional and is often part and parcel of the overarching behavioral goal, such as when dancing. Interpersonal rhythmic movement coordination also sometimes occurs spontaneously even when it is not intended or related in an obvious way to the behavioral goal.

Several studies have applied CRQA to a variant of a simple rhythmic coordination task that has served as a workhorse paradigm for understanding interpersonal coupling—coordinating the rhythmic movements of limb segments [8] or, more typically, hand-held pendulums (cf. [9]). As illustrated in Fig. 14.1, in the basic form of the wrist-pendulum coordination task [10] participants sit in a chair that has an armrest to support one hand and forearm. In the supported hand each participant holds a pendular object that they swing in the sagittal plane while, typically, watching the other participant swing the pendulum. The goal is to synchronize movements of the pendulum with the other person at the same frequency and at a specified phase relation—typically either 0° relative phase (pendulums are at the same points in their respective movement cycles, so that the movements are synchronous and in the same direction) or 180° relative phase (pendulums are at opposite points in their movement cycle, moving at the same frequency but in opposite directions), which are the two basic coordination patterns that are intrinsically stable (other phase relations can be learned but are unstable and difficult to produce).

The dynamics of this interpersonal rhythmic coordination task have been carefully studied, largely using measures of the average relative phase established by the interacting subjects (i.e., which coordination pattern they established) and the standard deviation (variability) of the coordination pattern, and also by identifying sudden changes (i.e., phase transitions) in the coordination pattern (see review by [11]). A very successful model of this type of coordination has also been identified, building on work in *intrapersonal* rhythmic coordination and the model of Haken et al. [12]. Richardson et al. [15] grounded CRQA in terms of constructs from this modeling framework, finding that *%cross-recurrence* maps onto the level of underlying noise in the coordination and that *cross-maxline* maps onto the strength of the attractors that govern stable states of interlimb coordination (see also [16]).

Fig. 14.1 Example of a common method used to study interpersonal interlimb coordination. Two participants swing hand-held pendula while looking at one another



Richardson et al. [17] used CRQA to compare interpersonal rhythmic coordination to *intrapersonal* interlimb rhythmic coordination (i.e., one person oscillates a pendulum in each hand—bimanual coordination). They found that attractor strength as indexed by *cross-maxline* was lower for interpersonal coordination than for intrapersonal coordination. *Cross-maxline* is sensitive to changes in attractor strength because *cross-maxline* refers to the duration of the longest movement sequences that the two times series being compared share with one another. Thus, when the attractor for a particular phase relation is weaker the system is unable to maintain the relation for as long as a stronger attractor. Interpersonal coupling of rhythmic movements thus differs from intrapersonal coupling in that the strength of the coupling is weaker, consistent with findings using regression-based techniques to estimate the coupling coefficients [18] and using the uncontrolled manifold approach which quantifies synergistic coupling by identifying how motor variability is structured so as to preserve the task goal (such variability can be left “uncontrolled”) or not (this variability must be restricted to achieve the task goal) [19]. Richardson et al. also found that noise magnitude, as indexed by *%cross-recurrence*, was not significantly different across interpersonal and intrapersonal coordination conditions. Increasing the magnitude of noise in a system reflects an increase in random perturbations to the dynamical state of the system. Thus, the number of shared configurations (indexed by *%cross-recurrence*) will decrease proportionally with the magnitude of noise in the system (see [15, 16], for examples of independent sensitivity of cross recurrence measures to attractor strength and noise magnitude). The specific contribution of recurrence methods in the study of this type of intentional interpersonal coordination task was thus to identify the reduced strength of coupling of interpersonal coordination compared to intrapersonal coordination.

14.3 Coordination and Communication

Richardson et al. [20] studied variations on the basic interpersonal rhythmic coordination task. Coordination was not an explicit goal for the subjects—they were not instructed to adopt any particular interpersonal movement pattern. Instead, subjects were instructed to perform a simple communication task of identifying differences between two cartoon faces that were attached to the pendulums in one condition or on a stand next to the subjects in another. The participants could not see their partner’s pictures—only their own—and so they had to find the differences by discussing what they saw with each. The faces on the pendulum was employed in the “visual” and “visual-verbal” conditions—subjects viewed each others’ pendulum motion in each of these but talked to each other to identify differences only in the latter—while faces on the stand constituted a “verbal” condition—subjects were coupled by verbal interaction but did not watch their partner’s pendulum movements. The instruction to oscillate the pendulums was described to subjects as a distractor task that served to make the puzzle task more challenging. Richardson et al. found that viewing the task partner’s pendulum promoted unintentional coordination—*%cross-recurrence* and *cross-maxline* were greatest in the visual condition—but verbal coupling was insufficient to result in substantial unintentional coordination on its own nor did it enhance coordination beyond visual coupling (both measures were nominally greater in the visual than in the visual-verbal condition, in fact).

The task and the verbal condition in Richardson et al. [20] was motivated in part by an earlier study on interpersonal coupling by Shockley et al. [21]. In one of the first applications of CRQA to interpersonal coupling, Shockley et al. studied coordination of *postural sway* between two people who were engaged in a similar find-the-differences task (see Fig. 14.2, left). Postural sway is the irregular, low-amplitude, continuous, and complex fluctuation of the body’s center of mass that always occurs when a person stands. Based on a large body of literature demonstrating various forms of behavioral synchrony between conversants (e.g., [22–24]), Shockley et al. [21] hypothesized that the postural sway of conversants would exhibit greater cross-recurrence and cross-maxline—i.e., greater and more stable coupling—than the postural sway of co-present but non-interacting subjects. They tracked the waist position of the participants over time as a global measure of postural sway while the participants completed the find-the-differences task. Because the waist positions embody the activity of a large number of variables (e.g., limb configurations, cardio-pulmonary dynamics, speech), the time series of waist motion was unfolded into a multidimensional space that was sufficient to capture the unfolding dynamics of movement (i.e., a reconstructed phase space; [25]) and thus each data point corresponded to a position in the multidimensional space, henceforth referred to as a (waist) configuration. Their findings confirmed this hypothesis—*%cross-recurrence* and *cross-maxline* of the waist configurations were greater for interacting than merely co-present participants (Fig. 14.2, right). In contrast to the results of Richardson et al. [20], vision of the task partner was not necessary for

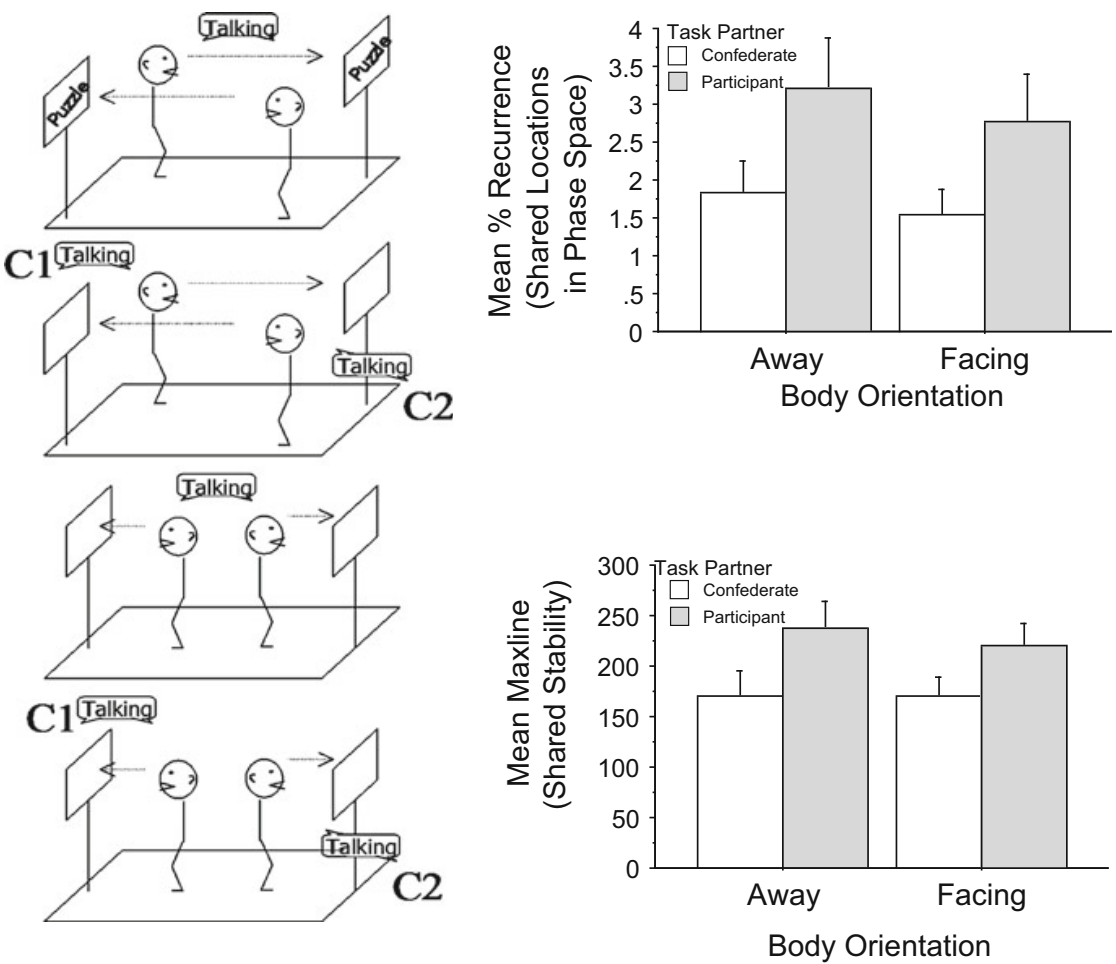


Fig. 14.2 (Left) Method of Shockley et al. [21]. (Right) Shared postural activity (%cross_recurrence) and coordination stability (cross_maxline) for the different experimental conditions. (From Shockley et al. [21], p. 329 [panel a], 330 [panel c]). Copyright 2003 by the American Psychological Association. Adapted with permission)

and did not enhance interpersonal coupling—%cross-recurrence and cross_maxline were not different between conditions in which task partners faced each other or faced away from each other while conversing to find the differences in the pictures.

Building on the understanding that conversants tend to coordinate their speech while engaging in cooperative conversation (e.g., [26–28]) and that the biomechanics of speaking influence postural sway dynamics [29–31], Shockley et al. [32] investigated whether the interpersonal postural coordination observed during cooperative conversation may simply be an artifact of how postural sway is affected similarly across members of a pair when their speech becomes coordinated. They had participants stand while uttering bisyllabic words that were presented on a computer monitor. The words were either presented simultaneously to each participant (in-phase) or in an alternating fashion (anti-phase) and the words were either the same for each participant (same word, same syllable; SS), different words with an emphasis on the same syllable (first or second) (DS), or different words with an emphasis on different syllables (i.e., one person had a word with

an emphasis on the first syllable, e.g., *donut*, while their task partner had a word with an emphasis on the second syllable, e.g., *about*, or vice versa) (DD). They found no influence of the phase manipulation. They did find, however, that there were greater shared postural configurations (*%cross-recurrence*) and greater coordination stability (*cross-maxline*) between members of a pair with increasing word similarity (i.e., SS > SD > DD). They also found, however, that by shuffling the pair arrangement such that postural sway of one participant is compared to a participant that completed the same task except with a different partner that there was no such increase in postural coordination with increases in word similarity. This suggested that although articulatory dynamics did influence the postural coordination patterns during conversation, there was still a social influence that impacted the interpersonal postural coordination between members of a pair (i.e., their partner had to be co-present for the enhanced coordination to occur).

Stoffregen and colleagues have further explored the spontaneous coordination that occurs during conversation that was observed by Shockley et al. [21]. In one study, Stoffregen et al. [33] investigated how the interpersonal postural coordination that occurs during conversation was influenced by the stability of the surface on which the interacting participants stood. They had pairs of participants either stand on the floor or stand on a mattress while they completed the task. They found that the enhanced coordination between members of a participant pair only occurred when they were talking to each other while standing on the floor, but not when standing on an unstable surface of support and not when talking to an experimenter. This suggested that the enhanced coordination when participants are talking to each other observed by Shockley et al. [21] is a subtle phenomenon, specifically that the constraints on postural control imposed by conversation are weaker than those imposed by the surface of support. This, again, hallmarks the sensitivity of CRQ measures to the degree of interpersonal coordination.

In a different study, Stoffregen et al. [34] investigated visual influences on interpersonal postural coordination. Shockley et al. [21] did not find an influence of whether the participant could see his/her task partner. However, they did not manipulate visual aspects of the puzzle that was inspected by participants. Given that previous studies have demonstrated an influence of visual constraints on postural sway dynamics of an individual (see [35], for a review), Stoffregen et al. [34] manipulated the size and distance of the visual targets used by participants in the find-the-differences task. In the first experiment they manipulated task partner (participant or experimenter; cf. [21]) and target distance (near vs. far; e.g., [36, 37]). They found more shared head configurations (greater *%cross-recurrence*) when the targets were near as compared to far. They did not, however, find enhanced coordination between participants when they were talking to each other. In their second experiment they manipulated target size (small vs. large) and task partner. In this experiment they did find an influence of task partner, replicating Shockley et al. [21]. However, they found that the greater shared head configurations (as opposed to the shared waist configurations observed by [21]) when participants were talking to each other compared to when participants were talking to an experimenter. They also found an influence of target size for both the head and the waist. Specifically,

there were greater shared head and waist configurations when participants were looking at larger targets than at smaller targets. They also found greater patterning of the coordination for head and waist (*%cross-determinism*) for larger targets than smaller targets. In a third experiment they manipulated the similarity in visual conditions between members of pair. In their first two experiments, although size and distance were manipulated, the conditions were always the same for the pair in a given trial. In their third experiment Stoffregen et al. crossed own target size (small vs. large) with partner target size (matched vs. mismatched). They found greater shared head configurations when participants' own target size was large. They also found that when the partner's target size was matched there were greater shared head configurations when the partner's target size was matched to one's own target size (i.e., when both targets were either small or large) compared to when the partner's target size was mismatched (i.e., when one target was large and the other was small or vice versa). They also found that the patterning of shared head and waist configurations were influenced by these manipulations. There was greater patterning in the shared head and waist configurations when viewing large targets as opposed to small targets. This study was the first demonstration of visual constraints on the postural coordination that occurs between conversants.

Shockley et al. [38] speculated that the coordination involved in conversation may reflect a functional organization that supports the joint goals of the individuals interacting. That is, the movement coordination observed during conversation may embody the coordinated cognition required to effectively communicate. While it is well-established that cognitive performance can interfere with concurrent motor performance, and vice versa (for a review see [39]), more recent investigations have revealed that in many cases action may not so much interfere with cognition, but instead may embody cognition. For example, the inhibition of motor activity [40] or an imposed inconsistency between motor activity and cognitive responses [41–44] can disrupt cognitive performance. In other circumstances, mental operations can be facilitated by action (e.g., [45]) and action-oriented tasks [46]. This interpretation is underscored by recent evidence that the time course of a cognitive process is reflected in the trajectory of action. For example, when moving to click 'yes' or 'no', mouse trajectories will travel directly towards 'yes' if the answer is clear (should you brush your teeth everyday?) but deviate towards no when the answer is ambiguous (is murder sometimes justified?), and the degree of motor deviation reflects the degree of cognitive certainty [47]. The discovery of the mirror neuron system (neurons that fire when observing another's actions; see reviews by [48, 49]) has underscored the growing recognition of the integral role of action in cognition, because it suggests a common neural mechanism for various forms of social perception, action, and cognition (see, e.g., [50–53]). If the movement coordination observed during communication indeed reflects the cognitive coordination required to effectively communicate, it stands to reason that constraining coordination may constrain communication. This was the focus of a recent study by [54].

Tolston et al. [54] manipulated movement coordination by restricting the movement of one or both members of a participant pair that completed the Shockley et al. [21] find-the-differences task while standing. Each pair of participants was asked

to look at a pair of pictures that were very to one another but had subtle differences. They could not see one another's pictures and so had to find the differences by discussing the pictures with each other. Either both participants were free to move (F–F), Person 1 was free to move while the hand movements of Person 2 were restrained by having them place their hands in the pockets of a waist apron (F–R), or both participants' hand movement were restricted (R–R). They hypothesized the weakest coordination in the asymmetrically restrained (F–R) condition given that participants were, by design, least able to coordinate their movements. They hypothesized that cognitive performance would increase with greater freedom to move. They found movement coordination of the head and waist to be reliably lower in the asymmetric (F–R) restraint manipulation as reflected in measures including *%cross_recurrence*, *%cross_determinism*, and *cross_maxline*. They did not find a significant influence on task performance as a direct function of the manipulations. They did, however, find that overall movement of the hands of Person 1 (the one with the most freedom to move overall) significantly correlated with task performance suggesting a relation between movement and communication. In spite of the fact that movement restraint did not influence task performance in the expected fashion, CRQ was, nonetheless, sensitive to the manipulations of movement coordination at both the head and the waist.

14.4 Interpersonal Coupling During Performance of Joint Precision Motor Tasks

Ramenzoni et al. [55] studied postural and manual interpersonal coupling during performance of a simple, precision motor task. In their task one subject held a target circle and the other subject extended a pointer so as to hold it inside the target circle without touching its sides (see Fig. 14.3). The size of the target circle was varied to manipulate task difficulty. Ramenzoni et al. recorded subjects' hand movements and torso motion (i.e., postural sway) with motion-capture sensors. *%cross-recurrence* systematically increased for both hand and torso movements as task difficulty increased. *Cross-maxline* was greater in more difficult task conditions for hand but not torso movements.

Ramenzoni et al. [56] used a slightly different approach to analyze performance of a similar interpersonal precision task. They measured 3-D displacements of the torso, upper arm, forearm, and hand of each participant. Rather than analyzing coordination between body segments (e.g., between subjects' hands), they submitted each subject's 12-dimensional data set to principal component analysis (PCA). PCA is a dimensional reduction technique that identifies covariation in complex data sets and effectively "collapses" redundant data along abstract dimensions that represent the directions along which most variation in the dataset occurs (i.e., the PCs). The first PC is that which accounts for most variance in the data, and subsequent PCs account for successively less and less variance. The original time series can then

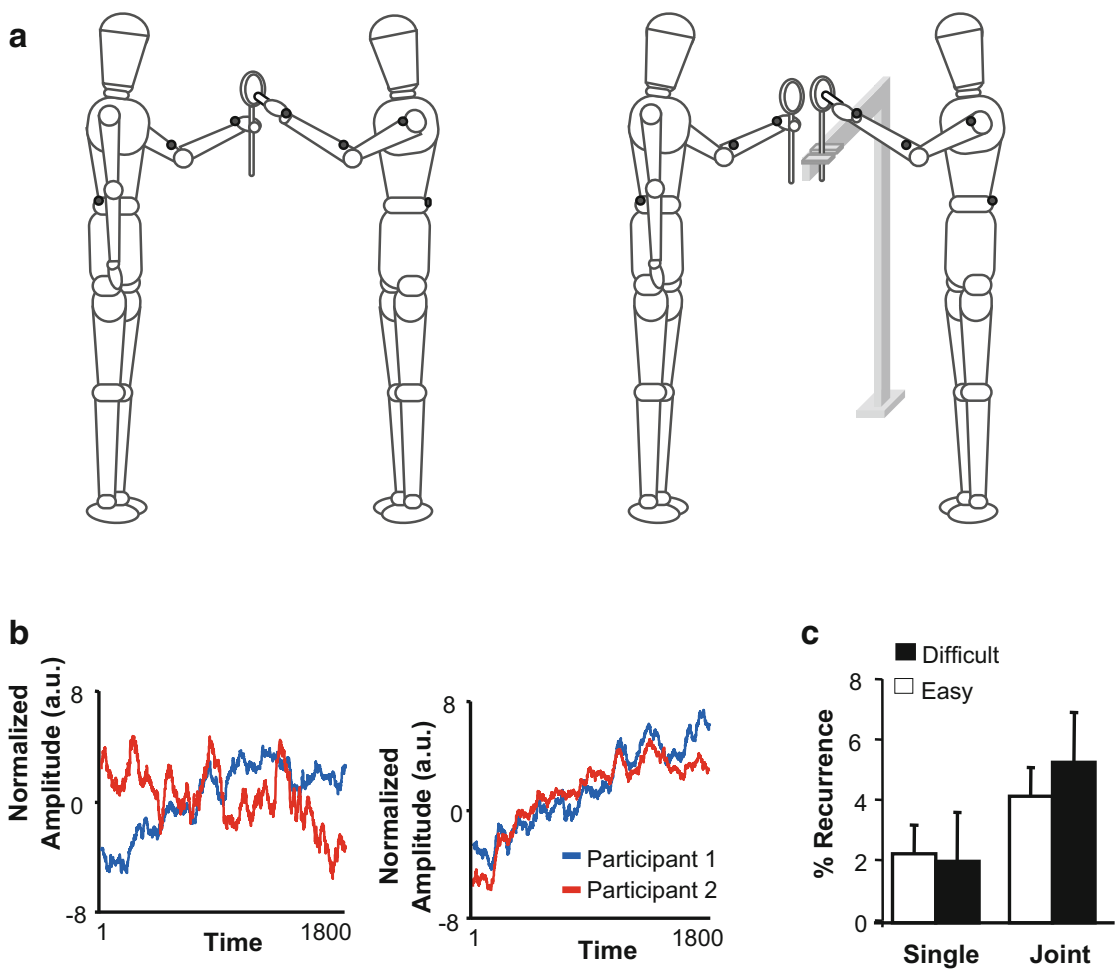


Fig. 14.3 (a) Depiction of the individual—(left) and interpersonal—(right) task conditions from Ramenzoni et al. [56]. (b) Time series of the data projected onto the intrapersonal principle components from the individual (left) and interpersonal-task (right) conditions. The striking coordination in the interpersonal-task condition was confirmed by cross-recurrence quantification analysis (c), which revealed a greater degree and higher stability of coupling in that condition. From Riley et al. [57]. Copyright 2011 by *Frontiers in Psychology*

be projected onto the axes of the principal components, resulting in time series that are abstract yet represent the overall or global pattern of movement—a data-driven “order parameter” [58] of sorts (see Fig. 14.3). Ramenzoni et al. then applied CRQA to analyze coupling between the time series of each subject’s first PC. Both %cross-recurrence (see Fig. 14.3) and cross-maxline were greater when subjects performed the interactive experimental task than a control task that did not require them to interact. Moreover, the CRQA measures were greater in the more difficult (smaller target) than the less difficult condition, but only during the experimental task. Interpersonal coupling occurred during performance of this task at the level of overall body movements, a level of description slightly more abstract than the movement of any individual body segment (see also, [57]).

14.5 Verbal Coordination

While there are many instances summarized above where speech and communication influence interpersonal motor coordination, communication itself has also become an object of inquiry using recurrence methods. Orsucci et al. [59] made an initial foray into quantifying conversational interaction by analyzing the transcriptions of conversations between two friends and between a therapist and client in order to illustrate the different dynamics of these two different types of verbal interaction. The transcribed text can be coded, letter by letter into numbers (e.g., a = 1, b = 2, etc.) yielding a numeric time series that represents the speech vector of each person involved in the conversation. Unlike the movement time series described above, however, the magnitude of the numbers in the time series have no significance. In other words, a value of 2 is not twice the value of 1 as it would be in ratio data. Rather, the data set is strictly nominal. When CRQA is used on this type of data it is referred to as *categorical* CRQA. As described in previous chapters of this book, the primary difference between continuous CRQA and categorical CRQA is that continuous CRQA typically involves time-delayed embedding of the two continuous time series into a reconstructed phase space and using a radius of inclusion that is greater than zero (i.e., the value in question must match the value to which it is compared within some radius in reconstructed phase space). Categorical CRQA, on the other hand, generally does not involve embedding, typically involves *data* series rather than *time* series (i.e., although the data are sequential, the spacing between observations is not necessarily equivalent across observations, and uses a radius of zero (i.e., the value in question must exactly match the value to which it is compared)). Orsucci et al. elected to use paired three-letter patterns (i.e., when both time series exhibited the same three letter sequence) as evidence of a recurring value between the two time series. They found that friends' conversational interaction exhibited greater synchrony between participants as indicated by greater shared sequences of utterances (i.e., greater *%cross_determinism*) as compared to the clinical interaction. This suggests that recurrence methods are sufficiently sensitive to pick up variations in the form of conversation.

Dale and Spivey [60] took recurrence-based investigations into speech coordination a step further by studying the similarity in word-class n-gram (bigram, trigram, quadrigram) sequences from three CHILDES corpora [61]. They performed categorical cross-recurrence of children and their caregivers. Dale and Spivey [60] explored both synchronous verbal coordination as well as leader–follower relationships. Synchronous coordination was evaluated by quantifying recurrence along the central diagonal region of the cross-recurrence plot (CRP), which corresponds to temporal coincidence between the two data series. Leader–follower relationships can be quantified by evaluating the upper and lower triangular regions separately. If time series A serves as the abscissa of a CRP and time series B serves as the ordinate, then recurrence in the upper triangular region refers to values in time series B that occurred in time series A at an earlier time (i.e., X led Y) and vice versa for the lower triangular regions (i.e., Y led X; Fig. 14.4, bottom). This is because,

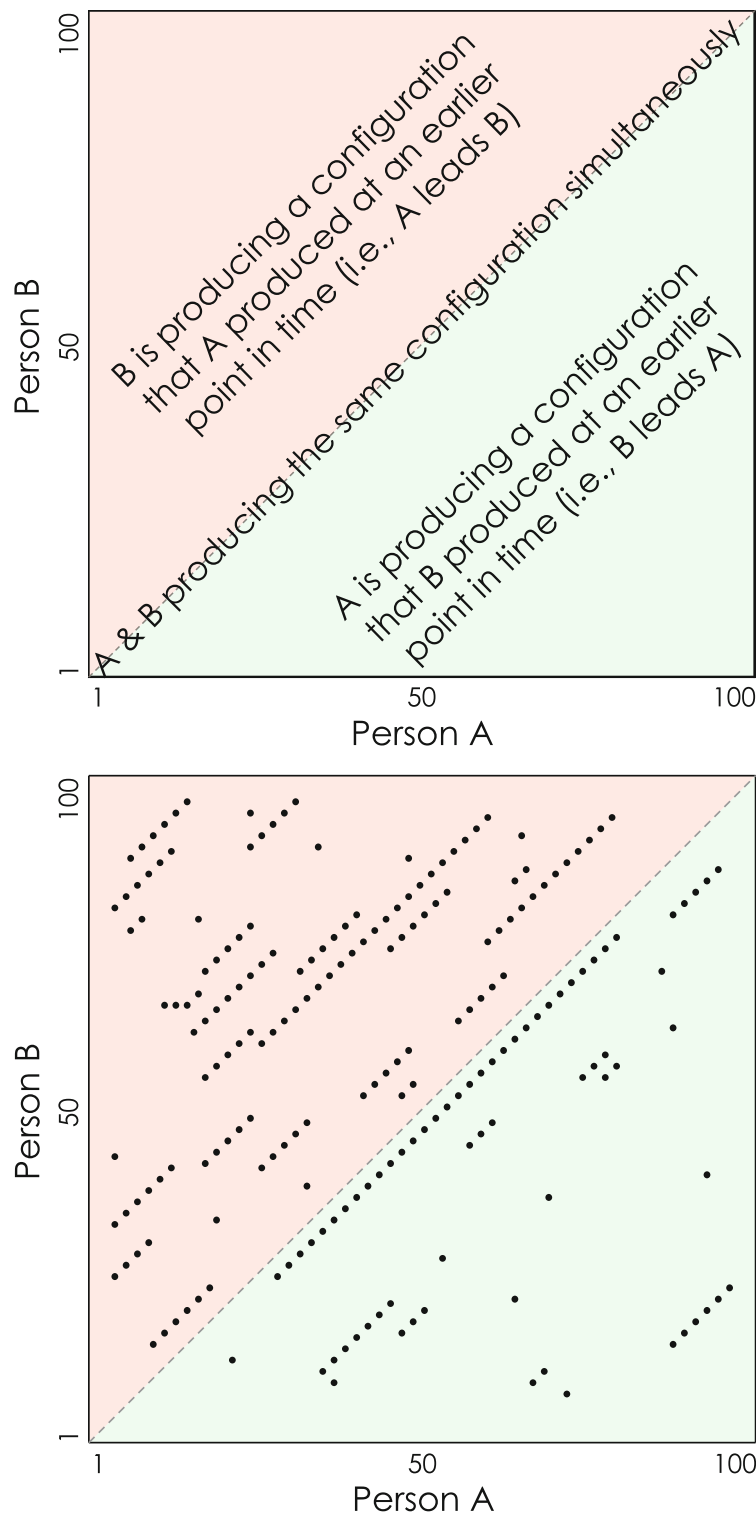
by definition, any index above the diagonal corresponds to an index of Y that is greater than the index of X (Fig. 14.4, top). Dale and Spivey demonstrated that there was indeed coordination between syntactic sequences of child and caregiver that was greater than chance (i.e., as compared to the coordination observed between pairs of randomly shuffled time series) suggesting that CRQA is a useful tool for capturing this type of verbal coordination. They also demonstrated that the leader–follower relationship between child and caregiver seemed to change over the course of development and that different children exhibit different lead–follower relations with their caregivers. This finding suggests that CRQA may be useful in differentiating different conversational interaction styles across interlocutors.

Warlaumont et al. [62] further explored verbal coordination between children and adults in a study investigating both autistic and typical control groups. They coded their data differently than in the previously described studies. They simply determined at a given time segment whether the child or adult was speaking. If one person was speaking that person received a 1 for the segment and the other received a null value. In a similar fashion to Richardson et al. [13], they found that using a 30-s window on either side of the central diagonal (i.e., coincidence in time) of the cross recurrence plot that autistic children tended to follow adults more and lead adults less than the control group.

Angus et al. ([63]; see also [64]) also investigated the similarity between utterances of dyads. However, rather than analyzing textual data (i.e., syntactic strings) as Dale and Spivey [60] did, they instead analyzed the conceptual similarity of the text samples in question. They used Salton's [65] strategy of building a semantic similarity model of the transcribed discourse under scrutiny based on the probability of the co-occurrence of terms in the text. Thus, the similarity of any two utterances can be assigned a number. The conceptual similarity of every utterance to every other utterance is determined yielding a matrix of similarity values. The strategy for quantifying the similarity in the discourse is a bit different than that used by Dale and Spivey [60]. Angus et al. created a single trajectory of utterances that included both interlocutors (i.e., a single vector of semantic similarity values). Because CRQA involves comparing two vectors, a semantic vector was the submitted to auto-recurrence quantification analysis (i.e., RQA) rather than CRQA. However, recurrence in this case reflects conceptual similarity between conversants rather than within a single speaker's utterance (see Fig. 14.6).

Angus et al. [66] used the same basic strategy. However, the investigators demonstrated the utility of their approach for quantifying multi-participant recurrence (i.e., more than two interacting individuals rather than the more conventional dyadic comparison) in the context of linguistic interaction. In this study they introduced new metrics for quantifying the recurrence in communication such as immediate topic repetition, topic consistency, and topic novelty based on what they referred to as utterance primitives including time scale (near, middle, and far from present utterance), direction (utterances forward or backward in time from the present utterance), and type of utterance (self or other). These metrics may be more meaningful to the study of conceptual similarity in the context of communication

Fig. 14.4 (Top) Illustration of how upper and lower triangular regions of a cross recurrence plot can capture leader–follower relationships in interpersonal interactions. By aligning any index on the abscissa with any index on the ordinate that is above the central diagonal (upper [red] triangular region) corresponds to an earlier point in time for Person A than Person B. (Bottom) Illustration of greater recurrence in upper triangular region than in lower triangular region indicating that, in general, B is following A



than the more conventional metrics used in recurrence-based analyses (e.g., *%cross-recurrence*, *%cross-determinism*, *cross-maxline*).

Gorman et al. [67] extended the analysis of discourse to three-member teams who were controlling an uninhabited air vehicle (UAV). They used categorical CRQA to evaluate communication data (i.e., the data were coded in terms of which team

member was speaking) under conditions where the team either changed composition or stayed the same following a retention interval. They found that after three missions the *%cross-determinism* of the mixed teams' communication interaction sequences did not change across three successive missions. The intact teams, however, exhibited greater *%cross-determinism* in their communication interaction sequences by the third mission. They found a similar pattern of results for pattern information as measured by mutual information. Intact teams exhibited greater pattern information by the third mission, whereas mixed teams did not show such an increase (and even showed a decrease in the second mission). They interpreted this enhanced patterning of interactions of intact teams as potentially reflecting a rigidity that may not be desirable given that control of UAVs may require adaptability and flexibility given the unpredictable dynamics of the task environment [68]. They argued that for a system to be flexible, it requires a mix of determinism and randomness that unchanging teams may move away from over time.

14.6 Interpersonal Coupling During Sport Performance

In many sports coupling of whole-body activity is a major component of the game. In American football, defensive backs must closely cover receivers to prevent them from catching a pass, which requires coupling to the receiver's movement across the field of play. In basketball, a defender must similarly couple to the movements of the player with the ball. Recurrence strategies have recently been applied to whole-body interpersonal coupling in sports environments such as these. Esteves and Araújo [69] found that CRQA can distinguish successful from unsuccessful performance (i.e., scoring a basket vs. not) during one-on-one scenarios in basketball. *%cross-recurrence* between the positions of the attacker and the defender is lower but *cross-maxline* is greater during successful attacker actions compared to unsuccessful ones.

Carvalho et al. [70] applied RQA to quantify coupling of opponents' movements during rallies in professional tennis matches. They focused on comparing opponent coupling as captured by time series of relative positional advantage. Positional advantage is determined by each player's position relative to the central line and to the net. Relative positional advantage is the difference between this quantity defined for each player. A player is considered to have greater positional advantage the closer the player is to the central line or to the net, relative to an opponent. The authors examined positional advantage before versus after a "break shot" (a critical shot that determined the outcome of a point). Prior to the break shot, *%recurrence* was lower and *maxline* was higher than after the break shot. These results are similar to the basketball findings presented above, suggesting that certain patterns of interpersonal coupling (a reduced overall likelihood of opponents sharing the same position, but at the same time exhibiting more stable patterns of coordination) characterize decisive moments related to scoring.

14.7 Interpersonal Coupling During Musical Performance

Music is a rhythmic means of communicating and conveying emotion, and as such might provide a natural medium for coupling rhythmic activities between individuals. Studies have used recurrence methods to describe coordination between time series of the music itself and movements of the musicians [71, 72] or have investigated the use of music sonification as a coupling medium during active listening to help people synchronize their movements when synchrony is an explicit task goal [73, 74]. Other studies described below have used recurrence methods to identify interpersonal coupling between musicians during musical performance, but in general very little work has been done in this area.

Gill et al. [75] used CRQA to quantify coordination of postural sway between participants who performed musical improvisations with shakers (percussion instruments consisting of tubes with objects inside them that make noise when they collide with each other or the tubular container). Similar to the findings of Shockley et al. [21], interpersonal coupling during this musical production task did not require that participants could see each other. This confirms the above intuition that music might serve as a medium for coupling musicians' actions.

Varni et al. [76] used a recurrence-based phase synchrony measure [77] to quantify interpersonal coupling of the head movements of two violinists performing a musical piece live with instructions given to one of the performers to accentuate different emotional states (anger, sadness, joy, or serenity) in the performance on a given trial (the same music was always used). The performances also occurred under conditions of visual plus musical (auditory) coupling or musical coupling alone. They were interested in the phase synchronization of the head movement involved in the two performances—the degree to which the time scales of the head movement of each the two musicians were related to one another. Varni et al. evaluated the probability distributions of distances between recurrence points for each of the head movements being compared to one another. Because the distances between recurrent points reflects the time scale of the system in question (e.g., the system revisits the same states every so often), then if another system conforms to a similar time scale (even if the particular states [e.g., head configurations] of one time series are entirely different than the states of the time series to which it is compared). Phase synchronization values were, overall, rather low and did not seem to depend on emotion or feedback conditions. However, Camurri et al. [78] reported that inducing positive emotions in one musician tended to increase head movement synchrony among groups of violinists compared to a no-emotional-induction control condition.

14.8 Interpersonal Cognitive Coordination

Many of the instances of movement coordination above imply some degree of cognitive coordination and it has even been suggested that the coordination of eye and body movements may reflect the cognitive coordination required to effectively

communicate [38]. However, cognitive coordination, in the sense of joint attention, has been explicitly studied in recent years. In a series of studies, Richardson et al. quantified joint attention by measuring the eye movements of two people discussing a common visual scene. This is particularly interesting from a cognitive standpoint because they demonstrated that gaze coordination reflects the shared knowledge of interacting dyads, their beliefs about each other, and the success of their communication. For example, Richardson and Dale [79] asked participants to talk about a television show while they tracked the gaze of the speaker who was looking at an array of characters' faces from the show in question. They then played back the recorded speech for a listener and tracked the listeners' eye movements toward the array of faces in the same fashion. They then evaluated the comprehension of the listener. They used CRQA to quantify the degree to which speaker and listener gaze coordinates overlapped at successive time lags (see Fig. 14.5). They found that two seconds after a speaker looked at a particular image, the listener was more likely than chance to be looking at that same image. They also found that the *%cross-recurrence* between the gaze positions of speaker–listener pairs was correlated with the listeners' comprehension of what the speakers said. Further, when the pictures at which a speaker was looking were flashed before the listener, this resulted in the listeners' gaze trajectories to look more similar to the speakers' and improved the listeners' comprehension, suggesting that gaze coordination is causally related to comprehension. In a different study, the gaze was tracked for pairs of conversants engaged in live dialogue while they discussed TV shows and paintings [13]. Conversants' eye movements were coupled as they looked at a shared display, peaking at a lag of 0 ms—the conversants were most likely to be looking at the same thing at the same point in time. They also found that when provided with the same background information about what they were looking at their gaze was more coupled. In other words, they demonstrated that shared knowledge enhanced joint attention. These studies demonstrate how CRQ of joint attention can be used to index shared knowledge in everyday verbal interchanges (see also [13]).

Jermann and Nüssli [80] investigated joint attention in a similar fashion to Richardson and Dale's studies, by having pairs of engineering students complete program understanding tasks while their gaze was recorded. They had pairs of participants jointly study code for a simple arithmetic game and their task was to explain the rules of the game based on their understanding of the code that was studied. The authors rated pairs on the interaction quality (e.g., collaboration flow, efforts to sustain mutual understanding, degree of division of labor) as well as level of understanding of the code following their interaction. They manipulated whether selection sharing (i.e., where one programmer can highlight text for the other programmer with which she is working) permitted (selection sharing and dual-selection sharing [i.e., where both programmers can select text that the other can see]) compared to an individual (i.e., no-sharing) condition. They found that within 200 ms of one programmer selecting text, the other programmer was more likely to look at the text selection during that time than if the pair was in the no-sharing condition. There was also an increase in speech when a selection was made.

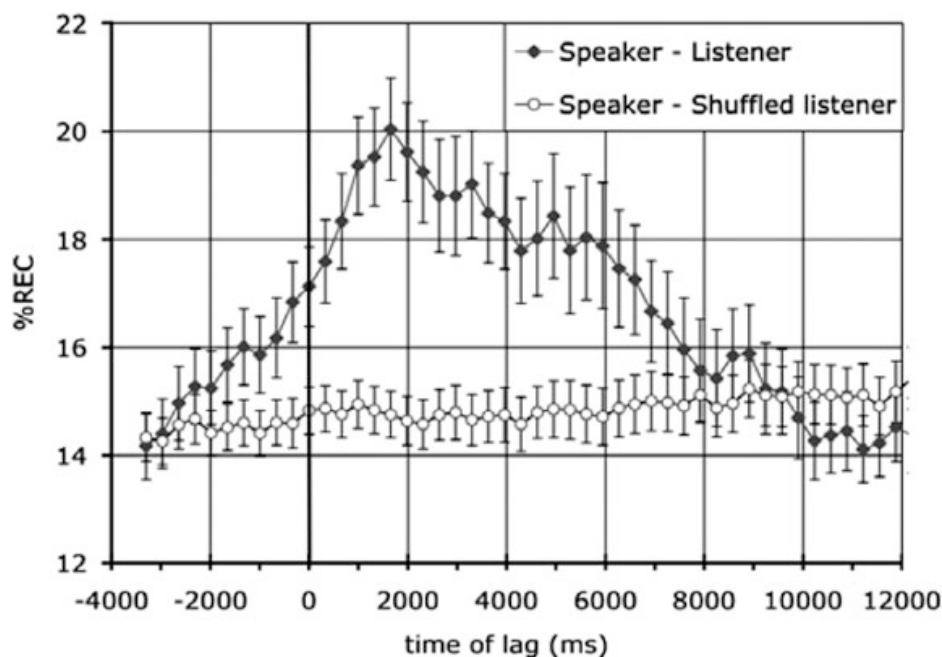


Fig. 14.5 *%cross_recurrence* of gaze coordinates as a function of the lag between speakers' and listeners' gaze. From Richardson and Dale [79]. Copyright 2005 by the Cognitive Science Society, Inc. Adapted with permission

Interestingly, this increase in speech was greater for the individual condition than for the sharing conditions. They found that both selection and speech increased gaze *%cross-recurrence* and that these also had an additive effect such that when selection was accompanied by speech this further increased *%cross-recurrence* of gaze. They also found that when the interaction quality was higher, gaze *%cross-recurrence* was also higher, but gaze *%cross-recurrence* did not vary with the level of understanding of the code. The relationship between gaze coordination and interaction quality echoes the findings of Richardson and Dale [79] who found that when interlocutors shared gaze more their comprehension of what the other person said increased (see Fig. 14.6).

14.9 Interpersonal Physiological Coordination

To our knowledge only one study to date has explored interpersonal physiological coordination using recurrence-based methods. Konvalinka et al. [81] investigated how synchronized the arousal was between performers of a fire-walking ritual and the related spectators of the ritual. Their study was motivated by previous studies that have shown synchronized behavior to enhance cooperation within groups [82, 83] and lead to increased rapport between group members [84, 85]. Konvalinka et al. speculated that enhanced group cohesion may emerge from the shared emotion that accompanies rituals rather than just from synchronized movements that may occur

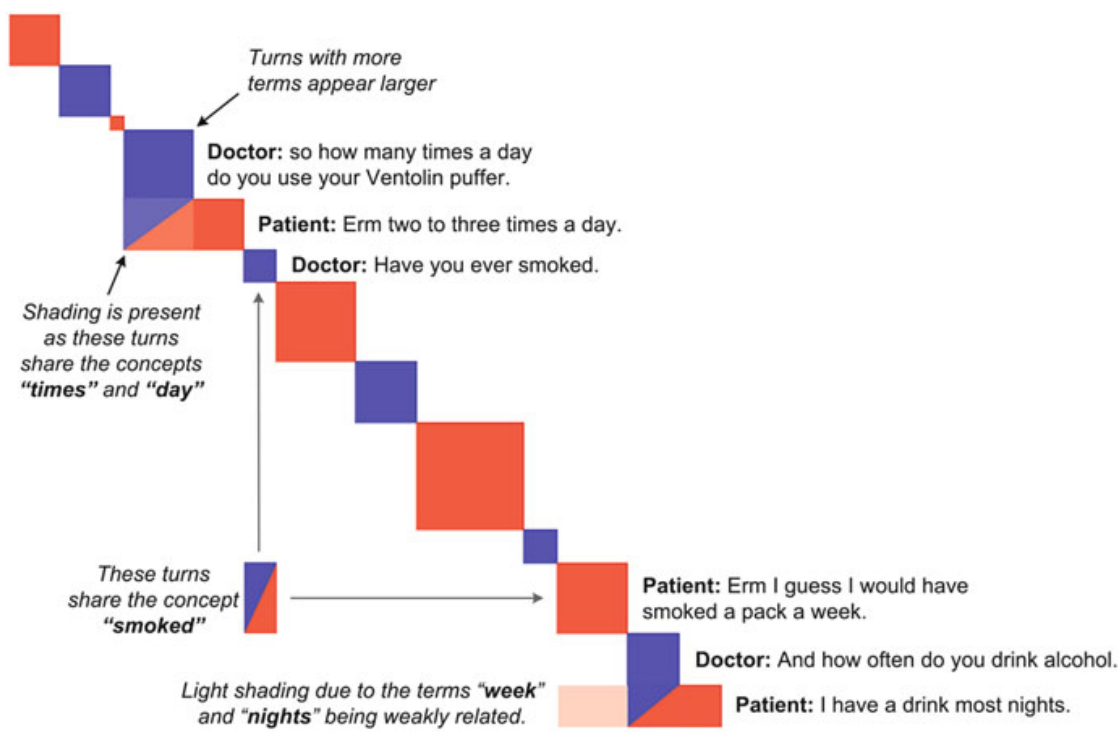


Fig. 14.6 Conceptual Recurrence Plot of 13 utterances and 4 corresponding recurrence elements from a Doctor/Patient consultation. The Patient is coloured *red* and the Doctor is coloured *blue*. Conceptual recurrence between the Patient and the Doctor is indicated by a half/half coloured *square*, and self-recurrence is in the speaker’s own colour From Angus et al. [63]

during rituals. They measured the heart rate of performers as well as spectators (both those related to the performers as well as general audience members). Konvalinka et al. found that all of the firewalkers had a distinctive signature to their heart rate dynamics, with a peak in heart rate around the firewalk itself. This same pattern was found for relatives of the fire walker. CRQA demonstrated that performers and those spectators to whom they were related or tangentially related maintained common heart rate trajectories longer (i.e., *cross-maxline*) than performers and those spectators to whom they were unrelated compared to a baseline level of arousal (i.e., heart rate when not observing the ritual). They proposed that this quantitative evidence of emotional synchrony (even in the absence of movement synchrony) may reflect the type of affective empathy that accompanies such rituals.

14.10 Conclusions

We have presented an overview of much of the work on interpersonal coordination that has quantified coordination using recurrence-based methods. Although this strategy for studying interpersonal coordination originated in movement coordination research, it is clear that it has made its way beyond the study of movement into a broad range of domains that continues to grow as research involving

these methods becomes more widely disseminated. It is also important to point out that the recurrence-based measures that have most commonly been used for quantifying coordination (e.g., *%cross-recurrence*, *%cross-determinism*, *cross-maxline*) are only a subset of the potential measures that could be developed. A number of researchers have branched out using new measures that may prove sensitive to different types of influences on interpersonal coordination. Angus et al. [63, 66], for example, have introduced a variety of new measures as described above. Lancia and Tiede [86] have likewise developed new strategies for quantifying deterministic structure in multi-signal CRPs (particularly in speech articulators) that avoid some of the assumptions of existing quantification tools. These also capture less conventional types of structure in CRPs, namely bowed diagonal lines that occur when two signals may not unfold on the same time scales (a form of non-stationarity). They applied an algorithm, based on a skeletonizing strategy often proposed by [87], which they use to calculate what Lancia and colleagues have subsequently referred to as *elastic determinism* [88, 89]. While their study did not explore interpersonal coupling, there is no barrier to using their strategy for doing so. The study of interpersonal coupling will continue to inform the development of recurrence strategies, just as recurrence strategies will continue to inform the theoretical and applied understanding of interpersonal coupling.

Acknowledgments This chapter is dedicated to Guy Van Orden (1952–2012) who foresaw the exceptional utility of recurrence-based methods for studying interpersonal coordination and encouraged our pursuit of these methods in behavioral research. We would also like to thank the American Psychological Association (APA) for sponsoring our Advanced Training Institute on Nonlinear Methods each summer since 2006 (taught at the University of Cincinnati in conjunction with John [Jay] Holden, Rick Dale, and Michael J. Richardson). APA's support, particularly the support of Nicolle Singer and Howard Kurtzman, has allowed us to share our knowledge of recurrence-based and other nonlinear methods with a greater number and variety of researchers than we could have hoped for. This work was supported by National Science Foundation grant: BCS-0926662.

References

1. J.P. Eckmann, S.O. Kamphorst, D. Ruelle, Recurrence plots of dynamical systems. *Eur. Lett.* **4**, 973–977 (1987)
2. C.L. Webber Jr., Rhythmogenesis of deterministic breathing patterns, in *Rhythms in Physiological Systems*, ed. by H. Haken, H.-P. Koepchen (Springer, Berlin, 1991), pp. 171–191
3. C.L. Webber Jr., J.P. Zbilut, Dynamical assessment of physiological systems and states using recurrence plot strategies. *J. Appl. Physiol.* **76**, 965–973 (1994)
4. C.L. Webber Jr., J.P. Zbilut, Assessing deterministic structures, in *physiological systems using recurrence plot strategies, in Bioengineering Approaches to Pulmonary Physiology and Medicine*, ed. by M.C.K. Khoo (Plenum Press, New York, 1996), pp. 137–148. Chapter 8
5. N. Marwan, J. Kurths, *Phys. Lett. A* **302**(5–6), 299 (2002)
6. K. Shockley, M. Butwill, J. Zbilut, C. Webber, Cross recurrence quantification of coupled oscillators. *Phys. Lett. A* **305**, 59–69 (2002)
7. J.P. Zbilut, A. Giuliani, C.L. Webber Jr., *Phys. Lett. A* **246**, 122 (1998)

8. R.C. Schmidt, C. Carello, M.T. Turvey, Phase transitions and critical fluctuations in the visual coordination of rhythmic movements between people. *J. Exp. Psychol. Hum. Percept. Perform.* **16**, 227–247 (1990)
9. P.N. Kugler, M.T. Turvey, *Information, Natural Law and the Self-Assembly of Rhythmic Movement* (Erlbaum, Hillsdale, 1987)
10. R.C. Schmidt, M.T. Turvey, Phase-entrainment dynamics of visually coupled rhythmic movements. *Biol. Cybern.* **70**, 369–376 (1994)
11. R.C. Schmidt, M.J. Richardson, Dynamics of interpersonal coordination, in *Coordination: Neural, Behavioral and Social Dynamics* (Springer, Berlin, 2008) pp. 281–308
12. H. Haken, J.A.S. Kelso, H. Bunz, A theoretical model of phase transitions in human hand movements. *Biol. Cybern.* **51**, 347–356 (1985)
13. D. Richardson, R. Dale, N.Z. Kirkham, The art of conversation is coordination: common ground and the coupling of eye movements during dialogue. *Psychol. Sci.* **18**, 407–413 (2007)
14. D.C. Richardson, R. Dale, M.J. Spivey, Eye movements in language and cognition, in *Empirical Methods in Cognitive Linguistics*, ed. by M.J. Spivey, M. Gonzalez-Marquez, I. Mittelberg, S. Coulson (John Benjamins, Amsterdam, 2007), pp. 323–344
15. M.J. Richardson, R.C. Schmidt, B.A. Kay, Distinguishing the noise and attractor strength of coordinated limb movements using recurrence analysis. *Biol. Cybern.* **96**(1), 59–78 (2007)
16. G.L. Pellicchia, K. Shockley, M.T. Turvey, Concurrent cognitive task modulates coordination dynamics. *Cognit. Sci.* **29**(4), 531–557 (2005)
17. M.J. Richardson, S. Lopresti-Goodman, M. Mancini, B. Kay, R.C. Schmidt, Comparing the attractor strength of intra- and interpersonal interlimb coordination using cross-recurrence analysis. *Neurosci. Lett.* **438**(3), 340–345 (2008)
18. R.C. Schmidt, M. Bienvenu, P.A. Fitzpatrick, P.G. Amazeen, A comparison of intra- and interpersonal interlimb coordination: coordination breakdowns and coupling strength. *J. Exp. Psychol. Hum. Percept. Perform.* **24**, 884–900 (1998)
19. D.P. Black, M.A. Riley, C.K. McCord, Synergies in intra- and interpersonal interlimb rhythmic coordination. *Motor Control* **11**, 348–373 (2007)
20. M.J. Richardson, K.L. Marsh, R.C. Schmidt, Effects of visual and verbal interaction on unintentional interpersonal coordination. *J. Exp. Psychol. Hum. Percept. Perform.* **31**, 62–79 (2005)
21. K. Shockley, M.V. Santana, C.A. Fowler, Mutual interpersonal postural constraints are involved in cooperative conversation. *J. Exp. Psychol. Hum. Percept. Perform.* **29**, 326–332 (2003)
22. W. Condon, W. Ogston, Speech and body motion synchrony of the speaker–hearer, in *The Perception of Language*, ed. by D. Horton, J. Jenkins (Charles E. Merrill, Columbus, 1971), pp. 150–184
23. M. LaFrance, Posture mirroring and rapport, in *Interaction Rhythms: Periodicity in Communicative Behavior*, ed. by M. Davis (Human Sciences Press, New York, 1982), pp. 279–298
24. D. Newtonson, The perception and coupling of behavior waves, in *Dynamical Systems in Social Psychology*, ed. by R.R. Vallacher, A. Nowak (Academic, San Diego, 1994), pp. 139–167
25. H.D.I. Abarbanel, *Analysis of Observed Chaotic Data* (Springer, New York, 1996)
26. J. Cappella, S. Planalp, Talk and silence sequences in informal conversations: III. Interspeaker influence. *Hum. Commun. Res.* **7**, 117–132 (1981)
27. M. Natale, Social desirability as related to convergence of temporal speech patterns. *Percept. Mot. Skills* **40**, 827–830 (1975)
28. R.L. Street Jr., Speech convergence and speech evaluation in fact-finding interviews. *Hum. Commun. Res.* **11**, 139–169 (1984)
29. B. Conrad, P. Schonle, Speech and respiration. *Arch. Psychiatr. Nervenkr.* **226**, 251–268 (1979)
30. M.C. Dault, L. Yardley, J.S. Frank, Does articulation contribute to modifications of postural control during dual-task performance? *Cogn. Brain Res.* **16**, 434–440 (2003)
31. L. Yardley, M. Gardner, A. Leadbetter, N. Lavie, Effect of articulation and mental tasks on postural control. *Neuroreport* **10**, 215–219 (1999)
32. K. Shockley, A.A. Baker, M.J. Richardson, C.A. Fowler, Articulatory constraints on interpersonal postural coordination. *J. Exp. Psychol. Hum. Percept. Perform.* **33**, 201–208 (2007)

33. T.A. Stoffregen, M.R. Givens, S. Villard, J. Yank, K. Shockley, Interpersonal postural coordination on rigid and non-rigid surfaces. *Motor Control* **13**, 471–483 (2009)
34. T.A. Stoffregen, M.R. Givens, S.J. Villard, K. Shockley, Effects of visual tasks and conversational partner on personal and interpersonal postural activity. *Ecol. Psychol.* **25**, 103–130 (2013)
35. M. Woollacott, A. Shumway-Cook, Attention and the control of posture and gait: a review of an emerging area of research. *Gait Posture* **16**, 1–14 (2002)
36. T.A. Stoffregen, L.J. Smart, B.G. Bardy, R.J. Pagulayan, Postural stabilization of looking. *J. Exp. Psychol. Hum. Percept. Perform.* **25**, 1641–1658 (1999)
37. T.A. Stoffregen, R.J. Pagulayan, B.G. Bardy, L.J. Hettinger, Modulating postural control to facilitate visual performance. *Hum. Mov. Sci.* **19**, 203–220 (2000)
38. K. Shockley, D.C. Richardson, R. Dale, Conversation and coordinative structures. *Top. Cogn. Sci.* **1**, 305–319 (2009)
39. L.W. Barsalou, W.K. Simmons, A.K. Barbey, C.D. Wilson, Grounding conceptual knowledge in modality-specific systems. *Trends Cogn. Sci.* **7**(2), 84–91 (2003)
40. D. Frick-Horbury, R.E. Guttentag, The effects of restricting hand gesture production on lexical retrieval and free recall. *Am. J. Psychol.* **111**, 43–62 (1998)
41. C. de'Sperati, N. Stucchi, Motor imagery and visual event recognition. *Exp. Brain Res.* **133**(2), 273–278 (2000)
42. R. Ellis, M. Tucker, Micro-affordance: the potentiation of components of action by seen objects. *Br. J. Psychol.* **91**(4), 451–471 (2000)
43. A.M. Glenberg, M.P. Kaschak, Grounding language in action. *Psychon. Bull. Rev.* **9**, 558–565 (2001)
44. M. Wexler, S.M. Kosslyn, A. Berthoz, Motor processes in mental rotation. *Cognition* **68**, 77–94 (1998)
45. D. Kirsh, P. Maglio, Some epistemic benefits of action: Tetris a case study, in *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society* (Morgan Kaufmann, New York, 1992)
46. D.H. Ballard, M.M. Hayhoe, J.B. Pelz, Memory representations in natural tasks. *Cogn. Neurosci.* **7**, 66–80 (1995)
47. C. McKinsty, R. Dale, M.J. Spivey, Action dynamics reveal parallel competition in decision making. *Psychol. Sci.* **19**, 22–24 (2008)
48. L. Fogassi, L. Simone, The mirror system in monkeys and humans and its possible motor-based functions, in *Progress in Motor Control: Neural, Computational, and Dynamic Approaches*, ed. by M.J. Richardson, M.A. Riley, K. Shockley (Springer, New York, 2013)
49. G. Rizzolatti, L. Craighero, The mirror-neuron system. *Annu. Rev. Neurosci.* **27**, 169–192 (2004)
50. V. Gallese, C. Keysers, G. Rizzolatti, A unifying view of the basis of social cognition. *Trends Cogn. Sci.* **8**, 396–403 (2004)
51. B. Hommel, J. Müsseler, G. Aschersleben, W. Prinz, The theory of event coding (TEC): a framework for perception and action planning. *Behav. Brain Sci.* **24**, 849–878 (2001)
52. P.L. Jackson, J. Decety, Motor cognition: a new paradigm to study self-other interactions. *Curr. Opin. Neurobiol.* **14**, 259–263 (2004)
53. W. Prinz, B. Hommel (eds.), *Common Mechanisms in Perception and Action: Attention and Performance XIV* (Oxford Press, Oxford, 2002)
54. M.T. Tolston, K. Shockley, M.A. Riley, M.J. Richardson, Movement constraints on interpersonal coordination and communication. *J. Exp. Psychol. Hum. Percept. Perform.* (in press)
55. V.C. Ramenzoni, T.J. Davis, M.A. Riley, K. Shockley, A.A. Baker, Joint action in a cooperative precision task: nested processes of intrapersonal and interpersonal coordination. *Exp. Brain Res.* **211**, 447–457 (2011)
56. V.C. Ramenzoni, M.A. Riley, K. Shockley, A. Baker, Interpersonal and intrapersonal coordinative modes for joint and single task performance. *Hum. Mov. Sci.* **31**, 1253–1267 (2012)
57. M.A. Riley, M.J. Richardson, K. Shockley, V.C. Ramenzoni, Interpersonal synergies. *Front. Psychol.* **2**, 1–7 (2011)

58. H. Haken, *Synergetics: An Introduction* (Springer, Berlin, 1977)
59. F. Orsucci, A. Giuliani, C. Webber Jr., J. Zbilut, P. Fonagy, M. Mazza, Combinatorics and synchronization in natural semiotics. *Phys. A Stat. Mech. Appl.* **361**(2), 665–676 (2006)
60. R. Dale, M.J. Spivey, Unraveling the dyad: using recurrence analysis to explore patterns of syntactic coordination between children and caregivers in conversation. *Lang. Learn.* **56**, 391–430 (2006)
61. B. MacWhinney, *The CHILDES Project: Tools for Analyzing Talk* (Erlbaum, Mahwah, 2000)
62. A.S. Warlaumont, D.K. Oller, R. Dale, J.A. Richards, J. Gilkerson, D. Xu, Vocal interaction dynamics of children with and without autism, in *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 121–126). Cognitive Science Society, Austin (2010)
63. D. Angus, B. Watson, A. Smith, C. Gallois, J. Wiles, Visualising conversation structure across time: insights into effective doctor–patient consultations. *PLoS One* **7**(6), e38014 (2012). doi:[10.1371/journal.pone.0038014](https://doi.org/10.1371/journal.pone.0038014)
64. D. Angus, A. Smith, J. Wiles, Human communication as coupled time series: quantifying multi-participant recurrence. *IEEE Trans. Vis. Comput. Graph.* **20**, 1795–1807 (2012)
65. G. Salton, *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer* (Addison-Wesley, Reading, 1989)
66. D. Angus, A. Smith, J. Wiles, Conceptual recurrence plots: revealing patterns in human discourse. *IEEE Trans. Vis. Comput. Graph.* **18**, 988–997 (2012)
67. J.C. Gorman, N.J. Cooke, P.G. Amazeen, S. Fouse, Measuring patterns in team interaction sequences using a discrete recurrence approach. *Hum. Factors: J. Hum. Factors Ergon. Soc.* **54**(4), 503–517 (2012)
68. J.C. Gorman, P.G. Amazeen, N.J. Cooke, Team coordination dynamics. *Nonlinear Dynamics Psychol. Life Sci.* **14**, 265–289 (2010)
69. P.T. Esteves, D. Araújo, Aplicação da cross recurrence quantification ao estudo da coordenação interpessoal no basquetebol. Paper presented at the 2nd international symposium on sports performance. CIDESD, Covilhã, Portugal, 2011
70. J. Carvalho, D. Araújo, B. Travassos, P. Esteves, L. Pessanha, F. Pereira, K. Davids, Dynamics of players' relative positioning during baseline rallies in tennis. *J. Sports Sci.* **31**(14), 1596–1605 (2013)
71. A. Demos, T. Frank, R. Chaffin, Understanding movement during performance: a recurrence quantization approach in *International Symposium on Performance Science* (2011)
72. A.P. Demos, T.D. Frank, T. Logan, Movement during performance: a hunt for musical structure in postural sway. Poster presented at the society for musical perception and cognition (SMPC), Rochester, New York, August 2011
73. G. Varni, M. Mancini, G. Volpe, A. Camurri, A system for mobile active music listening based on social interaction and embodiment. *Mobile Netw. Appl.* **16**(3), 375–384 (2011)
74. G. Varni, G. Dubus, S. Oksanen, G. Volpe, M. Fabiani, R. Bresin et al., Interactive sonification of synchronisation of motoric behaviour in social active listening to music with mobile devices. *J. Multimodal User Interf.* **5**(3–4), 157–173 (2012)
75. S. Gill, M.R. Thompson, T. Himberg, Body rhythmic entrainment and pragmatics in musical and linguistic improvisation tasks, in *Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music* ed. by E. Cambouropoulos, C. Tsougras, P. Mavromatis, K. Papiadakis (2012)
76. G. Varni, A. Camurri, P. Coletta, G. Volpe Emotional entrainment in music performance, in *Automatic Face & Gesture Recognition: Proceedings of the 8th IEEE International Conference (IEEE, 2008)* pp. 1–5
77. N. Marwan, M.C. Romano, M. Thiel, J. Kurths, Recurrence plots for the analysis of complex systems. *Phys. Rep.* **438**, 237–329 (2007)
78. A. Camurri, G. Varni, G. Volpe, Towards analysis of expressive gesture in groups of users: computational models of expressive social interaction, in *Lecture Notes in Computer Science*, vol. 5934, 122–133 (2010)

79. D. Richardson, R. Dale, Looking to understand: the coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognit. Sci.* **29**, 1045–1060 (2005)
80. P. Jermann, M.A. Nüssli, Effects of sharing text selections on gaze cross-recurrence and interaction quality in a pair programming task, in *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (ACM, 2012), pp. 1125–1134
81. I. Konvalinka, D. Xygalatas, J. Bulbulia, U. Schjødt, E.M. Jegindø, S. Wallot et al., Synchronized arousal between performers and related spectators in a fire-walking ritual. *Proc. Natl. Acad. Sci.* **108**(20), 8514–8519 (2011)
82. R.B. Van Baaren, R.W. Holland, K. Kawakami, A. Van Knippenberg, Mimicry and prosocial behavior. *Psychol. Sci.* **15**(1), 71–74 (2004)
83. S.S. Wiltermuth, C. Heath, Synchrony and cooperation. *Psychol. Sci.* **20**(1), 1–5 (2009)
84. E.E. Cohen, R. Ejsmond-Frey, N. Knight, R.I.M. Dunbar, Rowers' high: behavioural synchrony is correlated with elevated pain thresholds. *Biol. Lett.* **6**(1), 106–108 (2010)
85. M.J. Hove, J.L. Risen, It's all in the timing: interpersonal synchrony increases affiliation. *Soc. Cogn.* **27**(6), 949–960 (2009)
86. L. Lancia, M. Tiede, A survey of methods for the analysis of the temporal evolution of speech articulator trajectories, in *Speech Planning and Dynamics*, ed. by S. Fuchs, P. Perrier (Peter Lang, Frankfurt am Main, 2012), pp. 233–271
87. N. Marwan, M. Thiel, N. Nowaczyk, Cross recurrence plot based synchronization of time series. *Nonlinear Processes Geophys.* **9**(3–4), 325–331 (2002)
88. L. Lancia, H. Avelino, D. Voigt, Measuring laryngealization in running speech: interaction with contrastive tones in Yálalag Zapotec. Poster presented at 14th Annual Conference of the International Speech Communication Association Interspeech, Lyon (2013)
89. L. Lancia, S. Fuchs, M. Tiede, Application of concepts from cross-recurrence analysis in speech production: an overview and a comparison to other nonlinear methods. *J. Speech Lang. Hear. Res.* doi:[10.1044/1092-4388\(2013/12-0223\)](https://doi.org/10.1044/1092-4388(2013/12-0223))