International Journal of Bifurcation and Chaos, Vol. 28, No. 1 (2018) 1850007 (15 pages)
© The Author(s)
DOI: 10.1142/S0218127418500074

In Search of Determinism-Sensitive Region to Avoid Artefacts in Recurrence Plots

Dadiyorto Wendi*
Institute of Earth and Environmental Science, University of Potsdam, Potsdam, Germany
dwendi@gfz-potsdam.de
dadiyorto@outlook.com

Norbert Marwan
Potsdam Institute for Climate Impact Research (PIK), Potsdam, Germany
marwan@pik-potsdam.de

Bruno Merz
Section 5.4 Hydrology, GFZ German Research Centre for Geosciences, Potsdam, Germany
bmerz@gfz-potsdam.de

Received May 12, 2017; Revised November 7, 2017

Int. J. Bifurcation Chaos 2018.28. Downloaded from www.worldscientific.com by Dr. Norbert Marwan on 05/15/18. For personal use only.

As an effort to reduce parameter uncertainties in constructing recurrence plots, and in particular to avoid potential artefacts, this paper presents a technique to derive artefact-safe region of parameter sets. This technique exploits both deterministic (incl. chaos) and stochastic signal characteristics of recurrence quantification (i.e. diagonal structures). It is useful when the evaluated signal is known to be deterministic. This study focuses on the recurrence plot generated from the reconstructed phase space in order to represent many real application scenarios when not all variables to describe a system are available (data scarcity). The technique involves random shuffling of the original signal to destroy its original deterministic characteristics. Its purpose is to evaluate whether the determinism values of the original and the shuffled signal remain closely together, and therefore suggesting that the recurrence plot might comprise artefacts. The use of such determinism-sensitive region shall be accompanied by standard embedding optimization approaches, e.g. using indices like false nearest neighbor and mutual information, to result in a more reliable recurrence plot parameterization.

Keywords: Recurrence plot; phase space time delay embedding reconstruction; artefact avoidance.

1. Introduction

Recurrence is a fundamental property of many dynamical systems, which can be exploited to characterize the systems behavior in phase space, while a recurrence plot (RP) is the visualization tool for the analysis of this property. In this study, the phase space reconstruction method of time delay embedding [Packard *et al.*, 1980; Takens, 1981] is used [Eq. (1)]. Such a reconstruction is particularly useful when not all variables required to describe the

*Author for correspondence
This is an Open Access article published by World Scientific Publishing Company. It is distributed under the terms of the Creative Commons Attribution 4.0 (CC-BY) License. Further distribution of this work is permitted, provided the original work is properly cited.

system are available (i.e. data scarcity or limited set of observation variables), and where the topology of the system dynamics $\hat{\mathbf{x}}_i$ can still be created using only a single variable or observation u_i.

$$\hat{\mathbf{x}}_i = \sum_{i=1}^{m} u_{i+(j-i)\tau}\mathbf{e}_j, \tag{1}$$

where m is the embedding dimension and τ is the time delay. The vectors $(\mathbf{e}_j)$ are unit vectors and span an orthogonal coordinate system $(\mathbf{e}_i \cdot \mathbf{e}_j) = \delta_{i,j}$. The calculation of recurrence as elements of the RP is based on Eq. (2):

$$R_{i,j}(\varepsilon) = \Theta(\varepsilon - \|\mathbf{x}_i - \mathbf{x}_j\|), \quad i,j = 1,\ldots,N, \tag{2}$$

where N is the number of measured points $\mathbf{x}_i, \varepsilon$ is a threshold distance, $\|\cdot\|$ is a norm and $\Theta(\cdot)$ the Heaviside function.

The RP is basically the visual representation of the square matrix, in which the matrix elements correspond to those times at which a state of a dynamical system recurs (columns and rows correspond then to a certain pair of times). RPs are especially useful for nonstationary pattern in time series [Eckmann *et al.*, 1987; Marwan *et al.*, 2007]. Besides using RPs for the visual analysis of time series, RPs can also quantify structures hidden within the series through recurrence quantification analysis (RQA) [Zbilut & Webber, 1992; Marwan *et al.*, 2007]. In RQA, important elements are the diagonal and vertical/horizontal straight lines because they reveal typical dynamical features of the investigated system, such as range of predictability, chaos-order, and chaos-chaos transitions [Trulla *et al.*, 1996]. One of the prominent diagonal line measures is called determinism [DET, Eq. (3)], from which the system predictability can be inferred.

$$\mathrm{DET} = \frac{\sum_{l=l\min}^{N} lP(l)}{\sum_{i,j}^{N} R_{i,j}} \tag{3}$$

where $P(l) = \{l_i; i = 1,\ldots,N_l\}$ is the histogram of the lengths l of diagonal structures, and N_l is the absolute number of those diagonal lines.

For a deterministic signal (including chaos), many diagonal lines in the RP are typical, leading to high value of DET [Marwan, 2010]. However, single, isolated recurrence points can occur if states are rare, if they do not persist, or if they fluctuate heavily. For instance, stochastic or random signals would comprise such single points and result in a very low DET.

Since the use of RPs relies on the reconstructed phase space, its parameters' uncertainty includes those of the phase space reconstruction method, such as embedding dimension (m) and time delay (τ), in addition to the recurrence threshold (ε). Standard approaches for finding optimal embedding parameters are false nearest neighbors (FNN) for m, and auto-correlation or mutual information (MI) for τ [Kantz & Schreiber, 2005; Kennel *et al.*, 1992; Fraser & Swinney, 1986]. Other methods include wavering-products, fill-factor or integral local deformation [Buzug & Pfister, 1992]. Moreover, Marwan [2010] concludes that τ is sometimes overestimated by auto-correlation and mutual information, and that the choice of the embedding dimension has to be considered with care, as a wrong choice artificially increases diagonal lines, and hence DET, and leads to artefacts. For instance, a RP resulting from a random series should exhibit scattered or non-deterministic patterns (i.e. single points). However, when m increases to 2 and beyond with $\tau = 1$, the number and the length of diagonal lines start to increase and dominate the plot as artefacts. This may be misinterpreted as if the series was highly deterministic (Fig. 1).

In this study, we focus on the artefacts related to these embedding parameters. The impact of the recurrence threshold (ϵ) is not elaborated, since the selection of the optimal values of the recurrence threshold has been discussed earlier [Gao & Jin, 2009; Koebbe *et al.*, 1994; Zbilut & Webber, 1992; Zbilut *et al.*, 2002; Mindlin & Gilmore, 1992; Schinkel *et al.*, 2008; Thiel *et al.*, 2002]. Hence, the recurrence threshold is fixed to a 10% recurrence rate (recurrence points density). Supplementary information on the impact of changing this threshold is enclosed in the Appendix. The Appendix also includes the evaluation of DET values of a correlated random series, using an AR1 series as example, to showcase that high DET values are indeed associated with deterministic systems instead of its auto-correlation structures, although there are also cases at certain parameter values where the number/length of diagonal lines artificially increase. It is important to note that the proposed technique is not intended to be used as a new, independent method, but rather as an additional consideration during parameterization, when the dynamical system is known to be deterministic.

Int. J. Bifurcation Chaos 2018.28. Downloaded from www.worldscientific.com
by Dr. Norbert Marwan on 05/15/18. For personal use only.

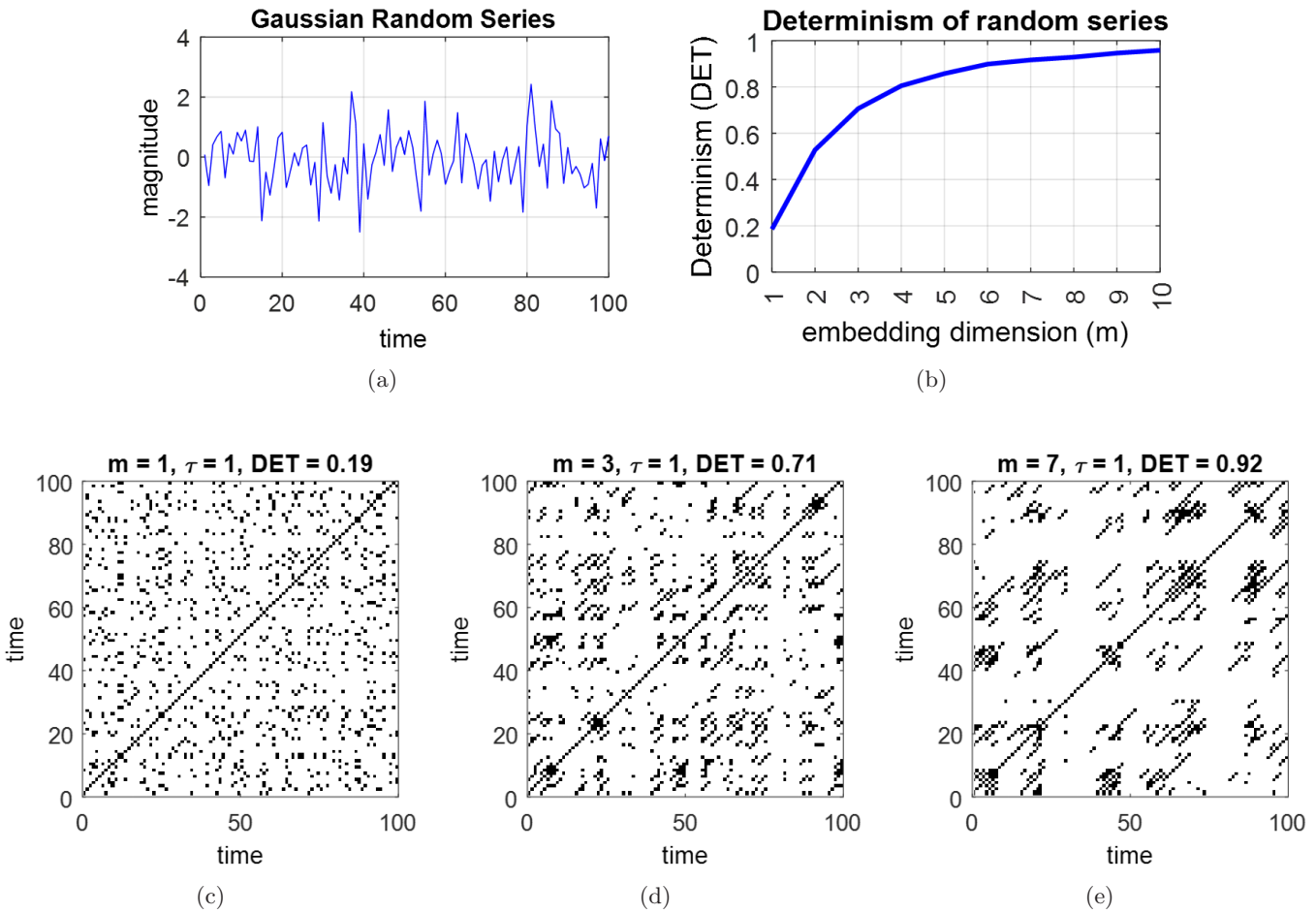


Fig. 1. (a) Misleading DET values of random series; (b) sub-figure shows the artificial increase of DET when embedding dimension (m) increases, while (c)–(e) show the corresponding RP with the increase of diagonal line structures at high embedding dimension ($m = 7$).

2. Methodology

Artificially biased line length distributions due to the embedding can overlay the true line length distributions and lead to wrong conclusions. Hence, it would be desirable to separate the contribution of the embedding induced line length distributions from the real underlying dynamics. However, separating both contributions is not possible without additional knowledge about the system (such as precise model or amount of observational noise). Therefore, we propose an approach that minimizes the contribution of the embedding. This approach is based on comparing the fraction of recurrence points that form diagonal lines in the RPs of the original time series (which includes both the real underlying dynamics as well as the embedding effect) with that of a random time series (which consists of the embedding effect only). As random time series we use simply shuffled versions of the original time series, because this preserves its value distribution and, thus, allows to use the same recurrence threshold and allows to compare the resulting RPs. As mentioned above, RPs of random time series should consist mainly of single points, but embedding artifacts would increase the fraction of recurrence points that form diagonal lines in the RP. Thus, this fraction measure is well suited for our purpose. Moreover, this measure is equal to the DET measure. Other measures that use the line length distribution (e.g. average and longest line length, entropy of the length distribution) would be possible but are less intuitive and interpretable. The advantage of the DET measure is that it considers the influence of scattered points that appear within the RP as well as in addition to just the diagonal lines, while the index of average and longest line length could easily suffer from large statistical uncertainty and can easily be influenced by a few extreme values.

Int. J. Bifurcation Chaos 2018.28. Downloaded from www.worldscientific.com
by Dr. Norbert Marwan on 05/15/18. For personal use only.

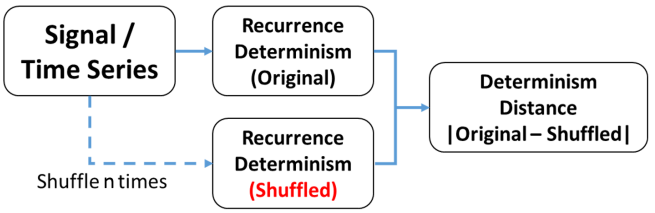


Fig. 2. Scheme of the proposed artefact avoidance method.

In order to compare the line length distributions of the original and the shuffled time series, we define DET_o for the original time series and DET_i for the shuffled version. For a number of shuffling iterations (i.e. n times), the resulting difference [called determinism distance, see Eq. (4)] is calculated for each shuffle (Fig. 2).

For nonoptimal embedding, we expect a rather high contribution of the embedding in the line length distributions in both, the original time series and in the shuffled version. Therefore, DET should have high values in both cases and should not differ so much from each other. For optimal embedding, and if there are deterministic structures in the RP of the original time series, the DET_i of the shuffled time series should be very low whereas DET_o of the original time series has still larger values. The distinctive high and low values of DET in deterministic and stochastic systems are exemplified in this paper using Lorenz and Gaussian random series. In this example, both the original and embedded Lorenz systems show DET values of around 0.8 to 0.9 with τ fixed at 3 following the first minimum of its auto-mutual information, with m varying from 1 to 10. In contrast, for the Gaussian random series, the DET values are shown to be between 0 to 0.2 (Fig. 3).

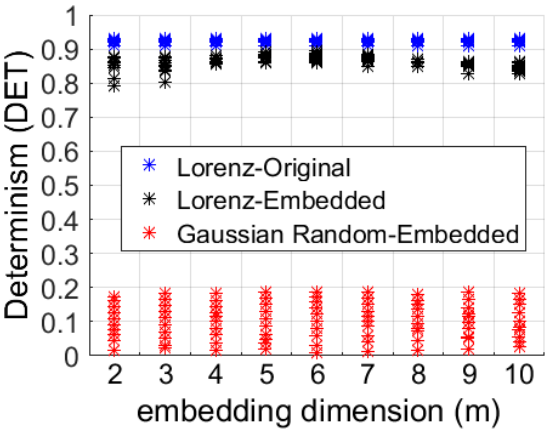


Fig. 3. Examples for high and low DET values from deterministic (Lorenz) and stochastic (Gaussian random) signals.

The resulting difference (determinism distance) between DET_i and DET_o would therefore be high. The undesired effect by the embedding should be minimal for the difference between DET_i and DET_o. Both median (M_d) and standard deviation (S_d) of these distances are used for identifying this determinism-sensitive region [Eqs. (4) and (5)]. The further (larger) the M_d of each parameter combination, the safer it is in terms of avoiding the mentioned artefacts, under the condition that S_d should be reasonably small (e.g. within 0.1)

$$M_d = \mathrm{Median}_{i=1,\ldots,n}(|\mathrm{DET}_o - \mathrm{DET}_i|), \tag{4}$$

$$S_d = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left[|\mathrm{DET}_o - \mathrm{DET}_i| - \frac{1}{n}\sum_{i=1}^{n}|\mathrm{DET}_o - \mathrm{DET}_i|\right]^2}, \tag{5}$$

where DET_o and DET_i are the recurrence determinism values of the original series and each shuffled iteration (i), and n is the total number of shuffling iterations.

3. Case Study Applications

This paper presents two application examples using Lorenz series derived from a mathematical model, and daily runoff observations from the station Burghausen at the Salzach River in south Germany. These signals are chosen for their nonlinear characteristics with known presence of determinism [Sivakumar, 2000; Martins *et al.*, 2011]. The resulting region of artefact-safe parameter set will be presented and discussed in Sec. 4. Caution should be taken when $\tau = 1$ because artificially high DET values can lead to misinterpretations [Figs. 1(b), 1(d) and 1(e)], and hence should be excluded. In addition to the resulting artefact-safe region as the boundary of the parameter sets, the final choice of the parameter set is still necessary to be optimal, i.e. being able to reconstruct the topology of system dynamics and minimal in the sense not to over-reduce data points in the signal. There are

Int. J. Bifurcation Chaos 2018.28. Downloaded from www.worldscientific.com
by Dr. Norbert Marwan on 05/15/18. For personal use only.

Int. J. Bifurcation Chaos 2018.28. Downloaded from www.worldscientific.com
by Dr. Norbert Marwan on 05/15/18. For personal use only.

many approaches to find optimal embedding parameters, such as the standard approaches mentioned in Sec. 1.

3.1. *Lorenz series*

The Lorenz system with known nonlinear, nonperiodic, three-dimensional and deterministic chaos behavior (i.e. with parameters $\alpha = 10$, $\rho = 28$, $\beta = 8/3$ and sampling time $\Delta t = 0.05$) is chosen as the first application example, following Eq. (6) [Lorenz, 1963; Sparrow, 1982]. Its RP and characteristics have also been studied by Marwan *et al.* [2007].

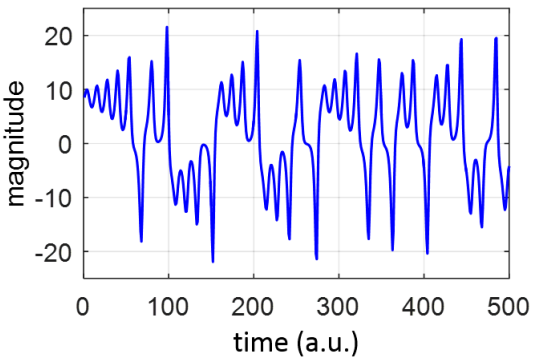
$$\begin{aligned} \frac{dx}{dt} &= \alpha(y - x); \\ \frac{dy}{dt} &= x(\rho - z) - y; \quad \text{and} \\ \frac{dz}{dt} &= xy - \beta z. \end{aligned} \tag{6}$$

This Lorenz system is described by three variables and integrated using the Euler scheme, and hence, we know the three-dimensional phase space that describes the topology of the system dynamics. In this study, the x variable is used as our Lorenz series test set [Fig. 4(a)] with its phase space reconstructed using the time delay embedding method. Thereafter, its DET is calculated. The reliability of these DET values is checked by using median and standard deviation of their determinism distance values (M_d and S_d) to qualitatively evaluate how much the constructed RP of a certain parameter set is influenced by artefacts.

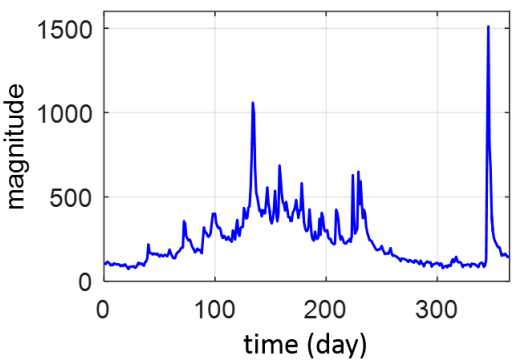
This Lorenz series is derived from a mathematical model with well-known phase space topology and recurrence characteristics, whereas real world observations are most likely contaminated by noise. Therefore, we also investigate the impact of noise on the method, i.e. in respect to the values of determinism and determinism distance. Gaussian white noise with a magnitude range corresponding to the standard deviation of the Lorenz signal is applied, i.e. added to the signal [Eq. (7)]

$$\tilde{x}(t) = x(t) + k\beta(t), \tag{7}$$

where, $\tilde{x}(t)$ is the resulting new series with the addition of noise and $x(t)$ is the original series (Lorenz); k is the noise level, while $\beta(t)$ is the Gaussian white noise with magnitude range corresponding to the standard deviation of $x(t)$. The noise levels used are 5%, 10%, 30% and 50%. For each of the noise-added signal, its determinism and determinism distance are calculated.



(a)



(b)

Fig. 4. Test applications of (a) Lorenz — x variable and (b) Burghausen daily runoff series (1961).

3.2. *River runoff series*

The second test application uses daily river runoff observations extracted from station Burghausen in south Germany for the year 1961. This station measures the streamflow of the Salzach River with a catchment area of 6600 km^2. The time series [Fig. 4(b)] is used as a test set representing real world data, i.e. it is potentially nonstationary and contaminated by noise and observation error.

4. Results and Discussion

This section presents the results of our proposed method for selecting an artefact-safe parameter region with the assumption of recurrence rate fixed at 10%. The range of embedding dimension (m) bounds from 1 to 10 and time delay (τ) from 1 to 20.

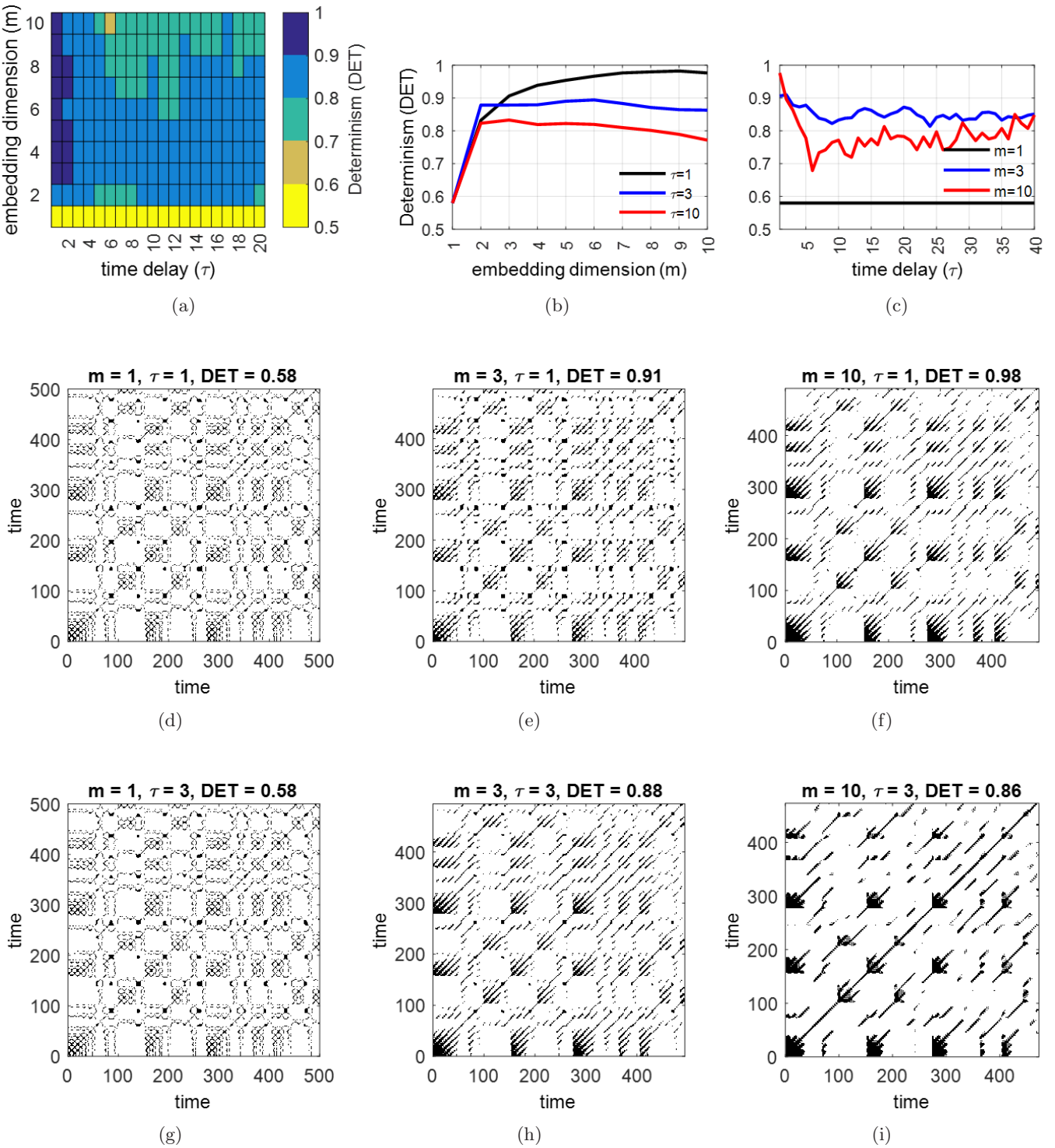


Fig. 5. Recurrence characteristics of the chaotic, deterministic Lorenz signal: (a) determinism corresponding to m and τ; (b) change in determinism corresponding to an increase of the embedding dimension (m) from the RP, with $\tau = 1, 3$, and 10; (c) change in determinism corresponding to an increase of the time delay (τ), with $m = 1, 3$, and 10; (d)–(f) RP of different embedding dimension with fixed $\tau = 1$ and (g)–(i) RP of different embedding dimension with fixed $\tau = 3$. All RPs and recurrence measures are calculated based on fixed 10% recurrence rate.

Int. J. Bifurcation Chaos 2018.28. Downloaded from www.worldscientific.com
by Dr. Norbert Marwan on 05/15/18. For personal use only.

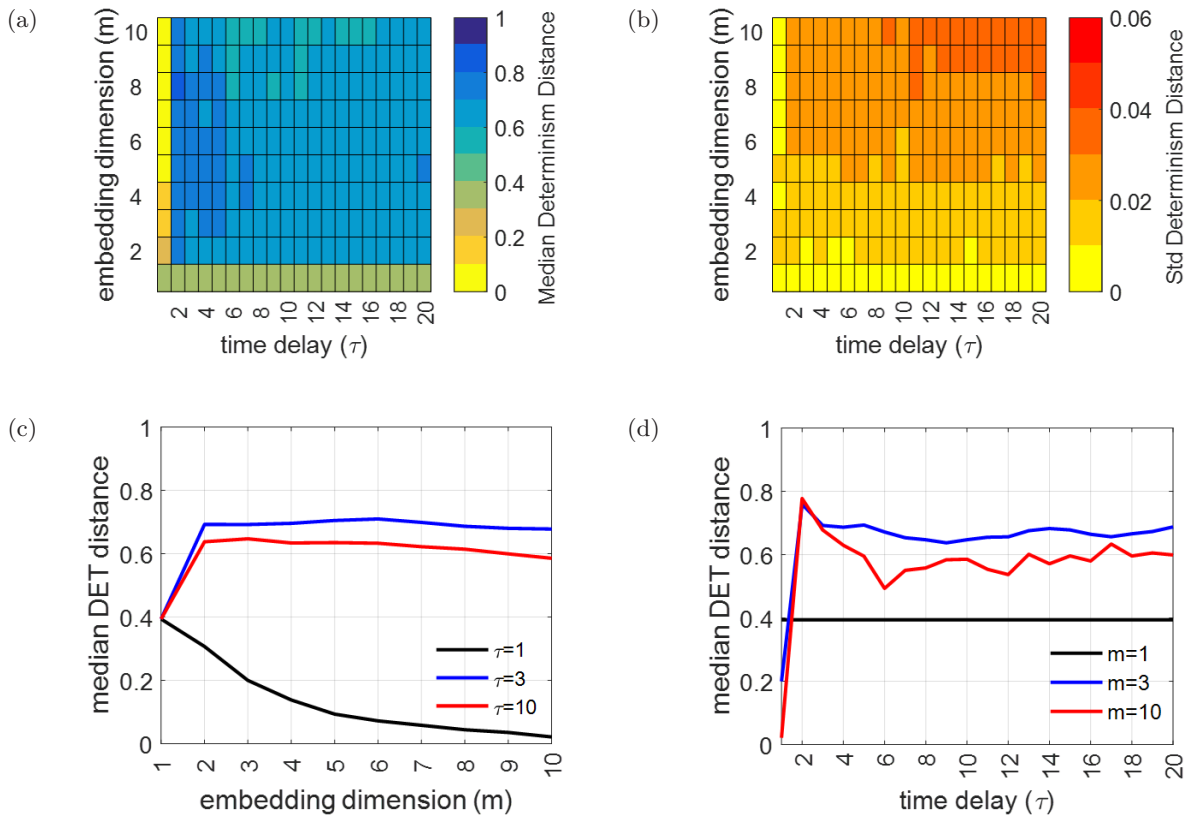


Fig. 6. Determinism distance of the Lorenz series: (a) median (M_d) and (b) standard deviation (S_d) of determinism distance between the RP of shuffled and original Lorenz series. (c) and (d) show the median determinism distance corresponding to $\tau = 1, 3, 10$ and $m = 1, 3, 10$.

4.1. *Lorenz series*

The Lorenz series is known for its deterministic feature, i.e. high determinism value, yet certain parameter combinations can give incorrect, low determinism values, e.g. when $m = 1$ or $m = 10, \tau = 6$ [Figs. 5(a)–5(c)]. Increasing the time delay at high embedding dimension is also seen to thicken the line structures of the RP [Fig. 5(i)]. Low determinism values reflect nonoptimal parameterization, and hence, misleading RP structures [Figs. 5(d) and 5(g)] with diagonal line structures as wobbly and perpendicular to the main diagonal [Marwan *et al.*, 2007].

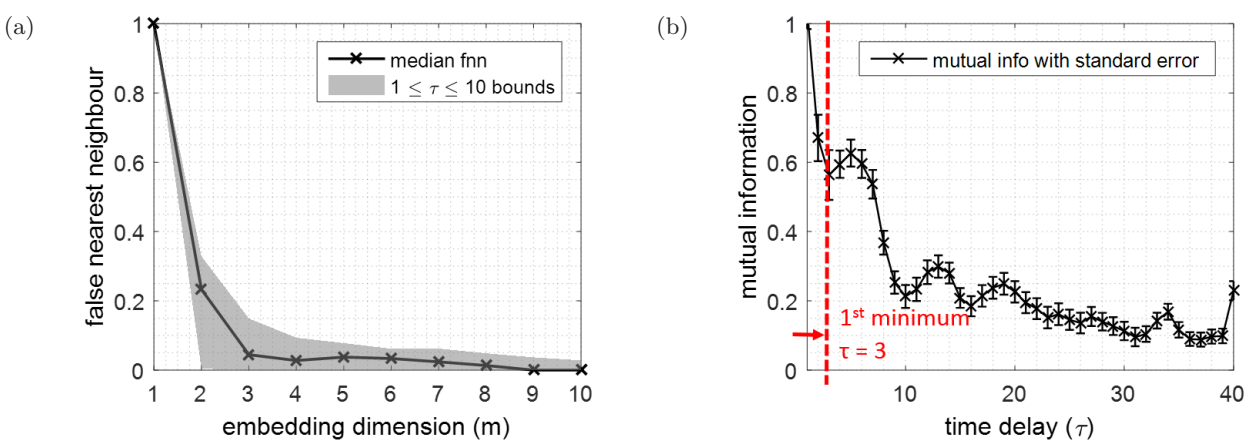


Fig. 7. Embedding parameters for the Lorenz series resulting from standard approaches: (a) false nearest neighbor (FNN) with median and bounds derived from parameter set $1 \leq \tau \leq 10$ and (b) mutual information (MI).

Int. J. Bifurcation Chaos 2018.28. Downloaded from www.worldscientific.com
by Dr. Norbert Marwan on 05/15/18. For personal use only.

In order to assess the reliability of the resulting RP corresponding to the m and τ parameter combinations, the proposed shuffling techniques is applied to find the determinism-sensitive region.

Using the proposed technique ($n = 100$ shuffles), M_d is low for the case without embedding ($m = 1$) as well as for $\tau = 1$, when $m > 1$ (Fig. 6). The latter suggests artefacts due to embedding. Those parameter values where M_d is high, e.g. for $\tau \geq 2$, when $m > 1$, can be considered to be less influenced by embedding artefacts. It can be noticed that when τ and m are higher, M_d starts to decrease

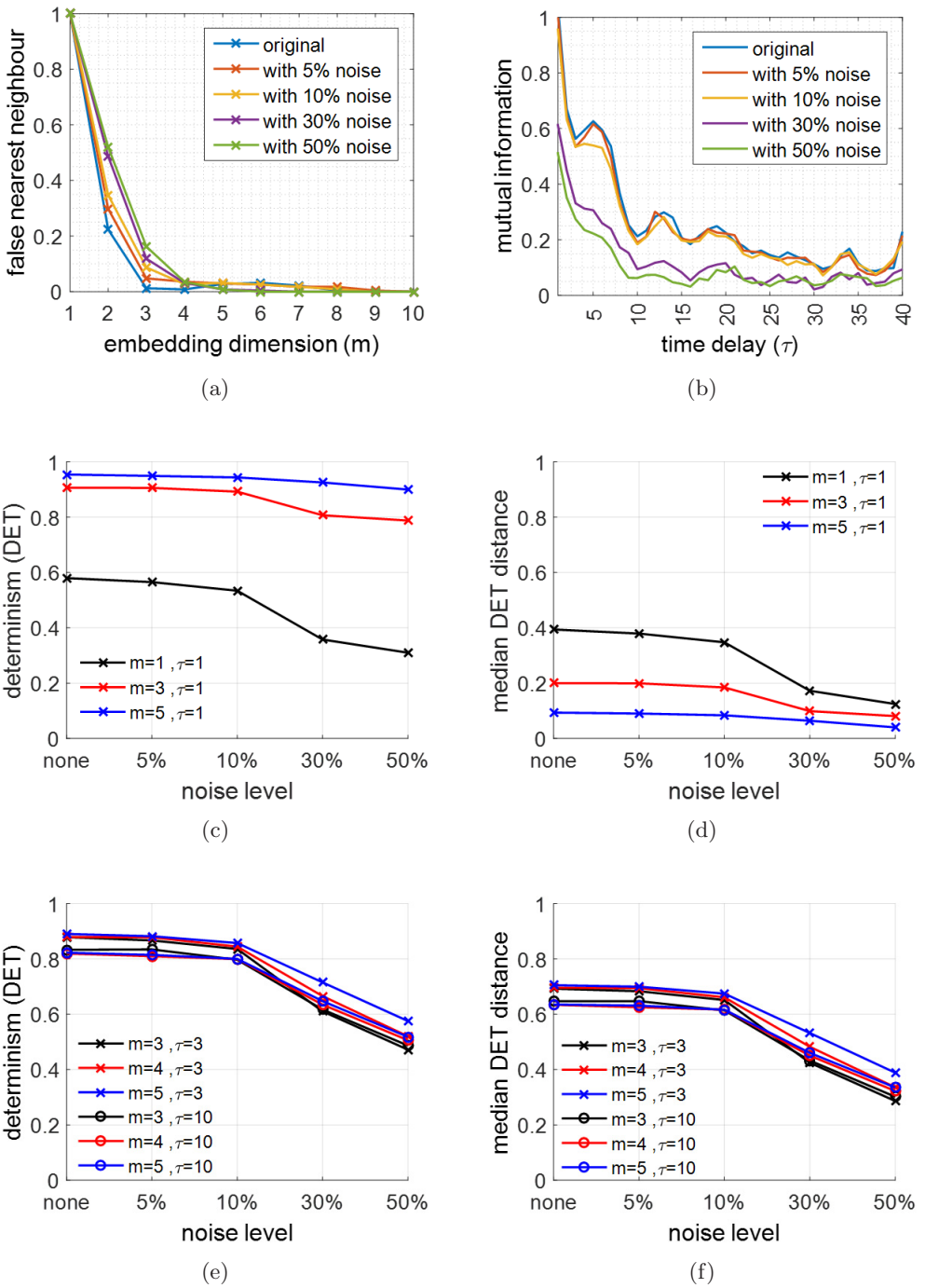
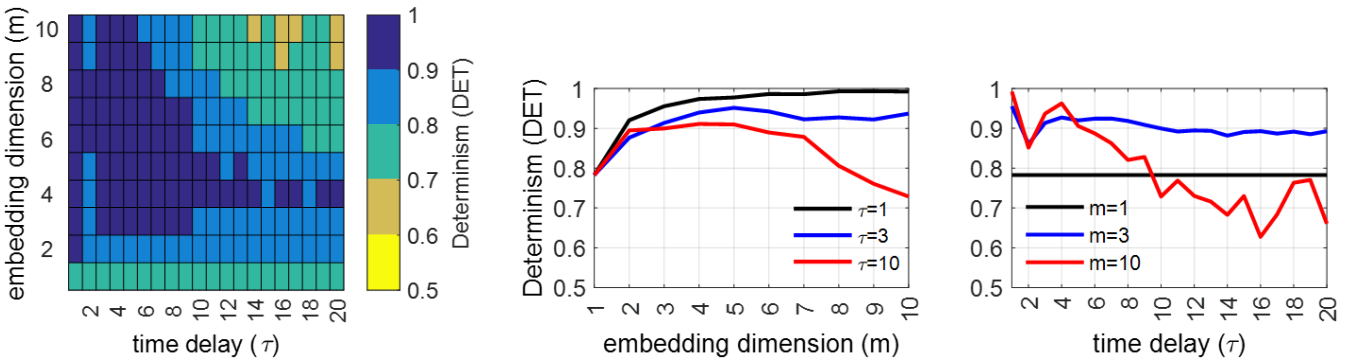


Fig. 8. Impact of noise levels of the Lorenz series on (a) false nearest neighbor, (b) mutual information, (c) determinism, and (d) median determinism distance extracted at $m = 1, 3, 5$ with $\tau = 1$. (e) and (f) present the extracted values with parameter bounds of $3 \leq m \leq 5$, $\tau = 3$ and 10, and 10% recurrence rate (ϵ). Noise added is Gaussian white noise with noise levels derived from the percentage of the signal standard deviation.

Int. J. Bifurcation Chaos 2018.28. Downloaded from www.worldscientific.com
by Dr. Norbert Marwan on 05/15/18. For personal use only.



(a) (b) (c)

Fig. 9. Recurrence characteristics of daily river runoff series: (a) determinism corresponding to m and τ, (b) change in determinism corresponding to increasing embedding dimension (m) from the RP, with $\tau = 1, 3, 7$; and (c) change in determinism corresponding to increasing time delay (τ), with $m = 1, 3, 7$.

and to fluctuate, as indicated by S_d. In this case, the use of the median is quite reliable due to the low S_d value (i.e. below 6%).

The identified determinism-sensitive region can be referenced with the standard approaches, such as FNN and MI, to find the optimal parameter set. This also serves to prevent the use of unnecessarily high parameter values that result in the reduction of data points (i.e. by $(m-1)\tau$). For instance, in the case of the Lorenz series, the optimal parameter set found by the standard approach is $m = 3$ and $\tau = 3$ (Fig. 7) which coincides well with the domain of high M_d values.

To investigate the impact of noise as in a real world scenario, Gaussian white noise with different noise levels is added to the signal as described in

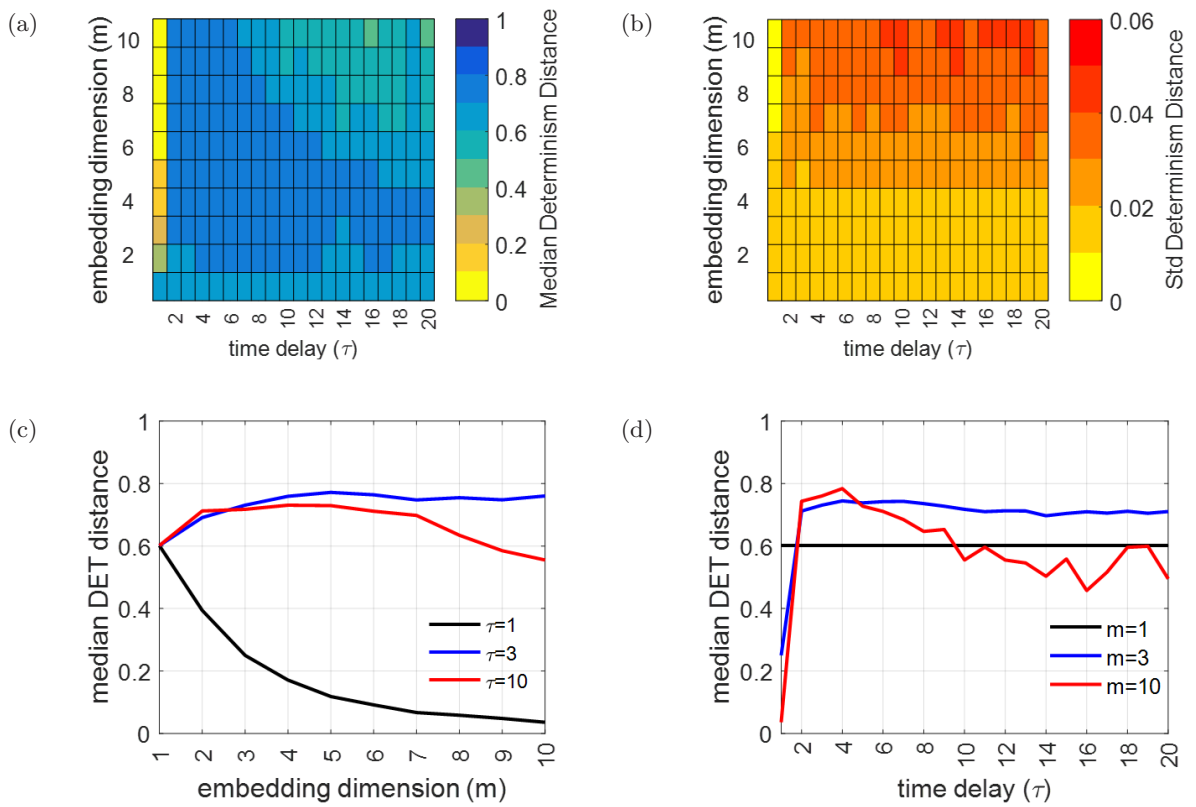


Fig. 10. Determinism distance of runoff series: (a) median and (b) standard deviation of the determinism distance between the RP of shuffled and original runoff series. (c) and (d) show the median determinism distance corresponding to $\tau = 1, 3, 10$ and $m = 1, 3, 10$.

Int. J. Bifurcation Chaos 2018.28. Downloaded from www.worldscientific.com
by Dr. Norbert Marwan on 05/15/18. For personal use only.

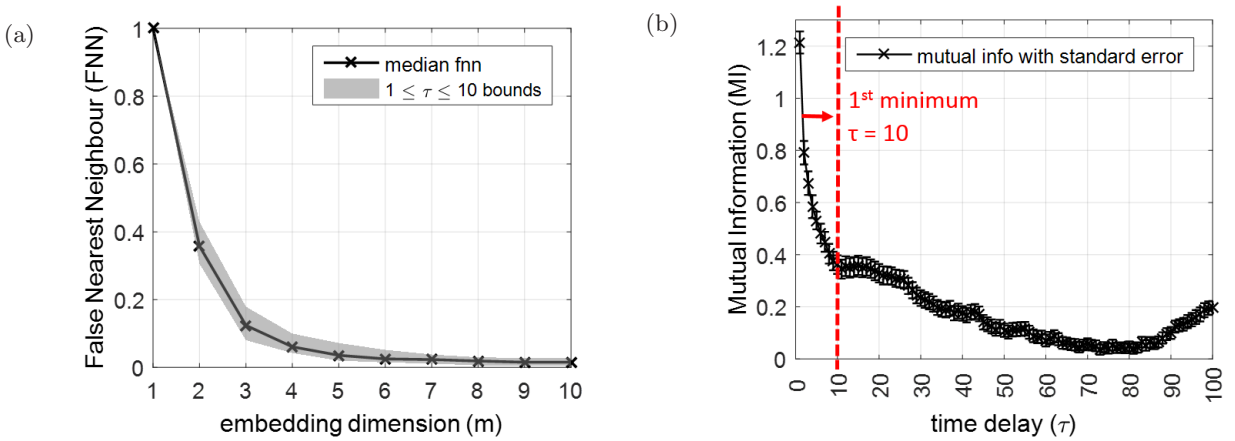


Fig. 11. Embedding parameter selection for daily river runoff using the standard approach: (a) false nearest neighbor with median and bounds derived from parameter set $1 \leq \tau \leq 10$ and (b) mutual information with first minimum found at $\tau = 10$.

Int. J. Bifurcation Chaos 2018.28. Downloaded from www.worldscientific.com
by Dr. Norbert Marwan on 05/15/18. For personal use only.

Sec. 4.1. Figures 8(a) and 8(b) show both false nearest neighbor and mutual information characteristics for the added-noise signal. The false nearest neighbor approach slightly increases at the optimal dimension of 3 causing a shift to the next dimension value, i.e. $m = 4$. When the noise level reaches 30% and 50%, the mutual information characteristics start to differ from the original, whereas the noise levels of 5% and 10% still preserve the original signal characteristics. Noise needs to be handled with care, as high level noise contamination potentially alters the determinism of the signal. It decreases in this case when Gaussian white noise is added, hence the determinism distance between the original and the shuffled series gets smaller.

4.2. *River runoff series*

A river runoff series is used to represent an example for field observations which are usually contaminated with noise. River runoff is typically a nonlinear deterministic series and exhibits chaos properties [Porporato & Ridolfi, 1997; Sivakumar, 2000; Martins *et al.*, 2011], hence, its DET is expected to be high. However, its recurrence determinism is low when parameter $m = 1$ and when both m and τ reach high values, e.g. $m > 8$ and $\tau > 9$ [Fig. 9(a)]. For instance, for $\tau = 10$ the DET value starts to decrease when $m > 7$ [Fig. 9(b)], while for $m = 10$ the increase of τ (i.e. above 4) also starts to reduce DET values [Fig. 9(c)].

When evaluated through 100 shuffles, the parameter set of $\tau = 1, m > 1$ should not be used due to the clear artefact potential suggested by its low determinism distance [see Fig. 10(a) — first column and Fig. 10(c) — black line]. The artefact-safe region could then be deduced from the high determinism distance domain corresponding to different combination parameter sets. For example, median determinism distance values above 0.8 imply high dissimilarity between the recurrence of the original signal and the shuffled ones [see Fig. 10(a)]. The S_d values in this case are also low to safely use the median values [see Fig. 10(b)].

As cross-check with the standard approach of parameter identification [Fig. 11(a)] suggested optimal embedding parameters in this case would be $\tau = 10$ days and $m = 5$.

5. Summary

We propose a method to identify a determinism-sensitive parameter region with minimal impact of artefacts due to embedding when constructing a Recurrence Plot (RP). The method utilizes both deterministic (incl. chaos) and stochastic characteristics of recurrence quantification, i.e. diagonal structures, as indicated by their determinism values. It is useful when the evaluated signal is known to be deterministic. The method involves randomly shuffling the time series for an abundant number of times in order to destroy its original characteristics and its determinism. Thereafter, determinism values are calculated for each shuffle iteration and compared with the determinism of the original signal for a range of parameters, resulting in a measure called determinism distance.

The matrix of the median values of this measure is plotted to depict the determinism-sensitive parameter region. The larger the determinism

distance, i.e. the closer to 1, the safer the parameter set is to avoid potential artefacts. The optimal parameter set can be selected by considering this artefact-safe region together with the standard approach of using false nearest neighbors and mutual information and auto-correlation.

Noise needs to be handled with care, since it affects the determinism structures of the signal or decreases the determinism values, therefore reducing the determinism distance between the original and shuffled series. One could apply this method as an artefact-precautionary measure especially when intending to choose high values of embedding parameters.

Int. J. Bifurcation Chaos 2018.28. Downloaded from www.worldscientific.com
by Dr. Norbert Marwan on 05/15/18. For personal use only.

Acknowledgment

This research was carried out within the Research Training Group "Natural Hazards and Risks in a Changing World (NatRiskChange; GRK 2043/1) funded by the Deutsche Forschungsgemeinschaft" (DFG).

References

Buzug, T. & Pfister, G. [1992] "Comparison of algorithms calculating optimal embedding parameters for delay time coordinates," *Physica D* **58**, 127–137.

Eckmann, J.-P., Kamphorst, S. O. & Ruelle, D. [1987] "Recurrence plots of dynamical systems," *Europhys. Lett.* **4**, 973.

Fraser, A. M. & Swinney, H. L. [1986] "Independent coordinates for strange attractors from mutual information," *Phys. Rev. A* **33**, 1134–1140.

Gao, Z. & Jin, N. [2009] "Flow-pattern identification and nonlinear dynamics of gas-liquid two-phase flow in complex networks," *Phys. Rev. E*, doi: 10.1103/PhysRevE.79.066303.

Kantz, H. & Schreiber, T. [2005] "Nonlinear time series analysis," *Technometrics* **47**, 381–381.

Kennel, M. B., Brown, R. & Abarbanel, H. D. I. [1992] "Determining embedding dimension for phase-space reconstruction using a geometrical construction," *Phys. Rev. A* **45**, 3403–3411.

Koebbe, M., Mayer-Kress, G. & Zbilut, J. [1994] "Use of recurrence plots in the analysis of time-series data," *Proc. SFI Studies in the Science of Complexity*, eds. Casdagli, M. & Eubank, S.

Lorenz, E. N. [1963] "Deterministic nonperiodic flow," *J. Atmosph. Sci.* **20**, 130–141.

Martins, O. Y., Sadeeq, M. A. & Ahaneku, I. E. [2011] "Nonlinear deterministic chaos in Benue river flow daily time sequence," *J. Water Resource and Protection* **3**, 747–757.

Marwan, N., Carmen Romano, M., Thiel, M. & Kurths, J. [2007] "Recurrence plots for the analysis of complex systems," *Phys. Rep.* **438**, 237–329.

Marwan, N. [2010] "How to avoid potential pitfalls in recurrence plot based data analysis," *Int. J. Bifurcation and Chaos* **21**, 1003–1017.

Mindlin, G. B. & Gilmore, R. [1992] "Topological analysis and synthesis of chaotic time series," *Physica D* **58**, 229–242.

Packard, N. H., Crutchfield, J. P., Farmer, J. D. & Shaw, R. S. [1980] "Geometry from a time series," *Phys. Rev. Lett.* **45**, 712–716.

Porporato, A. & Ridolfi, L. [1997] "Nonlinear analysis of river flow time sequences," *Water Resources Research* **33**, 1353–1367.

Schinkel, S., Dimigen, O. & Marwan, N. [2008] "Selection of recurrence threshold for signal detection," *Eur. Phys. J. Special Topics* **164**, 45–53.

Sivakumar, B. [2000] "Chaos theory in hydrology: Important issues and interpretations," *J. Hydrology* **227**, 1–20.

Sparrow, C. [1982] *The Lorenz Equations*: *Bifurcations, Chaos, and Strange Attractors*, Applied Mathematical Sciences, Vol. 41 (Springer, NY).

Takens, F. [1981] "Detecting strange attractors in turbulence," *Dynamical Systems and Turbulence, Warwick 1980*, Lecture Notes in Mathematics, Vol. 898 (Springer-Verlag), pp. 366–381.

Thiel, M., Romano, M. C., Kurths, J., Meucci, R., Allaria, E. & Arecchi, F. T. [2002] "Influence of observational noise on the recurrence quantification analysis," *Physica D* **171**, 138–152.

Trulla, L. L., Giuliani, A., Zbilut, J. P. & Webber, C. L. [1996] "Recurrence quantification analysis of the logistic equation with transients," *Phys. Lett. A* **223**, 255–260.

Zbilut, J. P. & Webber, C. L. [1992] "Embeddings and delays as derived from quantification of recurrence plots," *Phys. Lett. A* **171**, 199–203.

Zbilut, J. P., Zaldivar-Comenges, J.-M. & Strozzi, F. [2002] "Recurrence quantification based Liapunov exponents for monitoring divergence in experimental data," *Phys. Lett. A* **297**, 173–181.

Appendix A

Figure 12 describes the impact of recurrence threshold (i.e. recurrence rate) on the determinism distance between the original and all shuffled Lorenz time series. It can be seen that large recurrence threshold would lead the recurrence plot into artefact as implied by its low determinism distance [Figs. 12(b), 12(c), 12(e) and 12(f)]. Similarly when

Int. J. Bifurcation Chaos 2018.28. Downloaded from www.worldscientific.com
by Dr. Norbert Marwan on 05/15/18. For personal use only.

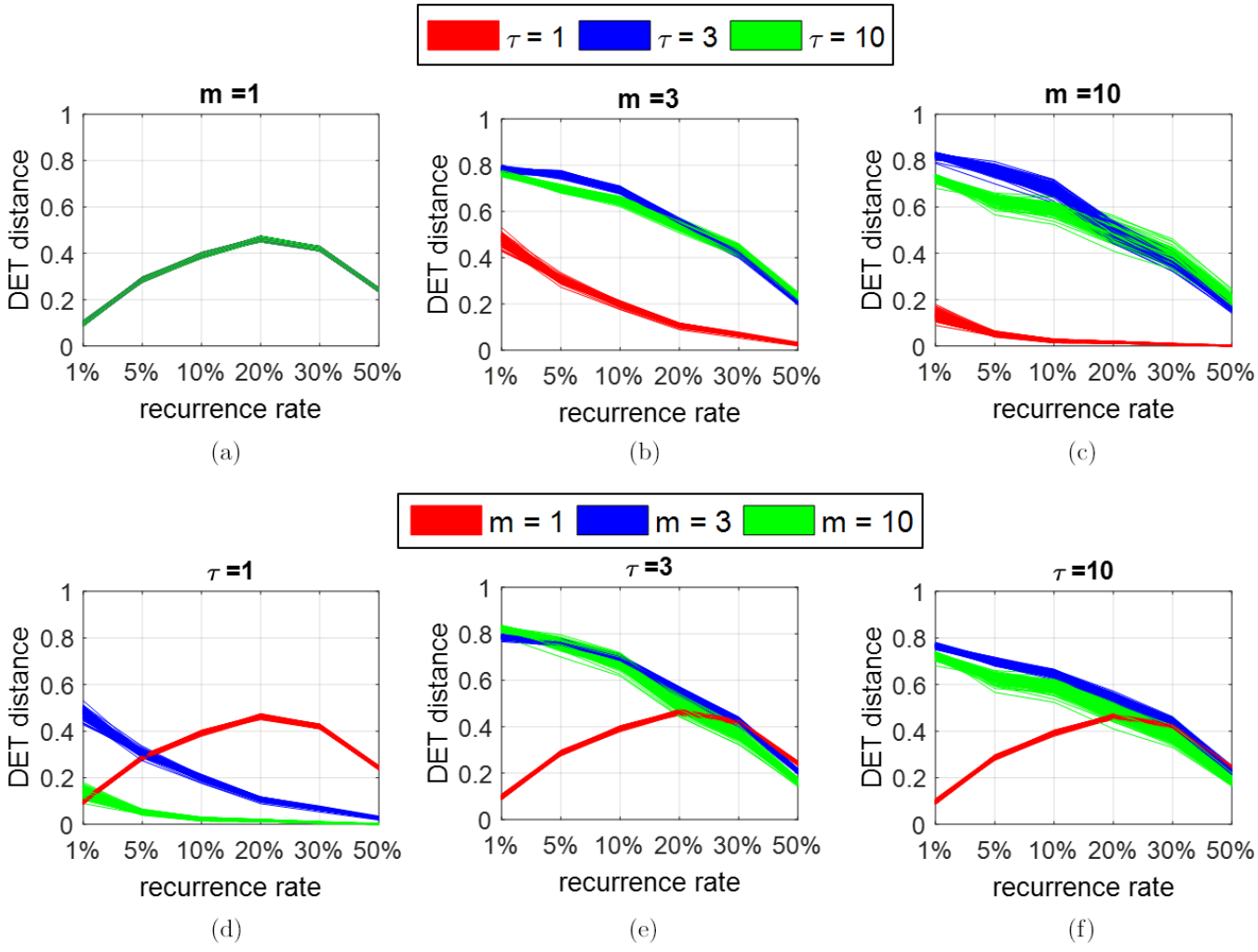


Fig. 12. Impact of recurrence threshold (i.e. recurrence rate) on determinism distances (100 shuffles) corresponding to different embedding parameters. Note in (a) red and blue band lines are overlapped by green bands, hence not visible.

Int. J. Bifurcation Chaos 2018.28. Downloaded from www.worldscientific.com
by Dr. Norbert Marwan on 05/15/18. For personal use only.

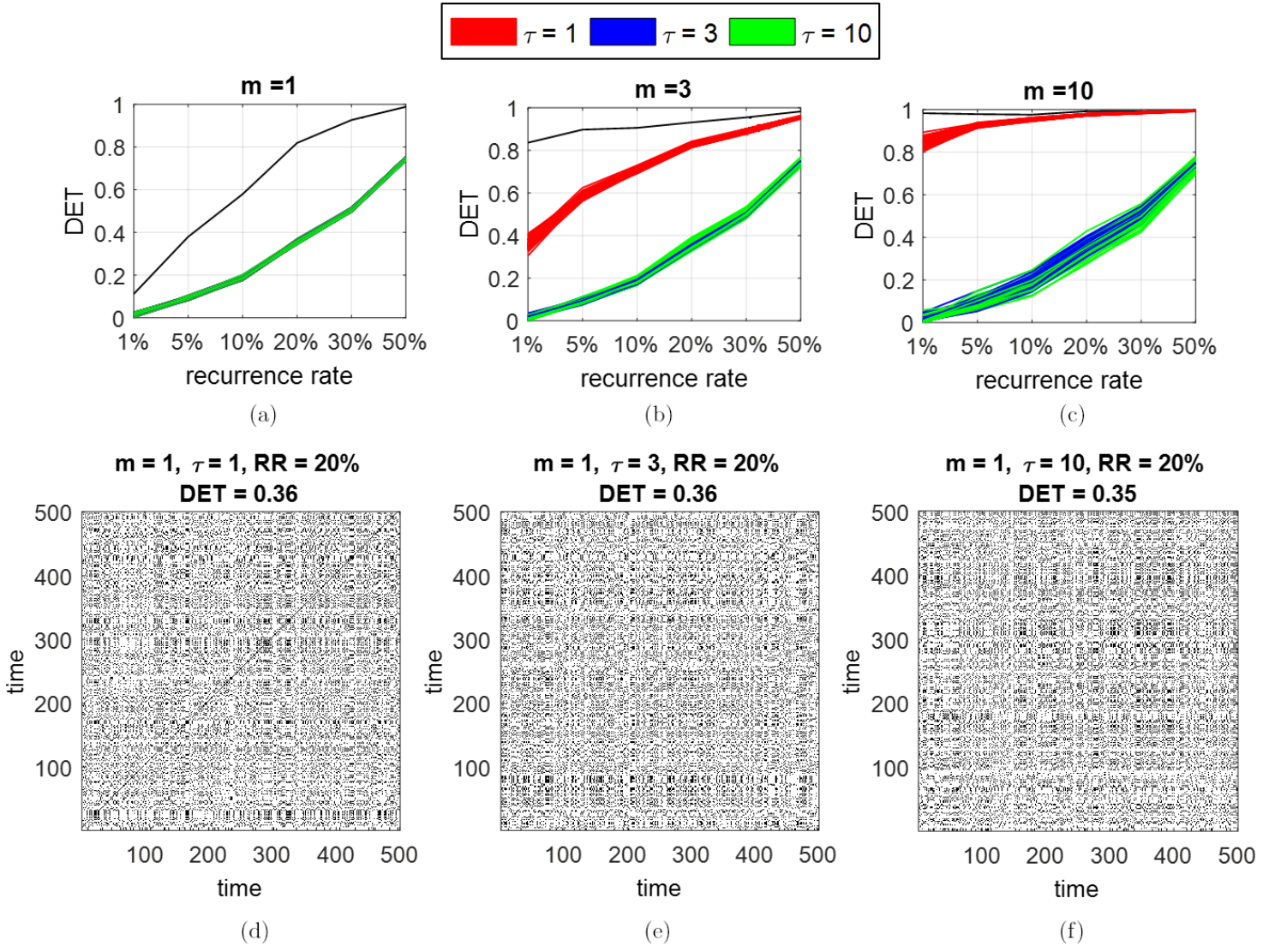


Fig. 13. Impact of recurrence threshold (i.e. recurrence rate) on DET (black line is resulted from original Lorenz time series, while each colored ones resulted from the shuffled series ($n = 100$), the colors correspond to different selection of τ) with $m = 1, 3$ and 10 respectively for (a)–(c). Subfigures (d)–(f) present an extracted sample of the shuffled recurrence plot with fixed $m = 1$ and $\tau = 1, 3$, and 10 respectively. Note in (a) red and blue band lines are overlapped by green bands, hence not visible.

$\tau = 1, m > 1$, the increase of recurrence rate further decreases the determinism distance [Fig. 12(d)]. This confirms that the use of such parameter value should be ignored regardless of the choice of recurrence threshold. However when $m = 1$, there is an increase of determinism distance i.e. peaking at 20% recurrence rate and decreases thereafter [Fig. 12(d)]. Despite the increase, the determinism distance is still regarded as low (i.e. below 0.5).

Both original and shuffled time series experience an increase of their DET values when recurrence rate is increased [Figs. 13(a)–13(c)]. However, when $m > 1$ the increase of DET values is rather sharp, changing significantly from low to high. The shuffled series recurrence plots with $m = 1$, and recurrence rate of 20% (i.e. at the peak of DET distance) still do not present any noteworthy deterministic features [Figs. 13(d)–13(f)]. In this case, users should avoid using large threshold values and special attention should be made on using such shuffling technique (i.e. when choosing $m = 1$).

Appendix B

In this Appendix, we include the evaluation of DET values of a correlated random series (exemplified using AR1 series) in contrast with uncorrelated random series and Lorenz to showcase that the high DET values is indeed associated with deterministic system instead of its auto-correlation structures (see Figs. 14 and 15). Although there are cases at certain embedding parameter set where diagonal lines of RP artificially increase, similar to random uncorrelated series, when τ equals 1, AR1 appears to artificially induce long diagonal lines and hence high DET when embedding dimension gets higher. Meanwhile, in addition to what was mentioned, there appear some (but rare) relatively high DET at certain parameter sets, e.g. when $6 \leq m \leq 10$ and $\tau = 7$ (see Figs. 13 and 14). However, there is a possibility that we have excluded some rare structures of a nondeterministic signal that also showcase high DET values. Therefore, it is important to note that this proposed technique is intended to be used when the user knows that the dynamical system is deterministic.

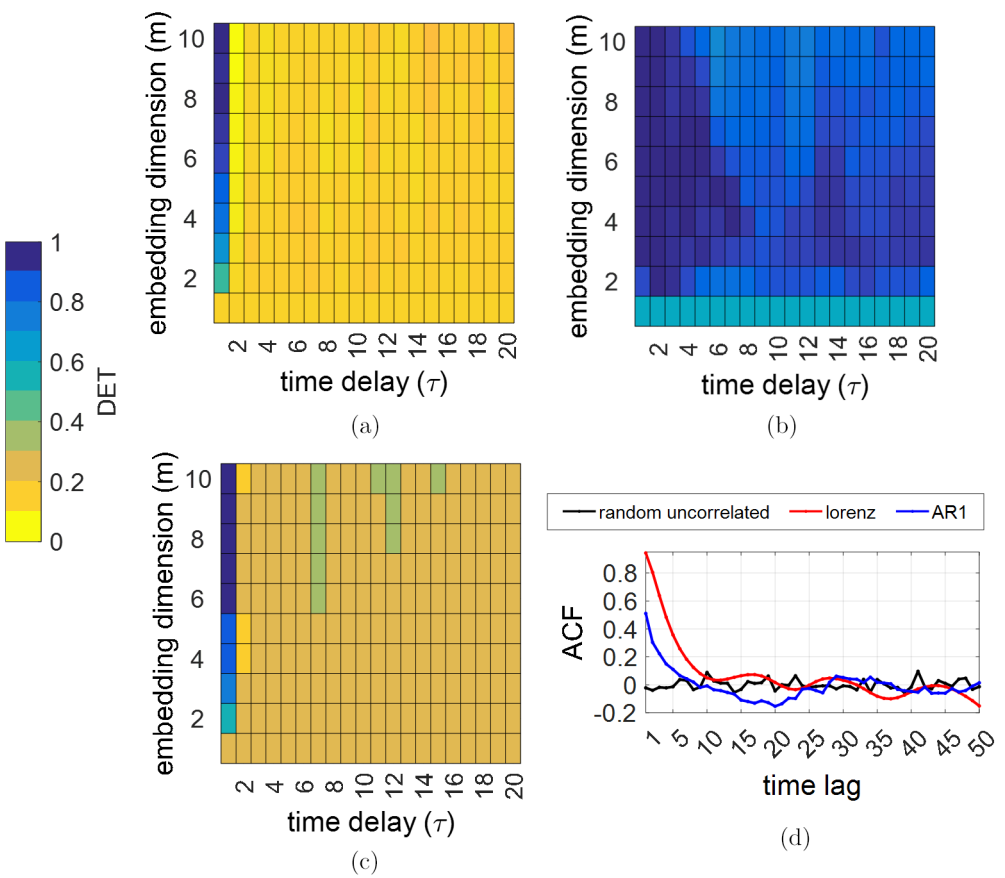


Fig. 14. DET of (a) random uncorrelated series, (b) Lorenz, (c) AR1 and (d) their corresponding autocorrelation.

Int. J. Bifurcation Chaos 2018.28. Downloaded from www.worldscientific.com
by Dr. Norbert Marwan on 05/15/18. For personal use only.

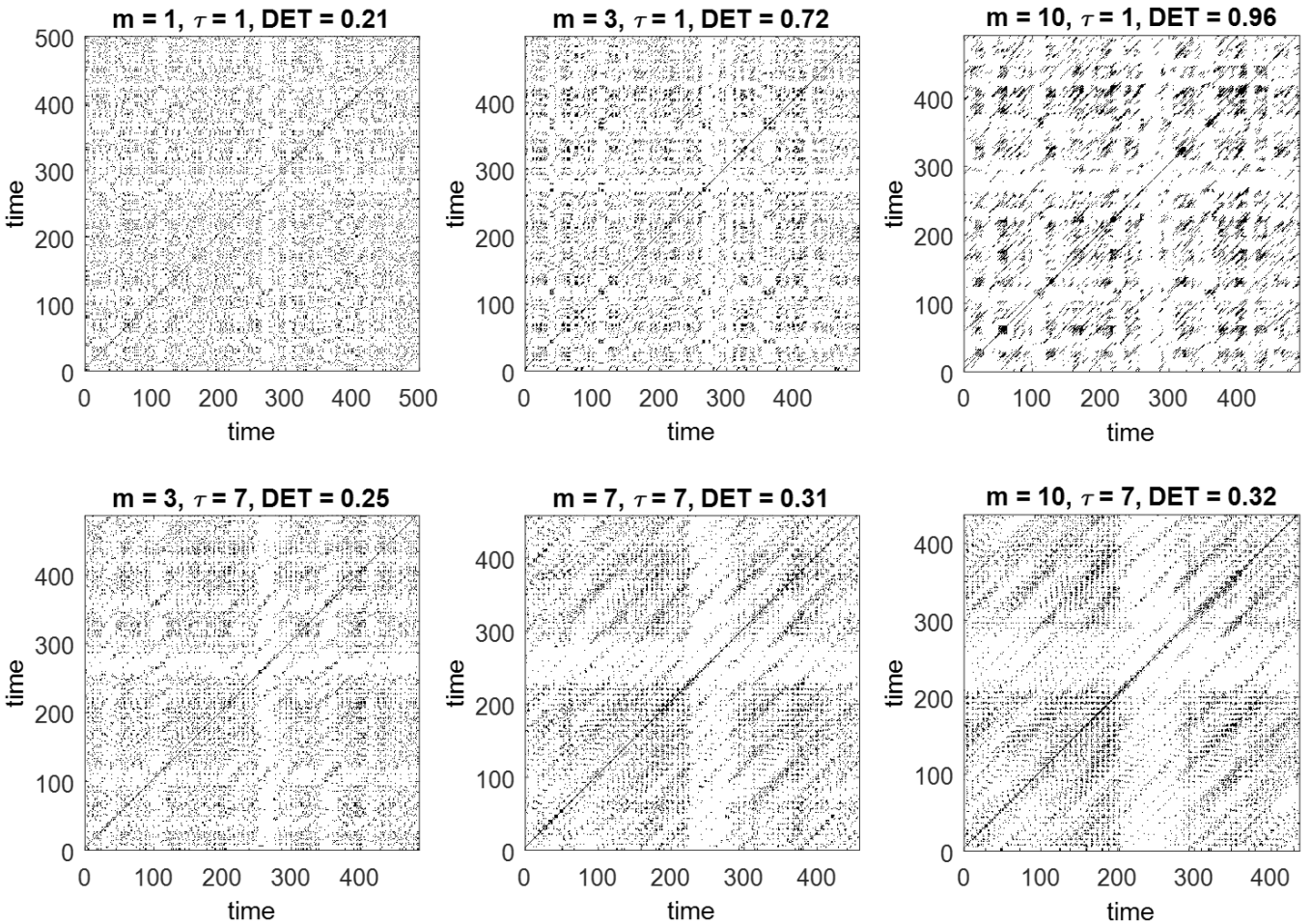


Fig. 15. Selected recurrence plots of AR1 with different embedding parameter sets to showcase the artificial increase of diagonal lines.

Int. J. Bifurcation Chaos 2018.28. Downloaded from www.worldscientific.com
by Dr. Norbert Marwan on 05/15/18. For personal use only.