Recurrence plots revisited

M.C. Casdagli
Scripps Institute of Oceanography, La Jolla, CA 92093, USA

Received 17 October 1996; accepted 31 January 1997
Communicated by A.M. Albano

Abstract

We show that recurrence plots (RPs) give detailed characterizations of time series generated by dynamical systems driven by slowly varying external forces. For deterministic systems we show that RPs of the time series can be used to reconstruct the RP of the driving force if it varies sufficiently slowly. If the driving force is one-dimensional, its functional form can then be inferred up to an invertible coordinate transformation. The same results hold for stochastic systems if the RP of the time series is suitably averaged and transformed. These results are used to investigate the nonlinear prediction of time series generated by dynamical systems driven by slowly varying external forces. We also consider the problem of detecting a small change in the driving force, and propose a surrogate data technique for assessing statistical significance. Numerically simulated time series and a time series of respiration rates recorded from a subject with sleep apnea are used as illustrative examples.

Keywords: Nonstationarity; Prediction; Reconstruction; Recurrence; Surrogates

1. Introduction

There has been much recent interest in nonstationary time series analysis for two reasons. First, many naturally occurring time series are nonstationary. For example, physiological systems tend to vary over multiple time scales. Also, time series from spatially extended physical systems have a tendency to exhibit dynamics over multiple time scales [1]. Thus nonstationary time series analysis is of wide applicability. Second, a fundamental assumption underlying almost all existing linear and nonlinear techniques of time series analysis is that the time series is stationary [11,44]. Misleading results can be obtained by applying these techniques to nonstationary time series. For example the dimension of an embedded random walk is equal to 2, and can be misidentified as deterministic chaos [27]. Thus the development of techniques for nonstationary time series analysis presents a major theoretical challenge.

\footnote{From a theoretical point of view a time series \(x_1, x_2, \ldots\) is nonstationary if, for some \(m\), the joint probability distribution of \(x_i, x_{i+1}, \ldots, x_{i+m-1}\) is dependent on the time index \(i\) [32]. From a practical point of view a time series \(x_1, x_2, \ldots, x_N\) is nonstationary if, for low \(m\), there are variations in the estimated joint distribution of \(x_i, x_{i+1}, \ldots, x_{i+m-1}\) that occur on time scales of order \(N\).}
Some principal objectives of nonstationary time series analysis are as follows:

1. **Characterization.** Suppose the dynamics underlying a time series is similar at the beginning and end of the time series, but different in the middle. The objective of a technique that characterizes nonstationarity is to extract this information in a convenient form. A straightforward technique for characterizing nonstationarity is to divide the time series into segments and monitor changes across segments of a quantity of interest. This approach is strongly dependent on the quantity chosen, and is somewhat arbitrary. More general techniques have been developed by Eckmann et al. [10], Manuca and Savit [25] and Schreiber [36]. For wavelet techniques see [5] and references therein.

2. **Prediction.** In order to obtain accurate predictions for a nonstationary time series it may be necessary to modify prediction algorithms developed for stationary time series. This objective is closely related to the characterization and modeling of nonstationarity. The prediction of nonstationary time series has been considered by Priestley [32], Stark [38], Weigend et al. [46] and Young [48].

3. **Change point detection.** In some cases there may be a single point at which the dynamics underlying a time series changes. The objective is to identify the change point. Change point detection has been considered by de Lima [6], Kennel [19], Lombard and Hart [22], Smith et al. [37] and Takanami and Kitagawa [41].

4. **Hypothesis testing.** In some cases there may be reason to believe that a time series is stationary. The objective is to develop hypothesis tests for the null hypothesis that the time series is stationary. If the null hypothesis is accepted, one can then have confidence in applying techniques of stationary time series analysis. This problem has been considered from a nonlinear perspective by de Lima [6], Isliker and Kurths [16] and Kennel [19]. There is also a large literature on “unit-root” tests that considers this problem from a linear perspective [7].

Objectives 3 and 4 can be combined to address the question: “When are two time series different?”, by concatenating the two time series and testing for nonstationarity. This question is of interest in signal classification and detection, and in model validation, and has been considered by Diggle [8], Fraser [13], Kadtke [17], Kantz [18], Moeckel and Murray [26] and Wright [47]. It is a variation on the question: “When are two probability distributions different?”. For a review of techniques that address the latter question see [30].

In this paper we show that techniques based on recurrence plots (RPs), introduced by Eckmann et al. [10] provide a unified framework for addressing objectives 1–4. We focus attention on time series generated by dynamical systems driven by slowly varying external forces. In Sections 2 and 3 we show that techniques based on RPs give detailed characterizations of nonstationarities in such time series as follows. In Section 2 we show that for deterministic dynamical systems RPs of the time series can be used to reconstruct the RP of the driving force if it varies sufficiently slowly. If the driving force is smooth and one-dimensional, we show that its functional form can then be inferred up to an invertible coordinate transformation. These results are supported by numerical simulation, scaling laws, and a heuristic argument that extends the concepts of embedding of Takens [42] and Sauer et al. [34] to parametrized families of dynamical systems. In Section 3 we show that the same results hold for stochastic systems if the RP of the time series is suitably averaged and transformed into a meta-recurrence plot (MRP). Techniques similar to MRPs have been proposed by Manuca and Savit [25] and by Schreiber [36]. We introduce a linearized version of the MRP to assess the extent to which the MRP characterizes nonstationarities in the mean, variance and autocorrelation of the time series. This is similar to a technique used by Palus et al. [28] to assess the role of linear correlations in the estimation of information-theoretic quantities for stationary time series.

The technique we use to predict nonstationary time series is a modification of local linear algorithms developed for predicting stationary time series [2,12] in which the prediction of the current point of a time series is based on information in the recent past. This approach is implicit in the recursive prediction algorithms of Priestley [32], Stark [38] and Young [48]. In Section 4 we show numerically that modified local linear prediction in general underperforms unmodified local linear prediction for time series generated by deterministic systems driven by slowly varying external forces. This result is supported by a scaling law and a heuristic argument based on RPs. In the case
of time series generated by stochastic systems driven by slowly varying external forces, we show that it is possible to obtain improved accuracy from modified local linear prediction. We conjecture that further improvements are possible by exploiting the MRP structure of the time series, and that this is implicit in the prediction algorithm of Weigend et al. [46].

The technique we use to detect change points in a time series is obtained by considering an average over an RP, and is related to techniques proposed by de Lima [6] and Kennel [19]. In Section 5 we show that, under certain conditions, the technique reliably detects change points for a wide range of time series. A surrogate data technique is used to assess the statistical significance of the change point. A similar surrogate data technique has been used in a different context by Politis and Romano [29].

The results in Sections 2–5 are illustrated by time series generated from dynamical systems with strong mixing properties. In Section 6 we consider a stationary time series generated from a dynamical system with very weak mixing properties. We show that MRPs characterize the time series as effectively nonstationary for very long time scales, and discuss the implications of this.

In Section 7 we apply the techniques of earlier sections to a time series of respiration rates recorded from a subject with sleep apnea. The data are described by Rigney et al. [33]. We do not claim any clinical relevance for the results. The analysis is intended to illustrate the use of the above techniques in a practical setting. In order to achieve clinical relevance a much wider range of data from different patients would have to be analyzed. In particular, the time series analyzed here is just one of several relevant physiological variables recorded in this patient.

We summarize our main results in Section 8, together with several open questions. Technical details are collected in Appendix A.

2. Recurrence plots and embedding

In this section we investigate the structure of RPs of time series generated from deterministic dynamical systems driven by slowly varying external forces. We present numerical evidence that RPs of the time series can be used to reconstruct the RP of the driving force. Scaling laws quantifying the accuracy of the reconstruction are developed. It is shown that, under certain conditions, RPs of the time series can be used to reconstruct the functional form of the driving force up to an invertible coordinate transformation.

2.1. Definitions

In this section we consider univariate time series \( x_1, x_2, \ldots, x_N \) generated by \( d \)-dimensional deterministic dynamical systems \( f : \mathbb{R}^{d+p} \rightarrow \mathbb{R}^d \) with \( p \) parameters and observation functions \( h : \mathbb{R}^d \rightarrow \mathbb{R} \) according to

\[
s_{i+1} = f(s_i, \gamma_i), \quad x_{i+1} = h(s_{i+1}),
\]

where \( s_i \in \mathbb{R}^d \) denotes the state of the system at time \( i \) and \( \gamma_i \in \mathbb{R}^p \) is an unobserved slowly varying driving force (i.e. \( \| \gamma_{i+1} - \gamma_i \| \) is small relative to typical distances between successive states of the system). We will assume that the functions \( f \) and \( h \) are smooth almost everywhere. In the absence of the driving force, a sequence of embedding vectors \( \psi_i, i = 1, 2, \ldots, N_m = N - (m - 1)\tau, \) with embedding dimension \( m \) and delay time \( \tau \) defined by

\[
\psi_i = x_i, x_{i+\tau}, x_{i+2\tau}, \ldots, x_{i+(m-1)\tau}
\]

can typically be used to reconstruct the dynamics on attractors of box-counting dimension \( D \) wherever \( m > 2D \) [34,42]. In practice it is necessary to choose \( \tau > 1 \) for finely sampled time series. In the examples considered in this paper it suffices to take \( \tau = 1 \), and we will assume this henceforth.
Eckmann et al. [10] introduced a graphical technique, called recurrence plots (RPs), that can be used to detect certain forms of driving forces. To construct RPs an array of dots, representing recurrences, is placed in an \( N_m \times N_m \) square at location \((i, j)\) if the embedding vector \(v_i\) is one of the \(k\)th nearest neighbors of \(v_j\), where \(k\) is a small preselected number. Driving forces that cause long term drift in the time series are identified by a fall off in recurrences away from the diagonal \(i = j\). Periodic driving forces are identified by periodic bands of recurrences parallel to the diagonal. Following [21], in this paper an alternative definition of recurrences is used. We define a recurrence to occur at \((i, j)\) with resolution \(r\) if \(i \neq j\) and \(\|v_i - v_j\| \leq r\). The resolution \(r\) is measured in units of the standard deviation of the time series. We use the maximum norm \(\|v_i - v_j\| = \max_{k=0,...,m-1} \mid x_{i+k} - x_{j+k}\|\), for computational efficiency; see Appendix A.2. Note that RPs as defined above are symmetric.

By definition, the probability that a recurrence occurs at a randomly selected location \((i, j)\) is equal to the correlation integral \(C_m(r)\) of the time series, and is given by

\[
C_m(r) = \frac{2}{N_m(N_m - 1)} \sum_{i=1}^{N_m} \sum_{j=1}^{i-1} \theta(r - \|v_i - v_j\|),
\]

where \(\theta\) is the step function, \(\theta(x < 0) = 0\), \(\theta(x \geq 0) = 1\). In order to infer properties of the driving force \(\gamma_1, \gamma_2, \ldots, \gamma_N\) from embedding vectors, it is desirable that whenever a recurrence occurs at location \((i, j)\), then with high probability \(\gamma_i\) and \(\gamma_j\) are close. This motivates us to define a false recurrence\(^2\) (resp. true recurrence) with resolution \((r, \varepsilon)\) to be a recurrence \((i, j)\) at resolution \(r\) satisfying \(\|\gamma_i - \gamma_j\| > \varepsilon\) (resp. \(\|\gamma_i - \gamma_j\| \leq \varepsilon\)). Denoting the probability that a false (resp. true) recurrence occurs at a randomly selected location \((i, j)\) by \(P_m^F(r, \varepsilon)\) (resp. \(P_m^T(r, \varepsilon)\)) we have \(P_m^F(r, \varepsilon) = 1 - P_m^T(r, \varepsilon)\) where

\[
P_m^T(r, \varepsilon) = \frac{2}{N_m(N_m - 1)} \sum_{i=1}^{N_m} \sum_{j=1}^{i-1} \theta(r - \|v_i - v_j\|)\theta(\varepsilon - \|\gamma_i - \gamma_j\|).
\]

2.2. Recurrence plot reconstruction

The results in this section were motivated by the following numerical experiment. Let \(f : [0, 1]^2 \to [0, 1]\) be the one parameter family of tent maps

\[
f(x, \gamma) = \begin{cases} 
2x + \gamma, & 0 \leq x \leq \frac{1}{2}(1 - \gamma), \\
-2x + 2 - \gamma, & \frac{1}{2}(1 - \gamma) < x \leq \frac{1}{2}(2 - \gamma), \\
2x + \gamma - 2, & \frac{1}{2}(2 - \gamma) < x \leq 1,
\end{cases}
\]

\(\gamma = 0\) gives the standard tent map of functional form "\(\land\)", and \(\gamma = 1\) gives an inverted tent map of functional form "\(\lor\)". As \(\gamma\) is varied from 0 to 1 the functional form varies continuously between these two extremes. This example was chosen because the dynamics is very similar at each fixed parameter \(\gamma\): for almost all initial conditions, iterations fill out the interval \([0,1]\) uniformly and have Lyapunov exponent \(\log 2\). Consequently it is difficult to monitor changes in the parameter \(\gamma\) by monitoring quantities such as dimensions, Lyapunov exponents and entropies.

Fig. 1(a) shows a phase portrait of embedding vectors for \(m = 2\) obtained from a time series of length \(N = 60,000\) generated by iterating (5) whilst slowly varying the parameter \(\gamma\) according to the driving force shown in Fig. 1(c). Two RPs of the time series are shown in Fig. 1(b) for \(m = 2\) (above the diagonal) and \(m = 3\) (below the diagonal).

\(^2\) False recurrences can be considered as an extension of the concept of false nearest neighbors introduced by Kennel and Abarbanel [20] to nonstationary systems.
Fig. 1. Recurrence plot (RP) reconstruction for time series generated by deterministic systems under the influence of a slowly varying driving force. (a) Phase portrait of a time series generated from the family of tent maps (5) whilst slowly varying the parameter \( \gamma \) according to the driving force shown in (c). (b) RP of the time series in (a) for \( r = 0.02 \) and \( m = 2 \) (above the diagonal) and \( m = 3 \) (below the diagonal). (c) The driving force used to generate (a). (d) Blowup of (b) for \( m = 3 \), \( i < 23000 \) and \( j < 2100 \) superimposed on a blowup of the driving force \( \gamma \) for \( i < 23000 \). (e) RP of the driving force shown in (c) for \( r = 0.02 \) and \( m = 1 \). (f) Phase portraits of time series generated from the Lorenz equations (6) at four fixed values of \( \gamma \). (g) RP of a time series generated from the Lorenz equation whilst varying the parameter \( \gamma \) according to (c), for \( r = 0.02 \) and \( m = 3 \) (above the diagonal) and \( m = 5 \) (below the diagonal).
The resolution was chosen to be $r = 0.02$. Fig. 1(e) shows the RP of the driving force $y_1, y_2, \ldots, y_N$ for $m = 1$ and $r = 0.02$. The RP of the time series for $m = 3$ is an accurate “reconstruction” of the RP of the driving force, even though the driving force was not used in its calculation. Moreover, the functional form of the driving force can be reconstructed directly from the RP of the time series. Fig. 1(d) shows a blowup of the portion of the RP in Fig. 1(b) for $m = 3$, $i < 23000$ and $j < 2100$, superimposed on a blowup of the driving force. Although the functional forms do not match exactly, the relative values of the minima and maxima are all preserved.

Figs. 1(f) and (g) show the results of a similar numerical experiment on a one-parameter family of Lorenz equations defined by

$$\dot{x} = (6 + 12\gamma(t))(y - x), \quad \dot{y} = 28x - y - xz, \quad \dot{z} = -\frac{8}{3}z + xy. \quad (6)$$

The standard Lorenz attractor is obtained by setting $\gamma = \frac{1}{3}$. A time series of length $N = 60000$ was generated by integrating these equations and sampling the $x$ component every 0.16 units of time, whilst varying the parameter $\gamma$ between 0 and 1 according to the driving force shown in Fig. 1(c). This example was chosen because an approximately two-dimensional attractor appears to exist over this range of parameters. Phase portraits in which $\gamma$ is fixed at four different values are shown in Fig. 1(f). Two RPs of the time series are shown in Fig. 1(g), for $m = 3$ (above the diagonal), and $m = 5$ (below the diagonal), with resolution $r = 0.02$. The RP of the time series for $m = 5$ gives a moderately good reconstruction of the RP of the driving force, although it is less accurate than for the RP of the tent map time series.

In Section 2.3 we present a theory of recurrence plot reconstruction that explains the above phenomena. In particular, scaling laws are developed describing the probability of false recurrences. A preliminary intuition for false recurrences can be obtained by considering the above examples. For the RP of the tent map time series with $m = 2$ shown in Fig. 1(b), these are a large number of false recurrences. This is because embeddings vectors in the upper left and lower right triangular regions of Fig. 1(a) do not uniquely specify the parameter $\gamma$: embedding vectors in these regions could have been generated by two distinct tent maps from the family (3). In the case of the RP of the Lorenz equations with $m = 5$ shown in Fig. 1(f), there are fewer recurrences because the time series is of higher dimension. Also, the fraction of recurrences that are false is higher. Intuitively, this is because the attractors reconstructed from the Lorenz equations at different values of $\gamma$ do not differ as much as those for the tent maps. These concepts will be made precise in Section 2.3.

### 2.3. Scaling laws

In order to make precise statements about RP reconstruction it is necessary to take the limit of a slowly varying driving force. We therefore assume that the driving force is obtained by sampling a continuous function $G : [0, 1] \rightarrow [0, 1]^p$ according to $\gamma_i = G(i/N)$, and then take the limit $N \rightarrow \infty$. Assume the set of points $\{(s_i, \gamma_i), i \in \mathbb{Z}^+\}$ fills out a homogeneous fractal\(^4\) of dimension $D$ lying in an $n$-dimensional manifold $M \subseteq \mathbb{R}^{d+p}$, where $n \leq D + 1$. Then, subject to genericity conditions on the dynamical system (1), the RP of the time series $x_1, x_2, \ldots$, for embedding dimension $m$ satisfies the following scaling laws as the resolution $r$ tends to zero.

The probability $C_m(r)$ of a recurrence at resolution $r$ satisfies

$$C_m(r) = \begin{cases} O(r^m), & m \leq D, \\ O(r^D), & m > D. \end{cases} \quad (7)$$

\(^3\) These assumptions can be relaxed. They are made to simplify the statement of the scaling laws and the heuristic argument supporting them.

\(^4\) We define a homogeneous fractal to be a fractal all of whose generalized dimensions are equal.
For any $\epsilon > 0$, the probability $P_m^F(r, \epsilon)$ of a false recurrence at resolution $(r, \epsilon)$ satisfies

$$P_m^F(r, \epsilon) = \begin{cases} O(r^m), & m \leq 2D - 1, \\ O(r^{2D-1}), & m > 2D - 1. \end{cases}$$

(8)

These scaling laws imply that, for embedding dimensions $m > D$, the fraction of recurrences that are false at resolution $(r, \epsilon)$ scales at most as $r^{m-D}$, and is vanishingly small as $r$ decreases, so that the RP of the time series is equal to the RP of the driving force to within resolution $\epsilon$. Note that as $r$ decreases the probability of a recurrence decreases as $r^D$; this imposes severe data constraints for high-dimensional systems. A heuristic argument supporting the above scaling laws is given in Appendix A.1. The idea is to consider embedding properties of the map $\varphi_m : \mathbb{R}^{d+p} \to \mathbb{R}^m$ defined by

$$\varphi_m(s, \gamma) = \hat{h}(s), \hat{h}(\tilde{f}(s, \gamma)), \hat{h}(\tilde{f}^2(s, \gamma)), \ldots, \hat{h}(\tilde{f}^m(s, \gamma)),$$

(9)

where $\hat{h}$ and $\tilde{f}$ are defined by $\hat{h}(s, \gamma) = h(s)$ and $f(s, \gamma) = (f(s, \gamma), \gamma)$. In the case of stationary dynamical systems rigorous embedding results have been obtained by Takens [42] and Sauer et al. [34].

We now investigate numerically the extent to which the above scaling laws apply at finite resolutions $r$ and sample sizes $N$ for the tent maps and Lorenz equations of Section 2.2. For simplicity we consider a driving force given by $y_i = i/N$ for $i = 1, 2, \ldots, N$, where $N = 60000$. RPs of the time series from the tent maps for $m = 2, 3$ (resp. Lorenz equations for $m = 3, 5$) are shown in Fig. 2(a) (resp. 2(b)) for $r = 0.02$. For this driving force a false recurrence at resolution $(r, \epsilon)$ occurs whenever there is a recurrence $(i, j)$ such that $|i - j|/N > \epsilon$. The properties of false recurrences in Figs. 2(a) and (b) are similar to those in Figs. 1(b) and (g): most false recurrences at a given resolution $(r, \epsilon)$ can be removed by increasing the embedding dimension. Also, the fraction of recurrences that are false is higher for the Lorenz time series.

The dotted lines in Fig. 2(c) show how the probability of a recurrence, $C_m(r)$, scales with $r$ for the tent map time series, where $m = 2$ and 3. The solid lines show how the probability of a false recurrence, $P_m^F(r, \epsilon)$, scales with $r$ for $\epsilon = \frac{1}{10}$. In this example the set of points $(x_i, y_i)$ fills out the square $[0, 1]^2$ uniformly, so the scaling laws (7) and (8) are expected to apply with $D = 2$. By inspection of Fig. 2(c) these scaling laws hold to a high degree of accuracy. Fig. 2(d) shows the corresponding results for the Lorenz equations with $m = 3, 4, 5$. As mentioned in Section 2.2, for a fixed $y \in [0, 1]$, the dimension of the attractor is approximately 2. Thus the set of points $(x_i, y_i, z_i, \gamma_i)$ fills out an approximately three-dimensional subset of $\mathbb{R}^3 \times [0, 1]$, so that the scaling laws (7) and (8) are expected to apply with $D = 3$. By inspection of Fig. 2(d) these scaling laws hold to a moderate degree of accuracy. The lack of accuracy may be due to difficulties in distinguishing attractors of this family at these resolutions (see Fig. 1(f)). However, it is clear that for $m > 3$ the probability of a false recurrence decreases more rapidly with $r$ than the probability of a recurrence, so that as $r$ tends to zero the fraction of recurrences that are false is vanishingly small.

Figs. 2(c) and (d) illustrated how the accuracy of an RP reconstruction depends on the dimensionality of the system and the resolution $r$. The accuracy of an RP reconstruction also depends on how small $r$ must be chosen relative to $\epsilon$ so that (8) is satisfied to good approximation. We conjecture that there exists a constant $K_m$, dependent on the dynamical system (1) and the embedding dimension $m$, such that (8) holds with $\epsilon = K_m r$. This is illustrated in Figs. 2(e) and (f) for the tent map and Lorenz time series, by plotting the fraction of recurrences that are false, $P_m^F(r, kr)/C_m(r)$, against $k$ for various $m$ and $r$. According to the above conjecture $P_m^F(r, kr)/C_m(r)$ should become vanishingly small as $r$ decreases for $m > D$ and $k > K_m$. In Fig. 2(e) (resp. 2(f)) it appears that $K_m < 1$ for $m > 2$ (resp. $K_m < 50$ for $m > 3$). In Appendix A.1 we argue that the constant $K_m$ depends on singular values of the derivatives $D\varphi_m$ of the map $\varphi_m$ defined by (9) evaluated on the set $\{(s_i, \gamma_i), i \in \mathbb{Z}^+\}$. The matrices $D\varphi_m$ quantify the information flow from $(s_i, \gamma_i)$ to embedding vectors $v_i$. For the Lorenz equations (6) the matrices $D\varphi_m$ are close to singular, causing a large fraction of false recurrences. A theory of information flow based on properties of $D\varphi_m$ for stationary dynamical systems has been developed by Casdagli et al. [3].
Fig. 2. Scaling laws for the probability of false recurrences in time series generated from the family of tent maps (5) and Lorenz equations (6) whilst varying the parameter $\gamma$ according to a linearly trending driving force $\gamma_i = i/N$. (a) (resp. (b)) Above the diagonal: RP of the tent map (resp. Lorenz) time series for $r = 0.02$ and $m = 2$ (resp. $m = 3$). Below the diagonal: RP for $m = 3$ (resp. $m = 5$). (c) (resp. (d)) Dotted lines: dependence of the correlation integral $C_m(r)$ on $r$ for the tent map (resp. Lorenz) time series with $m = 2, 3$ (resp. $m = 3, 4, 5$). Solid lines: dependence of the probability of generating a false recurrence $P_m^F(r, \epsilon)$ on $r$ for $\epsilon = 0.1$ and $m = 2, 3$ (resp. $m = 3, 4, 5$). Dashed lines: scaling laws for $P_m^F(r, \epsilon)$ predicted by (8). (e) (resp. (f)) Dependence of the fraction of recurrences that are false $P_m^F(r, kr)/C_m(r)$ on $k$ for the tent map (resp. Lorenz) time series with $m = 2, 3, 4, 5$ (resp. $m = 3, 4, \ldots, 8$) and $r = 0.02$ (Solid curves) and $r = 0.01$ (Dashed curves).
2.4. Reconstruction of the driving force

It was demonstrated above that for time series generated by deterministic systems driven by sufficiently slowly varying external forces, RPs of the time series can be used to reconstruct the RP of the driving force. We now investigate the extent to which the RP of the driving force can be used to reconstruct the functional form of the driving force.

As in Section 2.3 we assume the driving force is sampled from a continuous function $G : [0, 1] \rightarrow [0, 1]^p$ according to $y_i = G(i/N)$. In the limit as $N$ tends to infinity and $r$ tends to zero, the normalized RP of the driving force defined by $\{(i/N, j/N) \mid \|\gamma_i - \gamma_j\| < r\}$ converges onto the compact set of recurrences $RP(G)$ defined by

$$RP(G) = \{(s, t) \in [0, 1]^2 \mid G(s) = G(t)\}. \quad (10)$$

In this section we investigate the extent to which $RP(G)$ can be used to reconstruct the functional form of $G$. The most that can be hoped for is that $RP(G)$ determines $G$ up to an arbitrary invertible coordinate transformation $\phi : [0, 1] \rightarrow [0, 1]$. This is because if $\phi$ is invertible, $G(s) = G(t) \Rightarrow \phi G(s) = \phi G(t)$, so that from (10) we have $\phi G = \phi R(P(G))$.

In the case of a one-dimensional driving force $G : [0, 1] \rightarrow [0, 1]$ it is straightforward to show that $RP(G)$ determines $G$ uniquely up to an arbitrary invertible coordinate transformation as follows. Suppose that $F$ is another driving force satisfying $RP(F) = RP(G)$, and denote by $[G_{\min}, G_{\max}]$ (resp. $[F_{\min}, F_{\max}]$) the range of $G$ (resp. $F$). Then the set $\{(G(t), F(t)) \mid t \in [0, 1]\}$ forms the graph of an invertible function $F : [G_{\min}, G_{\max}] \rightarrow [F_{\min}, F_{\max}]$, since

$$F(s) = F(t) \Rightarrow (s, t) \in RP(F) \Rightarrow (s, t) \in RP(G) \Rightarrow G(s) = G(t).$$

By construction $\phi$ satisfies $F = \phi G$, and the result follows. Note that this proof does not give an explicit algorithm for reconstructing the functional form of $G$ from $RP(G)$.

In the case of a smooth one-dimensional driving force $G : [0, 1] \rightarrow [0, 1]$, an explicit reconstruction algorithm exists. Intuition for the algorithm can be obtained by considering the RP in Fig. 1(e) of the driving force in Fig. 1(c). A blowup of the RP for $i < 23000, j < 2100$ is given by the dots in Fig. 1(d). The solid curve in Fig. 1(d) is a blowup of the driving force, and appears to be related to the RP by an invertible coordinate transformation. The reason for this is as follows: Suppose the first turning point of the driving force $G(t)$ occurs at $t = t_1$, and that $G(t) \in [G(0), G(t_1)]$ for $t \in [0, t_2]$. Then the set $\Gamma(F)$ defined by

$$\Gamma(F) = RP(G) \cap [0, t_2] \times [0, t_1]$$

forms the graph of a well defined function $F : [0, t_2] \rightarrow [0, t_1]$. This is because if $(s, t), (s, t') \in \Gamma(F)$ then $G(s) = G(t) = G(t')$ so that $t = t'$ since $G$ is monotonic on $[0, t_1]$. For example in Fig. 1(d) the recurrences fill out the graph of a well-defined function. The function $F$ is a reconstruction of $G(t)$ for $t \in [0, t_2]$, in the sense that there exists an invertible function $\phi$ such that $\phi(F(t)) = G(t)$ for $t \in [0, t_2]$. In fact $\phi$ is given by $\phi(t) = G(t)$ for $t \in [0, t_1]$ because we have $(t, F(t)) \in RP(G)$ for $t \in [0, t_2]$ so that, by definition of $RP(G)$, $G(t) = G(F(t))$. The algorithm for reconstructing $G$ can be completed by extending the domain of the function $F$ from the interval $[0, t_2]$ to $[0, 1]$ by extracting additional subsets of $RP(G)$ and patching them onto $\Gamma(F)$. In practice, it should be possible to improve on this algorithm by constraining the function $F$ using all of the information in the RP.

We now consider RPs for a wider range of driving forces. The RP below the diagonal in Fig. 3(a) was obtained from the random walk driving force shown in Fig. 3(c), and exhibits a fractal structure. The RP above the diagonal in

---

5 The dots were obtained from the RP of the tent map time series in Fig. 1(b). However, ignoring a few false recurrences, identical results are obtained from the RP of the driving force in Fig. 1(e).
Fig. 3. Recurrence plot reconstruction for time series generated from the Lorenz equations (6) whilst varying the parameter $\gamma$ according to the random walk shown in (c). (a) Below the diagonal: RP of the driving force shown in (c) for $r = 0.02$ and $m = 1$. Above the diagonal: RP of the Lorenz time series for $r = 0.02$ and $m = 5$. (b) Below the diagonal: RP of the pair of driving forces shown in (d) for $r = 0.02$. Above the diagonal: RP of a time series constructed from a linear combination of the Lorenz time series analyzed in (a) with the second driving force in (d), for $r = 0.02$ and $m = 5$.

Fig. 3(a) was obtained from a time series generated by forcing the Lorenz equations (6) with this driving force, and gives a moderately good reconstruction of the RP of the driving force, as expected from the results of Section 2.3. It is possible to infer the box-counting dimension of the graph of the driving force from the RPs if the driving force varies sufficiently slowly, in the following sense. The box-counting dimension of the graph $\Gamma(G)$ of the driving force $G$ is equal to the box-counting dimension of $RP(G)$. This can be seen as follows. Let $d$ (resp. $d'$) be the box-counting dimension of $RP(G)$ (resp. $\Gamma(G)$). For fixed $t \in [0, 1]$, define the set $R_t = \{ s \in [0, 1] \mid G(s) = G(t) \}$. The set $R_t \times \{t\}$ (resp. $R_t \times \{G(t)\}$) is the intersection of a horizontal line with $RP(G)$ (resp. $\Gamma(G)$), which for typical $t$ has box-counting dimension $d - 1$ (resp. $d' - 1$). It follows that $d' = d$ because the box-counting dimensions of the sets $R_t \times \{t\}$ and $R_t \times \{G(t)\}$ are equal. The box-counting dimension of the graph of a random walk is $\frac{3}{2}$. Thus, in principle, the box-counting dimension of the RPs in Fig. 3(a) should be equal to about $\frac{3}{2}$, and it could be inferred that the graph of the driving force has box-counting dimension $\frac{3}{2}$. In practice, the most one could expect to infer is that the driving force exhibits structure on multiple time scales.

Finally, we consider the case of two driving forces, i.e. $\gamma_i = G(i/N)$, where $G = (G_1, G_2)$ is a continuous function from $[0, 1]$ to $[0, 1]^2$. We have $RP(G) = RP(G_1) \cap RP(G_2)$ because

$$G(s) = G(t) \iff G_1(s) = G_1(t) \text{ and } G_2(s) = G_2(t).$$
An example is shown in Fig. 3(b). The RP below the diagonal was obtained from the pair of driving forces shown in Fig. 3(d) according to \((i, j) \cap | \gamma_i - \gamma_j | < r\), where \(r = 0.02\). The RP above the diagonal is for a time series \(x_1', x_2', \ldots, x_N'\) obtained from the Lorenz time series \(x_1, x_2, \ldots, x_N\) analyzed in Fig. 3(a) and the second driving force shown in Fig. 3(d) according to \(x_i' = x_i + c\gamma_i^{(2)}\) where \(c\) is equal to 0.1 times the standard deviation of \(x_1, x_2, \ldots, x_N\) divided by the standard deviation of \(\gamma_1^{(2)}, \gamma_2^{(2)}, \ldots, \gamma_N^{(2)}\). The RP of the time series gives a moderately good reconstruction of the RP of the pair of driving forces. In principle, the box-counting dimension of the RPs in Fig. 3(a) should be equal to about 1, and inferences could be made concerning the fractal properties of the driving forces. In practice, the most one could expect to infer is that there is more than one driving force acting, and that the driving forces are not smooth functions.

2.5. Relationship to other techniques

The problem of reconstructing a slowly varying driving force from a deterministic time series is closely related to the problem of extracting a signal from a chaotic background. The latter problem has been addressed by Stark and Arumugam [39] and Taylor [43] with prediction techniques, in the case of signals added to a deterministic time series. The problem considered in this section is more general. However, the techniques used in this section will almost certainly be less accurate than prediction techniques for the special case of signals added to a deterministic time series.

3. Meta-recurrence plots

In this section it is shown that RPs can be modified to analyze time series generated from stochastic\(^6\) dynamical systems driven by slowly varying external forces. We first show that windowed recurrence plots (WRPs) can detect changes in the driving force which are not detected by RPs. Second, linearized recurrence plots (LWRPs) are computed that detect changes in the driving force which affect linear properties of the time series. Third, it is shown how to transform WRPs into meta-recurrence plots (MRPs) that can be used to reconstruct the RP of the driving force if it varies sufficiently slowly.

3.1. Windowed recurrence plots

The WRPs can be motivated by the following numerical experiment. Fig. 4(a) shows the RP for \(m = 2\) and \(r = 0.5\) of a time series \(x_1, x_2, \ldots, x_N\) of length \(N = 60000\) generated by iterating the stochastic system

\[
x_{i+1} = a_i x_i + (1 - a_i^2)^{1/2} n_i,
\]

where the \(n_i\) are independently normally distributed with unit variance, and \(a_i = 0.8(2\gamma_i - 1)\), where \(\gamma_i\) is the driving force shown in Fig. 1(c). Changes in the driving force are not detectable in the RP. A WRP for the time series is shown below the diagonal in Fig. 4(b) and exhibits nontrivial structure. A WRP is obtained from an RP by choosing a window width \(w\) and covering the RP with a grid of squares of size \(w \times w\). The density of recurrences in each square is then plotted as a function of square location \((I, J) \in [1, N_w]^2\), where \(N_w\) is the greatest integer less than \(N_m/w\). A grayscale is used in Fig. 4(b) in which denser locations are shaded more darkly; if possible a

---

\(^6\) For low-dimensional deterministic systems the techniques of Section 2 can be applied. However, if a times series is generated from a high-dimensional deterministic system, it will be necessary to project it to a lower dimension, inducing stochastic behavior.
color encoding is preferable. The density of the WRP at location \((I, J)\) is given by the cross-correlation integral, \(C_m(r, I, J)\), defined, for \(I \neq J\), as

\[
C_m(r, I, J) = \frac{1}{w^2} \sum_{i=1+(I-1)w}^{I} \sum_{j=1+(J-1)w}^{J} \theta(r - \|v_i - v_j\|).
\] (12)

In the case \(I = J\) a similar definition to (12) is used in which it is required that \(j < i\), and the normalization factor \(\frac{1}{2}w(w - 1)\) replaces \(w^2\), so that the \(C_m(r, I, I)\) are correlation integrals of segments of the time series. Cross-correlation integrals were introduced by Kantz [18] to quantify the distance between fractals.
WRPs have the disadvantage compared to RPs that an additional parameter, the window width $w$, must be determined. However, as shown in Figs. 4(a) and (b), there is much to be gained by analyzing time series with WRPs. In the above example it was necessary to use a relatively large resolution $r$ to obtain a WRP with distinctive structure. At large resolutions $r$ it is possible to display only a small fraction of the total number of recurrences in an RP of a time series of length 60,000. In Fig. 4(a) only 0.01% of the recurrences are shown, so that the displayed RP would not be fully covered with black dots. By contrast, WRPs summarize information in all the recurrences. An efficient algorithm for computing WRPs is outlined in Appendix A.2.

3.2. Linearized recurrence plots

It is possible to detect very general changes in a time series with WRPs. In many examples, such as that in Section 3.1, the detected changes may be due to changes in simple "linear" quantities, such as the windowed mean, variance or autocorrelation function. Consequently, it is desirable to compute a linearized version of WRPs in order to assess the extent to which changes detected by the WRP are due to nonlinear changes in the system.

LWRPs are defined by approximating successive windows of $w$ embedding vectors by multivariate Gaussian distributions $p_1, p_2, \ldots, p_{n_w}$, with means and covariances given by the means and covariances of the embedding vectors. Cross-correlation integrals between pairs of the multivariate Gaussian distributions are then computed analytically in the limit $r \to 0$ as follows. First, the double sum (12) is evaluated as an integral by computing the probability that two randomly chosen vectors $v_I, v_J \in \mathbb{R}^m$ from $p_I$ and $p_J$ lie within distance $r$ of each other. Suppose $v_I = v$; this occurs with probability $p_I(v)$. For small $r$, the probability that $v_J$ lies within distance $r$ of $v$ is given by the volume of the ball of radius $r$ at $v$ times $p_J(v)$. Using the maximum norm, and integrating over $v \in \mathbb{R}^m$, we obtain

$$C_m(r, I, J) = (2r)^m \int p_I(v)p_J(v) \, dv.$$  \hspace{1cm} (13)

Second, let $\mu_I, \Sigma_I$ and $\mu_J, \Sigma_J$ denote the means and covariances of $p_I$ and $p_J$ and let $|\Sigma|$ denote the determinant of $\Sigma$. The Gaussian distribution formula

$$p(v) = (2\pi)^{-m/2} |\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(v - \mu)^\prime \Sigma^{-1}(v - \mu)\right)$$

Suppose $v_I = v$; this occurs with probability $p_I(v)$. For small $r$, the probability that $v_J$ lies within distance $r$ of $v$ is given by the volume of the ball of radius $r$ at $v$ times $p_J(v)$. Using the maximum norm, and integrating over $v \in \mathbb{R}^m$, we obtain

$$C_m(r, I, J) = (2r)^m \int p_I(v)p_J(v) \, dv.$$  \hspace{1cm} (13)

Second, let $\mu_I, \Sigma_I$ and $\mu_J, \Sigma_J$ denote the means and covariances of $p_I$ and $p_J$ and let $|\Sigma|$ denote the determinant of $\Sigma$. The Gaussian distribution formula

$$p(v) = (2\pi)^{-m/2} |\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(v - \mu)^\prime \Sigma^{-1}(v - \mu)\right)$$

Suppose $v_I = v$; this occurs with probability $p_I(v)$. For small $r$, the probability that $v_J$ lies within distance $r$ of $v$ is given by the volume of the ball of radius $r$ at $v$ times $p_J(v)$. Using the maximum norm, and integrating over $v \in \mathbb{R}^m$, we obtain

$$C_m(r, I, J) = (2r)^m \int p_I(v)p_J(v) \, dv.$$  \hspace{1cm} (13)

Second, let $\mu_I, \Sigma_I$ and $\mu_J, \Sigma_J$ denote the means and covariances of $p_I$ and $p_J$ and let $|\Sigma|$ denote the determinant of $\Sigma$. The Gaussian distribution formula

$$p(v) = (2\pi)^{-m/2} |\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(v - \mu)^\prime \Sigma^{-1}(v - \mu)\right)$$
can be used to evaluate (13) to obtain

\[ C_m(r, I, J) = (2r)^m (2\pi)^{-m/2} |\Sigma_l + \Sigma_J|^{-1/2} \exp(-\frac{1}{2} \beta), \]

(14)

where \( \beta = 0 \) if \( \mu_l = \mu_J \). Otherwise \( \beta \) is given by

\[ \beta = \mu_J^{-1} - \mu_l^{-1} + \mu_J^{-1} - \mu_l^{-1} (\Sigma_l^{-1} + \Sigma_J^{-1})^{-1} (\Sigma_l^{-1} + \Sigma_J^{-1})^{-1} (\Sigma_l^{-1} + \Sigma_J^{-1})^{-1}. \]

An LWRP is identical to a WRP, except that the cross-correlation integral (12) defining the grayscale of the WRP at location \((I, J)\) is replaced by (14). An LWRP for the example in Section 3.1 is shown above the diagonal in Fig. 4(b). As expected for this example the LWRP is similar to the WRP, indicating that the changes in the driving force detected by the WRP are due to changes in the linear properties of the time series.

### 3.3. Meta-recurrence plots

The WRP and LWRP shown in Fig. 4(b) cannot be used directly to reconstruct the RP of the driving force. This is because the cross-correlation integrals on which they are based do not provide a good measure of distance between windows of embedding vectors for stochastic systems. Consequently the WRP and LWRP in Fig. 4(b) exhibit less structure than the RPs of the deterministic systems shown in Fig. 1. A simple transformation can be applied to WRPs and LWRPs that solves this problem.

An MRP is obtained from a WRP by replacing the cross-correlation integrals \( C_m(r, I, J) \) by the distance measure \( D_m(r, I, J) \) defined by

\[ D_m(r, I, J) = \frac{1}{(2r)^m} (C_m(r, I, I) + C_m(r, J, J) - 2C_m(r, I, J)). \]

(15)

\( D_m(r, I, J) \) is equal to the \( L^2 \) norm \( \int (p_I - p_J)^2 dv \) between the probability distributions \( p_I \) and \( p_J \) filled out by the pair of sequences of \( m \)-dimensional embedding vectors \( \{v_i, i \in [1 + (I - 1)w, Iw]\} \) and \( \{v_j, j \in [1 + (J - 1)w, Jw]\} \) in the limit of large \( w \) and small \( r \). This can be seen by combining (13) with the identity

\[ \int (p_I - p_J)^2 dv = \int p_I^2 dv + \int p_J^2 dv - 2 \int p_I p_J dv. \]

(16)

Since the MRP is based on a well-defined measure of distance, the thresholded MRP defined by

\[ \{ (I, J) | D_m(r, I, J) < \delta \} \]

(17)

can, in theory, be used to reconstruct the RP of the driving force if it varies sufficiently slowly, as follows. For small \( \delta \) and \( r \), and large \( w \), we have \( p_I \approx p_J \) whenever \((I, J)\) lies in the thresholded MRP. Therefore \( \gamma_{Iw} \approx \gamma_{Jw} \) if the driving force interacts nondegenerately with the time series (i.e., \( \gamma_{Iw} \neq \gamma_{Jw} \Rightarrow p_{I'} \neq p_{J'} \) for all \( I', J' \) at some embedding dimension \( m \)). Of course in practice, for stochastic systems, it is necessary to calculate \( D_m(r, I, J) \) by estimating cross-correlation integrals \( C_m(r, I, J) \) at low embedding dimensions \( m \) and large resolution \( r \). For example this is the case in Fig. 4(b).

An MRP for the example in Section 3.1 is shown below the diagonal in Fig. 4(c), and was thresholded\(^7\) by taking \( \delta = 10^{-3} \) in (17). Some of the structure of the RP of the driving force shown in Fig. 1(d) is revealed by the MRP. A more accurate\(^8\) reconstruction of the RP of the driving force is given by the linearized meta-recurrence plot (LMRP), shown above the diagonal in Fig. 4(c). The LMRP was obtained by applying the transformation (15) to the LWRP in

---

\(^7\) Negative values of \( D_m(r, I, J) \) that occur due to estimation errors are set to zero.

\(^8\) The concept of false recurrences can be extended to MRPs and LMRPs to quantify the accuracy of the reconstruction of the RP of the driving force. However, in this paper we will rely on visual inspection of MRPs and LMRPs.
Fig. 4(b). The accuracy of the LMRP is due to the superiority of the parametric estimate for $\int (p_I - p_J)^2 \, dv$ based on (14) over the nonparametric estimate based on (12) in this example, as follows. The Gaussian approximation used to derive (14) holds exactly for this example, so that errors in the estimate of $C_m(r, I, J)$ arise solely from errors in estimating $p_I$, $p_J$ and $x_I$, $x_J$. These quantities may be estimated with accuracy of order $w^{-1/2}$, where $w$ is the window width. By contrast, in the nonparametric estimate (12) much of the information in delay vectors is discarded, since contributions to the double sum are made only by delay vectors that lie within distance $r$ of each other.

3.4. Nonlinear examples

Fig. 4(d) shows the WRP and LWRP of the tent map time series of Section 2.2 for $m = 3$, $r = 0.02$ and $w = 200$. Observe that although this example is highly nonlinear, the LWRP detects several changes in the time series. This effect is even more pronounced in the thresholded LMRP shown in Fig. 4(e), which has an intriguing similarity to the RP of the driving force shown in Fig. 1(d). There are two reasons for this. First, the RP of the driving force is a subset of the thresholded LMRP in the following sense. If the driving forces acting in the $I$th and $J$th windows are similar, then $\mu_I \approx \mu_J$ and $\Sigma_I \approx \Sigma_J$, so that $(I, J)$ will be in the thresholded LMRP at small values of the threshold $\delta$. Second, although means and covariances cannot be used to uniquely specify nonlinear systems, in this example there is a two-to-one relationship: at a fixed parameter $\gamma$, the lag-1 autocorrelation of a time series generated from (5) is equal to $-\frac{1}{2} \gamma (1 - \gamma) \sqrt{3}$. Therefore, if the thresholded LMRP is used to reconstruct the RP of the driving force, half of the recurrences in the RP will be spurious.

Fig. 4(f) shows the thresholded LMRP of the Lorenz time series of Section 2.2, for $m = 5$, $r = 0.02$ and $w = 200$. The LMRP gives a recognizable reconstruction of the RP of the driving force. This is because, in this example, there is an approximately one-to-one relationship between the parameter $\gamma$ and the lag-1 autocorrelation of a time series generated from (6) at a fixed parameter $\gamma$. However, the reconstruction is not as accurate as that given by the RP shown below the diagonal in Fig. 1(g).

These two examples show that changes in nonstationary nonlinear time series can be detected with LMRPs. Examples for which this is not possible can be constructed by removing linear structure in these time series with a time varying linear filter. In these deterministic examples RPs could then be used to reconstruct the driving force. For stochastic examples MRPs would be required. The advantage of MRPs is that they are based on parametric estimates of distance requiring relatively few data points. The advantage of MRPs is that they are based on nonparametric estimates of distance that have a more general validity. However, nonparametric estimation requires large quantities of data, particularly in high dimensions. MRPs could be generalized using distance measures constructed from higher order correlations in the time series. For example Kadtke [17] uses polynomial regression to construct measures of distance; “semi-parametric” measures of distance could also be constructed with more general multivariate function approximation techniques [14,45]. Such “generalized LMRPs” are likely to combine the advantages of LMRPs and MRPs, but are beyond the scope of this paper.

3.5. Relationship to other techniques

Manuca and Savit [25] have proposed techniques for characterizing nonstationarity in time series based on computing arrays of distances between segments of the time series. MRPs are a simple example of such a technique, with the following theoretical advantages. First, we are able to make explicit the connection between MRPs and RPs; this allows us to make simple statements concerning the reconstruction of the RP of the driving force. Second, it is straightforward to compute LMRPs that measure the extent to which MRPs characterize nonstationarities due to
linear changes in the time series. On the other hand, the techniques of Manuca and Savit contain several innovations for efficiently characterizing nonstationarities of a form that are beyond the scope of this paper.

Schreiber [36] has proposed a technique for characterizing nonstationarity in time series based on computing arrays of cross-prediction errors between segments of the time series. Cross-prediction has also been used as a test for stationarity by Smith et al. [37] and Sugihara et al. [40]. MRPs have the practical advantage of being less computationally intensive, and the theoretical advantage of being based on a well-defined measure of distance. On the other hand, the cross-prediction technique is likely to be less data intensive and is more directly related to the nonlinear prediction algorithm discussed in Section 4.

4. Recurrence plots and prediction

In this section we investigate the accuracy of a local linear prediction algorithm for time series generated from dynamical systems driven by slowly varying external forces. A natural strategy to employ in predicting the current point of such a time series is to weight information from the recent past more heavily than that from the distant past. We investigate the advantages and limitations of this strategy numerically. RPs and MRPs are used both to interpret the results and to suggest improved prediction algorithms.

4.1. Local linear prediction

In this section we consider a one-step ahead local linear prediction algorithm in which the prediction of the current point of a time series is based on information in the most recent \( w \) embedding vectors, defined as follows. First a test set is selected, consisting of the last \( \mathcal{N}_l \) points of the time series \( x_1, x_2, \ldots, x_N \), and a window width \( w \) and an embedding dimension \( m \) are chosen. Second, for each point \( x_i \) in the test set, the \( k \) nearest neighbors \( v_{j_1-m}, v_{j_2-m}, \ldots, v_{j_k-m} \) to the embedding vector \( v_{i-m} \) are computed, where \( i-w \leq j_n < i \) for \( n = 1, 2, \ldots, k \). Third, a linear model with parameters \( a_0, a_1, \ldots, a_m \) is fitted by minimizing

\[
\sum_{n=1}^{k} \left( x_{j_n} - a_0 - \sum_{l=1}^{m} a_l x_{j_n-l} \right)^2.
\]

Fifth, the prediction \( \hat{x}_i \) for \( x_i \) is defined by

\[
\hat{x}_i = a_0 + \sum_{l=1}^{m} a_l x_{i-l}.
\]

Fifth, the prediction error \( E_m(k, w) \) is computed according to

\[
E_m(k, w)^2 = \mathcal{N}_l^{-1} \sum_{i=1}^{\mathcal{N}_l} (x_i - \hat{x}_i)^2 / \sigma^2,
\]

where \( \sigma^2 \) is the variance of the test set. The parameter \( k \) is allowed to vary from \( k = 2(m + 1) \) to \( k = w \), so that a range of prediction techniques from nonlinear deterministic to linear stochastic extremes is explored [2].

The prediction algorithm can be related to RPs because the \( k \) nearest neighbors \( v_{j_1}, v_{j_2}, \ldots, v_{j_k} \) used to predict an embedding vector \( v_{i} \) are related to the existence of recurrences at the locations \( (i, j_1), (i, j_2), \ldots, (i, j_k) \). Suppose the RP of the time series is an accurate reconstruction of the RP of the driving force. Then by definition,

9 The relationship can be made exact by using the \( k \) nearest neighbor definition of RPs of Eckmann et al. [10].
however large \( w \) is taken, very few of these recurrences will be false recurrences, and the \( k \) nearest neighbors will have the desirable property that the driving force acting at the times \( j_1, j_2, \ldots, j_k \) is approximately equal to the driving force acting at time \( i \). By contrast, if the RP of the time series is not an accurate reconstruction of the RP of the driving force, then in order to remove false recurrences from the \( k \) nearest neighbors a small \( w \) is required, of size the order of the time scale of variation of the driving force.

4.2. Deterministic time series

Fig. 5(a) shows the results of applying the local linear prediction algorithm to the Lorenz time series of Section 2.1. The test set was chosen to be of size \( N_f = 20000 \). The prediction errors \( E_m(k, w) \) are plotted against \( k \) for \( m = 3, 4, 5 \) and \( w = 40, 400, 4000, 39990 \). As expected for a deterministic system, predictions at the small \( k \) extreme are the most accurate. More surprisingly predictions at large \( w \), that use the most information in the past, are the most accurate. There are two reasons for this. First, the dynamics acting at the time a typical prediction is made occurs several times throughout the time series due to the form of the driving force shown in Fig. 1(c). All this information is exploited by using a large window width \( w \). Second, from the results of Section 2, the RP of this time series is an accurate reconstruction of the RP of the driving force for \( m > 3 \). Therefore predictions at large \( w \) are relatively unaffected by false recurrences.

In fact a scaling law of the form

\[
E_m(k, w) \sim (k/w)^{2/D}
\]  

(21)

appears to hold in Fig. 5(a) for \( m > 3 \), where \( D \approx 3 \). Scaling laws of this form apply to the local linear prediction of time series of length \( w \) generated from attractors of stationary deterministic systems, where \( D \) is one of the generalized dimensions of the attractor \([2, 12]\). The scaling law (21) applies in the case of deterministic systems driven by slowly varying external forces in the limit \( k \ll w \ll N \) for \( m > 2D_0 \), where \( D_0 \) (resp. \( D \)) is the box-counting dimension (resp. one of the generalized dimensions) of the set of points \( \{(s_i, \gamma_i), i \in \mathbb{Z}^+\} \), and is derived in Appendix A.3. The scaling law (21) indicates that the results obtained in Fig. 5(a) for the Lorenz time series are likely to be typical for a wide range of nonstationary deterministic systems. However, note that the scaling law (21) applies specifically to the prediction error defined by (20). We consider an alternative measure of predictability that does not satisfy the scaling law (21) in Appendix A.4.

4.3. Stochastic time series

Fig. 5(b) shows the results of applying the local linear prediction algorithm to the stochastic time series of Section 3.1. The test set was chosen to be of size \( N_f = 20000 \). The prediction errors \( E_m(k, w) \) are plotted against \( k \) for \( m = 1, 2, 3 \) and \( w = 40, 400, 4000, 39990 \). As expected for a time series generated by the slowly varying first-order linear system (11), linear (large \( k \)) prediction with \( m = 1 \) is optimal at small \( w \). The choice of an optimal value for \( w \) involves two conflicting effects. For small \( w \) the prediction error is high because there are not enough data points to average out the effects of the noise term \( n_i \) in (11). For large \( w \) the prediction error is high because the RP shown in Fig. 4(a) is not a good reconstruction of the RP of the driving force, so that predictions are affected by false recurrences. In this example the optimal value for \( w \) occurs near \( w = 400 \). In general, the optimal value for \( w \) will be larger for more complicated stochastic systems, in order to have a sufficient number of data points to estimate the dynamics.

In Section 3.3 it was shown that MRPs of time series generated from stochastic systems driven by slowly varying external forces give more accurate reconstructions of the RP of the driving force than RPs of the time series. For example the MRP in Fig. 4(c) is superior to the RP in Fig. 4(a). This raises the possibility of obtaining improved
predictions for such time series by using information in the MRP to reduce the effect of false recurrences, as follows. In order to predict the current point of the time series, the $k$ nearest neighbors are computed with a metric that combines the distance between embedding vectors with the "meta-distance" (15) used to define the MRP. Although we have not implemented such an algorithm, encouraging results have been obtained by Weigend et al. [46] with a similar strategy called "gated experts". In this strategy each expert is adapted to predict segments of the
time series with similar dynamics. Choosing between the experts is analogous to comparing meta-distances in the strategy proposed here.

4.4. Local prediction error

RPs and MRPs can be used to understand other phenomena that occur in the prediction of nonstationary time series. Fig. 5(c) (resp. 5(d)) shows the $i$-dependence of the local prediction error $E_l(i)$ defined by

$$E_l(i) = \frac{1}{l} \sum_{j=i-l+1}^{i} \frac{(\hat{x}_j - x_j)^2}{\sigma^2}$$

(22)

for the Lorenz (resp. stochastic) time series, with $l = 200$. The predictions $\hat{x}_j$ were obtained by fixing the parameters $m, k, w$ to give the smallest global prediction error $E_m(k, w)$. There are large fluctuations in the local prediction error, corresponding to extreme values of the driving force shown in Fig. 5(e). Prediction accuracy decreases at extreme values of the driving force for the Lorenz time series. This is because extreme values occurred infrequently in the past so that it is difficult to learn the dynamics at these times. Prediction accuracy increases at extreme values of the driving force for the stochastic time series. This is because, first, the predictions for the stochastic time series use a small $w$, so are unaffected by the fact that extreme values occur infrequently in the past. Second, the dynamics in (11) is less stochastic at extreme values of the driving force, and is thus easier to predict.

The properties of the local prediction error shown in Figs. 5(c) and (d) are easy to understand if it is possible to observe the driving force. However, the basic assumption in this paper is that the driving force is unobserved. In this case, as shown in Sections 2 and 3, one can use RPs and MRPs to reconstruct the functional form of the driving force. Fluctuations in the local prediction error can then be interpreted as being due to the recurrency properties of an underlying driving force. Moreover, one can attempt to predict future values of the local prediction error by predicting the driving force. For example, in the case of Fig. 5(e), it is likely that the next 1000 values of the driving force will not be extreme values, so that the local prediction errors in Fig. 5(c) (resp. 5(d)) will remain low (resp. high) during this period.

5. Change point detection and hypothesis testing

In this section we present a technique for detecting change points in a time series. A surrogate data technique is used to assess statistical significance.

5.1. Change point detection

The change point detection problem can be motivated by the following numerical experiment. Let $x_1, x_2, \ldots, x_i, \ldots, x_{5004}$ be a time series generated by integrating the Lorenz equations (6) and sampling the $x$ component every 0.16 units of time, with the parameter $\gamma = \frac{1}{3}$ for $i < 3000$ and $\gamma = \frac{5}{12}$ for $i \geq 3000$. A WRP of the time series is shown in Fig. 6(a) for $m = 5, r = 0.02$ and $w = 200$. Observe that large values of the cross-correlation integrals $C_m(r, I, J)$ only occur for $(I, J)$ in the two square regions $[1, 15]^2$ and $[16, 25]^2$. However, one might suspect this to be a random effect, and desire to test for the existence of a change point with a statistically robust technique.

The change point detector $d_m(r, k)$ considered in this section is defined by the conditional probability

$$d_m(r, k) = P(\|v_i - v_j\| < r | (i, j) \in [1, k]^2 \cup [k + 1, N_m]^2) / P(\|v_i - v_j\| < r)$$

(23)
Fig. 6. Change point detection for a time series generated by the Lorenz equations. (a) WRP for \( m = 5, r = 0.02 \) and \( w = 200 \). (b) Solid curve: dependence of the change point detector \( d_m(r, wK) \) on \( K \) for the time series analyzed in (a). Dotted curves: change point detectors for 40 surrogate data sets.

for \( 0 < k < N_m \), and by \( d_m(r, k) = 1 \) for \( k = 0 \) or \( k = N_m \). A change point \( k^* \) is detected by maximizing \( d_m(r, k) \) over \( k \). Geometrically, this is equivalent to maximizing the density of recurrences in the two squares \([1, k]^2\) and \([k+1, N_m]^2\) in the RP corresponding to Fig. 6(a). The change point detector can be calculated from the cross-correlation integrals shown in Fig. 6(a) according to

\[
d_m(r, wK) = (K^2 + (N_w - K)^2)^{-1} \left( \sum_{I,J=1}^{K} C_m(r, I, J) + \sum_{I,J=K+1}^{N_w} C_m(r, I, J) \right) / C_m(r),
\]

where \( N_w = 25 \) is the number of windows. The solid curve in Fig. 6(b) shows the \( K \) dependence of \( d_m(r, wK) \), and a change point is detected at \( K = 15 \). This quantifies our earlier observation that large values of \( C_m(r, I, J) \) occur only for \((I, J)\) in the two square regions \([1, 15]^2\) and \([16, 25]^2\) in Fig. 6(a). The magnitude of \( d_m(r, k^*) \) can be interpreted by labeling embedding vectors \( v_i \) by “0” if \( i \leq k^* \) and “1” if \( i > k^* \). Then it is straightforward to show that two randomly selected embedding vectors lying within distance \( r \) of each other are \( d_m(r, k^*) \) times as likely to have the same label as two randomly chosen embedding vectors.

Change points are accurately detected by \( d_m(r, k) \) for a wide range of time series if the time series is sufficiently long, in the following sense. Suppose the time series is a concatenation of two stationary time series of lengths \( \alpha N \) and \( (1 - \alpha)N \), with embedding vectors that converge onto two distinct probability distributions \( p_1 \) and \( p_2 \). Suppose that the following condition holds:

\[
C < \min(A, B),
\]

where \( C \) is the cross-correlation integral of \( p_1 \) and \( p_2 \), and \( A \) and \( B \) are the correlation integrals of \( p_1 \) and \( p_2 \), at a fixed resolution \( r \). Then as \( N \) tends to infinity, the change point detector \( d_m(r, \kappa N) \) converges onto a piecewise smooth function of \( \kappa \) with a unique maximum at \( \kappa = \alpha \) at which the derivative is discontinuous. For a proof see Appendix A.5. Condition (25) holds in most cases of interest, but exceptions can be found. For example it fails to hold if \( p_1 \) and \( p_2 \) are one-dimensional Gaussian distributions with distinct variances. It is shown in Appendix A.5 that if condition (25) fails to hold then the change point detector fails to converge to the correct change point in the limit of large \( N \), and an alternative change point detector is considered.
5.2. Hypothesis testing

For a stationary time series \( d_m(r, k) \approx 1 \) for all \( k \). It follows that if \( d_m(r, k^*) \approx 1 \) for a time series, the detected change point at \( k = k^* \) may be spurious. We use a surrogate data technique, related to the “stationary bootstrap” [29], to test the statistical significance of detected change points as follows. First, fix a time scale \( w \) and an embedding dimension \( w \) and let \( N_w \) be the greatest integer less than \( N_m / w \). Second, let \( \sigma \) be a random permutation of the set \( \{0, 1, \ldots, N_w - 1\} \), and shuffle embedding vectors in blocks of length \( w \) to produce a surrogate data set consisting of embedding vectors with the time ordering

\[
\begin{align*}
\sigma(0) + 1, & \quad \sigma(0) + 2, \ldots, \sigma(0) + w, \\
\sigma(1) + 1, & \quad \sigma(1) + 2, \ldots, \sigma(1) + w, \\
\ldots & \\
\sigma(N_w - 1) + 1, & \quad \sigma(N_w - 1) + 2, \ldots, \sigma(N_w - 1) + w.
\end{align*}
\]

Third, compute a statistic for the surrogate data set. To test the significance of detected change points the statistic is chosen to be the function \( d_m(r, k) \) or its maximum value \( d_m(r, k^*) \). Fourth, repeat the above procedure several times and compare the statistic for the original time series to those for the surrogate data sets. If there is a significant difference, we can conclude that the original time series has structure on time scales of order \( w \) or greater, i.e. is effectively nonstationary at these time scales. This is because, by construction, the surrogate data sets preserve all the structure in the original time series on time scales of order less than \( w \), so that a significant difference in the statistic must be due to the existence of structure on time scales of order \( w \) or greater.

The dotted curves in Fig. 6(b) show change point detectors \( d_m(r, k) \) computed for 40 surrogate data sets, with \( m = 5, r = 0.02 \) and \( w = 200 \). Note that the surrogate data technique is not numerically intensive: it is equivalent to shuffling time indices on the cross-correlation integrals used to compute \( d_m(r, k) \) for the original time series. We conclude from Fig. 6(b) that, with high probability, the original time series is nonstationary on time scales of order \( w = 200 \). It cannot necessarily be concluded that there is a change point in the time series, because other forms of nonstationarity can cause rejection of the null hypothesis. In order to obtain further evidence for a change point the time series can be split into two pieces at the conjectured change point, and each piece tested for stationarity using the above procedure.

5.3. Relationship to other techniques

de Lima [6] has proposed a statistic for change point detection based on recursively updated correlation integrals. The technique has the advantage that condition (25) is not required, but the limitation that the statistic does not achieve an optimum at the change point. Also, the null hypothesis that the time series is independent and identically distributed is tested. This has the advantage that the distribution of the statistic can be computed analytically under the null hypothesis; also the time scale \( w \) is not required. However, for this null hypothesis to be interesting, a model for the stationary structure in the time series must be fitted and residuals computed. Model fitting can be time consuming, so the surrogate data technique proposed here will be advantageous in preliminary analyses of time series.

Takanami and Kitagawa [41] have proposed a technique for change point detection based on recursively fitted linear models. The technique is optimal for change point detection in time series generated by linear systems but cannot be used to detect nonlinear changes. However, as shown by the examples in Section 3, many nonlinear systems driven by external forces undergo changes that can be detected with linear techniques. A linear change point detector is useful in such cases.

Smith et al. [37] have proposed a technique for change point detection based on a nonlinear prediction algorithm. This approach is intermediate between a linear technique and the nonparametric technique of Section 5.1.
Smith et al. [37] showed the nonlinear prediction technique to be superior to a nonparametric change point detector for deterministic time series. We suspect this will be true in general, because there are several sophisticated function approximation techniques (for example see [14, 45]) that are less prone to the "curse of dimensionality" than non-parametric techniques such as those based on correlation integrals. However, function approximation techniques tend to be numerically intensive, so that the technique proposed here will be advantageous in the case of long time series consisting of several thousand data points.

Kennel [19] has proposed a statistic for change point detection similar to ours, but uses a different approach for assessing statistical significance. The technique of Section 5 has the advantage of being supported by convergence theorems, but numerical experiments will be required to compare the practical advantages of the two techniques. Kilwell [19] has developed other statistics that are advantageous in testing for nonstationarities of a more general form than change points. The surrogate data technique of Section 5.2 is applicable to these statistics.

6. Barriers to transport

Many autonomous nonlinear dynamical systems have a tendency to exhibit structure on long time scales [1]. The results of the previous sections must be carefully interpreted to take into account the existence of such phenomena. We illustrate this with a time series generated by a dynamical system with strong barriers to transport. The time series is generated by iterating the following area preserving dynamical system and recording $\theta_i, i = 1, 2, \ldots, 2.5 \times 10^6$, holding the parameter $\gamma_i$ fixed at $\gamma_i = 1.05/2\pi$:

$$
\theta_{i+1} = \theta_i + r_i + \gamma_i \sin 2\pi \theta_i \pmod{1}, \quad r_{i+1} = r_i + \gamma_i \sin 2\pi \theta_i.
$$

(26)

A WRP (resp. LWRP) of the time series is shown below (resp. above) the diagonal in Fig. 7(a) with $m = 2$, $r = 3 \times 10^{-4}$ and $w = 5000$. Structure on multiple time scales is apparent in the WRP, and to a lesser extent in the LWRP. By comparison with Fig. 3(a), it appears that there is an underlying driving force which varies according to a random walk at the beginning and end of the time series, and that takes on two distinct values in the middle of the time

Fig. 7. Barriers to transport. (a) A WRP (resp. LWRP) is shown below (resp. above) the diagonal for a time series generated by the standard map (26) for $m = 2$, $r = 3 \times 10^{-4}$ and $w = 5000$, and a logarithmic grayscale from $3.9 \times 10^{-8}$ to $3.2 \times 10^{-5}$ (resp. $2.5 \times 10^{-6}$ to $3.8 \times 10^{-6}$). (b) Phase portraits of three segments of the time series corresponding to the windows $l = 1, 201$ and 251.
series. At first this may seem surprising because the parameter \( \gamma_i \) was held fixed throughout the iteration. However, the WRP accurately measures changes in the probability distribution filled out by sequences of embedding vectors from the time series. This is illustrated in Fig. 7(b) for three sequences of embedding vectors length \( w = 5000 \) from the time series, that fill out three distinct probability distributions. The effect is due to the existence of strong barriers to transport in the dynamical system (26), a well-known phenomenon in conservative dynamical systems. We have obtained similar WRPs for a dissipative dynamical system consisting of a randomly driven particle moving in a potential well with several local minima.

We now evaluate the results of previous sections within the context of the above phenomenon. First, the WRP shown in Fig. 7(a) is not a good reconstruction of the RP of the driving force, which in this example is uniform: every location \((i, j) \in [1, N_m]^2\) satisfies \( \gamma_i = \gamma_j \). However, the reconstruction is not poor due to false recurrences, and there is no conflict with the results of Section 2.3. The reconstruction is poor because, unlike the examples in Section 2.2, the dynamical system (26) has very weak mixing properties. In such cases, for deterministic time series, the WRP of the time series is a reconstruction of a proper subset of the RP of the driving force. The same remark holds true for MRPs of stochastic time series. Second, the WRP shown in Fig. 6(a) appears to identify nonstationarity in the time series. A WRP generated with the surrogate data technique of Section 5.2 with a large time scale \( w \) would look very different to Fig. 6(a), even though the underlying dynamical system is stationary. Again, there is no conflict, because the surrogate data technique must be interpreted as identifying nonstationarity in the time series at time scales of order \( w \) or greater.

Further insight into time series with barriers to transport can be obtained by considering an alternative surrogate data technique due to Manuca and Savit [24]. In this technique, which we refer to as the Markov surrogate data technique, an embedding dimension \( m \) and a resolution \( r \) are chosen, and a surrogate data set of embedding vectors \( \{v_i\} \) is generated iteratively, as follows. Embedding vectors \( \{v_{ij}\} \) within distance \( r \) of \( v_i \) are computed, and \( v_{i+1} \) is chosen at random from the embedding vectors \( \{v_{ij+1}\} \). If no such embedding vector exists, \( v_{i+1} \) is chosen to be \( v_{j+1} \), where \( j \) satisfies \( v_j = v_i \). If this technique is applied to the above example, the surrogate data sets will exhibit structure on very long time scales. The Markov surrogate data technique is suited to testing the null hypothesis that the time series is generated by a Markov process of order \( m \), which in this example will not be rejected. The Markov surrogate data technique is thus complementary to the surrogate data technique of Section 5.2.

7. Applications

In this section we analyze a time series of respiration rates recorded from a subject with sleep apnea. We were motivated to analyze this time series by results of Schreiber [36], who reported evidence for a change point occurring about half way through the time series. Sleep apnea is a potentially life threatening disorder in which the subject stops breathing during sleep. The time series was obtained from the overnight monitoring of the subject for a duration of about 5 h at a sampling rate of 2 Hz. The subject was briefly awake during this period; also some artifacts appear in the time series. We eliminated these data by sampling the points 1–13 001, 18 201–19 200, 22 001–27 000, 28 001–34 000, to obtain a time series of length \( N = 26 001 \) for analysis.

Fig. 8(a) shows RPs of the time series for \( r = 0.1 \) and \( m = 2 \) (below the diagonal) and \( m = 3 \) (above the diagonal). Evidence for a change point occurring about half way through the time series is apparent. A WRP (resp. LWRP) of the time series is shown below (resp. above) the diagonal in Fig. 8(b), for \( m = 2, r = 0.1 \) and \( w = 200 \). The cross-correlation integrals shown in the WRP vary over a much wider range of values than in the LWRP, indicating that the time series is highly nonlinear. Fig. 8(c) shows a thresholded MRP (resp. LMRP) obtained from the WRP (resp. LWRP), and indicates that a particularly sharp change point occurs exactly half way through the time series. Moreover, the MRP indicates that the second half of the time series is nonstationary. Schreiber obtained
Fig. 8. Characterization of nonstationarity in physiological data. (a) RP for \( r = 0.1 \) and \( m = 2 \) (below the diagonal) and \( m = 3 \) (above the diagonal). 1\% of the recurrences are shown. (b) WRP (resp. LWRP) for \( m = 2, r = 0.1 \) and \( w = 200 \) corresponding to (a) is shown below (resp. above) the diagonal, with logarithmic grayscale from \( 10^{-4} \) to 0.3 (resp. \( 2 \times 10^{-4} \) to \( 2 \times 10^{-2} \)). (c) Thresholded MRP and LMRP obtained from (b) with a grayscale varying from 0 to 0.25. (d) Thresholded MRP and LMRP for the normalized time series, with grayscale from 0 to 0.25. (e) Dependence of the windowed mean \( \mu_j \) on the window index \( j \) for the unnormalized time series. (f) Dependence of the correlation integrals \( C_m(r, l, l) \) on \( l \) for the normalized time series. (g) Sequence of phase portraits of the unnormalized time series. (h), (i) Time series segments of duration 100 s.
a similar result with a nonlinear cross-prediction technique. However, the LMRP indicates that nonstationarities identified by nonlinear techniques may be largely due to changes in the linear properties of the time series. We investigated this possibility by computing the mean of the time series in nonoverlapping windows of width $w$, as shown in Fig. 8(e). The jump in the mean occurring half way through the time series is so clear that we are lead to conjecture that it is due to an artifact in the recording. In order to proceed with the analysis, we therefore normalized the time series to have zero mean and unit variance in each of the windows.

An MRP (resp. LMRP) of the normalized time series is shown below (resp. above) the diagonal in Fig. 8(d), for $m = 2$, $r = 0.1$ and $w = 200$. The LMRP shows some evidence for a change point. The MRP shows clearer evidence for a change point, as well as evidence for nonstationarity in the second half of the time series. Additional insight is obtained by monitoring the $I$-dependence of the correlation integrals $C_m(r, I, I)$, computed for nonoverlapping windows of width $w = 200$, with $m = 2$ and $r = 0.1$, and shown by the solid curve in Fig. 8(f). The nonstationarities identified by the MRP appear to correspond to increases in the correlation integrals. The dotted line in Fig. 8(f) shows the $I$-dependence of linearized correlation integrals, obtained from (14), indicating that these increases are not due to changes in the linear properties of the time series.

Fig. 8(g) shows a sequence of phase portraits obtained by dividing the unnormalized time series into 12 nonoverlapping segments of duration 1083 s. The increase in the correlation integrals in Fig. 8(f) can be interpreted as quantifying the increase in clustering in the phase portraits in the second half of Fig. 8(g). The phase portraits in the second half of Fig. 8(g) also exhibit a marked asymmetry with respect to reflections through the diagonal: phase portraits $\{(x_i, x_{i+1})\}$ obtained from time reversed data would be easily distinguishable from the phase portraits $\{(x_i, x_{i+1})\}$ of the original data. This indicates that the second half of the time series is irreversible. Note that irreversibility is a nonlinear phenomenon.\footnote{Linear dynamical systems are reversible, and this property is preserved under nonlinear observation functions, see [9] and references therein.} Fig. 8(h) (resp. 8(i)) shows a segment of duration 100 s sampled from the first (resp. second) half of the time series. Both segments show abnormal breathing, and are characteristic of apnea. Such segments occur throughout the time series, interspersed with segments showing normal breathing. However, there are two qualitative differences between the segments shown in Figs. 8(h) and (i). The second segment is smoother than the first segment. Also, the second segment is more irreversible than the first segment: it takes longer to fall from a high than to rise from a low in the second segment. This effect shows up clearly in the asymmetries of the phase portraits shown in the second half of Fig. 8(g).

Ignoring the jump in the mean, the dominant form of nonstationarity in these data can be characterized as: “The second half of the time series is smoother and more irreversible than the first half of the time series”. There is also some nonstationarity in the second half of the time series. It is difficult to tell whether the nonstationarities are due to the influence of an external driving force with multiple time scales as in the example of Section 2.4, or are self-generated as in the example of Section 6.

We now apply the prediction algorithm of Section 4 to the unnormalized time series. The test set was chosen to be of size $N_t = 16000$. In Fig. 9(a) the prediction errors $E_m(k, w)$ are plotted against the number of neighbors $k$ for window widths $w = 100, 1000, 9990$ and embedding dimensions $m = 2, 3, 4$. The most accurate predictions are obtained for $w = 1000$, at which nonlinear forecasts with $k \approx 50$ are superior to linear forecasts ($k \approx 1000$).

Less accurate predictions are obtained for the largest window width $w = 9990$. We conclude from the scaling law (21) that the time series is not likely to have been generated by a low-dimensional deterministic system driven by slowly varying external forces. The choice of a large $w$ is suboptimal for the time series because RPs do not provide an accurate reconstruction of the driving force characterized by MRPs. For example, consider the RP for $m = 3$ shown above the diagonal in Fig. 8(a). The recurrences in the square region $[1, \frac{1}{2} N] \times [\frac{1}{2} N, N]$ are identified as false recurrences by the MRP shown below the diagonal in Fig. 8(c). As argued in Section 4.1, for large $w$ several...
Fig. 9. Prediction and change point detection in physiological data. (a) Dependence of the prediction error $E_m(k, W)$ on $k$ for the unnormalized time series with $w = 100, 1000, 9990$, and $m = 2$ (dotted curves), $m = 3$ (solid curves) and $m = 4$ (dashed curves). Curves towards the right correspond to larger $w$. (b) Dependence of the change point detector $d_m(r, W K)$ on $K$ for the first half of the normalized time series, with $m = 2$, $w = 200$ and $r = 0.1$. Dotted curves: change point detectors for 40 surrogate data sets. (c) Dependence of the local prediction error $E(i)$ on the time index $i$ for the unnormalized time series form $I = 3$. Solid curve: $w = 1000$ and $k = 50$. Dotted curve: $w = 9990$ and $k = 100$. Dashed line: prediction error $E_m(k, W)$ for $m = 3$, $k = 50$ and $w = 1000$.

of these false recurrences will be used in computing nearest neighbors, resulting in poor predictions. This effect can be seen in Fig. 9(c) in which the time dependence of the local prediction error $E(i)$ is shown for $w = 9990$ and $w = 1000$. Poor predictions are obtained for $w = 9990$ immediately after the change point at $i = 13000$.

Finally, we apply the change point detection technique of Section 5 to the first half of the normalized time series. The results are shown in Fig. 9(b) for $m = 2$, $r = 0.1$, $w = 200$ and 40 surrogate data sets. The change point detector $d_m(r, W K)$ for the time series is distinguishable from that of the surrogate data sets for $k \approx 32$, indicating that there is significant nonstationarity on time scales of order $w = 200$. The nonstationarity is probably not due to a change point because, in contrast to the example in Fig. 6(b), the maximum value of $d_m(r, W K)$ is not sharp. The absence of a clear change point is consistent with a visual inspection of the MRP in Fig. 8(d) for $I < 65$. On the other hand, the surrogate data technique identifies significant nonstationarity that is difficult to identify in the MRP for $I < 65$. In general we have found MRPs to be useful in preliminary "exploratory" analyses of time series, and change point detectors to be useful in identifying small effects that escape visual detection.

8. Conclusions

In this paper we have investigated techniques based on recurrence plots (RPs) for the analysis of time series generated from dynamical systems driven by slowly varying external forces. The techniques provide a unified
The concept of a false recurrence was introduced to quantify the accuracy with which the RP of the driving force can be reconstructed from the RP of the time series. For deterministic systems, it was shown that the fraction of recurrences that are false can be made arbitrarily small if the driving force varies sufficiently slowly. This was shown by deriving scaling laws from an embedding theory for parameterized families of dynamical systems. In the case of smooth one-dimensional driving forces an explicit algorithm was given for extracting the functional form of the driving force from the reconstructed RP. However, many dynamical systems of interest are stochastic, or high-dimensional. It was shown that external driving forces acting on such systems can be characterized using averaged and transformed RPs, referred to as meta-recurrence plots MRPs. Linearized MRPs were introduced to assess the extent to which these characterizations reflect changes in the linear properties of the time series. RPs and MRPs were used to gain insight into the nonlinear prediction of time series generated from dynamical systems driven by slowly varying external forces. In particular, for deterministic systems, a scaling law was derived implying that unmodified algorithms developed for the nonlinear prediction of stationary time series are optimal. Finally, RPs were used to construct change point detectors with well-defined convergence properties, and a surrogate data technique was introduced for assessing statistical significance.

Several directions for further research were raised by our investigations as follows: First, most of the results in this paper are based on numerical simulation and heuristic arguments. It would be desirable to obtain rigorous mathematical proofs of the results. Second, the time series analyzed in this paper are of long duration. It would be desirable to develop techniques that require less data. We expect that techniques based on nonlinear cross-prediction will be less data intensive. Also, MRPs based on measures of cross-prediction may lead to improved prediction algorithms for nonstationary stochastic systems. Third, nonstationarities characterized by MRPs may be due to external driving forces or to intrinsic properties of the dynamical system, such as barriers to transport. In the latter case, nonstationarity is traditionally characterized by low frequency properties of the power spectrum [1]. MRPs provide a more general characterization of nonstationarity, and may prove useful for investigating the existence of barriers to transport in time series generated from high-dimensional, or spatially extended, dynamical systems. Fourth, the definition of MRPs used in this paper is based on an $L^2$ measure of distance between probability distributions. In information theory the natural measure of distance between two probability distributions is the relative entropy [4]. An information theoretic version of MRPs would have the advantage of being interpretable in terms of coding and communication theory. Also, in statistics, the best known techniques used to discriminate between probability distributions are the $\chi^2$ test and the Kolmogorov–Smirnov test [30]. The Kolmogorov–Smirnov test has the advantage that a resolution parameter $\tau$ need not be chosen. We expect that a Kolmogorov–Smirnov version of an MRP would have some important statistical advantages. Finally, the time series in Section 7 used to illustrate the techniques in this paper is of limited scope. In some applications spatio-temporal time series are available, recorded under a variety of conditions. It would be desirable to apply the techniques in this paper to investigate spatial anisotropy and classification problems in such data.

Acknowledgements

We thank Radu Manuca and David Wolpert for useful conversations and Jonathan Raz for referring us to work on the stationary bootstrap [29]. We are grateful for support from the Office of Naval Research under grant ONR:N00014-95-1-0034.

For example spatio-temporal data in one spatial dimension can be analyzed for spatial anisotropy using the techniques presented above by segmenting the data in space instead of time, and computing cross-correlation integrals between the segments.
Appendix A

A.1. Scaling laws for recurrence plots

A heuristic argument supporting the scaling law (7) can be given as follows. The probability of a recurrence, \( C_m(r) \), scales as \( r^{D_2} \) where \( D_2 \) is the correlation dimension of the image of the set \( \{(s_i, \gamma_i), i \in \mathbb{Z}^+\} \) of correlation dimension \( D \) under the piecewise smooth map \( \Phi_m \) defined by (9). Generically, the correlation dimension \( D \) is invariant under piecewise smooth maps into \( m \)-dimensional spaces if \( m > D \), and it follows that \( D_2 = D \). A similar argument shows that \( C_m(r) \) scales as \( r^m \) for \( m < D \).

We now give a heuristic argument supporting the scaling law (8). Let \( \alpha \) denote the closure of the set \( \{(s_i, \gamma_i), i \in \mathbb{Z}^+\} \), and define the set \( \Sigma_m \) of "self-intersections" of \( \Phi_m(\alpha) \) to be the set of \( \nu \in \Phi_m(\alpha) \) such that the equation \( \Phi_m(s, \gamma) = \nu \) does not have a unique solution \((s, \gamma) \in \alpha\). The probability \( P_m^\epsilon(r, \epsilon) \) that a false recurrence occurs at a randomly selected location \((i, j)\) satisfies

\[
P_m^\epsilon(r, \epsilon) \leq (P(\text{dist} (\nu_i, \Sigma_m) < r) + Q_{m,F}(r, \epsilon)) P(\|\nu_i - \nu_j\| < r),
\](A.1)

where \( Q_{m,F}(r, \epsilon) \) is defined by

\[
Q_{m,F}(r, \epsilon) = P(\|\gamma_i - \gamma_j\| > \epsilon \|\nu_i - \nu_j\| < r \text{ and dist} (\nu, \Sigma_m) \geq r).
\](A.2)

The scaling properties of \( P(\text{dist} (\nu_i, \Sigma_m) < r) \) can be computed by covering \( \mathbb{R}^m \) with boxes of radius \( r \). The number of boxes covering \( \Phi_m(\alpha) \) (resp. \( \Sigma_m \)) scales as \((1/r)^D\) (resp. \((1/r)^{\text{dim}(\Sigma_m)}\)). Therefore, assuming \( \alpha \) is a homogeneous fractal, we have

\[
P(\text{dist} (\nu_i, \Sigma_m) < r) = O((1/r)^{D - \text{dim}(\Sigma_m)}),
\](A.3)

\( \text{dim}(\Sigma_m) \) can be calculated as follows. Generically, the intersection of two sets of dimension \( D \) in a space of dimension \( m \) is empty for \( m > 2D \), and a set of dimension at most \( 2D - m \) for \( m \leq 2D \). Thus one might conjecture that, generically, \( \Sigma_m \) is empty for \( m > 2D \), and has dimension at most \( 2D - m \) for \( m \leq 2D \). This has been shown to be true in the case of stationary dynamical systems by Sauer et al. [34]. However, in the case of parametrized families of dynamical systems, fixed points and low-order periodic orbits must be taken into account. Define the set of fixed points \( \alpha_F \) by

\[
\alpha_F = \{(s, \gamma) \in \alpha | s = f(s, \gamma)\}.
\](A.4)

Then \( \Phi_m(\alpha_F) \subset \Delta \) where \( \Delta \), the diagonal of \( \mathbb{R}^m \), has dimension 1, so that \( \text{dim}(\Phi_m(\alpha_F)) \leq 1 \). It can be shown that self-intersections due to low-order periodic orbits generically have dimension at most zero\(^{12}\) It follows that \( \text{dim}(\Sigma_m) \leq \max(2D - m, 1) \). Thus the scaling law (8) follows from (A.3) and (A.1) using \( P(\|\nu_i - \nu_j\| < r) = O(r^D) \) if it can be shown that \( Q_{m,F}(r, \epsilon) \leq O(r^{m-D}) \).

The scaling properties of \( Q_{m,F}(r, \epsilon) \) can be computed using Taylor’s theorem. For fixed \( \delta \), define \( \Sigma_m'(\delta) \) to be the set of points \( (s, \gamma) \in \mathbb{R}^{d+p} \) such that \( \lambda_m(s, \gamma) \leq \delta \), where \( \lambda_m(s, \gamma) \) denotes the singular value of \( D\Phi_m \) at \((s, \gamma)\) of smallest size. Since \( \Phi_m \) is invertible on \( \Phi_m(\alpha) - \Sigma_m \), if \( \|\nu_i - \nu_j\| < r \) and \( \text{dist}(\nu_i, \Sigma_m) \geq r \) then for sufficiently small \( r \) we have

\[
\|\gamma_i - \gamma_j\| \leq \|\Phi_m^{-1}(\nu_i) - \Phi_m^{-1}(\nu_j)\| < 2r \sqrt{m \lambda_m^{-1}(s_i, \gamma_i)} < \epsilon
\](A.5)

\(^{12}\) For example, there is an unstable period-2 orbit \( \frac{1}{4}, \frac{3}{4} \) for the family of tent maps (5) for \( \gamma = \frac{1}{4} \) and \( \gamma = \frac{3}{4} \), so that \( \Sigma_m \) contains at least two points for all embedding dimensions \( m \). This can be shown to persist under small perturbations.
if \((s_i, \gamma_i) \notin \Sigma'_m(2r/\sqrt{m}/\epsilon)\). Thus \(P_r^i(m)(r, \epsilon) < P((s_i, \gamma_i) \in \Sigma'_m(2r/\sqrt{m}/\epsilon))\), and in order to complete the argument supporting the scaling law (8) is sufficient to show that \(P((s_i, \gamma_i) \in \Sigma'_m(r)) \leq O(r^{m-D})\).

From the assumption in Section 2.2 that \(\{(s_i, \gamma_i), i \in \mathbb{Z}^+\}\) lies in an \(n\)-dimensional manifold \(M\) with \(n \leq D + 1\), we have \(\dim(\Sigma'_m(0)) \leq 2D - m\) as follows. The derivative of \(\Phi_m\) restricted to \(M\) at \((s, \gamma)\) is an \(n \times m\) matrix, so is generically singular on a subset \(S_m\) of \(M\) of dimension at most \(2n - m - 1\). Since \(\Sigma'_m(0)\) is the intersection of \(S_m\) and \(\alpha\) in \(M\), generically we have

\[
\dim(\Sigma'_m(0)) \leq (2n - m - 1) - D - n \leq 2D - m.
\]

For \((s_i, \gamma_i)\) close to \(\Sigma'_m(0)\), \(\lambda_m(s_i, \gamma_i)\) can be evaluated using Taylor's theorem to give

\[
\lambda_m(s_i, \gamma_i) \approx 0 + a(s, \gamma)\|s - \gamma\|,
\]

where \((s, \gamma)\) is the closest point in \(\Sigma'_m(0)\) to \((s_i, \gamma_i)\). Fixing a constant \(C > 0\) such that \(|a(s, \gamma)| > C\) outside an arbitrarily small subset of \(\Sigma'(0)\), we have

\[
P((s_i, \gamma_i) \in \Sigma'_m(r)) \leq P(\text{dist}(s_i, \gamma_i), \Sigma'_m(0)) < r/C).
\]

Since \(\dim(\Sigma'_m(0)) \leq 2D - m\), substituting \(\Sigma'_m(0)\) for \(\Sigma_m\) in (A.3), it follows that \(P((s_i, \gamma_i) \in \Sigma'_m(r)) \leq O(r^{m-D})\) and the heuristic argument is complete.

The above argument shows that false recurrences at resolution \((r, \epsilon)\) occur due to two effects: lack of invertibility of \(\Phi_m\), and small singular values of the matrices \(D\Phi_m\). For \(m > D\) and small \(\epsilon\) the second effect dominates the first. For \(m > 2D\) the above argument shows that \(D\Phi_m\) is generically invertible on \(\{(s_i, \gamma_i), i \in \mathbb{Z}^+\}\) so that from (A.5), the probability of a false recurrence at resolution \((r, kr)\) satisfies \(P_r^i(m)(r, kr) \approx 0\) for \(k > L_m\) where

\[
L_m = 2\sqrt{m} \max(\lambda_m^{-1}(s, \gamma) \mid (s, \gamma) \in \alpha)\).
\]

This supports the conjecture made at the end of Section 2.3 that there exists a constant \(K_m\) such that the scaling law (8) holds with \(\epsilon = K_m r\). Note that inequality (A.5) gives a very weak\(^{13}\) upper bound on \(\|\gamma_i - \gamma_j\|\) for most \(i, j\). Thus in numerical examples we expect \(K_m \ll L_m\). This can be confirmed for the tent map example of Fig. 2(e) for which \(L_m\) can be computed analytically.

A.2. Computational considerations

The algorithms used in this paper for computing RPs, WRPs and change point detectors are based on a data structure developed by Grassberger [15] for the efficient calculation of correlation integrals. We use this data structure to compute the time indices \((i, j)\) of pairs of \(m\)-dimensional embedding vectors satisfying \(\|v_i - v_j\| \leq r\) in the maximum norm, where \(m \geq 2\). The computation time is of order \(N^2 C_2(r)\) and is strongly dependent on the resolution \(r\). Time series of length a few thousand points can be analyzed with WRPs in a few seconds on a fast workstation. Time series of length a few million points can be analyzed in a few hours at small resolutions \(r\). We use a different algorithm for \(m = 1\) in which cross-correlation integrals \(C_1(r, I, J)\) are calculated separately for each \((I, J)\). Segments of window width \(w\) from the time series are first sorted. The natural ordering of the reals can then be exploited to compute each \(C_1(r, I, J)\) in time of order \(w\), resulting in a total computation time of order \(N^2 w\).

An alternative definition of RPs that leads to a much faster algorithm can be obtained by "quantizing" \(\mathbb{R}^m\) into boxes of volume \(r^m\), and defining recurrence to occur at location \((i, j)\) if \(v_i\) and \(v_j\) fall in the same box. There

\(^{13}\)From the results of Casdagli et al. [3] for stationary dynamical systems, we expect that sharp upper bounds on \(\|\gamma_i - \gamma_j\|\) can be obtained by adding a small level of Gaussian observational noise to the time series \(x_1, x_2, \ldots\), and considering the noise amplification properties of the map \(\Phi_m^{-1}\).
are theoretical similarities between this definition and the definition of RPs given in Section 2.1. The original definition is usually to be preferred because information in the embedding vectors is more efficiently exploited, particularly at small resolutions \( r \). The alternative definition is appropriate for time series that are already strongly quantized, i.e. that consist of a sequence of symbols drawn from a small alphabet. The alternative definition leads to a box-counting technique for the estimation of cross-correlation integrals used to compute WRPs. For a discussion of the advantages and limitations of box-counting techniques in the context of correlation integrals, generalized dimensions and entropies, see [31] and references therein.

A.3. Prediction of nonstationary deterministic time series

The scaling law (21) can be established as follows. For simplicity, suppose the set of points \( \{(s_i, \gamma_i), i \in \mathbb{Z}^+\} \) is homogeneous, i.e. that all the generalized dimensions are equal to \( D \). For example for the Lorenz time series of Section 2.2, \( D \approx 3 \) for all the generalized dimensions. From the definition of \( \Phi_m \) in (9) we have
\[
  x_i = h(f(\Phi_m^{-1}(v_{i-m})�).
\]
From Appendix A.1, \( \Phi_m^{-1} \) is well-defined and smooth for \( m > 2D \). By Taylor’s theorem, it follows that the prediction error \( |x_i - \hat{x}_i| \) obtained with a \( k \) nearest neighbor local linear technique satisfies
\[
  |x_i - \hat{x}_i| \sim d(k, w)^2,
\]
where \( d(k, w) = \|v_{i-m} - v_{j_k-m}\| \) denotes the distance to the \( k \)th nearest neighbor of \( v_{i-m} \). From (7), embedding vectors lie on a \( D \)-dimensional set for \( m > D \), so that for a fixed \( w \) and small \( k \) we have
\[
  d(k, w) \sim C_w k^{1/D}.
\]
Now suppose that \( k \) is fixed and \( w \) is decreased from \( w = N \). At first, \( d(k, w) \) will increase depending on the recurrency properties of the driving force. As \( w \) is decreased to the time scale of variation of the driving force, the time indices \( i - j_1, i - j_2, \ldots, i - j_k \) of the \( k \) nearest neighbors relative to that of the current embedding vector \( v_{i-m} \) will be approximately uniformly distributed. If \( w \) is then decreased by a factor of \( \alpha \), it follows that
\[
  d(\alpha k, \alpha w) \approx d(k, w). \tag{A.9}
\]
Combining (A.8) and (A.9) we have
\[
  d(k, w) \sim (k/w)^{1/D} \tag{A.10}
\]
and the scaling law (21) is obtained from (A.7) and (A.10).

A.4. Probabilistic prediction

For deterministic systems, the scaling law (21) shows that it is advantageous to use large values of \( w \) to obtain small prediction errors \( E_m(k, w) \). The opposite is true for the measure of predictability \( S_m(r, w) \) defined by
\[
  S_m(r, w) = P(|x_i - x_j| < r \mid \|v_{i-m} - v_{j-m}\| < r \text{ and } i - w \leq j < i). \tag{A.11}
\]
\( S_m(r, w) \) can be interpreted as the average probability that a prediction \( x_j \) for the point \( x_i \) lies within distance \( r \) of \( x_i \), where \( j \) is chosen by randomly selecting one of embedding vectors \( v_{j-m} \) that lies within distance \( r \) of \( v_{i-m} \) (if no such embedding vector exists, no prediction is made). For stationary systems \( S_m(r, w) \) is independent of \( w \), and is identical to a measure of predictability introduced by Savit and Green [35]. For nonstationary systems \( S_m(r, w) \)
is in general a decreasing\textsuperscript{14} function of $w$, since conditional probabilities such as (A.11) in general decrease as the conditions are relaxed. There is no conflict with the scaling law (21) because the quantity $S_m(r, w)$ does not account for the fact that at large $w$ there are more points available for estimating the underlying dynamics, a crucial feature of local linear prediction.

From a theoretical point of view there are reasons to prefer the measure of predictability $S_m(r, w)$ over the measure $E_m(k, w)$. For example, in the case of stationary dynamical systems, $S_m(r, w)$ is a special case of a generalized conditional entropy quantifying the information flow in the underlying dynamical system [31]. However, from a practical point of view, $S_m(r, w)$ has serious limitations. For example $S_m(r, w)$ is maximized at small values of $w$ but very few predictions are actually made in the underlying prediction scheme at small values of $w$. We therefore decided to focus attention on the measure $E_m(k, w)$ in this paper.

A.5. Convergence of change point detectors

In this section we prove the assertions made in Section 5.1 concerning the convergence of the change point detector $d_m(r, k N)$ as $N$ tends to infinity.

In the limit of large $N$, $d_m(r, k N)$ can be evaluated analytically from definition (23) using the stationarity assumptions in Section 5.2 to obtain $d_m(r, k N) = d(k) C_m(r)$, where $d(k)$ is given by

$$d(k) = \begin{cases} 
(k^2 + (\alpha - k)^2)A + 2(1 - \alpha)(\alpha - k)C + (1 - \alpha^2)B/(k^2 + (1 - \alpha^2)) & \kappa < \alpha, \\
(\alpha^2 A + 2\alpha(\kappa - \alpha)C + (\kappa - \alpha)^2 + (1 - \kappa)^2 B)/(\kappa^2 + (1 - \kappa)^2) & \kappa \geq \alpha,
\end{cases} \quad (A.12)$$

where $C$ is the cross-correlation integral of $p_1$ and $p_2$, and $A$ and $B$ are the correlations integrals of $p_1$ and $p_2$, at resolution $r$.

Suppose that condition (25) holds. We first show that $d(\alpha + \delta) < d(\alpha)$ for $0 < \delta \leq 1 - \alpha$, by proving the following inequality:

$$(d(\alpha + \delta) - d(\alpha))(\alpha + \delta)^2 + (1 - \alpha - \delta)^2(2\alpha^2 + (1 - \alpha)^2) < 0. \quad (A.13)$$

From the second equation in (A.12) the left-hand side of (A.13) is given by

$$2\delta((\alpha C - (1 - \alpha)B)(\alpha^2 + (1 - \alpha^2)) - (\alpha^2 A + (1 - \alpha)^2 B)(2\alpha - 1) + 2\delta(\alpha^2 - B - A)). \quad (A.14)$$

For the case $A \geq B$, using $C < B$, (A.14) is strictly less than

$$2\delta(\alpha B - (1 - \alpha)B)(\alpha^2 + (1 - \alpha^2) - (\alpha^2 B + (1 - \alpha)^2 B)(2\alpha - 1) = 0.$$

For the case $A \leq B$, using $C < A$ and $\delta < 1 - \alpha$, (A.14) is strictly less than

$$2\delta(\alpha A - (1 - \alpha)B)(\alpha^2 + (1 - \alpha^2) - (\alpha^2 A + (1 - \alpha)^2 B)(2\alpha - 1) + (1 - \alpha)a^2(B - A))$$

$$= 2\delta(\alpha^2 - A - B) \leq 0.$$

A similar calculation shows that $d(\alpha - \delta) < d(\alpha)$ for $0 < \delta \leq \alpha$. Thus $d(k)$ has a unique maximum at $k = \alpha$.

Similar calculations show that the sum of the terms of order $\delta$ in (A.14) is nonzero, so that the derivative of $d(k)$ at $k = \alpha$ is nonzero. The proof is completed by observing that the function $d(k)$ in (A.12) is smooth except at $k = \alpha$.

Suppose that condition (25) fails. We consider the case $B < C < A$; other cases can be dealt with similarly. The change point detector $d_m(r, k)$ fails to converge to the correct change point if it can be shown that $d(\alpha + \delta) > d(\alpha)$

\textsuperscript{14} This is straightforward to verify numerically for the examples above.
for some $\delta$. It suffices to show that the left-hand side of (A.13) is positive for small $\delta$. The left-hand side of (A.13) is given by (A.14), which, to order $\delta$ is strictly greater than

$$2\Delta((aB - (1 - \alpha)B)(a^2 + (1 - \alpha)^2) - (\alpha^2 B - (1 - \alpha)^2)B)(2\alpha - 1) = 0,$$

where we have used $C > B$ and $B < A$. We have experimented with an alternative change point detector defined by

$$d_m(r, k) = (P_{11} + P_{22} - 2P_{12})/2C_m(r),$$

(A.15)

where $P_{11}$ (resp. $P_{22}$, $P_{12}$) denote the probability that a recurrence $(i, j)$ is located in $[1, k]^2$ (resp. $[k + 1, N_m]^2$, $[1, k] \times [k + 1, N_m]$). With similar calculations to the above it can be shown that $d_m(r, kN)$ has a unique maximum at $k = \alpha$ in the limit $N \to \infty$ if $C < 1/2(A + B)$. The later condition is weaker than condition (25), and always holds for sufficiently small resolutions $r$ if the probability distributions $p_1$ and $p_2$ are distinct. Thus $d_m(r, k)$ has superior theoretical properties to $d_m(r, k)$. However, in practice, $d'_m(r, k)$ can exhibit unstable behavior near $k = 1$ and $k = N_m$, and is not considered further in this paper.

References