Simpler methods do it better: Success of Recurrence Quantification Analysis as a general purpose data analysis tool

Charles L. Webber Jr. [a,*], Norbert Marwan [b], Angelo Facchini [c], Alessandro Giuliani [d]

[a] *Department of Cell and Molecular Physiology, Loyola University Medical Center, Maywood, IL, USA*
[b] *Potsdam Institute for Climate Impact Research (PIK), 14412 Potsdam, Germany*
[c] *Center the Study of Complex Systmes and Department of Information Enginering, University of Siena, 53100 Siena, Italy*
[d] *Environment and Health Department, Istituto Superiore di Sanità, Roma, Italy*

A R T I C L E   I N F O

A B S T R A C T

Over the last decade, Recurrence Quantification Analysis (RQA) has become a new standard tool in the toolbox of nonlinear methodologies. In this Letter we trace the history and utility of this powerful tool and cite some common applications. RQA continues to wend its way into numerous and diverse fields of study.

Mathematical tools adopted in science can be partitioned into two general classes. One class is a collection of methods stemming from some specific theory or tailored around a given phenomenon. Examples include renormalization group theory in statistical physics and *S* matrix formalism in quantum mechanics. Alternately, another class combines under a single rubric all those methods designated as statistical. Such methods are independent of any specific application, but rather refer to very basic epistemological concepts like the mean (barycentre) of a distribution or the distance between two points. Examples in this class include Pearson correlation coefficient and the Fisher's exact test.

In some cases, rare to be sure, methods rooted in the first class (specific applications) can push into the second class (general purpose). Such methodologies are incredibly valuable in science because, beyond mere hypothesis testing submitted to statistical inquiries, these strategies provide an underlying framework or theoretical modeling applicable to the particular system under study in the given field of interest. Probably the most famous of these "interface" methods is principal component analysis (PCA) for the simple reason that it diffused into numerous different field of science. The invasions came not simply as some "statistical" onslaughts, but the infiltration was so complete that different disciplines claimed PCA as "field specific" giving rise to multiple aliases: Singular Value Decomposition, Singular Spectrum Analysis,

or Essential Dynamics [1]. This diffusive spreading of PCA spanned centuries and was possible by the essential simplicity of the core concepts the technique. Such simplicity imparted sufficient flexibility in application that various quantitative needs in fields as diverse as meteorology and molecular biology were adequately met.

At the close of the twentieth century, another "interface" technique was born, Recurrence Quantification Analysis (RQA). RQA is grounded in theory, but possesses statistical utility in dissecting and diagnosing nonlinear dynamic systems across multiple fields of science. Like its predecessor, PCA, RQA has the potential of becoming endemic to all scientific fields [2].

RQA was first introduced as a purely qualitative technique, the Recurrence Plot, by Eckmann, Kamphorst and Ruelle [3]. The interest of these mathematicians was mainly descriptive as they discovered a way to visualize dynamical systems as recurrences of phase space trajectories in higher dimensions. A state recurs if it has neighbors in the phase space. There are several definitions of neighborhoods possible [4]; the most frequent definition considers the spatial distance between two state vectors $(\vec{x}_i - \vec{x}_j)$ in phase space and tests whether this distance is smaller than some threshold or radius (ε) according to the fundamental formula for the recurrence matrix. Members of the recurrence set include only those vector pairs that fall within the specified radius. Plotting the intersections of the i and j indices of all vector pairs within the set generates the two-dimensional recurrence plot.

$$\Re_{i,j} = \Theta\left(\varepsilon - |\vec{x}_i - \vec{x}_j|\right). \tag{1}$$

* Corresponding author.
E-mail addresses: cwebber@lumc.edu (C.L. Webber), marwan@pik-potsdam.de (N. Marwan), a.facchini@unisi.it (A. Facchini), alessandro.giuliani@iss.it (A. Giuliani).

The implicit mathematical foundation of the recurrence plot (RP) is deceptively simple and is no more than the Pythagorean theorem applied to rows of the embedding matrix of a specific time series. Such plots give very vivid and impressive images of dynamical system trajectories. For example, different dynamics exhibits typical patterns in the RP which characterize both global and local patterns (Fig. 1). For the Lorenz system of equation (below), recurrences are computed on the first variable (y_1) which serves as the surrogate for the other two variables (y_2 and y_3) by embedding in three space ($m = 3$).

$$dy_1 = 10 \bullet (y_1 - y_2), \tag{2}$$

$$dy_2 = y_1 \bullet y_3 + 28 \bullet y_1 - y_2, \tag{3}$$

$$dy_3 = y_1 \bullet y_2 - 8 \bullet y_3/3. \tag{4}$$

The global (large-scale) appearance of a RP can give hints on stationarity and regularity. Eckmann et al. [3] immediately noted that the local (small-scale) patterns are related to important dynamical properties, like determinism and divergence. They also stated that the lengths of the diagonal line structures in the RP are related to the positive Lyapunov exponent. Moreover, the study of recurrences of a state allows for a point-wise comparison of states, whereas other comparison techniques (e.g. correlation, mutual information) require sequences of data points.

The *quantum leap* that allowed the recurrence plots to extend beyond their initial theoretical realm and to *invade* the other fields of science occurred when RPs were rendered quantitative by Joseph Zbilut and Charles Webber [5,6]. Interestingly, this quantitative jump occurred not in the field of mathematics or physics, but in the much less quantitative field of physiology. Thus, recurrence quantifications were born out of the life sciences, but the underlying recurrence plots were conceived in the womb of mathematical physics.

It is a matter of fact that physicists commonly put mathematics at the very beginning of a scientific enterprise. That is, the initial theory is posited prior to the execution of the experiment which is run to either confirm or falsify the theory. Any *a posteriori* data analysis is normally considered much less a mathematical affair, something that is normally left to standard techniques or purely graphical evidence. But such is not the case with physiology (and other biomedical sciences) where experimental data rule the day and drive biological theories. And indeed, these theories can be mathematical or non-mathematical in nature. The challenge is to be able to extract meaningful information from dynamical systems that have no supporting mathematical theory or conceptualization (or bias). But RQA can do this very task on any data set with differing degrees of intrinsic deterministic information or lack thereof.

Zbilut and Webber started a productive research collaboration in 1988 that lasted until the untimely death of Zbilut more than twenty years later. Both had mathematical bents and functioned as physiologists in different medical schools, but the two were as different as night and day. Zbilut was the big concept and novel idea person whereas Webber was the methodical computer programmer and data handler. In the late 1980s Zbilut and Webber were frustrated with the inability of standard time series measures to adequately process physiological signals that were notoriously nonlinear, nonstationary and noisy. Zbilut's interests were with the cardiovascular system whereas Webber was studying the pulmonary system. In their hands, Fourier transforms and chaotic mathematics had only limited applicability to physiological systems.

It was a wonderful discovery for Zbilut and Webber to find the short physics paper of Eckmann et al. [3] on recurrence plots. Immediate applicability was seen for cardiac and respiratory systems. Recurrence plots gave beautiful pictorial, but qualitative represen-
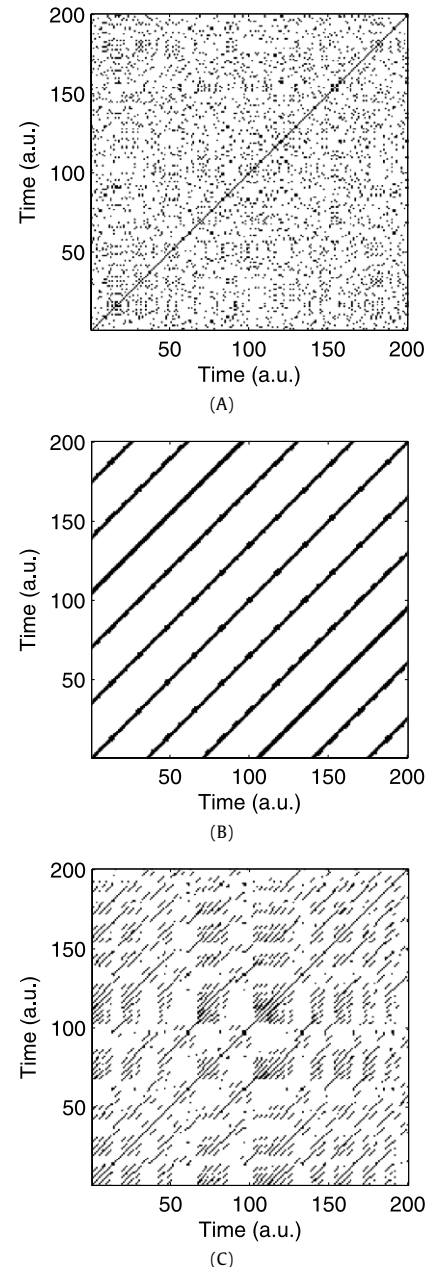

Fig. 1. Prototypical examples of RPs exhibiting different global and local recurrence patterns. (A) normally distributed noise from random number generator ($m = 1$), (B) sine wave function ($m = 2$), and (C) Lorenz oscillator variable y_1 ($m = 3$).

tations of dynamical systems. Interpretations were "in the eye of the beholder" as it were. Seeing the power of recurrence plots in general alongside their qualitative restrictions, Zbilut and Webber gave explicit mathematical definition to distinct features in recurrence plots as explained in two fundamental papers [5,6]. Early programs were dubbed Recurrence Plot Analysis (RPA) routines, but soon the software was renamed as Recurrence Quantification Analysis (RQA), the basic concepts of which are continuing to be expanded upon worldwide. The last paper of Zbilut and Webber [7] discusses RQA in the context of unstable singularities in biological systems. Such systems are piecewise deterministic meaning that physiological systems are best characterized by selectable and deterministic trajectories and interspersed with dynamical pauses or singularities from which the system can be steered.

One early but very instructive example illustrating the importance of recurrence analysis (RQA) over spectral analysis (FFT) was
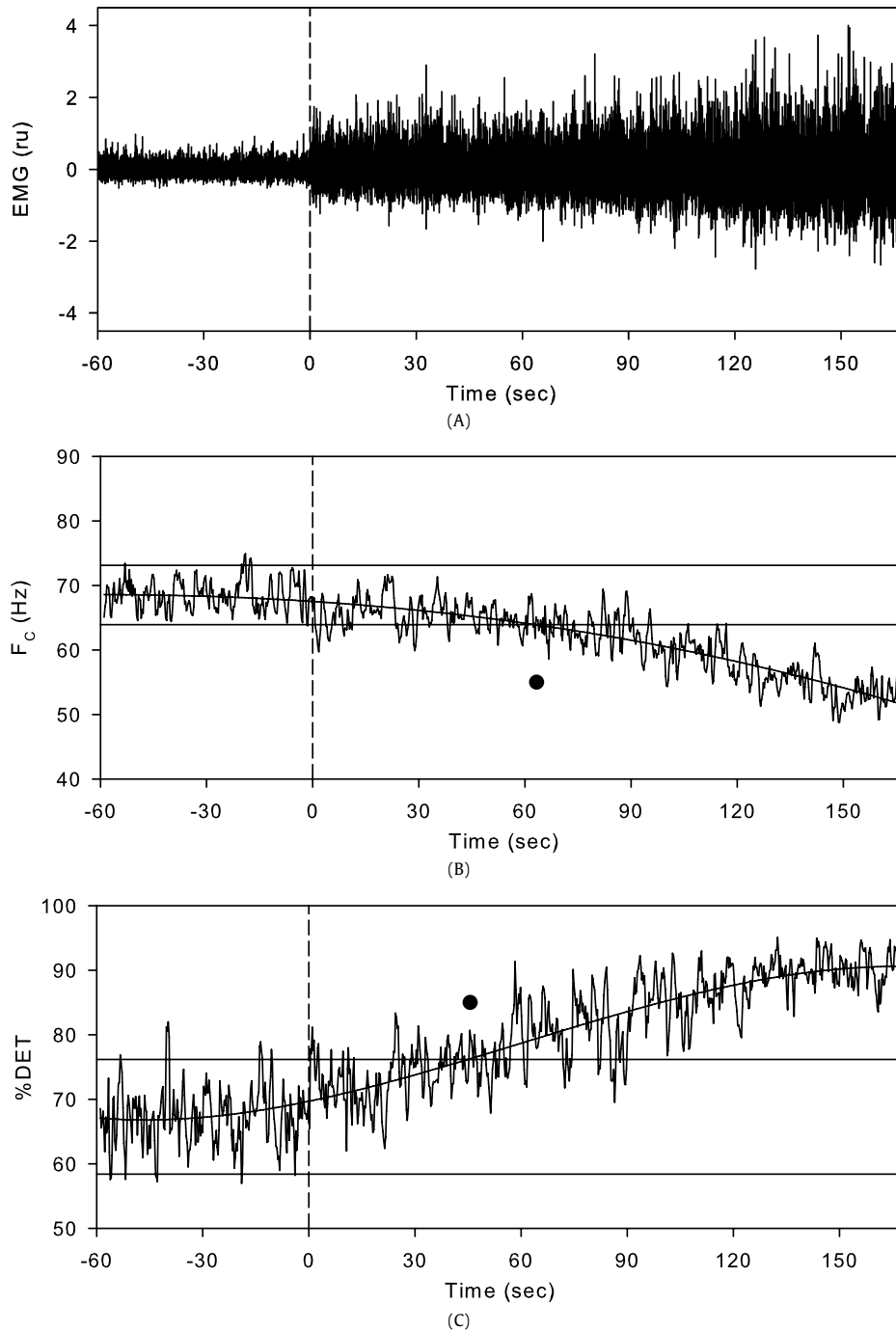
Fig. 2. Comparison of spectral analysis and recurrence analysis in detecting muscle fatigue. (A) Biceps EMG during light (left panel) and heavy (right panel) isometric muscle loading. (B) Slow fall in FFT spectral center frequency during heavy loading. (C) Fast rise in RQA determinism during heavy loading. Parallel lines designate the 95% confidence limits of variables during light loading. Smooth curved lines indicate polynomial fitting of the variables. Dots indicate when the polynomial crosses outside the 95% confidence limits.

successfully carried out by comparing performances of the two methodologies on the exact same, real-world time signal. The research question being asked was how sensitive are RQA and FFT in detecting muscle fatigue [5]? In this case, biceps EMGs were recorded from normal individuals during light and heavy isometric loading of the muscle as shown, respectively, before and after the time zero line in Fig. 2A (reproduced from Fig. 3 of Ref. [8] with permission). The test was run until the subject could no longer support the hand-held weight without moving. FFT and RQA data were computed across the signal within a 1.024 s moving window at 0.256 s intervals. As shown, the falling center frequency

variable (f_C) of FFT and the rising determinism variable (%DET) of RQA were both able to detect fatigue. However, the key is that the %DET rose sooner and higher than the f_C fell (compare dots in Figs. 2B and 2C). It was concluded that RQA had a higher sensitivity than FFT in detecting dynamical changes in EMG signals. This was not surprising since recurrence analysis is a nonlinear, multi-dimensional tool whereas spectral analysis is a linear, single-dimensional analysis.

From a general computational point of view, one big strength of RQA allowing for its great spread was the possibility to coalesce two apparently very distant approaches to data analysis: the

nonlinear dynamics one in which the interest is focused on the temporal dimension, and the multivariate statistics one in which the interest is focused on the relative similarities between data points. This was very apparent when RQA enlarged its field of application outside time series to go into the study of biological polymers where the role of time was played by the order of aminoacid residues along a protein [9]. In the case of protein sequences the theoretical constructs of quasi-attractors, sojourn points of trajectories were automatically translated into their biochemical counterparts of hydrophobic cores and protein–protein interaction. Moreover, the fact a protein has not only a primary structure but a three-dimensional folding allowed to consider a recurrence plot as a contact matrix of residues (points) in the space so allowing to recognise the substantial equivalence of RQA descriptors and network invariants: graph theory coalesced into the dynamical perspective [10,11]. The consequent continuation of this approach is the generalization of the RQA to 2- or 3-dimensional spatial data [12,13]. This technique was successfully applied to study, e.g., complex spatial structures of trabecular bone, and more recently to the analysis of two-dimensional, spatially distributed systems [14].

RQA has become a general purpose technique allowing for generating models endowed with a theoretical appeal in virtually any science fields, starting with cardiology and other life sciences, over engineering, economics, astrophysics and up to Earth sciences [2,15]. A clear example is Earth and environmental science. RQA can be fruitfully exploited when dealing with ecological systems. In general, when trying to model the temporal evolution of physicochemical variables of ecological systems, one must cope with the fact that the class of realistic models is often wide and complicated, and the identification of the right one often needs some a-priori knowledge about the model [13]. Furthermore, identification techniques need the recording of time series, which in the case of ecological systems are often short, non-stationary and corrupted by noise. Under this point of view, providing a pre-analysis tool able to catch visual and quantitative information can become very useful. The problem of dealing with such time series is easily solved by RPs and RQA, as showed for the case of the Orbetello Lagoon [16]: the application of RQA revealed a transition in the oscillations of dissolved oxygen just before the onset of an anoxic crisis. This method was also proposed for monitoring regime shifts in environmental time series, like lake eutrophication and sea water oxygen variability in coastal regions [17]. Based on improvements on the neighborhood criterion for recurrences, the calling-behavior of Japanese tree frogs was analyzed [18]. This study revealed an anti-phase synchronization of the frogs' calls and transitions between different calling behaviors.

The rising interest in applications was for long time steps ahead the interest of theoretical scientists. Meanwhile, several studies on the understanding of the patterns and recurrence structures [19–21] as well as their mathematical relationship to dynamical properties, like K_2 entropy, information or correlation dimensions, of the system have been published [4,22–24]. Furthermore, extensions to analyze periodic, spatial or multivariate data have been developed, yielding cross and joint recurrence plots and their quantification [12,13,25–27], and to an application of the Wiener–Khinchin theorem on the RQA [28].

All in all, the story of RQA is a paradigmatic case of a technique allowing recognizing the substantial unity of all the science fields. This can be made much more efficiently than by simply making scientists of different background to work together, the common appreciation of a very simple technique whose output can be immediately translated into different scientific languages turns the interdisciplinary dream into an immediately operational reality. A great part in the fulfilling of this dream came from the passionate work of our friend Joseph P. Zbilut who passed away the 10th of January of 2009. The incredible ability of RQA to provide a theoretical and statistical frame that can be adapted to any science field is an image of the Joe's ability to make connections among apparently diverse fields in his mind.

Looking back over the years one can weave together the magnificent contributions of Joseph P. Zbilut to science. As a free thinker, Zbilut was gifted with mathematic skills, was filled with physiological insights and savvy in the laboratory as well as clinic, and had an unbelievable grasp of multiple languages and cultures. He was an integrationist seeking and seeing connections between systems and among disciplines that others commonly overlooked. As much as Zbilut was served by RQA over the last two decades, however, his fundamental scientific tool and gift to us all was his brain and unbelievable, nay uncanny way of thinking outside the established scientific box. Like a spilt fluid spill that soaks everything in its path, so Zbilut moved with fluidity within and between and fields. His personality and enthusiasm for what he was doing was infectious, influential, stimulating and inspiring. He will be sorely missed by the scientific community, not to mention his dear wife and two children whom he leaves behind.

References

[1] A. Giuliani, J.P. Zbilut, Complexity 3 (5) (1998) 23.
[2] N. Marwan, Eur. Phys. J. Special Topics 164 (1) (2008) 3.
[3] J.-P. Eckman, S.O. Kamphorst, D. Ruelle, Europhys. Lett. 4 (1987) 973.
[4] N. Marwan, M.C. Romano, M. Thiel, J. Kurths, Phys. Rep. 438 (5–6) (2007) 237.
[5] C.L. Webber Jr., J.P. Zbilut, J. Appl. Physiol. 76 (1994) 965.
[6] J.P. Zbilut Jr., C.L. Webber, Phys. Lett. A 171 (1992) 199.
[7] J.P. Zbilut Jr., C.L. Webber, Eur. Phys. J. Special Topics 164 (2008) 55.
[8] C.L. Webber Jr., J.P. Zbilut, Int. J. Bifur. Chaos 17 (2007) 3467.
[9] A. Giuliani Jr., R. Benigni, J.P. Zbilut, C.L. Webber, P. Sirabella, A. Colosimo, Chem. Rev. 102 (2002) 1471.
[10] A. Krishnan, J.P. Zbilut, M. Tomita, A. Giuliani, Current Protein Peptide Sci. 9 (1) (2008) 28.
[11] C.L. Webber Jr., A. Giuliani, J.P. Zbilut, A. Colosimo, Proteins: Struct. Funct. Genet. 44 (2001) 292.
[12] N. Marwan, J. Kurths, P. Saparin, Phys. Lett. A 360 (4–5) (2007) 545.
[13] D.B. Vasconcelos, S.R. Lopes, R.L. Viana, J. Kurths, Phys. Rev. E 73 (2006) 056207.
[14] A. Facchini, C. Mocenni, A. Vicino, Physica D 238 (2009) 162.
[15] N. Marwan, A. Facchini, M. Thiel, J.P. Zbilut, H. Kantz, Eur. Phys. J. Special Topics 164 (1) (2008) 1.
[16] A. Facchini, C. Mocenni, N. Marwan, A. Vicino, E. Tiezzi, Ecological Modeling 203 (3–4) (2007) 339.
[17] J.-M. Zaldívar, F. Strozzi, S. Dueri, D. Marinov, J.P. Zbilut, Ecological Modeling 210 (1–2) (2008) 58.
[18] I. Aihara, S. Horai, H. Kitahata, K. Aihara, K. Yoshikawa, IEICE Trans. Fund. Electronic Commun. Comput. Sci. E90-A 10 (2007) 2154.
[19] A. Facchini, H. Kantz, Phys. Rev. E 75 (2007) 036215.
[20] N. Marwan, J. Kurths, Phys. Lett. A 336 (455) (2005) 349.
[21] Y. Zou, D. Pazó, M.C. Romano, M. Thiel, J. Kurths, Phys. Rev. E 76 (1) (2007) 016210.
[22] P. Faure, H. Korn, Physica D 122 (1–4) (1998) 265.
[23] J.B. Gao, H.Q. Cai, Phys. Lett. A 270 (1–2) (2000) 75.
[24] M. Thiel, M.C. Romano, P.L. Read, J. Kurths, Chaos 14 (2) (2004) 234.
[25] N. Marwan, J. Kurths, Phys. Lett. A 302 (5–6) (2002) 299.
[26] N. Marwan, M. Thiel, N.R. Nowaczyk, Nonlinear Process. Geophys. 9 (3–4) (2002) 325.
[27] M.C. Romano, M. Thiel, J. Kurths, W. von Bloh, Phys. Lett. A 330 (3–4) (2004) 214.
[28] J.P. Zbilut, N. Marwan, Phys. Lett. A 372 (44) (2008) 6622.