In: Mathematical Physics Research at the Cutting Edge
Editor: C.V. Benton, pp. 101-139
ISBN 1-59033-939-8
© 2004 Nova Science Publishers, Inc.

Cross Recurrence Plots and Their Applications

Norbert Marwan* and Jürgen Kurths
*Nonlinear Dynamics Group, Institute of Physics,
University of Potsdam, 14415 Potsdam, Germany,
marwan@agnld.uni-potsdam.de

Abstract

Cross recurrence plots are a new tool for nonlinear data analysis. They exhibit characteristic structures which can be used for the study of differences between two processes or for the alignment and search for matching sequences of two data series, even in the case when cross-correlation techniques fails or when the data are nonstationary. Selected applications of the introduced techniques to various kind of data demonstrate their potential.

1 Introduction

A major task in bi- or multivariate data analysis is to compare or to find interrelations between different time series. Often, these data are gained from natural systems, which show generally nonstationary and complex behaviour. Furthermore, these systems are often observed by rather few measurements providing short data series. Linear approaches of time series analysis are often not sufficient to analyse this kind of data. In the last two decades a great variety of nonlinear techniques has been developed to analyse data of complex systems (cf. Abarbanel et al., 1993; Kantz and Schreiber, 1997). Most popular are methods for estimation of fractal dimensions, Lyapunov exponents or mutual information (Kantz and Schreiber, 1997; Kurths and Herzel, 1987; Mandelbrot, 1982; Wolf et al., 1985). However, most of these methods need rather long data series. The uncritical application of these methods especially to natural data often leads to pitfalls.

In the last decade a new method based on nonlinear data analysis has become popular: recurrence plots (Eckmann et al., 1987). Recurrence is a fundamental property of dissipative dynamical systems. Even though small disturbations of

such a system cause exponential divergence of its state, after some time the system will reach a state that is arbitrary close to the former state and pass through a similar evolution. Recurrence plots visualize such recurrent behaviour of dynamical systems. Although they are not completely theoretically understood, practitioners of this method claim its relevance for short and nonstationary data. These features are indeed the crucial advantage of recurrence plots. Zbilut and Webber Jr. (1992) have made an important further step by introducing a quantification analysis based on recurrence plots, which became well known in the analysis especially of physiological data. Hundreds of works and publications using this quantification analysis can be found in literature. It seems that the reason for this amazing growth in the popularity of recurrence plots is not only the technical aspect. Recurrence plots can be very decorative and attract attention.

An extension of the method of recurrence plots to *cross recurrence plots* (CRPs) enables investigating the time dependent behaviour of two processes which are both recorded in a single time series (Zbilut et al., 1998; Marwan and Kurths, 2002; Marwan et al., 2002a). The basic idea of this approach is to compare the phase space trajectories of two processes in the same phase space. CRPs can be used in order to study the similarity of two different phase space trajectories.

In order to understand the methods based on CRPs, we will first review the concepts of phase space trajectories and recurrence plots. In the following sections we will expand these concepts to CRPs and develop further measures of complexity, which are based on CRPs and evaluate the similarity of the considered systems. This nonlinear approach enables us to identify epochs when there are linear and even nonlinear interrelations between both systems. In the last section we will apply the methods of CRPs to data from geology.

2 Review on Recurrence Plots

2.1 Phase Space

The analysis of phase space trajectories is a basic concept of nonlinear data analysis. In this subsection a brief introduction of this concept is given. For further reading see for example Eckmann and Ruelle (1985), Abarbanel et al. (1993) or Ott (1993).

The states of systems in nature or engineering typically change in time. The

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

$$x_1(t), x_2(t), \ldots, x_d(t), \tag{1}$$

which form a vector $\vec{x}(t)$ 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

$$\dot{\vec{x}}(t) = \partial_t \vec{x}(t) = \vec{F}(x). \tag{2}$$

The temporary succession of the phase space vectors forms a trajectory (phase space trajectory, orbit). The velocity field $\vec{F}(x)$ 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 $\vec{F}(x)$ is known, the state at a given time can be determined by integrating the equation system (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 respective 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. 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 (Packard et al., 1980; Takens, 1981). A method frequently used for reconstructing such a trajectory $\hat{\vec{x}}(t)$ is the time delay method: $\hat{\vec{x}}_i = (u_i, u_{i+\tau}, \ldots, u_{i+(m-1)\tau})^T$, 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 (Takens, 1981).

Both embedding parameters, the dimension m and the delay τ, have to be chosen appropriately. Different approaches are possible for the determination of these embedding parameters (cf. Cao, 1997; Kantz and Schreiber, 1997). A rather common method for the estimation of the embedding dimension is the investigation

of the changes in the neighbourhood of phase space points under changes of the embedding dimension. Such a dimension has to be taken where the neighbourhood of phase space points (e.g. the amount of false nearest neighbours) does not change. Possible means of determining the delay are the *autocovariance function* or the *mutual information.*

A phase space trajectory or its reconstruction can be used in order to estimate characteristic properties of the dynamical system, like Lyapunov exponents, various types of dimensions etc. (cf. Eckmann and Ruelle, 1985; Abarbanel et al., 1993; Ott, 1993). Besides, the phase space reconstruction is the starting point for the construction of a recurrence plot.

2.2 Recurrence Plots

The recurrence plot (RP) is a tool for visualizing the dynamics of phase space trajectories. Usually, a phase space does not have a dimension (two or three) which allows it to be pictured. Higher dimensional phase spaces can only be visualized by projection into the two or three dimensional sub-spaces. However, an RP enables us to visualize the m-dimensional phase space trajectory through a two-dimensional representation of its recurrences (Fig. 1).

The recurrence plot is defined as

$$\mathbf{R}(t_1, t_2) = \Theta\left(\varepsilon(t_1) - \|\vec{x}(t_1) - \vec{x}(t_2)\|\right), \tag{3}$$

or for time-discrete variables ($t = i\Delta t$)

$$\mathbf{R}_{i,j} = \Theta\left(\varepsilon_i - \|\vec{x}_i - \vec{x}_j\|\right), \tag{4}$$

where $\varepsilon(t)$ or ε_i is a predefined cut-off distance, $\|\cdot\|$ is the norm (e. g. the Euclidean norm) and $\Theta(x)$ is the Heaviside function. The values *one* and *zero* in this matrix can be simply visualized by the colours black and white. Depending on the kind of the application, ε_i can be a fixed value or it can be changed for each i in such a way that in the ball with the radius ε_i a predefined amount of nearest neighbours occurs. The choice of a fixed amount of nearest neighbours will provide a constant density of recurrence points in each column of the RP. The radius ε_i of such a neighbourhood changes for each $\vec{x}_i$ ($i = 1 \ldots N$) and $\mathbf{R}_{i,j} \neq \mathbf{R}_{j,i}$ because the vicinity of $\vec{x}_i$ does not have to be the same as that of $\vec{x}_j$. This property leads to an asymmetric RP, but all columns of the RP have the same recurrence density. For the use of this neighbourhood criterion, we make the following statement: ε is the parameter for

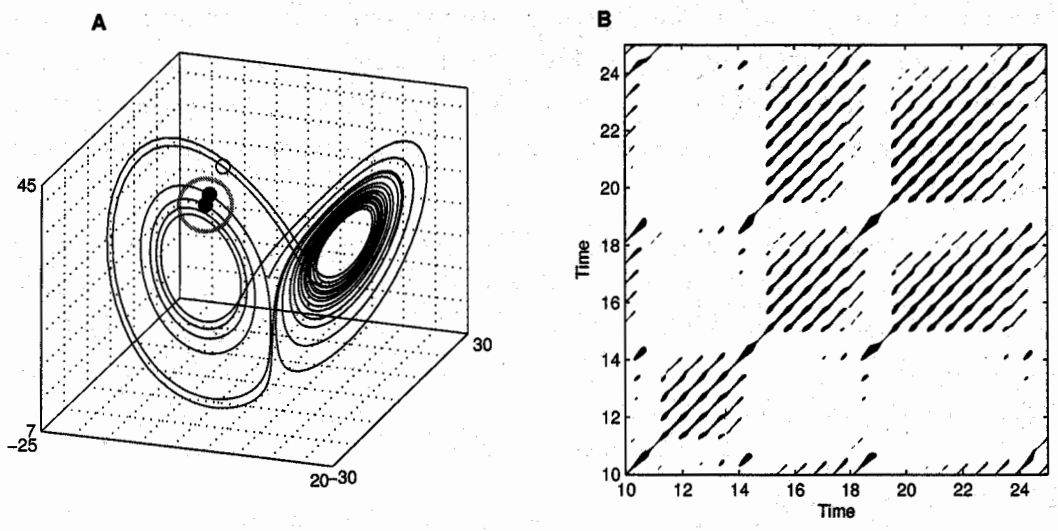

Figure 1: (A) Segment of the phase space trajectory of the Lorenz system (for standard parameters $r = 28$, $\sigma = 10$, $b = \frac{8}{3}$; Lorenz, 1963) computed by using its three components and (B) its corresponding recurrence plot. A point of the trajectory at j which falls into the neighbourhood (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 predefinition of the recurrence density, but does not mean the radius of the vicinity. This means that with a given $\varepsilon = 0.15$ the real, locally chosen ε_i is adjusted in such a way that the neighbourhood covers 15% of all phase space vectors, and thus the recurrence density is 0.15. We denote this neighbourhood as *fixed amount of nearest neighbours (FAN)*.

Since $\mathbf{R}_{i,i} = 1$ $(i = 1 \ldots N)$ 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, it is possible to reconstruct the properties of the data from the totality of all recurrence points. McGuire et al. (1997) have shown the preservation of main dynamical properties for the distance matrix $\mathbf{D}^m_{i,j} = \|\vec{x}_i - \vec{x}_j\|$. Nevertheless, the phase space trajectory can also be reconstructed from the binary RP, where the information about the absolute length of the phase space vectors is

lost. The RP provides information for reordering the indices of the phase space vectors, such that the vectors are sorted by their norm. If the cumulative distribution of the lengths of the phase space vectors is known, the restored phase space trajectory will recover its amplitude by equating the sorted indices with this distribution.

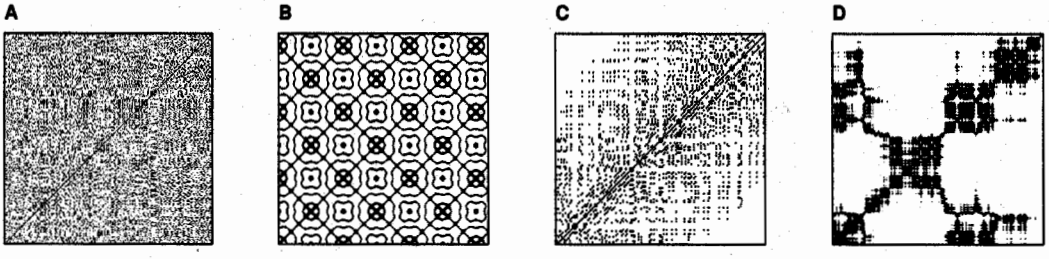

Figure 2: Characteristic typology of recurrence plots: (A) homogeneous (uniformly distributed noise), (B) periodic (super-positioned harmonic oscillations), (C) drift (logistic map $x_{i+1} = 4x_i(1 - x_i)$ corrupted with a linearly increasing term) and (D) disrupted (Brownian motion). These examples 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 $\varepsilon = 0.2$ (A, C, D) and $\varepsilon = 0.4$ (B).

An RP exhibits characteristic large-scale and small-scale patterns which are caused by typical dynamical behavior (Eckmann et al., 1987; Marwan et al., 2002b; Webber Jr. and Zbilut, 1994), e. g. diagonals or horizontal and vertical black lines. The large-scale patterns were denoted by Eckmann et al. (1987) as *typology* and the small-scale patterns as *texture*. The typology offers a global impression which can be characterized as *homogeneous*, *periodic*, *drift* and *disrupted* (Fig. 2).

Single, isolated recurrence points 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 single recurrence point contains no information about the state itself.

A *diagonal line* $\mathbf{R}_{i+k,j+k} = 1$ (for $k = 1 \ldots l$, where l is the length of the diagonal line) occurs when the trajectory visits the same region of the phase space at different times and a segment of the trajectory runs parallel to another segment. 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). Since the definition of the maximal Lyapunov exponent uses the time of the parallel running of trajectories, the relationship between the diagonal lines in an RP and the maximal Lyapunov exponent is obvious. There exist different attempts to relate the diagonal lines with the maximal Lyapunov exponent (Trulla et al., 1996; Choi et al., 1999; Gao and Cai, 2000).

A *vertical (horizontal) line* $\mathbf{R}_{i,j+k} = 1$ (for $k = 1 \ldots v$, 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).

In a more general sense the line structures in an RP exhibit locally the time relationship between the current trajectory segments. A line structure in an RP of length l corresponds to the closeness of the segment $f(T_1(t))$ to another segment $f(T_2(t))$, where $T_1(t)$ and $T_2(t)$ are the local time scales which preserve that $f(T_1(t)) \approx f(T_2(t))$ for some time $t = 1 \ldots l\Delta t$. The local slope $\mathrm{m}(t)$ of a line in an RP represents the ratio between the local time derivatives of these time transformations $T_i(t)$

$$\mathrm{m}(t) = \frac{\partial_t T_1(t)}{\partial_t T_2(t)}. \tag{5}$$

We will consider here an illustrative example. A further explanation of the relationship between the slope of the lines and the trajectories is given in Subsec. 3.3. Let us consider a function $f(T) = T(t)$ with a section of a monotonical, linear increase $T_{lin} = t$ and another (hyperbolic) section which follows $T_{hyp} = -\sqrt{r^2 - t^2}$ (Fig. 3A). Since the ratio between the derivatives of the functions T_i of the linear and the hyperbolic sections

$$\mathrm{m} = \frac{\partial_t T_{lin}(t)}{\partial_t T_{hyp}(t)} = \frac{t}{\sqrt{r^2 - t^2}} \tag{6}$$

corresponds to the derivative of a circle line with a radius r, a bowed line structure with the form of a circle occurs in the RP (Fig. 3C).

The small scale structures, diagonal and vertical lines, are the base of a quantitative analysis of the RPs, which is known as *recurrence quantification analysis (RQA)*. This analysis has been developed to assess the RPs and to detect transitions (e. g. bifurcation points) in complex systems (Trulla et al., 1996; Webber Jr. and Zbilut, 1994; Zbilut and Webber Jr., 1992; Marwan et al., 2002b). The quantification tech-

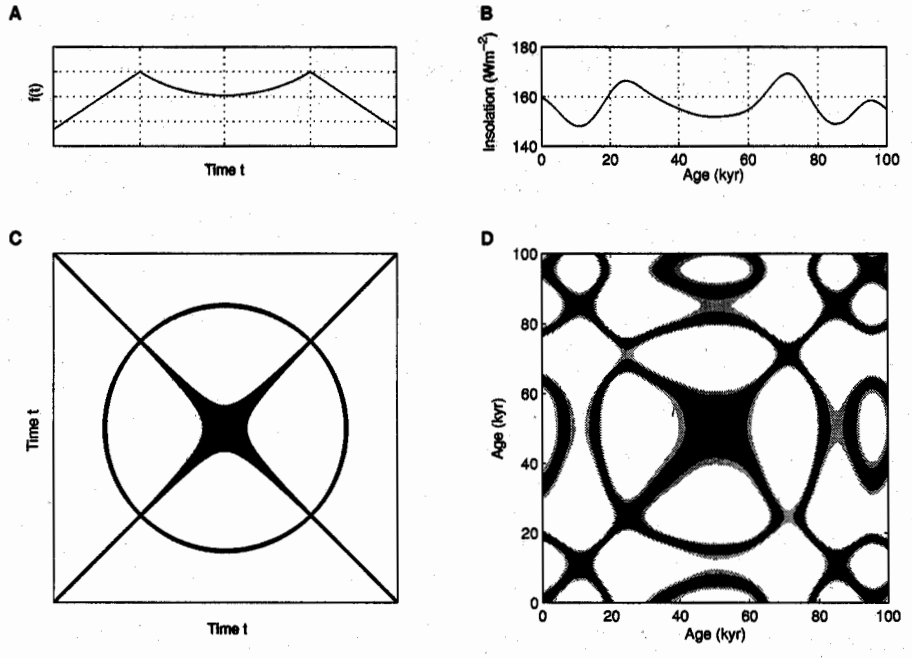

Figure 3: (A, C) Illustrative example of the relationship between the slope of lines in an RP and the local derivatives of the involved trajectory segments. Since the ratio between the local derivative of the linear and the hyperbolic sections corresponds to the derivative of a circle line, a circle occurs in the RP. (B, D) A corresponding structure found in nature: the solar insolation on the latitude 44°N for the last 100 kyr (data from Berger and Loutre, 1991). RPs created without embedding; in (D) an RP for a larger threshold is additionally shown in gray.

nique contains numerous measures of complexity as *recurrence rate, determinism, laminarity, trapping time* etc.

The RPs test the distance between all points of the same phase space trajectory. However, why should it not be possible to test each point of one trajectory with each point of another trajectory in the same phase space? This leads us to the concept of *cross recurrence plots (CRP)*, which we will focus on in the next section.

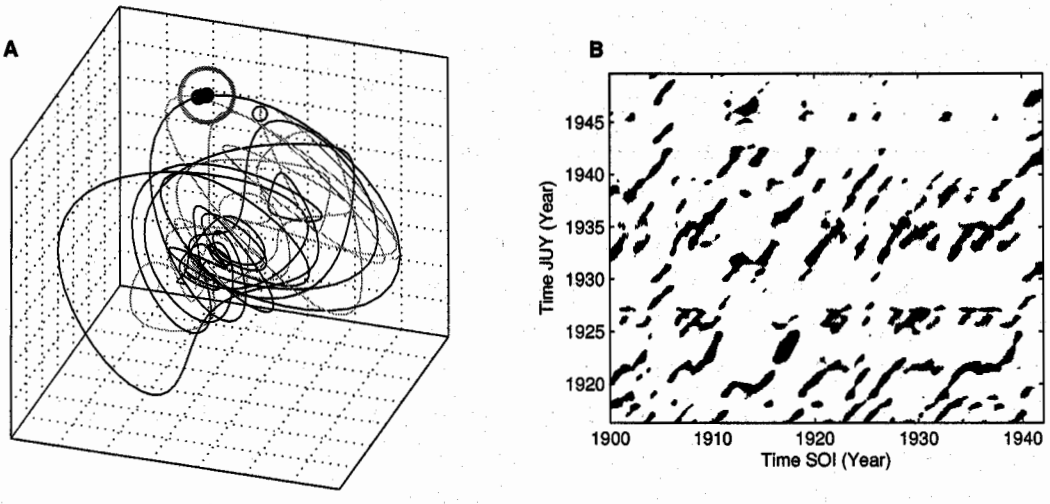

Figure 4: (A) Segments of the phase space trajectories of the El Niño Southern Oscillation Index (SOI, black line; data from the Climate Server of NOAA http://ferret.wrc.noaa.gov) and a precipitation time series of San Salvador de Jujuy (Argentina, gray line; data from Bianchi and Yañez (1992)) by using time delay embedding (smoothed, monthly data; $\tau = 7$ months). In (B) the corresponding cross recurrence plot is shown. If a point of the precipitation trajectory at j (black point on the gray line in (A)) falls into the neighbourhood (gray circle in (A)) of a point of the SOI trajectory at i, in the CRP at the location (i, j), a black point will be marked. A point outside the neighbourhood (small circle in (A)) is marked as a white point in the CRP. For creating this CRP the FAN criterion with $\varepsilon = 0.15$ is used.

3 Cross Recurrence Plots

3.1 Definition of Cross Recurrence Plots

Starting with the concept of RPs we regard a phase space with one trajectory $\vec{x}_i$ of length N_x. Now we add a second trajectory $\vec{y}_j$ with the length N_y into the same phase space (Fig. 4). The test between all points of the first trajectory with all points of the second trajectory leads to the *cross recurrence plot (CRP)*

$$\mathbf{CR}_{i,j}^{m,\varepsilon_i} = \Theta\left(\varepsilon_i - \left\|\vec{x}_i - \vec{y}_j\right\|\right), \quad \vec{x}_i, \vec{y}_j \in \mathrm{R}^m, \\ i = 1 \ldots N_x, \; j = 1 \ldots N_y. \tag{7}$$

The notation is analogous to the definition of RPs (Eq. 4). If in the second trajectory a state at time j is close to a state on the first trajectory at time i, a black dot will be assigned to the matrix **CR** at location (i, j). This occurrence of neighbours in both trajectories is not a "recurrence" of states, hence, the matrix (Eq. 7) does not represent recurrences but the conjunctures of states of both systems. Therefore, this representation is not really a "cross *recurrence* plot". Nevertheless we call it "cross recurrence plot" in order to follow the way of a generalization of RPs and because of the occurrence of the term "cross recurrence quantification" in the literature for the parallel concept of the generalization of the RQA (Zbilut et al., 1998). The vectors $\vec{x}$ and $\vec{y}$ do not need to have the same length, hence the matrix **CR** is not necessarily square. This extension of RPs was firstly used by Zbilut et al. (1998) for the *cross recurrence quantification*. Independently of their work, the concept of cross recurrence plots also surfaces in Marwan (1999). If the second vector $\vec{y}_j$ is the same as the vector $\vec{x}_j$, the CRP will correspond to the RP of $\vec{x}_j$.

Both trajectories for the creation of a CRP have to represent the same dynamical system with equal state variables because they are in the same phase space. This must be taken into account if time series of different measurements (e. g. temperature and pressure) are involved. For the embedding and the following CRP analysis the time series can be taken from different measurements if they are components or state variables of the same system. The time series can also be from different subsystems, which belong to the same system (e. g. the sub-systems El Niño Southern Oscillation and North Atlantic Oscillation within system of the global atmospheric oscillation). A precedent data normalization may solve the problems of different units and amplitude scaling. However, the application of CRPs to measurements of completely different systems, which cannot be regarded as observations of the same dynamical system (e. g. a stocks index and the precession of the Earth's rotation), makes no sense. For such different kinds of data the presently developed concept of intersected RPs can be used (Romano et al., 2003).

Assuming that both trajectories come from the same process but have different absolute values, the CRP will not yield the expected RP if a fixed threshold ε is chosen. Therefore, it is necessary to adapt both trajectories to the same range of values, e. g. by using a normalization to the standard deviation. However, the application of a fixed amount of nearest neighbours (FAN), i. e. ε_i changes for each state x_i, solves this problem automatically, and a modification of the amplitudes is not necessary. This choice of a neighbourhood has the additional advantage of

working well for slowly changing trajectories (e. g. drift).

Since the values of $\mathbf{CR}_{i,i}$ $(i = 1 \ldots N)$ are not necessarily one, the black main diagonal usually vanishes. As we will discuss in Subsec. 3.3, the line of identity (LOI) can be replaced by the *line of synchronization (LOS)*, may be bowed and may ultimately not have the angle $\pi/4$. However, the lines which are more or less diagonally oriented are also of interest. They represent segments of both trajectories, which run nearly parallel for some time. The frequency and lengths of these lines are obviously related to a certain similarity between the dynamics of both subsystems.

A time dilatation or time compression of one of the trajectories causes a distortion of the main diagonal line. This case will be discussed in subsection 3.3. In the following subsection we presume the situation that both systems have the same time scale (equal length N and sample time Δt), hence, the CRP is an $N \times N$ array.

3.2 Measures for Similarities Between Two Observed Processes

The long diagonal structures in the CRP reveal similar time evolution of the trajectories of both processes. It is obvious that a progressively increased similarity between both processes causes an increase of the recurrence point density along the main diagonal $\mathbf{CR}_{i,i}$ $(i = 1 \ldots N)$ until a black straight main diagonal line occurs (which would be in fact the LOI) and the CRP becomes an RP. Thus, the occurrence of diagonal lines in a CRP can be used in order to benchmark the similarity between the considered processes.

In order to quantify this similarity, some quantitative measures have to be defined. Since we are interested in the occurrence of the more or less discontinuous diagonal lines, these measures should be diagonalwise applied (Marwan and Kurths, 2002).

Let us consider a diagonal $\mathbf{CR}_{i,j}$ $(j - i = k = \text{const.})$ which is parallel to the main diagonal and has a time distance $t = k\,\Delta t$ from the main diagonal. The recurrence points in this diagonal correspond with tests between the time delayed trajectories (delay t). In the following, some RQA measures will be redefined for these diagonals. Hence, these measures will be functions of the distance k from the main diagonal. Using this approach it is possible to assess the similarity in the dynamics depending on a certain time delay.

Following this procedure, we need to define the frequency distributions of the diagonal line lengths $P_k^{\varepsilon}(l) = \{l_i;\ i = 1 \ldots N_l\}$ (N_l is the absolute number of diag-

onal lines) for each diagonal parallel to the main diagonal $\mathbf{CR}_{i,j}^{m,\varepsilon}$ $(j-i=k)$. For $k=0$ this line is the LOI, $k>0$ diagonals above and $k<0$ diagonals below the LOI, which represent positive and negative time delays, respectively.

The recurrence rate RR is now modified to

$$RR_k = RR_*(t) = \frac{1}{N-k} \sum_{j-i=k} \mathbf{CR}_{i,j}^{m,\varepsilon} = \frac{1}{N-k} \sum_{l=1}^{N-k} l\, P_k^\varepsilon(l) \tag{8}$$

and reveals the probability of the occurrence of similar states in both systems with a certain delay $t = k\,\Delta t$. A high density of recurrence points in a diagonal results in a high value of RR_*. This is the case for systems whose trajectories often visit the same phase space regions.

Analogous to the RQA the determinism

$$DET_k = \frac{\sum_{l=l_{min}}^{N-k} l\, P_k^\varepsilon(l)}{\sum_{l=1}^{N-k} l\, P_k^\varepsilon(l)} \tag{9}$$

is the proportion of recurrence points forming long diagonal structures to all recurrence points, but here it is constrained to the considered diagonal. Smooth trajectories with long autocorrelation times will result in a CRP with long diagonal structures, even if the trajectories are not linked to each other (this effect corresponds to the tangential motion of one trajectory). In order to avoid the counting of such "false" diagonals, the lower limit for the diagonal line length l_{min} should be of the order of the autocorrelation time.

Stochastic as well as heavily fluctuating processes cause none or only short diagonals, whereas deterministic processes cause longer diagonals. As mentioned above, if two deterministic processes have the same or similar time evolution, i. e. parts of the phase space trajectories meet the same phase space regions for certain times, the amount of longer diagonals will increase and the amount of shorter diagonals will decrease. The average diagonal line length

$$L_k = \frac{\sum_{l=l_{min}}^{N-k} l\, P_k^\varepsilon(l)}{\sum_{l=l_{min}}^{N-k} P_k^\varepsilon(l)} \tag{10}$$

quantifies the duration of such a similarity in the dynamics. A high coincidence of both trajectories increases the length of these diagonals. Besides, the entropy of the probability $P_k^\varepsilon(l)$ can also be defined. Still, we focus here on the first three measures.

High values of RR_* represent high probabilities of the occurrence of the same state in both processes, high values of DET_* and L_* represent a long time span of the occurrence of a similar dynamics in both processes. Whereas DET_* and L_* are sensitive to fast and highly fluctuating data, RR_* measures the probabilities of the occurrence of the same states in spite of these high fluctuations (noisy data). It is important to emphasize that these parameters are statistical measures and that their validity increases with the size of the CRP, i. e. with the observation length.

An additional CRP

$$\mathbf{CR}^{-}_{i,j} = \Theta\left(\varepsilon_i - \left\|\vec{x}_i + \vec{y}_j\right\|\right) \tag{11}$$

with opposite signed second trajectory $-\vec{y}_j$ allows to distinguish positive and negative relations between the considered trajectories (Marwan and Kurths, 2002). In order to recognize the measures for both possible CRPs, we add the superscript index $+$ to the measures for the positive linkage and the superscript index $-$ for the negative linkage, e. g. RR^{+}_{k} and RR^{-}_{k}.

Another approach used to study the positive and negative relations between the considered trajectories involves the composited measures for the recurrence rate

$$RR^{c}_{k} = \frac{1}{N-k} \sum_{j-i=k} \left(\mathbf{CR}^{+}_{i,j} - \mathbf{CR}^{-}_{i,j}\right), \tag{12}$$

the determinism

$$DET^{c}_{k} = DET^{+}_{k} - DET^{-}_{k}, \tag{13}$$

and the average diagonal length

$$L^{c}_{k} = L^{+}_{k} - L^{-}_{k}, \tag{14}$$

where $P^{\pm}_{k}(l)$ is the histogram of the diagonal line lengths in $\mathbf{CR}^{\pm}_{i,j}$ $(j-i=k)$, as it is used in Marwan et al. (2003). This representation is similar to those of the cross correlation function and is more intuitive than the separate representation of RR^{+}_{*}, RR^{-}_{*} etc. However, for the investigation of interrelations based on even functions, these composited measures are not suitable.

A substantial advantage of this method is its capability of finding nonlinear similarities in short and nonstationary time series with high noise levels as they typically occur, e. g., in biology or earth sciences.

However, the shortness and nonstationarity of data limit this method as well. As mentioned above, one way to reduce problems accompanying nonstationary data is the alternative choice of a neighbourhood with a fixed amount of neighbours.

3.3 Time Scale Alignment of Time Series

In data analysis one is often faced with time series measured on varying time scales. These could be, for example, sets from borehole or core data in geophysics or tree rings in dendrochronology. Sediment cores might have undergone a number of coring disturbances such as compression or stretching. Moreover, cores from different sites with differing sedimentation rates would have different temporal resolutions. All these factors require a method of synchronizing or aligning the time scales.

Regarding the conventional RP (Eq. 4), a black main diagonal line (LOI) can always be found in the plot because of the identity of the (i,i) states. The RP can be considered as a special case of the CRP which usually does not have a main diagonal because the (i,i) states are not identical.

Assuming two identical trajectories, the CRP is the same as the RP of one trajectory and contains an LOI. If we slightly modify the values of the second trajectory, the LOI will become somewhat disrupted. This leads to the situation discussed in Subsec. 3.2. However, if we do not modify the amplitudes but stretch or compress slightly the second trajectory, the LOI will be kept continuous but not as a straight line with an angle of $\pi/4$. Rather this line can be bowed (Fig. 5). The local slope of lines in an RP as well as CRP corresponds to the transformation of the time axes of the two considered trajectories (Marwan et al., 2002a). 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.

For illustration let us consider two sine functions where we rescale the time axis of the second sine function in the following way

$$\sin(\varphi t) \longrightarrow \sin\big(\varphi t + a\sin(\psi t)\big). \tag{15}$$

The terms rescaling and synchronization are used here in the meaning of the rescaling of the time scale. The rescaling of the second sine function with different parameters a results in a deformation of the main diagonal (Fig. 5). The distorted line contains the information on the rescaling, which we constructively use in order to re-synchronize the two functions. Therefore, this distorted diagonal is called *line of synchronization (LOS)*.

In the following, we present a toy model in order to explain the relation between the time series $f(t_1), g(t_2)$ and the LOS $t_2 = \phi(t_1)$. In a one dimensional situation, the CRP is simply

$$\mathbf{CR}(t_1,t_2) = \Theta\big(\varepsilon - \|f(t_1) - g(t_2)\|\big). \tag{16}$$

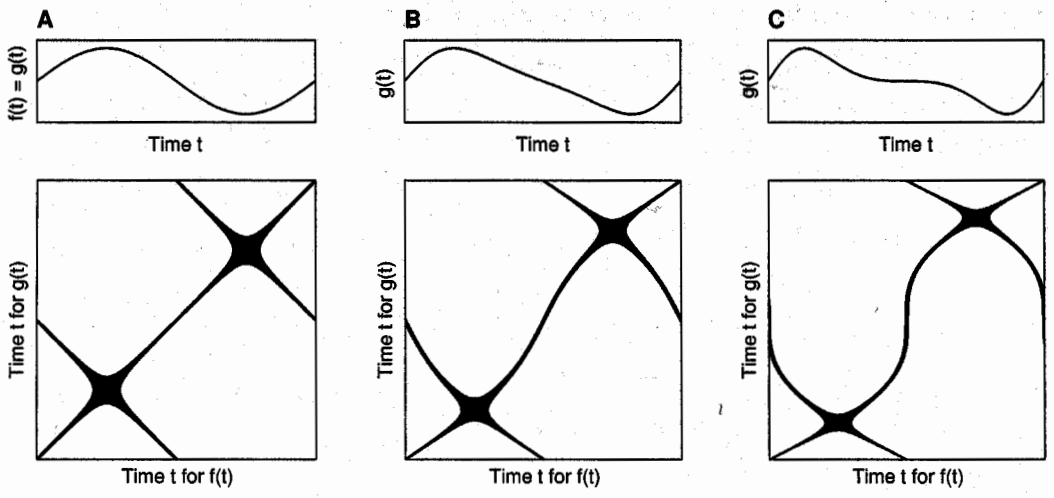

Figure 5: Cross recurrence plots of sine functions $f(t) = \sin(\varphi t)$ and $g(t) = \sin(\varphi t + a\sin(\psi t))$, whereas (A) $a = 0$, (B) $a = 0.5$ and (C) $a = 1$. The variation in the time domain leads to a deforming of the synchronization line. The CRPs are computed without embedding.

Provided that we set $\varepsilon = 0$ to simplify the condition, (Eq. 16) will deliver a recurrence point if

$$f(t_1) \equiv g(t_2). \tag{17}$$

In general, this is an implicit condition that links the variable t_1 to t_2. Considering the physical examples above, it can be assumed that the time series are essentially the same; i. e. $f \equiv g$ up to a rescaling function of time. So we can state

$$f(t_1) \equiv f\big(\phi(t_1)\big). \tag{18}$$

In some special cases (18) can be resolved with respect to t_1. An example of such a special case is a system of two sine functions with different frequencies

$$f(t) = \sin(\varphi \cdot t + \alpha), \quad g(t) = \sin(\psi \cdot t + \beta) \tag{19}$$

Using (17) and (18) we find

$$\sin(\varphi\, t_1 + \alpha) - \sin(\psi\, t_2 + \beta) = 0, \tag{20}$$

and one explicit solution of this equation is

$$\Rightarrow \quad t_2 = \phi(t_1) = \left(\frac{\varphi}{\psi} t_1 + \gamma\right) \tag{21}$$

with $\gamma = \frac{\alpha-\beta}{\psi}$. In this special case the slope m of the main line in the corresponding cross recurrence plot represents the frequency ratio, and the distance between the origin of the axes and the intersection of the LOS with the ordinate reveals the phase difference. Considering the time transformation functions $T_1 = \varphi \cdot t + \alpha$ and $T_2 = \psi \cdot t + \beta$ whithin the (Eqs. 19), we get the same result for the slope of the LOS by the ratio of the derivatives (cf. Eq. 5)

$$\mathrm{m} = \frac{\partial_t T_1(t)}{\partial_t T_2(t)} = \frac{\varphi}{\psi}. \tag{22}$$

The function $t_2 = \phi(t_1)$ is the transfer or rescaling function which allows to rescale the second system to the first system. If the rescaling function is not linear, the LOS will also be curved.

If the functions $f(\cdot)$ and $g(\cdot)$ are not identical, our method will in general not be capable of deciding whether the difference in the time series is due to different dynamics or to simple rescaling. So the assumption that the dynamics remain equal up to a rescaling in time (the underlying systems must be the same) is essential. Nevertheless, for some cases where $f \neq g$, the method can be applied in the same way. If we consider the functions $f(\cdot) = a \cdot \bar{f}(\cdot) + b$ and $g(\cdot) = \bar{g}(\cdot)$, whereby $f(\cdot) \neq g(\cdot)$ are the observations and $\bar{f}(\cdot) = \bar{g}(\cdot)$ are the states, normalization, with respect to the mean and the standard deviation, will allow to use our method,

$$f(\cdot) = a \cdot \bar{f}(\cdot) + b \longrightarrow \tilde{f}(\cdot) = \frac{f(\cdot) - \langle f(\cdot) \rangle}{\sigma(f(\cdot))} \tag{23}$$

$$\tilde{g}(\cdot) = \frac{g(\cdot) - \langle g(\cdot) \rangle}{\sigma(g(\cdot))}. \tag{24}$$

With $\bar{g}(\cdot) = \bar{f}(\cdot)$ the functions $\tilde{f}(\cdot)$ and $\tilde{g}(\cdot)$ are the same after the normalization, hence, our method can be applied without any further modification.

For application one has to determine the LOS – usually non-parametrically – and then rescale one of the time series by using this transfer function. This connection between the local slope of the LOS and the relation between the segments of the trajectories also applies to other line structures in CRPs as well as RPs.

This technique can also be used in order to find the closest matching segments in two data series. For example, in the geological framework there could be a long reference data series which has a time scale and a second but short profile with the same physical measurement. The task lies in finding the section in the reference data which matches to the second profile in order to yield the corresponding

time scale for the profile. This section can be found by looking for a more or less continuous black line in the CRP (the dislocated LOS).

The CRP based alignment of time series has conspicuous similarities with the method of sequence slotting described by Thompson and Clark (1989). The first step in their method is the calculation of a distance matrix ($\mathbf{D}^m_{i,j} = \|\vec{x}_i - \vec{x}_j\|$), which allows the use of multivariate data sets. Thompson and Clark (1989) referred to the distance measure as dissimilarity. It is used to determine the alignment function in such a way that the sum of the dissimilarities along a path in the distance matrix is minimized. This approach is based on dynamic programming methods which were mainly developed for speech pattern recognition in the 1970's (e. g. Sakoe and Chiba, 1978). In contrast, RPs were developed to visualize the phase space behaviour of dynamical systems. Therefore, a threshold was introduced to make recurrent states visible. The involvement of a FAN in the phase space and the possibility of increasing the embedding dimensions, which enhance the quality of the transfer function, distinguishes the CRP approach from the sequence slotting method.

4 Current Developments of Recurrence Plots

During the last five years a rather promising development of recurrence plots has been in progress. These new findings work toward a better understanding of the structures found in RPs. An RP can be used in order to obtain some properties of dynamical systems, such as the Rényi entropy, the correlation dimension or the information dimension (Faure and Korn, 1998; Gao, 1999; Thiel et al., 2003). The most recent development proposes intersections of RPs and time shifted RPs

$$\mathbf{R}^{m,\hat{\varepsilon}}_{i,j} \cdot \mathbf{R}^{m,\tilde{\varepsilon}}_{i+\tau,j+\tau} \tag{25}$$

which can be used for the estimation of the generalized mutual information (Thiel et al., 2003). Furthermore, this approach can also be applied to different phase space trajectories, which leads to a completely new concept of cross recurrence plots (Romano et al., 2003, this kind of cross recurrence plot is denoted as XRP). Based on this new approach, the cross mutual information and Rényi entropy can be estimated. In addition, the XRP can be used to study phase synchronization in complex systems (Rosenblum et al., 1996). The XRP can be applied to measurements of different systems whose observations cannot be considered as state variables of

the same system. XRPs are not restricted to only two systems; it is a multivariate analysis tool. In contrast to CRPs, XRPs can only be applied to time series of equal time scale, length and sample resolution.

The development of RP based methods is not yet concluded. The very last years in particular have shown that large potential lies in the analysis of RPs.

5 Applications

The high potential for the analysis based on recurrence or cross recurrence plots arises with their applicability. Hundreds of applications of recurrence plots and recurrence quantification analysis, especially to physiological data, represent the increasing importance of these methods. In this section selected applications of the newest strategies based on cross recurrence plots to geological data are presented. Methods of linear and nonlinear data analysis mostly fail in these applications because of the rather short length of the time series and their nonstationarity.

5.1 Similarities Found within Present and Past Climatic Data Series

Cross recurrence plots can be used for studying a similar time evolution of phase space trajectories and hence to assess the similarity or interrelation between the underlying processes. The following example illustrates the functioning of this technique on natural data from geology.

Data from geology are often characterized as short and nonstationary. The unique character of outcrops or drilling cores does not usually allow to repeat or refine a measurement. Therefore, data analysis of geological data is often confronted with problems regarding the length, nonstationarity or gaps in the data. In the previous application of CRPs we have seen that this method can be used for this kind of data. Therefore, in this subsection the application of CRPs will be used for the analysis of palaeo-climatology data that are of short length and nonstationarity (Marwan et al., 2003; Trauth et al., 2003).

Higher variability in rainfall and river discharge could be of major importance in landslide generation in the northwestern Argentine Andes. A potential cause of such variability is the El Niño/ Southern Oscillation (ENSO). Annual layered deposits of a landslide-dammed lake in the Santa Maria Basin (site El Paso, Province Salta, NW Argentina) with an age of 30 000 ^{14}C years provide an archive of precipitation variability during this time (Fig. 6). The annual cycle of wet summers

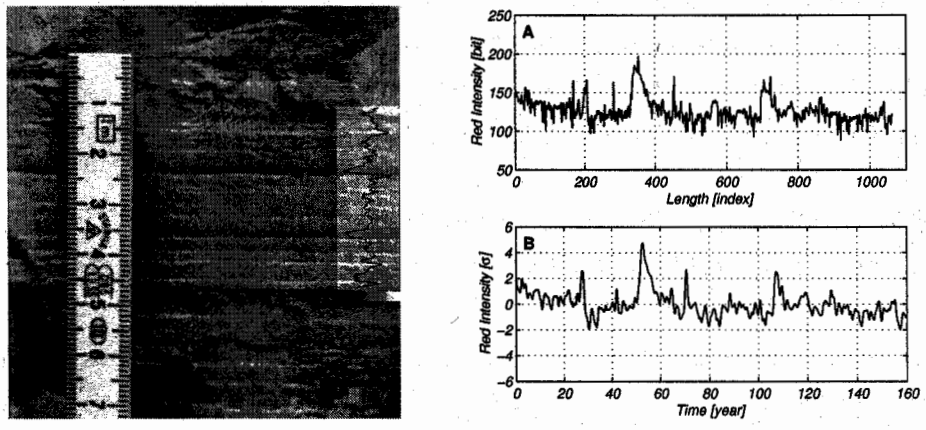

Figure 6: Left: Photograph of varved lake sediments from the El Paso site in the Santa Maria Basin with cyclic occurrence of dark red colourations recording more precipitation and sediment flux with ENSO-like periodicities (Trauth and Strecker, 1999). The overlayed curve shows a representative red colour intensity transect of the deposits. Right: Red intensity values of the lake sediments of site EP160 on (A) a length scale and on (B) a time scale and after smoothing and normalization; the unit of raw data is one bit, the unit of transformed and smoothed data is the standard deviation σ.

and dry winters caused significant changes in the lake's sedimentation. During the rainy season mainly ocher coloured silty sediments were deposited; during the subsequent dry season a thin white layer consisting of the skeletons of silica algae (diatoms) was deposited. Due to its white colour, the diatomaceous layer can be used to identify single years in these sediments. Recurring intense red colouration of the silty part of the annual layers comes from reworked older sediments which are eroded and deposited only during extreme rainfall events. Therefore, the intensity of red colour in the varved deposits can be interpreted as a proxy for precipitation variation in the Santa Maria Basin (Trauth and Strecker, 1999; Trauth et al., 2000). The more intense red colouration is evidence of more precipitation during the rainy season. The estimate of the power spectrum of the red colour intensity reveals significant peaks within the ENSO frequency band of two to four years, suggesting an ENSO-like influence (Trauth et al., 2000). Because of the nonstationarity of these data (the sedimentation process in a lake is not stationary, which results in nonstationary proxy variables for the in-lake processes) linear correlation analysis is unsuitable. Therefore, the CRP analysis is applied to these data.

Our research includes the quantification analysis of CRPs of an index data series of the ENSO (Southern Oscillation Index, SOI) and the modern as well as palaeo-precipitation data in order to compare the magnitude and causes of rainfall variability in the NW Argentine Andes today and during the time of enhanced landsliding around 30 000 ^{14}C years ago (Marwan et al., 2003; Trauth et al., 2003). For the assessment of the modern ENSO influence on local rainfall in NW Argentina, the monthly precipitation data from the three stations San Salvador de Jujuy (JUY), Salta (SAL) and San Miguel de Tucuman (TUC) are analyzed (Figs. 7, 8). These locations are influenced by different local winds (Fig. 7); Jujuy and Salta mainly receive northeasterly and easterly moisture-bearing winds during the summer rainy season, whereas Tucuman is characterized by southerly and south-westerly winds (Prohaska, 1976).

An appended surrogate test provides an evaluation of the results of the CRP analysis. The assumption for the surrogate data is that the considered processes are linearly independent and do not have any similar dynamics. These surrogates should reveal some features like in our original data but also features caused by the randomness of a possible correlation (stochastic processes). Linear correlated noise is a paradigmatic example for such processes (Kantz and Schreiber, 1997). We calculate a surrogate time series based on this class of processes with the following

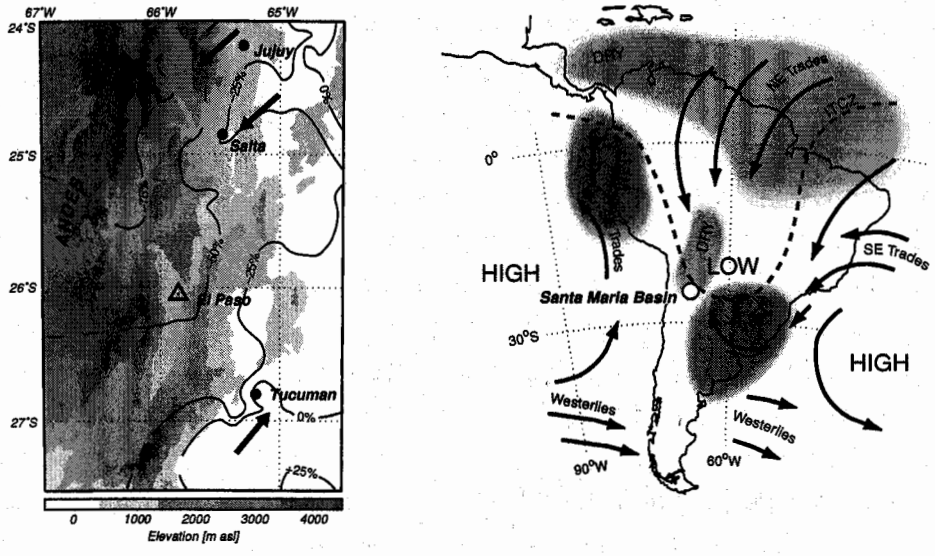

Figure 7: Left: Study area in the Santa Maria Basin with the locality of annual layered lake deposits from the El Paso site, the relative precipitation anomaly during the El Niño 1965/66 compared to mean annual precipitation (annual precipitation as a mean from July to June; data from Bianchi and Yañez, 1992) and the prevailing wind directions during January (black arrows; wind directions from Prohaska, 1976). Right: Present-day airflow pattern during the summer rainy season and principal areas of rainfall anomalies during El Niño events in South America (modified after Trauth et al., 2000).

recursive function, an autoregressive process of order p,

$$x_n = \sum_{i=1}^{p} a_p \, x_{n-p} + b \, \xi_n,$$

where ξ is white noise and a_i are coefficients which determine the auto-correlation of the system and allow to adapt this stochastic system to our natural processes. We fit the model to the precipitation series of the station Tucuman. Then we perform the CRP analysis using the SOI data and the ensemble of, e. g. 10,000 realizations of precipitation series produced by the AR model. Using the distributions of the RR and L measures obtained from these CRPs, we estimate their empirical confidence bounds (we will use the 2σ-bounds which approximately correspond with the 95% confidence level).

With these confidence bounds we evaluate the obtained measures of CRP and the relations of the natural processes. Since the surrogates are from a stationary

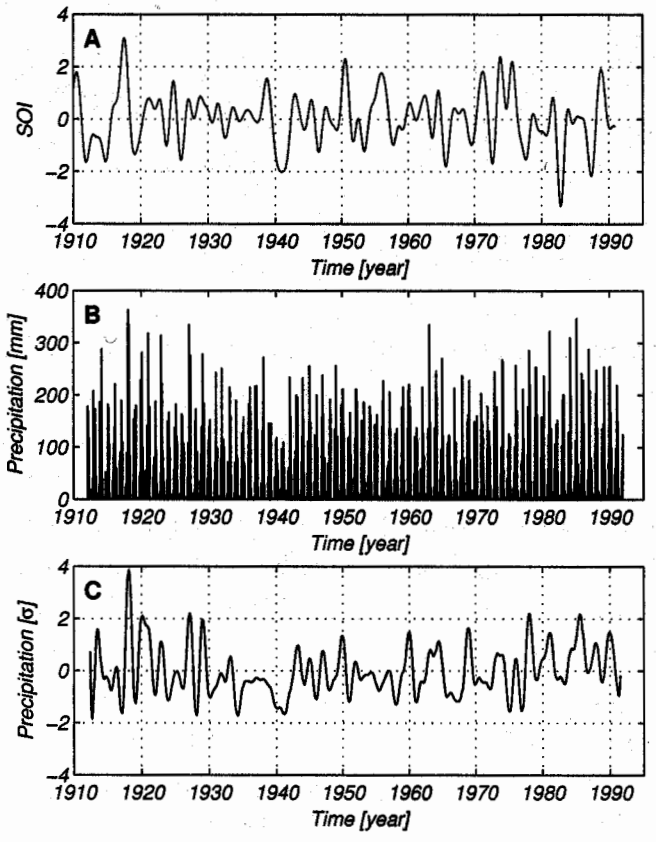

Figure 8: Smoothed and σ-normalized time series of the Southern Oscillation Index (A), monthly precipitation data of Salta (B) and its smoothed and σ-normalized time series (C). SOI based on COADS data from the NOAA Live Access Server (http://ferret.wrc.noaa.gov).

system and the natural data are nonstationary, we have further applied this kind of evaluation to more stationary segments in the natural data and got the same results. This kind of surrogates is a special realization, which is prototypical for linear stochastic processes. There are, of course, a lot of other possibilities to construct surrogates (cf. Kantz and Schreiber, 1997).

The CRPs of pairs of SOI and present-day precipitation as well as SOI and palaeo-precipitation show similar features (Fig. 9). These significant similarities indicate that the red colour intensity records from the varved lake sediment do reflect rainfall in NW Argentina. First, we discuss the CRP of Salta precipitation (data series SAL) vs. the Southern Oscillation Index (SOI) and the CRP of red colour intensity of varves (data series EP160) vs. SOI. The x-axis represents time along the phase space trajectory of the SOI, whereas the y-axis represents the time along the phase space trajectory of SAL or EP160, respectively. The CRP of SAL vs. SOI exhibits longer diagonal lines in two to four year intervals, which matches the same frequency band obtained by the power spectral analysis (Fig. 9). This indicates that some parts of the phase space trajectory of the SOI recur in epochs of the phase space trajectory of SAL after relocating by the time of two to four years. Vertical white bands in the CRP represent less frequent states in SOI, whereas horizontal white bands represent less frequent states in SAL. The latter occurs with intervals of more than ten years. The CRP between EP160 and SOI shows similar characteristics as the CRP described above (Fig. 9). Longer diagonal lines have spacings of about two to four years. White bands occur at time scales of more than ten years. Some linkages in both CRPs are obvious through visual inspection. Next, a quantitative analysis of the CRPs is performed in order to study statistically these relations and to allocate the predefined causality patterns to certain localities.

We find that the parameter RR^c_* of the CRPs between TUC and SOI has small negative values which do not exceed the 2σ-bounds and do not show preferences for a distinct lag. The parameter L^c_* also has small values, but it has rather small maxima and minima at delays of -1, 4 and 8 months. These results indicate that the precipitation in Tucuman is not strongly influenced by ENSO. If there is a weak influence, the rainfall will increase during El Niño (Fig. 10 A, E). However, the analysis of JUY and SOI reveals clear positive values around a lag of zero and negative values after 8 – 12 months, which suggests a significant link between Jujuy rainfall and ENSO (Fig. 10 C, G). The measures for the analysis of SAL versus SOI show smaller maxima for a delay of about zero and minima after a lag of 8

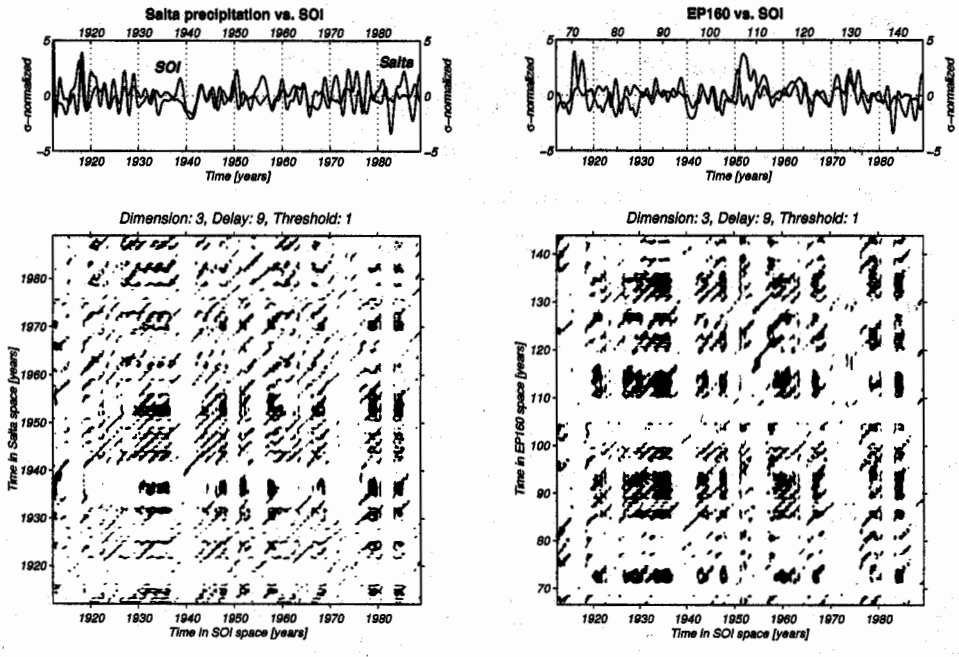

Figure 9: Left: Cross recurrence plot of SOI vs. precipitation data from the city of Salta (data shown at top). Black points represent the occurrence of similar states in both processes. Diagonal lines correspond with epochs of similar dynamics in both processes. The amount and length of these lines can be used as measures of the similarity of both processes. Right: Cross recurrence plot of SOI vs. the best matching section of palaeo-precipitation (EP160, data shown at top). The x-axis shows the time along the phase space trajectory of the SOI and the y-axis that of SAL and EP160, respectively.

– 12 months. Therefore, we infer a weaker link between Salta rainfall and ENSO (Fig. 10 B, F; the disrupted minima in the L_*^c parameter at around ten months is due to the short data length and a resulting nonstationarity in the CRP). The measures for both SAL and JUY exceed the 2σ-bounds.

The 30 000 ^{14}C year old precipitation data are not simply comparable with present-day data, because there is no information available about how to synchronize the rainfall records with modern climate indices. Therefore, we seek the time window in these data showing the highest coincidence in the dynamics using maximum values for RR_*^c and L_*^c as the key criterion. The linear correlation coefficients could be used to find such a sequence, but this results in numerous ambiguous possibilities.

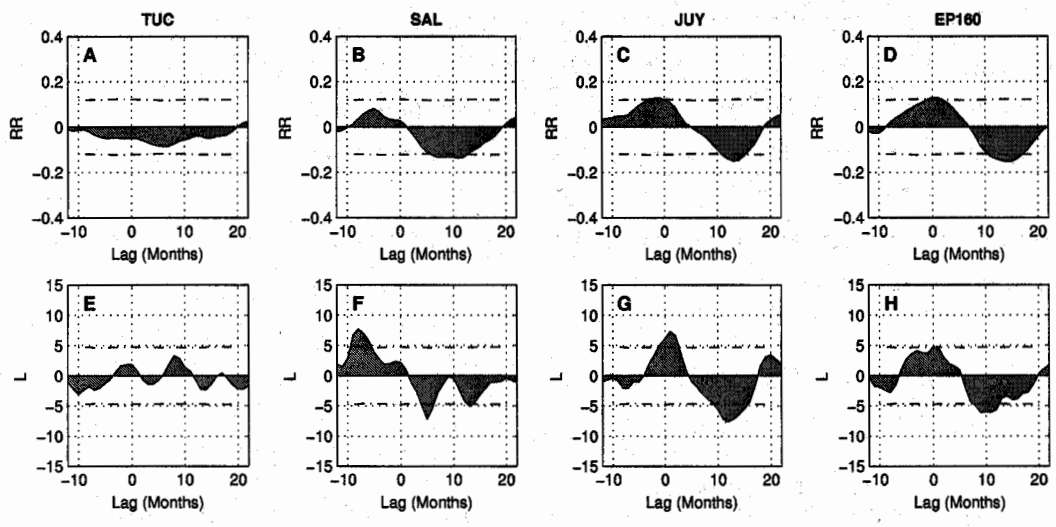

Figure 10: RR^c_* and L^c_* measures of the cross recurrence plots between SOI and precipitation in Tucuman (A, E), Salta (B, F), Jujuy (C, G) and palaeo-precipitation (D, H). Extreme values reveal high similarity between the dynamics of the rainfall and the ENSO. The dash-dotted lines are the empirical 2σ-bounds from the distributions of an ensemble of data based on a 5th-order AR-model.

The complexity measures based on CRP provide a differentiated search that also considers time based features of the signal. This method reveals indeed a clearer result. The measures presented herein are not the only measures used. For maintaining clarity, the further measures are not presented in this application, although they are used to find the sequence in the sediment data. Even though the observed coincidence is not very high, it yields the time section in the palaeo-precipitation record EP160 which can be in best accordance with modern data. In our palaeoclimate data EP160 we find such a section represented by maximum and minima values for RR^c_* and L^c_* for delays of about zero and ten months, similar to those found for JUY and SAL (Fig. 10D, H). The RR^c_* and L^c_* measures also exceed the 2σ-bounds.

The similarities between the time series of the present-day rainfall data and the palaeo-precipitation record from the lake sediments suggest that an ENSO-like oscillation was active around 30 000 ^{14}C years ago (roughly corresponding to 34 000 cal. years BP), which corresponds with the results of the investigation of Coccolithophores production (Beaufort et al., 2001). In the semiarid basins of the NW Argentine Andes, the ENSO-like variation could have caused significant fluctua-

tions in local rainfall around 30 000 ^{14}C years ago similar to modern conditions. Together with generally higher moisture levels, as indicated by lake balance modeling results (Bookhagen et al., 2001), this mechanism could help to explain more frequent landsliding approximately 34 000 years ago in the semiarid basins of the Central Andes. For the comparison of the past and modern climate conditions, the CRP analysis has been used because a linear correlation analysis would reveal ambiguous results.

5.2 Time Scale Alignment of Borehole Data

The CRP contains information about the time synchronization of data series (in the following the terms synchronization and rescaling refers to the alignment of the time scales). This is revealed by the distorted main diagonal, the LOS. A nonparametric rescaling function is provided by isolating this LOS from the CRP and can be used for the re-alignment of the time scales of the considered time series. We expect that this approach will open a wide range of applications, such as scale alignment and pattern recognition, e. g. in geology, molecular biology or ecology.

In this application we consider geophysical measurements of two sediment cores from the Makarov Basin, in the central Arctic Ocean, PS 2178-3 and PS 2180-2 (Marwan et al., 2002a) and compare the method of cross recurrence plot matching with the conventional method of visual wiggle matching (interactive adjustment). The task is to align the data of the PS 2178-3 core (data length $N = 436$) with the scale of the PS 2180-2 (data length $N = 251$) in order to get a depth-depth-function which that to synchronize both data sets (Fig. 11).

The phase space trajectories are formed by the following normalized six measures: low field magnetic susceptibility (κ_{lf}), anhysteretic remanent magnetization (ARM), ratio of anhysteretic susceptibility to κ_{lf} (κ_{ARM}/κ_{lf}), relative palaeointensity (PJA), median destructive field of ARM (MDF_{ARM}) and inclination (INC). Each measure is used as one component of the phase space vector. However, this embedding can be combined with the time delay method according to Takens (1981) in order to further increase the dimension of the phase-space.

Using an embedding of $m = 3$ (absolute dimension is $3 \times 6 = 18$), $\tau = 1$ and a recurrence criterion of FAN with $\varepsilon = 0.05$, the resulting CRP shows a clear LOS and some clustering of black patches (Fig. 12). Black patches arise whenever the variation in the data is smaller than the used vicinity threshold ε for a given time (plateau). The next step is to fit a nonparametric function (the desired depth-depth-

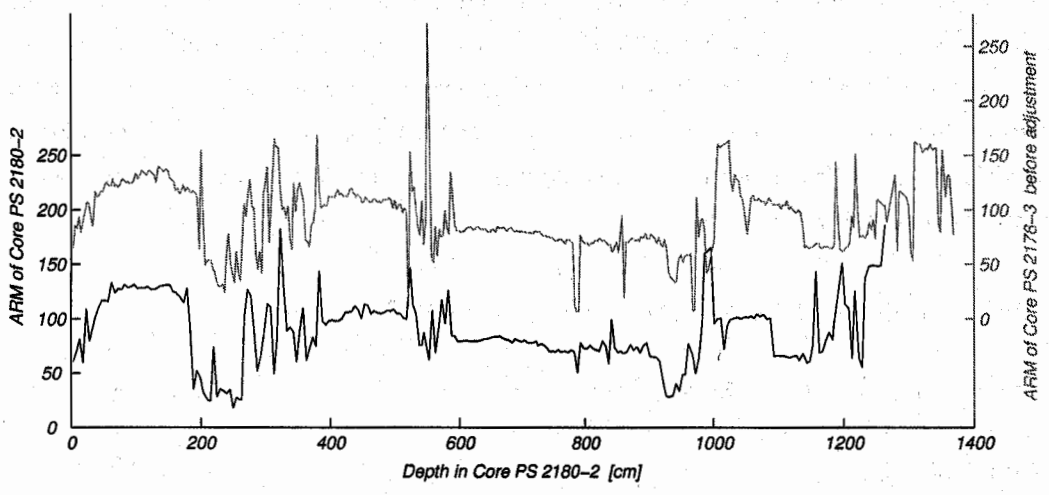

Figure 11: ARM data of the boreholes PS 2178-3 GPC and PS 2180-2 GPC in the Central Arctic Ocean before adjustment.

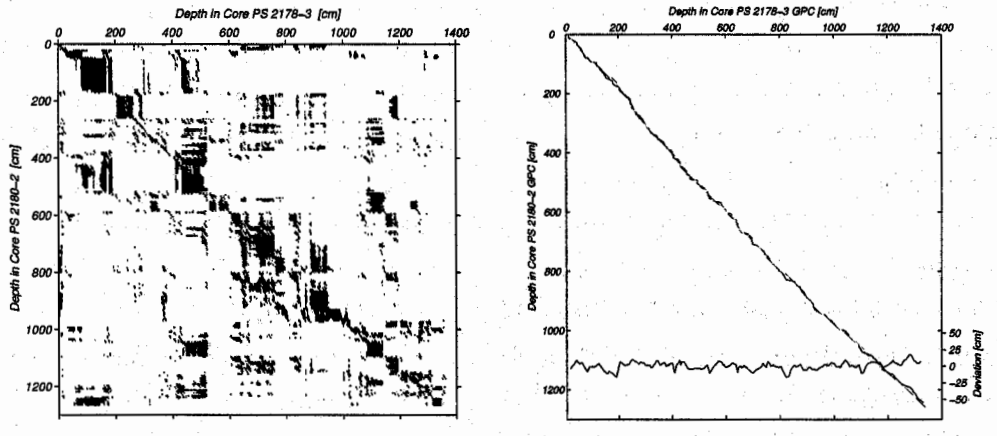

Figure 12: Left: Cross recurrence plot based on six normalized sediment parameters and an additional embedding dimension of $m = 3$ ($\tau = 1$, $\varepsilon = 0.05$) Right: Depth-depth-curves. In black the curve gained with the CRP, in gray the manually matching result. The horizontal curve shows the deviation between both results.

curve) to the LOS in this CRP (Fig. 12). Different approaches can be considered for this procedure. However, they could have to be chosen appropriately because they have a large effect on the quality of the found LOS. In our example a two step algorithm is chosen that is able to tend locally towards the direction of the centre of gravity of clustered black points. A detailed explanation is given in Marwan et al. (2002a). With so determined LOS we are able to align the scale of the PS 2178-3 core to that of PS 2180-2 (Fig. 13).

The determination of the depth-depth-function with the conventional method of visual wiggle matching is based on the interactive and parallel searching for the same structures in the different parameters of both data sets. If the adjustment does not work in a section of the one parameter, one can use another parameter for this section, which allows the multivariate adjustment of the data sets. The recognition of the same structures in the data sets requires a degree of experience. However, human eyes are usually better in the visual assessment of complex structures than a computational algorithm.

Our depth-depth-curve differs slightly from the curve which was gained by the visual wiggle matching (Fig. 12). However, despite our rather simple algorithm used to fit the non-parametric adjustment function to the LOS, we obtained a good result of adjusted data series. If they are well adjusted, the correlation coefficient between the parameters of the adjusted data and the reference data should not vary so much. The correlation coefficients between the reference and adjusted data series is about $0.70-0.80$, where the correlation coefficients of the interactive rescaled data varies from $0.71-0.87$ (Tab. 1). The χ^2 measure of the correlation coefficients emphasizes more variation for the wiggle matching than for the CRP rescaling. The results obtained with the CRP rescaling can be further improved by a more suitable algorithm for the search of the LOS.

The comparison of the CRP aligned geophysical measurements with the conventional visual matching (wiggle matching) shows an acceptable reliability level of the new method (Marwan et al., 2002a). The advantage is the automatic, objective and multivariate alignment. Moreover, further attempts exist to align geological data automatically. They either use parametrical approaches (minimal cost functions, Fourier series estimation of the mapping function and others; Martinson et al., 1982; Brüggemann, 1992) or they have to fit a large number of parameters and apply trial-and-error algorithms (sequence slotting; Thompson and Clark, 1989).

Table 1: Correlation coefficients ρ_1 (wiggle matching) and ρ_2 (CRP matching) between adjusted data and reference data and their χ^2 deviation. The correlation of the interactive adjusted data varies more than the automatic adjusted data. The data length is $N = 170$ (wiggle matching) and $N = 250$ (CRP matching). The difference between the both correlation coefficients ρ_1 and ρ_2 is significant at a 99 % significance level, when the test measure $\hat{z}$ is greater than $z_{0.01} = 2.576$.

Parameter	ρ_1	ρ_2	$\hat{z}$
ARM	0.8667	0.7846	6.032
MDF_{ARM}	0.8566	0.7902	4.791
κ_{LF}	0.7335	0.7826	2.661
κ_{ARM}/κ_{LF}	0.8141	0.8049	0.614
PJA	0.7142	0.6995	0.675
INC	0.7627	0.7966	1.990
χ^2	141.4	49.1	

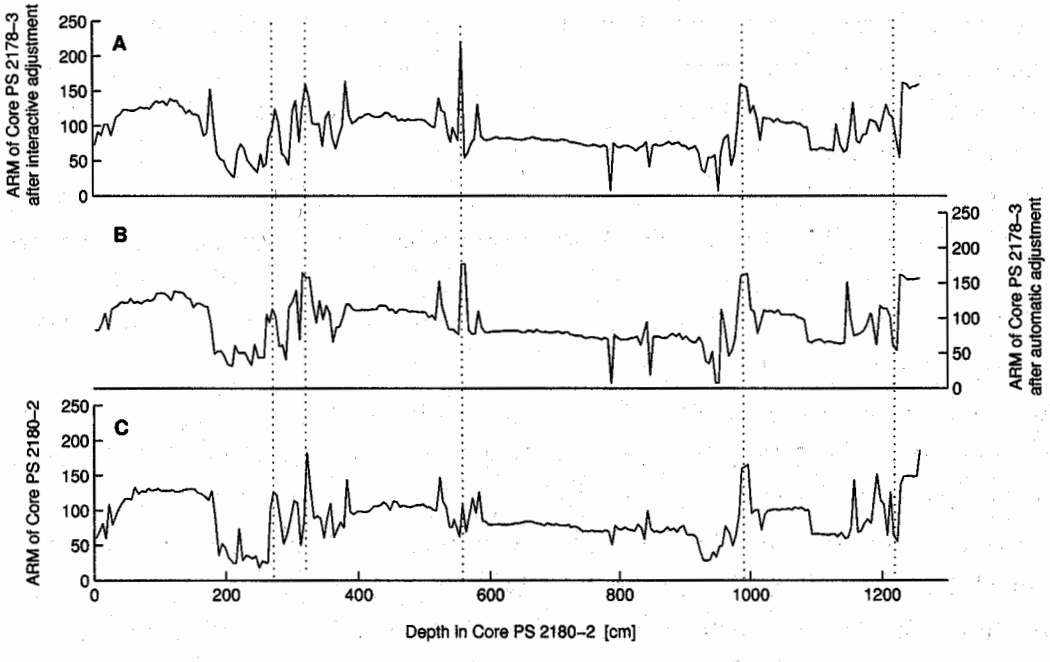

Figure 13: The ARM data are exemplary shown after alignment by wiggle matching (A) and by automatic alignment (B) using the LOS from Fig. 12. Plot (C) shows the reference data.

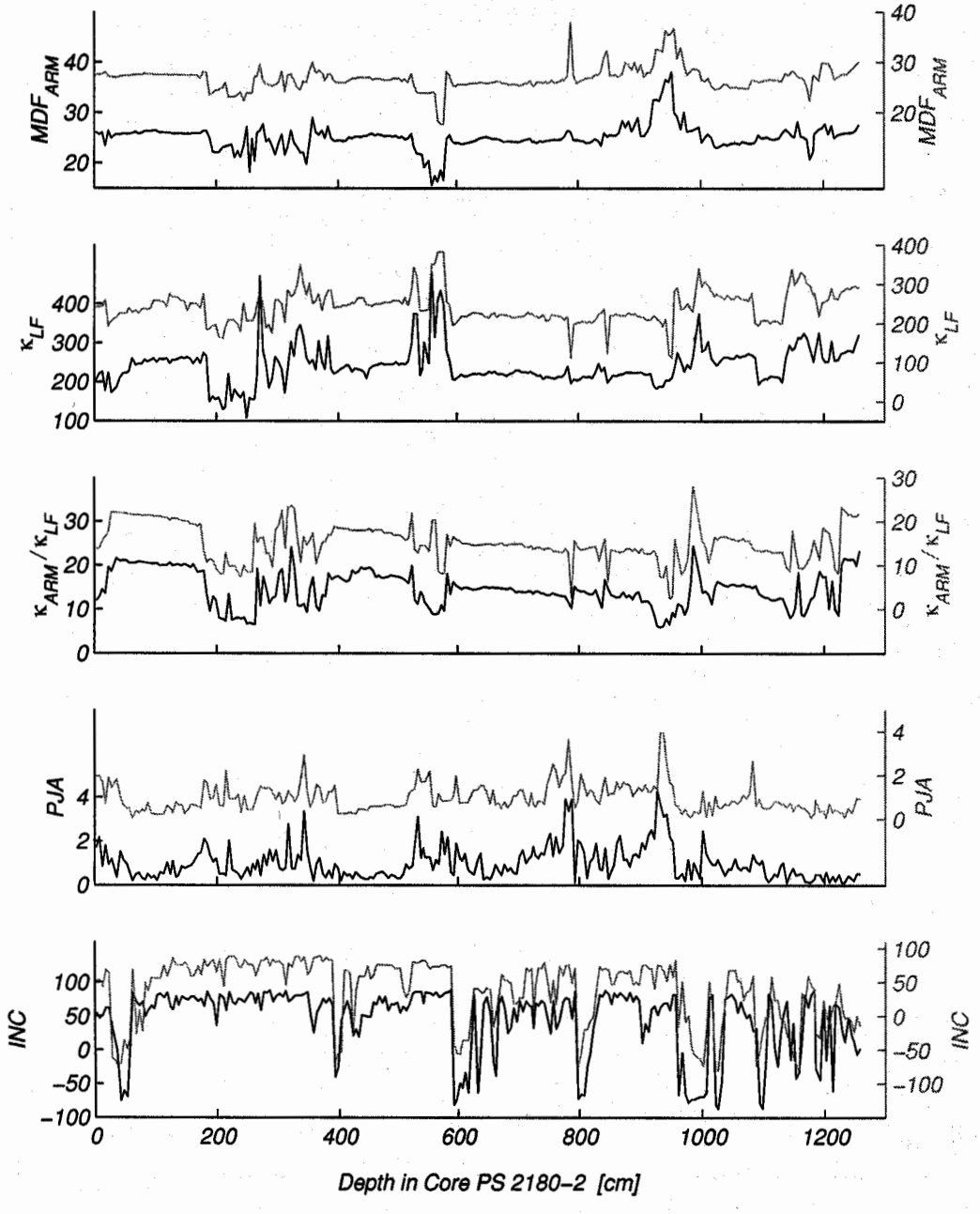

Figure 14: The adjusted marine sediment parameters. The construction of the CRP was done with the normalized parameters. In this plots we show the parameters, which are not normalized.

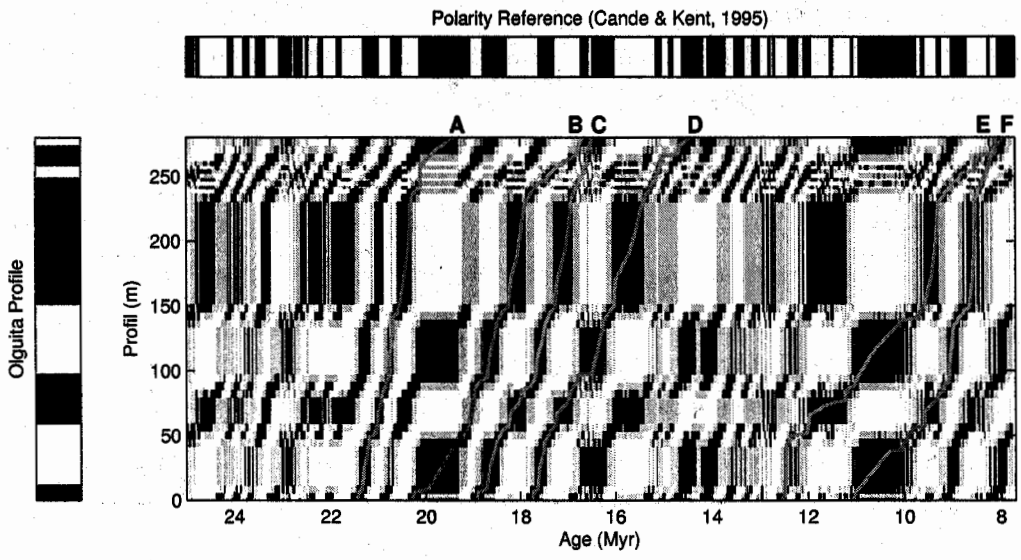

Figure 15: CRP between the polarity data of the Olguita profile and the reference data according to Cande and Kent (1995). In the polarity data the white colour marks a polarity of the Earth's magnetic field in the present, whereas the black colour marks a reversal. Six potential LOS are marked with gray lines (A–F, corresponding to the potential LOS given in Fig. 16).

5.3 Search for an Appropriate Sequence in a Geological Profile (Magnetostratigraphy)

In the following application of CRP the possibility of finding an appropriate sequence in a given data series relating to a reference series (and vice versa, respectively) is presented. For this task the LOS in the CRP must be found. Although we present only applications in geology here, the methods based on CRPs can also be applied to data from other scientific field. The search of sequences in reference data based on CRPs has been also successfully applied in gene sequencing and speech recognition (Marwan, 2003).

From a sediment profile (Olguita profile, Patagonia, Argentina; Warkus, 2002) a measurement of the palaeo-polarity of the Earth's magnetic field (along with other measurements) is available. The starting point for any geological investigation of such a profile is determining the time at which these sediments were deposited. By

applying the magnetostratigraphic approach and a geomagnetic polarity reference with known time scale, the polarity measurements can be used to determine a possible time scale for the profile. Cande and Kent (1995) provide such a geomagnetic polarity reference, which covers the last 83 Myr. The Olguita profile contains seven reversals. The polarity data consist of the values one, for the polarity direction as today, and the values zero, for the inverse polarity. Unfortunately, this data series is too short (only 16 measurements) for a credible analysis. Nevertheless, for our purpose of demonstration we will enlarge this data by interpolation. The Olguita profile is transformed to an equidistant scale of 300 data points and the reference data is transformed to an equidistant scale of 1 200 data points.

A CRP is created from these two data series by using an embedding dimension $m = 4$, a delay of $\tau = 6$ and a neighbourhood criterion of FAN (30% recurrence rate). Varying degrees of continuous lines between 21 and 16 Myr BP and between 12 and 8 Myr BP occur in the CRP, which can be interpreted as the desired LOS (Fig. 15). We will analyse six of these possibilities to estimate the LOS. The search for the potential LOS is conducted using the same algorithm described in Marwan et al. (2002a). Moreover, we can evaluate the quality of these potential LOS by introducing a quality factor that takes into consideration the amount of gaps $N_\circ$ and black dots $N_\bullet$ on this line

$$Q = \frac{N_\bullet}{N_\bullet + N_\circ} 100\%. \tag{26}$$

A larger Q is a better LOS; $Q = 100\%$ stands for an absolute continuous line. Moreover, the obtained LOS can be interpreted as the sedimentation rate (Fig. 16). Abrupt changes in the sedimentation rate are not expected, thus, the potential LOS should not change abruptly. As an evaluator for this criterion we can use the averaged second derivative with respect to the time $\langle \partial_t^2 \rangle$.

The potential LOS differs slightly in the Q factor, but more in the occurrence of abrupt changes in their slope (Fig. 16 and Tab. 2). The LOS in Fig. 16C has the smallest $\langle \partial_t^2 \rangle$ and could be, therefore, a good LOS for the dating of the Olguita profile. Regarding this result, the Olguita profile would have an age between 16.5 and 18.9 Myr and an age-depth-relation as it is represented by the possibility of a LOS in Fig. 16C. Warkus' investigation reveals the same result (Warkus, 2002), although he also mentioned that the dating based on the polarity data is ambiguous.

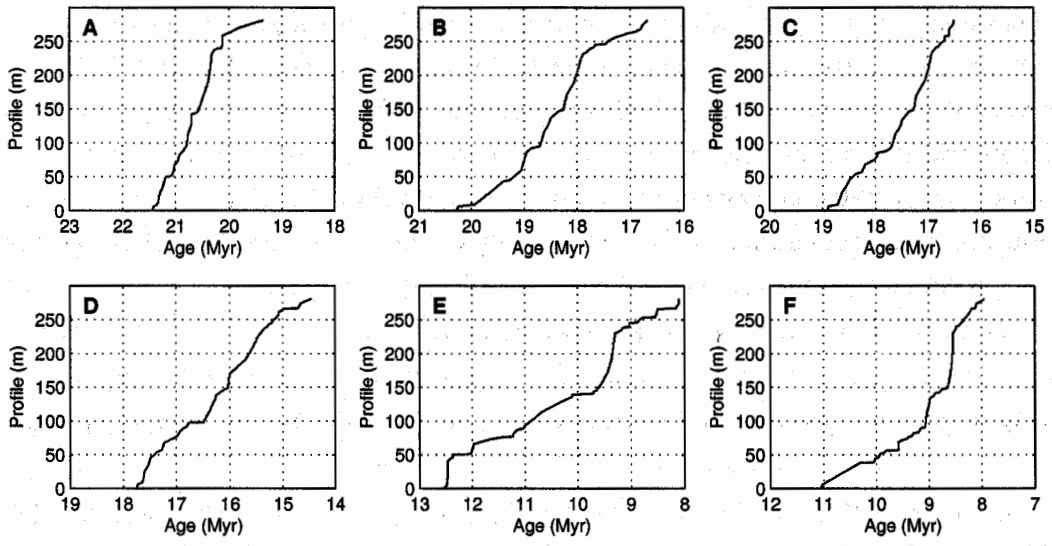

Figure 16: Potential LOS of the CRP presented in Fig. 15. They correspond to the potential sedimentation rates of the Olguita profile and mark sequences in the polarity reference, which match with the Olguita profile. Due to this matching, the Olguita profile can be dated.

Table 2: Possible ages of the Olguita profile, which are based on the found potential LOS (Fig. 16) and characteristics of these potential LOS.

Plot	Age (Myr)	$N_\bullet$	$N_\circ$	Q (%)	$\langle \partial_t^2 \rangle$
A	19.4–21.4	345	23	93.8	5.5
B	16.7–20.3	407	43	90.4	12.5
C	16.5–18.9	351	15	95.9	5.0
D	14.4–17.8	392	16	96.1	20
E	8.1–12.6	482	16	96.8	23
F	7.9–11.1	399	13	96.8	18

6 Conclusion

The extension of the concept of recurrence plots to a test for interrelations between two different phase space trajectories leads to cross recurrence plots (CRPs). From the point of view of a CRP, an RP can be considered as a special case of a CRP for two identical processes. However, if these two processes become progressively different, typical structures of the RP, like the main diagonal (*line of identity, LOI*), will dissolve. A quantification of these structures has been used in order to assess the variation or similarity between the dynamics of both processes.

The CRP analysis provides another useful application. The orientation of the line structures in the CRP is related to the time relation between the corresponding segments of the phase space trajectories. In the case of two sufficiently similar processes with different time dilatations, the CRP shows a bowed line of identity which is called *line of synchronization (LOS)*. This line corresponds to the transfer function between the time scales of the considered time series. A nonparametrical function fitted to LOS can be used in order to align the two processes to the same time scale.

Considering two processes, where epochs of the second are partly contained in the first, the CRP facilitates finding the location of these epochs in the first process.

Applications to real data from geology have revealed the applicability of the methods based on CRPs. Applying CRPs we were able to investigate palaeoclimate conditions, to align geophysical data from different boreholes to the same time/ depth scale and to find possible time-scales of a geological profile by using a reference data set.

A Matlab® toolbox for the application of recurrence plots as well as cross recurrence plots is available through the WorldWideWeb. An online and printable manual with illustrative examples as well as an extensive bibliography of applications of RPs and CRPs can also be found there. The current address is `http://www.agnld.uni-potsdam.de/~marwan/toolbox`.

Acknowledgment

We are grateful to Marco Thiel, Carmen M. Romano and Udo Schwarz for new ideas, fruitful discussions and suggestions. We also acknowledge Martin Trauth

(Institute of Geosciences, University of Potsdam), Norbert Nowaczyk (GeoForschungsZentrum Potsdam) and Frank Warkus (Institute of Geosciences, University of Potsdam) for cooperation and providing data. Some part of this research was supported by the Priority Programme 1097, *Geomagnetic Variations, Spatio-temporal structure, processes and effects on system Earth* of the Deutsche Forschungsgemeinschaft as well by the ESA-Project MAP AO-99-030, *2D and 3D Quantification of Bone Structure and its Changes in Microgravity Condition by Measures of Complexity.*

References

ABARBANEL, H. D. I., BROWN, R., SIDOROWICH, J. J., TSIMRING, L. S., The analysis of observed chaotic data in physical systems. Review of Modern Physics 65 (4), 1993, 1331–1392.
DOI: 10.1103/RevModPhys.65.1331

BEAUFORT, L., DE GARIDEL-THORON, T., MIX, A. C., PISIAS, N. G., ENSO-like Forcing on Oceanic Primary Production During the Late Pleistocene. Science 293, 2001, 2440–2444.

BERGER, A., LOUTRE, M., Insolation values for the climate of the last 10 million years. Quaternary Science Reviews 10, 1991, 297–317.

BIANCHI, A. R., YAÑEZ, C. E., Las precipitaciones en el noroeste Argentino. Instituto Nacional de Tecnologia Agropecuaria, Estacion Experimental Agropecuaria Salta, 1992.

BOOKHAGEN, B., HASELTON, K., TRAUTH, M. H., Hydrological modelling of a Pleistocene landslide-dammed lake in the Santa Maria Basin, NW Argentina. Palaeogeography, Palaeoclimatology, Palaeoecology 169, 2001, 113–127.
DOI: 10.1016/S0031-0182(01)00221-8

BRÜGGEMANN, W., A Minimal Cost Function Method For Optimizing the Age-Depth Relation of Deep-Sea Sediment Cores. Paleoceanography 7 (4), 1992, 467–487.

CANDE, S. C., KENT, D. V., Revised calibration of the geomagnetic polarity timescale for the late Cretaceous and Cenozoic. Journal of Geophysical Research 100, 1995, 6093–6095.

CAO, L., Practical method for determining the minimum embedding dimension of a scalar time series. Physica D 110 (1–2), 1997, 43–50.
DOI: 10.1016/S0167-2789(97)00118-8

CHOI, J. M., BAE, B. H., KIM, S. Y., Divergence in perpendicular recurrence plot; quantification of dynamical divergence from short chaotic time series. Physics Letters A 263 (4–6), 1999, 299–306.
DOI: 10.1016/S0375-9601(99)00751-3

ECKMANN, J.-P., KAMPHORST, S. O., RUELLE, D., Recurrence Plots of Dynamical Systems. Europhysics Letters 5, 1987, 973–977.

ECKMANN, J.-P., RUELLE, D., Ergodic theory of chaos and strange attractors. Review of Modern Physics 57 (3), 1985, 617–656.
DOI: 10.1103/RevModPhys.57.617

FAURE, P., KORN, H., A new method to estimate the Kolmogorov entropy from recurrence plots: its application to neuronal signals. Physica D 122 (1–4), 1998, 265–279.
DOI: 10.1016/S0167-2789(98)00177-8

GAO, J. B., Recurrence Time Statistics for Chaotic Systems and Their Applications. Physical Review Letters 83 (16), 1999, 3178–3181.
DOI: 10.1103/PhysRevLett.83.3178

GAO, J. B., CAI, H. Q., On the structures and quantification of recurrence plots. Physics Letters A 270 (1–2), 2000, 75–87.
DOI: 10.1016/S0375-9601(00)00304-2

KANTZ, H., SCHREIBER, T., Nonlinear Time Series Analysis. University Press, Cambridge, 1997.

KURTHS, J., HERZEL, H., An attractor in a solar time series. Physica D 25, 1987, 165–172.
DOI: 10.1016/0167-2789(87)90099-6

LORENZ, E. N., Deterministic Nonperiodic Flow. Journal of the Atmospheric Sciences 20, 1963, 120–141.

MANDELBROT, B. B., The fractal geometry of nature. Freeman, San Francisco, 1982.

MARTINSON, D. G., MENKE, W., STOFFA, P., An Inverse Approach to Signal Correlation. Journal of Geophysical Research 87, 1982, 4807–4818.

MARWAN, N., Untersuchung der Klimavariabilität in NW Argentinien mit Hilfe der quantitativen Analyse von Recurrence Plots. Master's thesis, Dresden University of Technology, Dresden University of Technology, October 1999.

MARWAN, N., Encounters With Neighbours – Current Developments Of Concepts Based On Recurrence Plots And Their Applications. Ph.D. thesis, University of Potsdam, 2003.

MARWAN, N., KURTHS, J., Nonlinear analysis of bivariate data with cross recurrence plots. Physics Letters A 302 (5–6), 2002, 299–307.
DOI: 10.1016/S0375-9601(02)01170-2

MARWAN, N., THIEL, M., NOWACZYK, N. R., Cross Recurrence Plot Based Synchronization of Time Series. Nonlinear Processes in Geophysics 9 (3/4), 2002a, 325–331.

MARWAN, N., TRAUTH, M. H., VUILLE, M., KURTHS, J., Comparing modern and Pleistocene ENSO-like influences in NW Argentina using nonlinear time series analysis methods. Climate Dynamics (in press), 2003.
DOI: 10.1007/s00382-003-0335-3

MARWAN, N., WESSEL, N., KURTHS, J., Recurrence Plot Based Measures of Complexity and its Application to Heart Rate Variability Data. Physical Review E 66 (2), 2002b, 026702.
DOI: 10.1103/PhysRevE.66.026702

MCGUIRE, G., AZAR, N. B., SHELHAMER, M., Recurrence matrices and the preservation of dynamical properties. Physics Letters A 237 (1–2), 1997, 43–47.
DOI: 10.1016/S0375-9601(97)00697-X

OTT, E., Chaos in Dynamical Systems. University Press, Cambridge, 1993.

PACKARD, N. H., CRUTCHFIELD, J. P., FARMER, J. D., SHAW, R. S., Geometry from a Time Series. Physical Review Letters 45 (9), 1980, 712–716.
DOI: 10.1103/PhysRevLett.45.712

PROHASKA, F. J., The climate of Argentina, Paraguay and Urugay. Vol. 12 of World Survey of Climatlogy. Elsevier, Amsterdam, Oxford, New York, 1976. pp. 13–73.

ROMANO, M., THIEL, M., KURTHS, J., A new definition of Cross Recurrence Plots. Physics Letters A (subm.), 2003.

ROSENBLUM, M. G., PIKOVSKY, A. S., KURTHS, J., Phase Synchronization of Chaotic Oscillators. Physical Review Letters 76, 1996, 1804.
DOI: 10.1103/PhysRevLett.76.1804

SAKOE, H., CHIBA, S., Dynamic programming algorithm optimization for spoken word recognition. IEEE Trans. Acoustics, Speech and Signal Proc. 26, 1978, 43–49.

TAKENS, F., Detecting Strange Attractors in Turbulence. Vol. 898 of Lecture Notes in Mathematics. Springer, Berlin, 1981. pp. 366–381.

THIEL, M., ROMANO, M. C., KURTHS, J., Analytical description of Recurrence Plots of white noise and chaotic processes. Chaos (subm.), 2003.

THOMPSON, R., CLARK, R. M., Sequence slotting for stratigraphic correlation between cores: theory and practice. Journal of Paleolimnology 2, 1989, 173–184.

TRAUTH, M. H., ALONSO, R. A., HASELTON, K., HERMANNS, R., STRECKER, M. R., Climate change and mass movements in the northwest Argentine Andes. Earth and Planetary Science Letters 179 (2), 2000, 243–256.
DOI: 10.1016/S0012-821X(00)00127-8

TRAUTH, M. H., BOOKHAGEN, B., MARWAN, N., STRECKER, M. R., Multiple landslide clusters record Quaternary climate changes in the northwestern Argentine Andes. Palaeogeography Palaeoclimatology Palaeoecology 194 (1–3), 2003, 109–121.
DOI: 10.1016/S0031-0182(03)00273-6

TRAUTH, M. H., STRECKER, M. R., Formation of landslide-dammed lakes during a wet period between 40,000 and 25,000 yr B.P. in northwestern Argentina. Palaeogeography, Palaeoclimatology, Palaeoecology 153 (1–4), 1999, 277–287.
DOI: 10.1016/S0031-0182(99)00078-4

TRULLA, L. L., GIULIANI, A., ZBILUT, J. P., WEBBER JR., C. L., Recurrence quantification analysis of the logistic equation with transients. Physics Letters A 223 (4), 1996, 255–260.
DOI: 10.1016/S0375-9601(96)00741-4

WARKUS, F., Die neogene Hebungsgeschichte der Patagonischen Anden im Kontext der Subduktion eines aktiven Spreizungszentrums. Ph.D. thesis, University of Potsdam, Institute of Earth Sciences, 2002.
URL: urn:nbn:de:kobv:517-0000555

WEBBER JR., C. L., ZBILUT, J. P., Dynamical assessment of physiological systems and states using recurrence plot strategies. Journal of Applied Physiology 76 (2), 1994, 965–973.

WOLF, A., SWIFT, J. B., SWINNEY, H. L., VASTANO, J. A., Determining Lyapunov Exponents from a Time Series. Physica D 16 (3), 1985, 285–317.
DOI: 10.1016/0167-2789(85)90011-9

ZBILUT, J. P., GIULIANI, A., WEBBER JR., C. L., Detecting deterministic signals in exceptionally noisy environments using cross-recurrence quantification. Physics Letters A 246 (1–2), 1998, 122–128.
DOI: 10.1016/S0375-9601(98)00457-5

ZBILUT, J. P., WEBBER JR., C. L., Embeddings and delays as derived from quantification of recurrence plots. Physics Letters A 171 (3–4), 1992, 199–203.
DOI: 10.1016/0375-9601(92)90426-M

Errata

Norbert Marwan and Jürgen Kurths:
Cross Recurrence Plots And Their Applications
Mathematical Physics Research at the Cutting Edge, Nova Schience Publishers 2004, pp. 101–139

page 107, paragraph and equations (5) and (6)

In a more general sense the line structures in an RP exhibit locally the time relationship between the current trajectory segments. A line structure in an RP of length l corresponds to the closeness of the segment $f(T_1(t))$ to another segment $f(T_2(t))$, where $T_1(t)$ and $T_2(t)$ are the local time scales (or transformations of an imaginary absolute time scale t) which preserve that $f(T_1(t)) \approx f(T_2(t))$ for some time $t = 1 \dots l$. Under some assumptions (e. g. piecewise existence of an inverse of the transformation $T(t)$) the local slope $\mathfrak{m}(t)$ of a line in an RP represents the local time derivative of the inverse second time scale $T_2^{-1}(t)$ applied to the first time scale $T_2(t)$

$$\mathfrak{m}(t) = \partial_t T_2^{-1}\left(T_1(t)\right). \tag{5}$$

We will consider here an illustrative example. A further explanation of the relationship between the slope of the lines and the trajectories is given in the Subsec. 3.3. Let us consider a function $f(T) = T(t)$ with a section of a monotonical, linear increase $T_{lin} = t$ and another (hyperbolic) section which follows $T_{hyp} = -\sqrt{r^2 - t^2}$ (Fig. 3A). Since the inverse of the hyperbolic section is $T_{hyp}^{-1} = \sqrt{r^2 - t^2}$, the derivative

$$\mathfrak{m} = \partial_t T_{lin}^{-1}\left(T_{hyp}(t)\right) = \frac{t}{\sqrt{r^2 - t^2}} \tag{6}$$

corresponds to the derivative of a circle line with a radius r, a bowed line structure with the form of a circle occurs in the RP (Fig. 3C).

page 116, paragraph and equation (22)

... In this special case the slope $\mathfrak{m}$ of the main line in the corresponding cross recurrence plot represents the frequency ratio, and the distance between the origin of the axes and the intersection of the LOS with the ordinate reveals the phase difference. Considering the time transformation functions $T_1 = \varphi \cdot t + \alpha$ and $T_2 = \psi \cdot t + \beta$ whithin the (Eqs. 19) and the inverse $T_2^{-1} = \frac{t-\beta}{\psi}$, we get the same result for the slope of the LOS by using the derivative (cf. Eq. 5)

$$\mathfrak{m} = \partial_t T_2^{-1}\left(T_2(t)\right) = \frac{\varphi}{\psi}. \tag{22}$$