Recurrence quantification analysis as a tool for nonlinear exploration of nonstationary cardiac signals

Joseph P. Zbilut a,*, Nitza Thomasson a, Charles L. Webber b

a Department of Molecular Biophysics and Physiology, Rush University, 1653 W. Congress, Chicago, IL 60612, USA
b Loyola University Chicago, Stritch School of Medicine, 2160 S. First Ave., Maywood, IL 60153, USA

Received 20 December 2000; received in revised form 22 May 2001; accepted 12 November 2001

Abstract

The complexity, nonlinearity and nonstationarity of the cardiovascular system typically defy comprehensive and deterministic mathematical modeling, except from a statistical perspective. Living systems are governed by numerous, continuously changing, interacting variables in the presence of noise. Cardiovascular signals can be shown to be discontinuous alternations between deterministic trajectories and stochastic pauses (terminal dynamics). One promising approach for assessing such nondeterministic complexity is recurrence quantification analysis (RQA). As reviewed in this paper, strategies implementing quantification of recurrences have been successful in diagnosing changes in nonstationary cardiac signals not easily detected by traditional methods. It is concluded that recurrence quantification analysis is a powerful discriminatory tool which, when properly applied to cardiac signals, can provide objectivity regarding the degree of determinism characterizing the system, state changes, as well as degrees of complexity and/or randomness. © 2002 IPEM. Published by Elsevier Science Ltd. All rights reserved.

Keywords: Recurrence quantification; Nonstationarity; Electrocardiograms; Heart rate; Dynamics

1. Introduction and historical context

The last decade has witnessed curious developments in the analysis of biological signals: the original hope that “chaos theory” would help elucidate the complexities of biology are being questioned. Initially, it was hoped that chaotic invariants could capture subtle nonlinear aspects of dynamical biological systems. But as more investigators become aware of the mathematical requisites (and limitations) of chaotic measures such as Liapunov exponents and dimensions, they have recognized that new tools are needed. An important recognition in this respect is that biological signals, in addition to being nonlinear, tend to be nonstationary, noisy and high dimensional [1]. Certainly such a statement is not revolutionary, however, during a time when new, exciting concepts are emerging, it sometimes becomes easy to overlook basic facts, and to ignore fundamental assumptions.

Our own doubts about the ability of many chaotic measures to clarify physiologic processes surfaced with our work to understand heart rate dynamics. Even cursory examinations of plots of heart rate demonstrate frequent, often sudden transitions. To reduce the putative degrees of freedom, correlation dimensions were calculated for heart transplant recipients, whose hearts, by virtue of the surgery, were denervated [2]. The subjects were equilibrated to a quiet environment and resting for several minutes prior to recording. Surprisingly, stationarity was exceedingly difficult to obtain, and, moreover, the calculated dimensions were inconsistent and relatively high. In an attempt to gain greater control, experiments were performed on isolated, perfused rat hearts [3]. Again, difficulties were observed in gaining the requisite stationarity, and dimensions and entropies exhibited error bars which could not confirm chaotic dynamics. More importantly, a piecewise linear map was successful at modeling the dynamics, but only with the addition of a small amount of noise to force the dynamics [4] (see below).
It became clear to us that low dimensional chaos could not explain the phenomena. Certainly others were beginning to come to similar conclusions, and efforts were initiated to explore the use of methods such as wavelets, surrogate testing, and other so-called nonlinear methods [5]. Our perception of these methods, however, was that they were inadequate in that they were still based on linear systems theory, or required stationarity. Even efforts aimed at developing hypothesis testing, such as surrogate analysis, still require stationarity, which, in our experience, are exceedingly rare in biological systems [6].

In this context, a rather short, simple paper by Eckmann, Kamphorst and Ruelle was published [7]. In evaluating a physical experiment, the authors embedded the time series in a higher dimensional space, and then plotted the recurrences in a matrix according to a rule defining an error tolerance. To their surprise patterns were viewed which were previously not apparent in the original series. What is remarkable about this method is that the algorithm requires no mathematical transformations or assumptions. Indeed, one of the purported uses for such plots originally pointed out by Eckmann et al. [7]. Hence, the question of recurrences leads to similar conclusions, and efforts were not explain the phenomena. Certainly others were beginning to come to similar conclusions, and efforts were initiated to explore the use of methods such as wavelets, surrogate testing, and other so-called nonlinear methods [5]. Our perception of these methods, however, was that they were inadequate in that they were still based on linear systems theory, or required stationarity. Even efforts aimed at developing hypothesis testing, such as surrogate analysis, still require stationarity, which, in our experience, are exceedingly rare in biological systems [6].

In this context, a rather short, simple paper by Eckmann, Kamphorst and Ruelle was published [7]. In evaluating a physical experiment, the authors embedded the time series in a higher dimensional space, and then plotted the recurrences in a matrix according to a rule defining an error tolerance. To their surprise patterns were viewed which were previously not apparent in the original series. What is remarkable about this method is that the algorithm requires no mathematical transformations or assumptions. Indeed, one of the purported uses for such plots originally pointed out by Eckmann et al. [7]. Hence, the question of recurrences leads to similar conclusions, and efforts were not explain the phenomena. Certainly others were beginning to come to similar conclusions, and efforts were initiated to explore the use of methods such as wavelets, surrogate testing, and other so-called nonlinear methods [5]. Our perception of these methods, however, was that they were inadequate in that they were still based on linear systems theory, or required stationarity. Even efforts aimed at developing hypothesis testing, such as surrogate analysis, still require stationarity, which, in our experience, are exceedingly rare in biological systems [6].

Although the visual features of such plots are appealing, a drawback was their qualitative nature. As a result, we set out to see if some of these features could be meaningfully quantified. The remainder of this paper presents the results of these efforts, with emphasis on practical application as well as brief examples of their use.

2. Theory

2.1. Recurrences

Recurrence times are certainly not new [8]. Poincaré is perhaps the most famous for describing them in the context of dynamical systems as points which visit a small region of phase space. Also, the statistical literature points out that recurrences are the most basic of relations [9]. In this respect, it is important to reiterate the fact that calculation of recurrence times, unlike other methods such as Fourier, Wigner-Ville or wavelets, requires no transformation of the data, and can be used for both linear and nonlinear systems [10,11]. Because recurrences are simply tallies, they make no mathematical assumptions. Given a reference point, \( X_0 \), and a ball of radius \( r \), a point is said to recur if

\[
B_r(X_0) = \{ X | |X - X_0|| \leq r \} 
\]

A trajectory of size \( N \) falling within \( B_r(X_0) \) is denoted as

\[
S_i = \{ X_1, X_2, ..., X_i, ... \} 
\]

with the recurrence times defined as

\[
T_i = t_{i+1} - t_i, \quad i = 1, 2, ..., N 
\]

If the dynamics is stationary and ergodic on an attractor, the reference point can be chosen arbitrarily. In most biological contexts, this cannot be assumed, yet, as Eckmann et al. have pointed out [7], this very condition is revealed by contextual changes seen in a plot of recurrences.

2.2. Recurrence plots

Given a scalar time series \( \{ x(i) = 1, 2, 3, ... \} \) an embedding procedure will form a vector, \( X_r(i), x(i+L), x(i+(m-1)L) \) with \( m \) the embedding dimension and \( L \) the lag. \( \{ X = 1, 2, 3, ..., N \} \) then represents the multi dimensional process of the time series as a trajectory in \( m \)-dimensional space. Recurrence plots (RP) are symmetrical \( N \times N \) arrays in which a point is placed at \( (i,j) \) whenever a point \( X_i \) on the trajectory is close to another point \( X_j \). The closeness between \( X_i \) and \( X_j \) is expressed by calculating the Euclidian distance between these two normed vectors, i.e., by subtracting one from the other:

\[
|X_i - X_j| \leq r \text{ where } r \text{ is a fixed radius. If the distance falls within this radius, the two vectors are considered to be recurrent, and graphically this can be indicated by a dot (Fig. 1).}
\]

An important feature of such matrixes is the existence of short line segments parallel to the main diagonal, which correspond to sequences \( (i,j), (i+1,j+1), ... (i+k,j+k) \) such that the piece of \( X(i), X(i+1), ..., X(i+k) \), is close to \( X(i), X(i+1), ..., X(i+k) \) in series which are deterministic. The absence of such patterns suggest randomness [7].

2.3. Recurrence quantification

Because graphical representation may be difficult to evaluate, RQA was developed to provide quantification of important aspects revealed by the plot. Recurrent points which form diagonal line segments are considered to be deterministic (as distinguished from random points which form no patterns). Unfortunately, beyond general impressions of drift and determinism, the plots of themselves provide no quantification. As a result, Zbilut and Webber [1] developed several strategies to quantify features of such plots originally pointed out by Eckmann et al. [7]. Hence, the quantification of recurrences leads to the generation of five variables including: \%REC (percent of plot filled with recurrent points), \%DET (percent of recurrent points forming diagonal lines, with a minimum of two adjacent points), ENT (Shannon information entropy of the line length distribution), MAXLINE, length of longest line segment (the reciprocal of which is an approximation of the largest positive Ljapunov exponent and is a measure of system divergence [12]); and TREND (measure of the paling of
Fig. 1. Example of recurrence plot (RP) for a sine wave (a) (RP below) and the logistic equation (b) (detail of RP below) in the chaotic regime. The period of a of the sine wave can be calculated by the distance of points between diagonal lines. Note the short diagonal line segments of the logistic RP indicative of deterministic processes.

recurrent points away from the central diagonal. These five recurrence variables quantify the deterministic structure and complexity of the plot.

In order to follow changes of these variables in time, a “windowed” version of RQA can be performed, such that for a time series \((s_1, s_2, ..., s_n)\), where \(s_j = s_j(t)\) and \(t_s = \text{sampling time}\). For an \(N\) point long series

\[
E_1 = (s_1, s_2, ..., s_N) \\
E_2 = (s_{1+w}, s_{2+w}, ..., s_{N+w}) \\
E_3 = (s_{1+2w}, s_{2+2w}, ..., s_{N+2w}) \\
: \\
E_p = (s_{1+(p-1)w}, s_{2+(p-1)w}, ..., s_{N+(p-1)w})
\]

with \(w = \text{the offset, and the number of epochs (windows)}, E_p, \text{satisfies the relation, } N+(p-1)w \leq n\).

Analogous to cross power spectral analysis, cross recurrence analysis is also possible. For the series \(X_i = (x(i), x(i+L), ..., x(i+(m-1)L))\), another series, \(Y_i = (y(i), y(i+L), ..., y(i+(m-1)L))\) can be compared for recurrences by the relation \(||X_i - Y_j||\) for a given \(r\) [13]. Windowed versions are similarly possible.

The data obtained can also be used to obtain estimations of local Liapunov exponents, information entropy, or simply plotted as \(N_{\text{recurrences}}\) vs. period; i.e., a histogram of recurrence times. In the case of histograms, strictly periodic points demonstrate instrumentally sharp peaks; whereas chaotic or nonlinear systems reveal more or less wider peaks depending upon the radius chosen and noise effects. RQA can also be combined with other statistical techniques such as principal components analysis to gain more information [14–16].

2.4. Determining parameters for nonstationary series

As has been emphasized, RQA is useful for understanding nonstationary time series. Yet, since a given system may be changing state; i.e., as the relevant degrees of freedom may change, the choice of \(m, L\) and \(r\) can become confounding. Unfortunately, most algorithms for such choices are based upon computer simulations of well-known, stationary examples. Typically, as has been pointed out, biological systems are rarely stationary, and often exhibit rather sudden changes of state. Nonetheless, some guidelines can be established, based upon available research, and a careful consideration of the import of nonstationarity.

2.4.1. Choice of embedding

In the context of nonstationarity, the notion of a “correct” embedding or delay is inappropriate as has been demonstrated by Grassberger and Schreiber [17]. Instead it becomes important to remember that a sufficiently large embedding be chosen which will “contain” the relevant dynamics (as it may change from onedimensionality to another) as well as account for the effects of noise, which tend to inflate dimension. There are no clear guidelines relative to this question, except from what can be inferred from studies of noise. In this respect Ding et al. [18] have indicated that noise will tend to require
higher dimensions, even in the case of stationary dynamics. Gao and Cai [19] have studied this question in the context of a noisy Lorenz attractor, and concluded that an embedding of 6 is required to provide reasonable clarification of the dynamics. Because of the high complexity of biological systems, we have empirically embedded in 10. Care, however, must be made to make sure that the system is not excessively noisy, since embedding will amplify such noise to the detriment of the real dynamics.

2.4.2. Choice of lag

Choice of lag is governed by similar considerations. As a system changes from one dimension to another the effects of the lag are performed changed. Thus, a so-called “optimal” lag in one embedding, becomes less so as the relevant dimension changes [17].

Although there have been numerous proposals for choice of lag, chief among them the first local minimum of the autocorrelation or mutual information, they all are presented with the assumption of stationarity [20,21]. What would appear to be more important is an understanding of the data acquisition apparatus (typically A/D), as well as the system studied. Specifically, what is the sampling, the bandwidth, the precision, the nonlinearity?

Since there is no fully developed theory of nonlinear systems as there is for linear systems, the problem would seem to be complex. This is to say that a nonlinear system, is often sampled by guidelines developed for linear systems; i.e., the sampling theorem. As a matter of fact, the delay reconstruction theorem indicates that practically any delay is appropriate [22]. In practice, actual sampling determines whether a nonlinear system is sufficiently sampled to capture the important dynamics. Thus, it would seem to be reasonable to “over” sample a system to determine if there are important small, but putatively significant features in the time series. Clearly, however, care must be made to balance this need for adequate sampling against the possibility of introducing greater amounts of noise.

Once this has been determined, use of autocorrelation or mutual information as a guide may be used to provide a reconstruction which may maximally “unfold” the dynamics, while keeping in mind the limitations. (An alternative would be to ramp up the embedding and delay and check if the number of recurrences plateau, but this most likely would not happen in the case of nonstationarity [11].) If the embedding is high enough to contain the nonstationary dynamics, and the lag constant, the change in recurrences simply fulfill the objective of identifying state changes. The critical point is that the method of choice of lag be consistent for a given method of data acquisition.

2.4.3. Choice of radius

The object of RQA is to view the recurrences in a locally defined (linear) region of phase space. Practically speaking, however, because of intrinsic and extrinsic noise, too small a value of r results in quantification of the noise only; whereas too large a value captures values which can no longer be considered recurrent. To get to the dynamics proper, a strategy is to calculate %REC for several increasing values of r and to plot the results on a log-log plot to determine a “scaling” region; i.e., where the dynamics are found. Fig. 2 demonstrates such a strategy.

If the data are extremely nonstationary, a scaling region may not be found. The guide then is the percent recurrence. A critical factor is that there be sufficient numbers of recurrences so as to make sense for computation of the other variables. A value of 1% recurrence tends to fulfill this criterion. Further verification can be obtained from an inspection of the recurrence plot: too sparse a matrix suggests a modest increase. Windowed RQA is especially informative. If a given window fails to achieve 1% recurrence, the radius should be increased, or the window enlarged.

3. Examples

There have been numerous articles substantiating the utility of RQA in cardiovascular physiology, especially in circumstances where standard FFT methods are found to be wanting [23–29]. The following brief examples below were chosen with a view to illustrate this.

3.1. Singularities of the heartbeat

The first example demonstrates that careful inspection of the plot, can yield important information not readily apparent by other means. Specifically, the plot of recurrences reveals that the ECG signal is not really continuous, but characterized by singularities. To appreciate this, the basic ECG signal itself needs to be analyzed, and not just the R-R measurements. It is also important to note that such singularities can be easily missed if careful consideration is not given to choice of radius.

Both physiologically and numerically, a pause before the P wave has been described in the context of a resetting mechanism, or “integrate and fire” process. The full implication of such a description suggests a singular event between the T and P waves. Inspection of the ECG by an RP does demonstrate a discontinuity at this location (Fig. 3). The full import of this discontinuity is beyond the scope of the present paper, but briefly, it does imply that analysis by linear and stationary nonlinear methods is not justified [see 30–34]. (We are currently evaluating quantification of these singularities.) It may be better modeled by non-Lipschitz differential equa-
Fig. 2. Plot of %REC vs Radius. An appropriate choice of radius is the beginning of the scaling region above the noise floor.

Fig. 3. RP of an ECG sampled at 500 Hz. Arrow points to discontinuity occurring between the T and P waves. Note how the trajectory from left to right is broken. Thus the dynamics are not continuous (and not stationary), but are broken by such singularities. The singularities are variable, accounting for much of the “randomness” of the ECG. It may also account for some of the physiological ramifications of cardiac control, such as the so-called “R-on-T” phenomenon, and methods of scanning to pace/control the heart often used in electrophysiological laboratories. FFTs and wavelet analysis do not reveal these singularities.

Fig. 4 shows the ECG (top panel) with the beginning of the seizure indicated by the arrow. The remaining panels depicting the RQA variables demonstrate transients (between dotted lines) with an apparent change in the values subsequently. This is compared with a control sample from the same subject (Fig. 5). In neither of these situations whereby there is no single solution to the equation. Instead, the problem becomes a combinatorial one with a resultant “stochastic attractor.” This is to say that each singularity has an associated probability distribution regulated by both intra (electrical events, stretch) and extra cardiac factors (autonomic nervous system, hormones). Thus each beat (and each subsequent beat) is conditioned by the probabilities. Based upon this phenomenon as a paradigm, a dynamical system whose solutions are stochastic processes with a prescribed joint density can be developed.

3.2. ECG in epilepsy

The next example demonstrates the utility of windowed RQA to localize in time specific events. Choice of window size is based upon considerations similar to those for spectro-temporal FFTs; i.e., the larger the window, the smaller the resolution, and vice versa:

It has been observed for some time that heart rate changes and arrhythmias are often concomitants to or antecedents of seizures [35,36]. The assumption is that the neural processes which result in ictal activity are also transmitted to the heart via the autonomic nervous system. Preliminary research using the windowed form of RQA supports such a hypothesis [37].

Fig. 4 shows the ECG (top panel) with the beginning of the seizure indicated by the arrow. The remaining panels depicting the RQA variables demonstrate transients (between dotted lines) with an apparent change in the values subsequently. This is compared with a control sample from the same subject (Fig. 5). In neither of these
examples does the raw ECG provide any clear hint to change. These are contrasted with a spectro-temporal FFT of the same samples (Fig. 6). In all cases the windows were 128 points in length and overlapped by one point at a time. The spectro-temporal analysis shows none of the transients or changes found in the RQA of the pre-seizure data.

4. Conclusion

Although there is the temptation to totally dismiss linear techniques in the analysis of cardiac signals, this would be unfortunate. In many circumstances, methods such as the FFT can be very useful when there is a large, clear effect such as in the case of approximate period-
icities of the ECG. However, when there are nonstationarities and/or nonlinearities, such as are often encountered in ECGs, which result in important, but perhaps subtle effects, methods such as RQA should be considered. The advantage of RQA, and the basis of its wide applicability, is its simplicity. RQA requires no mathematical assumptions, and needs only to count similar events in an embedded space.

References

[22] Packard NH, Crutchfield JP, Farmer JD, Shaw RS. Geometry from a time series. Phys Rev Lett 1980;45:712–6. We note the idea was based on a suggestion by D. Ruelle.


