



POTSDAM-INSTITUT FÜR
KLIMAFOLGENFORSCHUNG

Originally published as:

Krämer, K.-H., Donner, R. V., Heitzig, J., Marwan, N. (2018): Recurrence threshold selection for obtaining robust recurrence characteristics in different embedding dimensions. - Chaos, 28, 8, 085720.

DOI: <https://doi.org/10.1063/1.5024914>

This article may be downloaded for personal use only. Any other use requires prior permission of the author and AIP Publishing.

Recurrence threshold selection for obtaining robust recurrence characteristics in different embedding dimensions

K. Hauke Kraemer,^{1,2, a)} Reik V. Donner,¹ Jobst Heitzig,¹ and Norbert Marwan¹

¹⁾*Potsdam Institute for Climate Impact Research, Telegrafenberg A31, 14473 Potsdam, Germany, EU*

²⁾*Institute of Earth and Environmental Science, University of Potsdam, Karl-Liebknecht-Str. 24-25, 14476 Potsdam-Golm, Germany, EU*

(Dated: 13 August 2018)

The appropriate selection of recurrence thresholds is a key problem in applications of recurrence quantification analysis (RQA) and related methods across disciplines. Here, we discuss the distribution of pairwise distances between state vectors in the studied system's state space reconstructed by means of time-delay embedding as the key characteristic that should guide the corresponding choice for obtaining an adequate resolution of a recurrence plot. Specifically, we present an empirical description of the distance distribution, focusing on characteristic changes of its shape with increasing embedding dimension. Our results suggest that selecting the recurrence threshold according to a fixed percentile of this distribution reduces the dependence of recurrence characteristics on the embedding dimension in comparison with other commonly used threshold selection methods. Numerical investigations on some paradigmatic model system with time-dependent parameters support these empirical findings.

PACS numbers: 05.45.Tp, 05.90.+m, 89.75.Fb

Recurrence plots provide an intuitive tool for visualizing the (potentially multi-dimensional) trajectory of a dynamical system in state space. In case only univariate observations of the system's overall state are available, time-delay embedding has become a standard procedure for qualitatively reconstructing the dynamics in state space. The selection of a threshold distance ε , which distinguishes close from distant pairs of (reconstructed) state vectors, is known to have a substantial impact on the recurrence plot and its quantitative characteristics, but its corresponding interplay with the embedding dimension has not yet been explicitly addressed. Here, we point out that the results of RQA and related methods are qualitatively robust under changes of the (sufficiently high) embedding dimension only if the full distribution of pairwise distances between state vectors is considered for selecting ε , which is achieved by consideration of a fixed recurrence rate.

world applications, however, inferring complete dynamical information from observations is hampered by the fact that only some of the dynamically relevant variables are directly observable. In such cases, it has been demonstrated¹ that it is possible to qualitatively reconstruct representations of the unobserved components of a higher-dimensional system by means of embedding techniques applied to a suitably chosen individual component². Specifically, time-delay embedding has become a widely utilized method in nonlinear time series analysis, where a series of univariate observations $\{x_i\}$ (the actual time series at hand) is unfolded into a sequence of m -dimensional state vectors $\{\vec{x}_i\}$ ^{1,3} defined as $\vec{x}_i = (x_i, x_{i-\tau}, \dots, x_{i-(m-1)\tau})^T$, where m and τ denote the chosen embedding dimension and embedding delay, respectively.

Introduced by Eckmann et al.⁴, recurrence plots (RPs) provide a versatile tool for visualizing and quantitatively analyzing the succession of dynamically similar states in a time series. For this purpose, dynamical similarity is measured in terms of some metric distance $d_{i,j} = \|\vec{x}_i - \vec{x}_j\|$ defined in the underlying system's (reconstructed) state space. Based on the resulting distance matrix $\mathbf{d} = (d_{i,j})$, a recurrence matrix $\mathbf{R} = (R_{i,j})$ is defined as a thresholded version such that its entries assume values of 1, if the distance between the two associated state vectors is smaller than or equal to a threshold ε , and 0 otherwise:

$$R_{i,j}(\varepsilon) = \begin{cases} 1 & d_{i,j} \leq \varepsilon \\ 0 & d_{i,j} > \varepsilon, \end{cases} \quad i, j = 1, \dots, N. \quad (1)$$

Equivalently, we can write

$$R_{i,j}(\varepsilon) = \Theta(\varepsilon - d_{i,j}), \quad i, j = 1, \dots, N, \quad (2)$$

I. INTRODUCTION

A vector time series $\{\vec{x}_i\}_{i=1}^N$ (with $\vec{x}_i = \vec{x}(t_i)$) provides an approximation of a specific trajectory of a given dynamical system in finite-time and (for time-continuous dynamical systems) finite-resolution. In many real-

^{a)}Electronic mail: hkraemer@pik-potsdam.de, hkraemer@uni-potsdam.de

where $\Theta(\cdot)$ is the Heaviside function. In this definition, the threshold ε is fixed with respect to all pairwise distances contained in \mathbf{d} , and we will focus only on this kind of threshold application throughout this paper. An alternative definition of the recurrence matrix^{4,5}, which shall not be further considered in this study, replaces the global, fixed recurrence threshold ε applied to all state vectors \vec{x}_i by an adaptive local one that is defined such that the number of recurrences (i.e., close state vectors) is the same for each \vec{x}_i (fixed amount of nearest neighbors)⁴, leading to a constant local recurrence rate.

According to the above definition, for a given time series the recurrence matrix depends on the chosen recurrence threshold ε together with the selected norm $\|\cdot\|$ used for defining pairwise distances between the state vectors. In this work, we will restrict ourselves to two of the most commonly used norms: the Euclidean (L_2) and maximum (L_∞ , supremum, Chebychev) norms. Specifically, we will study how the distributions of pairwise L_2 and L_∞ distances depend on the embedding dimension.

Previous studies have provided various complementary suggestions for (i) selecting the right method of determining the recurrence threshold and (ii) choosing its actual value in some automatic way based on the specific properties of the system under study. Corresponding approaches include the spatial extent of the trajectory in the (reconstructed) state space^{6,7}, signal to noise ratio⁷⁻¹⁰, the specific dynamical system underlying the time series under investigation^{8,11}, or properties of the associated recurrence network^{12,13} with adjacency matrix $A_{i,j} = R_{i,j} - \delta_{i,j}$ (with $\delta_{i,j}$ being the Kronecker symbol) like the percolation threshold^{14,15}, second smallest eigenvalue of the graph's Laplacian¹⁶, breakdown of ε^{-1} scaling of the average path length¹⁴, or information-theoretic characteristics¹⁷. In practice, the appropriate choice of the method for determining the recurrence threshold, as well as its resulting value itself, can depend on the specific problem under study and take any of the above criteria or even some multiple-objective considerations based on different criteria into account. To this end, a general solution to the second problem of selecting a specific value of ε has not yet been obtained, and we will also not address this problem specifically in the course of the present paper. Instead, we are attempting to provide some further insights into the first, more conceptual problem setting (i.e., which type of approach for selecting recurrence thresholds should be taken in case of varying situations such as different embedding dimensions).^{12-14,17}

As we will further detail in the course of this paper, some previously suggested approaches^{6,7,10,18} to link a recurrence threshold to a certain percentage of the maximal or mean distance of all pairwise distances of state vectors (i.e., a given fraction of the attractor's diameter in the reconstructed state space) cause the resulting recurrence characteristics to strongly depend on the embedding dimension. The reason for this behavior is that in addition to a general increase of distances¹⁹ (depending on the chosen norm)¹⁸, the shape of the distance distribution

also changes with increasing embedding dimension (see Fig. 1 and further discussions in Section II).

It should be noted that embedding a time series with $m \sim \mathcal{O}(10^1)$ or even larger can become necessary when the correlation dimension D_2 of the attractor is rather large. This is due to the fact that Takens' theorem (and several extensions thereof) guarantee the existence of a diffeomorphism between the original and the reconstructed attractor if m satisfies $m \geq 2D_2 + 1$.^{1,5,20} Hegger et al.²¹ emphasize that it is also advisable to choose a rather high value of m when dealing with time series originating from a D -dimensional deterministic system that is driven by P slowly time dependent parameters. An appropriate yet conservative choice for m then fulfills $m \geq 2(D + P)$. Concerning practical applications of nonlinear time series analysis, one commonly deals with signals originating from complex, non-stationary systems and, therefore, high embedding dimensions can become necessary, requiring threshold selection methods which lead to robust results of RQA and related state space based techniques that are robust under different choices of the embedding dimension.

In the following Section II, we study the influence of an increasing embedding dimension on the shape of the distance distribution in more detail. We deduce that, in order to avoid problems arising due to an unfavorable fixed recurrence threshold when varying m , we could choose ε as a certain percentile of the distance distribution rather than a certain percentage of the maximum or mean phase space diameter. Successively, Section III presents a numerical example of a classical Lorenz-63 system with a time-dependent parameter, illustrating that the changes in some recurrence characteristics with varying embedding dimension are particularly small under a fixed recurrence rate in comparison with other strategies. Finally, the main results of this study are summarized in Section IV.

II. INFLUENCE OF EMBEDDING DIMENSION ON THE DISTANCE DISTRIBUTION

Let us consider a univariate time series $\{x_i\}$ of length N . As an overarching question, we study the effect of time-delay embedding on the distribution of all pairwise distances of its reconstructed state vectors. The variation of this distribution with increasing embedding dimension m is expected to depend on the chosen norm used for the calculation of distances. Note that the effective number of state vectors $N_{\text{eff}}(m) = N - (m - 1)\tau$ available for estimating the probability distribution of distances in m dimensions will decrease with m . In order to avoid sample size effects in comparing the results for different m , we therefore choose N sufficiently large so that $1 - N_{\text{eff}}(m_f)/N \ll 1$, where m_f is the largest considered embedding dimension.

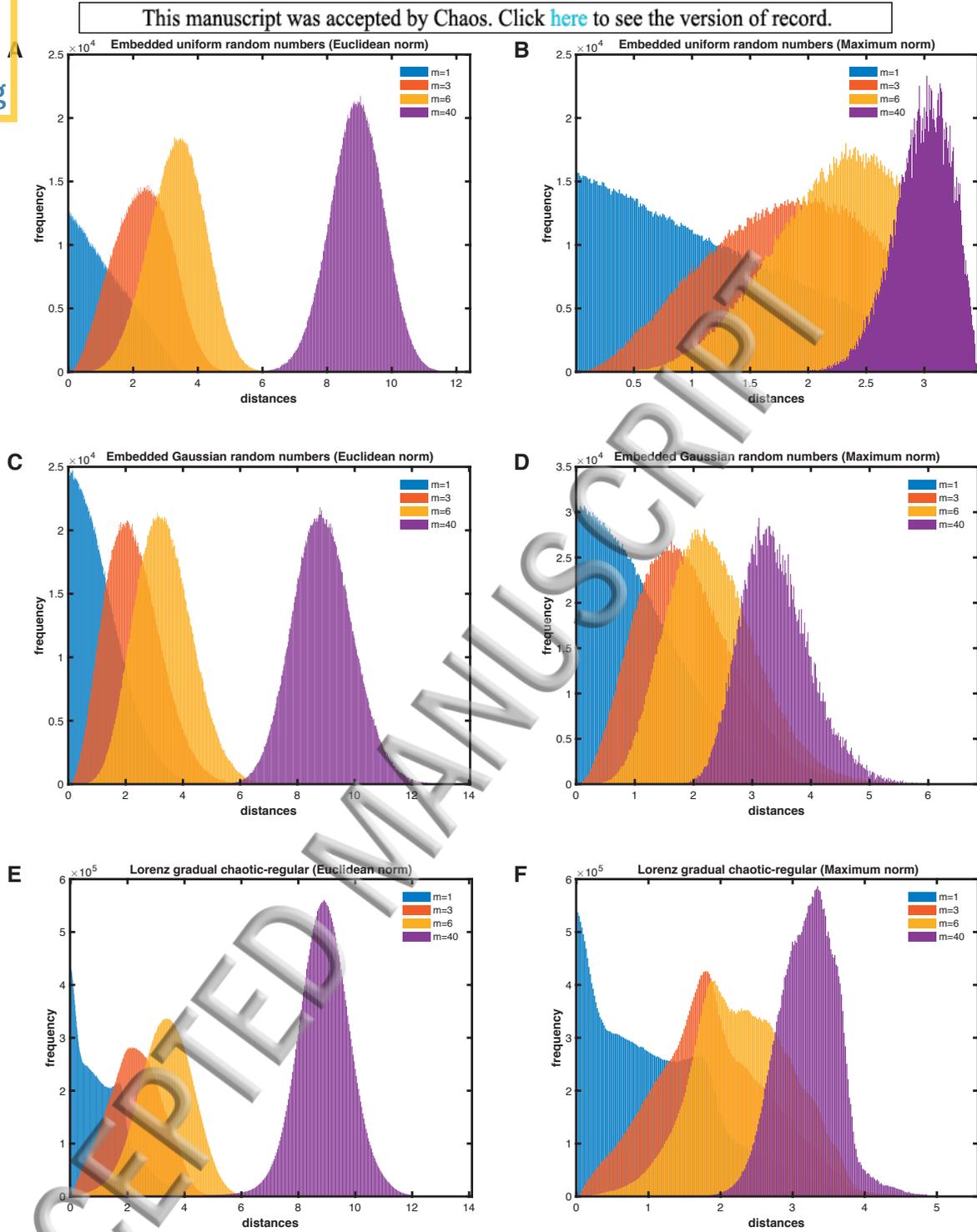


FIG. 1. Selected histograms of the L_2 (A,C,E) and L_∞ (B,D,E) distances of $N = 1,500$ independent random numbers with uniform (A,B) and Gaussian (C,D) distribution as well as (E,F) for the y component of the Lorenz-63 system (Eq. (11), $N = 6,000$, see Section III) with control parameters $\sigma = 10$, $\beta = 8/3$ and r linearly increasing from 180 (chaotic regime) to 210 (periodic regime), for different embedding dimensions m .

A. Maximum norm

Numerical results for different types of systems demonstrate (see Appendix A) that the largest of all pairwise

L_∞ distances, $d_{max}^{(\infty)}$, stays constant with increasing embedding dimension, whereas the mean of all pairwise L_∞ distances, $d_{mean}^{(\infty)}$, monotonically increases with m (Fig. 3). In order to understand this observa-

tion, recall that the L_∞ distance between two embedded state vectors $\vec{x}_i = (x_{i,1}, x_{i,2}, \dots, x_{i,m})^T$ and $\vec{x}_j = (x_{j,1}, x_{j,2}, \dots, x_{j,m})^T$ is

$$\|\vec{x}_i - \vec{x}_j\|_\infty = \max_{k=1, \dots, m} |x_{i,k} - x_{j,k}| = d_{i,j}^{(\infty)}(m) \quad (3)$$

For $m = 1$ (i.e., no embedding), the distance between two observations at times t_i and t_j therefore is simply $d_{i,j}^{(\infty)}(1) = |x_i - x_j|$. For $m = 2$, we find

$$\begin{aligned} d_{i,j}^{(\infty)}(2) &= \max\{|x_i - x_j|, |x_{i+\tau} - x_{j+\tau}|\} \\ &= \max\{d_{i,j}^{(\infty)}(1), |x_{i+\tau} - x_{j+\tau}|\} \geq d_{i,j}^{(\infty)}(1). \end{aligned} \quad (4)$$

By induction, we can easily show that

$$d_{i,j}^{(\infty)}(m) = \max\{d_{i,j}^{(\infty)}(m-1), |x_{i-(m-1)\tau} - x_{j-(m-1)\tau}|\}$$

and therefore

$$d_{i,j}^{(\infty)}(m) \geq d_{i,j}^{(\infty)}(m-1) \quad \forall m > 1. \quad (5)$$

Hence, considering all possible pairs of state vectors (\vec{x}_i, \vec{x}_j) from the time series, the largest L_∞ distance

$$d_{max}^{(\infty)}(1) = \max_{i,j} [d_{i,j}^{(\infty)}(1)] = \max_{i,j} [d_{i,j}^{(\infty)}(m)] = d_{max}^{(\infty)}(m), \forall m$$

cannot change with m , since the largest maximum distance will already appear for $m = 1$. The mean distance

$$d_{mean}^{(\infty)}(m) = \frac{1}{N_{\text{eff}}^2(m)} \sum_{i,j=1}^{N_{\text{eff}}(m)} d_{i,j}^{(\infty)}(m),$$

however, necessarily increases with m or stays at most constant. More specifically, as m increases, smaller distances systematically disappear, so that the entire distribution is systematically shifted towards its (constant) maximum, thereby becoming narrower and exhibiting an increasing mean along with decreasing variance. We conjecture that, for large m , the distribution of $d^{(\infty)}(m)$ will converge to a limiting distribution (see below) possibly depending on the embedding delay τ .

B. Euclidean norm

In case of the L_2 (Euclidean) norm, both mean and maximum of all pairwise distances ($d_{mean}^{(2)}(m)$ and $d_{max}^{(2)}(m)$, respectively) monotonically increase with rising m (Appendix A, Fig. 4). This can be understood as follows: The L_2 distance between two points in an m -dimensional state space, \vec{x}_i and \vec{x}_j , is given as

$$\|\vec{x}_i - \vec{x}_j\|_2 = \left(\sum_{k=1}^m |x_{i,k} - x_{j,k}|^2 \right)^{\frac{1}{2}} = d_{i,j}^{(2)}(m) \quad (6)$$

For the squared L_2 distance, this implies:

$$\begin{aligned} [d_{i,j}^{(2)}(1)]^2 &= (x_i - x_j)^2 \\ [d_{i,j}^{(2)}(2)]^2 &= (x_i - x_j)^2 + (x_{i-\tau} - x_{j-\tau})^2 \\ &= [d_{i,j}^{(2)}(1)]^2 + (x_{i-\tau} - x_{j-\tau})^2 \\ &\geq [d_{i,j}^{(2)}(1)]^2 \\ &\vdots \\ [d_{i,j}^{(2)}(m+1)]^2 &\geq [d_{i,j}^{(2)}(m)]^2 \geq \dots \geq [d_{i,j}^{(2)}(1)]^2, \end{aligned} \quad (7)$$

which explains the observed behavior of both mean and maximum distance using the L_2 norm. Specifically, unlike for L_∞ , the maximum L_2 distance between two points is not bound by the largest pairwise distance in one dimension.

In a similar way, we may argue for all L_p distances ($p \in (0, \infty)$) defined as

$$\|\vec{x}_i - \vec{x}_j\|_p = \left(\sum_{k=1}^m |x_{i,k} - x_{j,k}|^p \right)^{\frac{1}{p}} = d_{i,j}^{(p)}(m) \quad (9)$$

that, by the same argument as above,

$$[d_{i,j}^{(p)}(m+1)]^p \geq [d_{i,j}^{(p)}(m)]^p, \quad (10)$$

implying again a monotonic increase of mean and maximum distances with rising embedding dimension (recall the positive semi-definiteness of distances and p).

C. Changing shape of distance distribution with increasing embedding dimension

Building upon our previous considerations and numerical results, a mathematically more specific yet challenging question is how exactly an increasing embedding dimension m is affecting the shape of the distribution of all pairwise distances rather than just its central tendency (mean).

For the maximum norm, one may argue that the individual components of each embedded state vector are commonly constructed such that they are as independent as possible²². Accordingly, for a system without serial correlations (i.e., uncorrelated noise), the absolute differences $d = d^{(\infty)}(1)$ between the components of two state vectors are also independent, identically distributed (i.i.d.) and lie within the interval $[0, d_{max}]$. In such case, for sufficiently large m , the pairwise L_∞ distance between two state vectors can be interpreted as the maximum of m i.i.d. variables that are bounded from above, which should follow a reversed Weibull distribution according to the Fisher-Tippett-Gnedenko theorem from extreme value statistics. Note, however, that this expectation is

valid only if m is sufficiently large and the i.i.d. assumption is (approximately) fulfilled, both of which do not necessarily have to be the case for real-world time series. Moreover, it is not guaranteed that the given distance distribution in one dimension lies within the domain of attraction of the reversed Weibull class²³, which calls for further theoretical investigation in each specific case.

For other L_p norms including the Euclidean norm, the aforementioned considerations do not apply. Instead, for any L_p norm with $p < \infty$,

- the pairwise distances d^p are of the form $(\sum_i z_i^p)^{1/p}$ ($i = 1, \dots, m$) as given in Eq. (9) with approximately i.i.d. variables z_i .
- From the central limit theorem it follows that the distribution of d^p is approximately a normal distribution with mean and standard deviation growing proportionally with m and \sqrt{m} , respectively, for large m .
- The coefficient of variation of d^p thus declines approximately as $\sim 1/\sqrt{m}$.
- For large m , also $d = (d^p)^{1/p}$ is approximately normally distributed with mean and standard deviation growing approximately as $\sim m^{1/p}$ and $\sim \sqrt{m} \frac{dz^{1/p}}{dz} |_{z=m} \sim \sqrt{m} m^{1/p-1} = m^{1/p-1/2}$.
- The coefficient of variation of d thus behaves approximately as $\sim m^{1/p-1/2}/m^{1/p} = 1/\sqrt{m}$, just as for d^p .
- As a consequence, the relative variability of d narrows in the same fashion for all $p < \infty$ as m grows, and only the growth of the absolute scale of d with m depends on p (“curse of dimensionality”¹⁹).

The considerations made above do explain the numerical results in Fig. 1, showing histograms of the distances of three different time series for selected values of the embedding dimension m and for the L_2 and L_∞ norms. In addition to time series fulfilling the i.i.d. assumption (Fig. 1 A,B,C,D), here we are also interested in deterministic systems. As an illustrative example, we choose the Lorenz-63 system (Eq. (11), Fig. 1 E,F) in some non-stationary (drifting parameter) setting, which will be further studied in Section III.

In this regard, it is confirmed that the expectation value of the distance distribution takes higher values with increasing m . The probability to find small distances therefore decreases. In case of the L_∞ norm (Fig. 1 B,D,F), this growth is bounded and we can identify a convergence of the distribution, in some cases eventually towards the aforementioned reversed Weibull distribution. In turn, for the L_2 norm (Fig. 1 A,C,E) the convergence towards a normal distribution is discernible. Considering the Lorenz-63 time series (Fig. 1 E,F), the empirical expectations are approximately met by the observations, even though the distribution of L_∞ distances

exhibits a slightly more complex (i.e., less symmetric) shape than for the two noise series. Specifically, for the L_2 norm the resulting distance distribution is left-skewed with a pronounced lower tail (see Fig. 1 E), whereas for the L_∞ norm we observe a disturbed Weibull-like shape. Notably, the i.i.d. assumption is violated when dealing with such a deterministic dynamical system. For a more detailed characterization of the shape of the empirically observed pairwise distance distributions shown in Fig. 1, see Appendix B.

In general, we emphasize that it is not straightforward to analytically describe the shape of the distance distribution of an embedded time series stemming from an arbitrary dynamical system with potentially nontrivial serial correlations. Regarding our overarching question how we could automatically choose a recurrence threshold such that the resulting recurrence characteristics are as independent as possible of the embedding dimension and chosen norm, we need to consider both,

- (i) the general increase of distances together with their successive concentration and
- (ii) the varying shape of the distribution of distances

with increasing embedding dimension. The first aspect could be accounted for by relating the threshold selection to the spatial extent of the state space object (attractor), similar as, for instance, suggested by Abarbanel²⁴ in the context of the false nearest neighbor algorithm. However, our findings suggest that accounting for the second point is key to an appropriate recurrence threshold selection method that relieves the effects of the embedding dimension on the recurrence properties as much as possible. As a simple possible solution, we recommend to use a numerical estimate of a certain (sufficiently low) percentile of the distance distribution as threshold^{12–14,17}. This approach considers both above mentioned effects and leads to a constant global recurrence rate (which equals the chosen percentile). As a result, the recurrence properties become much less dependent on the embedding dimension and chosen norm than when using other methods, as we will exemplify in the following section.

We emphasize that in addition, by conserving the recurrence rate, possible dependences of RQA characteristics on the density of recurrences for different m are omitted, and corresponding residual changes of these measures upon varying m could rather point to either insufficiently low embedding dimension (missing essential factors contributing to the system’s dynamics, in a similar spirit as, e.g., for the false nearest neighbor method) or spurious recurrence structures arising from overembedding²⁵. These ideas should be further studied in future work.

III. NUMERICAL EXAMPLE

In this section, we will demonstrate the effect of the varying shape of the distance distribution with increas-

ing embedding dimension on different threshold selection approaches working with a globally fixed value of ε . In order to mimic a practically relevant test case of a non-stationary low-dimensional dynamical system, where we should use some higher embedding dimension (following Hegger *et al.*²¹) instead of a more moderate choice, we consider the classical Lorenz-63 system²⁶

$$\begin{aligned}\dot{x} &= \sigma(y - x) \\ \dot{y} &= x(r - z) - y \\ \dot{z} &= xy - \beta z.\end{aligned}\quad (11)$$

Depending on the parameters σ , β and r , the system exhibits either regular or chaotic dynamics. Here, we consider a transitory setting, where the parameter r gradually increases from 180 to 210 while keeping $\beta = 8/3$ and $\sigma = 10$ fixed. In this case, the system undergoes a transition from a chaotic regime into a regular (limit cycle) phase as r rises before it exhibits again a chaotic behavior. Note again that instead of studying the stationary Lorenz-63 system for different values of r , we intentionally employ a gradual parameter change leading to a non-stationary system which calls for a systematic overembedding when performing nonlinear time series analysis²¹. Specifically, we implement a linear variation of r as

$$r(t_{is}) = 180 + 2.5 \cdot 10^{-2} t_{is}.\quad (12)$$

For numerically solving this system of equations, we use a fourth-order Runge-Kutta integrator with an integration step of $t_{is} = 0.001$ and a total of 1,300,000 iterations. Therefore, we simulate the system's evolution over 1,300 time units (t.u.). By using a sampling interval of $\delta t = 0.2$ t.u. we obtain 6,500 samples forming our time series for the three components x , y and z . We remove the first 500 points ($\cong 100$ t.u.) that could be affected by transient dynamics and retain the remaining 6,000 points ($\cong 1200$ t.u.) of the y component for further analysis.

We integrate the Lorenz-63 equations, Eq. (11), with the linear parameter change, Eq. (12), 1,000 times with randomly chosen initial conditions, embed the y component time series using a delay $\tau = 4$, consistent with the first local minimum of the mutual information²², and assess the resulting RPs. For each of these 1,000 RPs, we use a running window along their main diagonal with a window size of $w = 400$ and mutual shift of $ws = 40$ data points, i.e., 90% overlap between consecutive windows, to quantitatively study the time-dependence of the resulting recurrence characteristics. We repeat this procedure for embedding dimensions ranging from $m = 3$ to $m = 10$ and for four different threshold selection methods: (i) a fixed percentile of the distance distribution (as recommended by our theoretical considerations in Section II) as well as some fixed percentage of the (ii) maximum, (iii) mean and (iv) median pairwise distances between all state vectors in the reconstructed state space, respectively.

Since we are aiming to study the change of recurrence properties associated with a transition between chaotic

and periodic dynamics and vice versa, we choose the *recurrence time entropy (RTE)*. Here, instead of using the diagonal or vertical “black” (recurrence) lines in the RP as in most “conventional” RQA measures, we use “white” (non-recurrence) vertical lines with lengths t_w , as they correspond to recurrence times. In general, such recurrence times can be estimated directly from the RP in different ways²⁷, among which the vertical non-recurrence lines offer a particularly simple estimator. The normalized entropy of the distribution of recurrence times, referred to as the *recurrence period density entropy*²⁸ and originally introduced without any direct link to RPs, is given as

$$RTE = \frac{1}{\ln T_{max}} \sum_{t_w=1}^{T_{max}} p(t_w) \ln p(t_w) \in [0, 1] \quad (13)$$

with $p(t_w)$ being the probability of a recurrence time t_w and T_{max} the largest recurrence time. Using RPs, it is possible to estimate $p(t_w)$ from the histogram of recurrence times, $h(t_w)$, as $p(t_w) = \frac{h(t_w)}{\sum_{t_w} h(t_w)}$, i.e., as the probability to find a white vertical line of exactly length t_w in the RP. It can be shown that *RTE* is closely linked to the Kolmogorov-Sinai (KS) entropy of the system under study²⁹.

We choose the actual recurrence threshold for each threshold selection method (i)-(iv) such that a global recurrence rate of $RR \approx 4\%$ is achieved in all four cases for $m = 3$. Therefore, for each embedding dimension we obtain a distribution of 1,000 *RTE* time series and show the mean (blue lines in Fig. 2) together with the two-sided 90% confidence interval ([5%, 95%], gray shaded areas). In order to put these time dependent *RTE* estimates of the non-stationary Lorenz-63 system into a context, we consider a reference reflecting the time-dependent *RTE* values directly computed from the true three-dimensional state vectors without embedding, using otherwise the same analysis strategy (window size and overlap) as for the embedding scenario. Thus, for each point in time we obtain 1,000 reference measurements and consider the mean (red line) and the two-sided 90% confidence interval (red shaded area in Fig. 2).

The robustness of the observed time-dependence of *RTE* with respect to the chosen embedding dimension when using a fixed percentile of the distance distribution (i.e., a fixed recurrence rate) is shown in Fig. 2A (here we used the L_2 norm, but the results are similar when using the L_∞ norm). For any embedding dimension larger than $m = 4$, the variations of the *RTE* estimates originating from the embedding procedures match the red reference time series within its uncertainties for times $t \gtrsim 200$. For adequately revealing the chaotic regime in the first part until $t \approx 160$, an embedding dimension larger than $m = 7$ seems to be inevitable, whereas results from any embedding dimension coincide with the reference estimate within its uncertainties at the limit cycle regime ($1,000 \leq t \leq 1,080$). In case of not using the recommended threshold selection method, this robustness

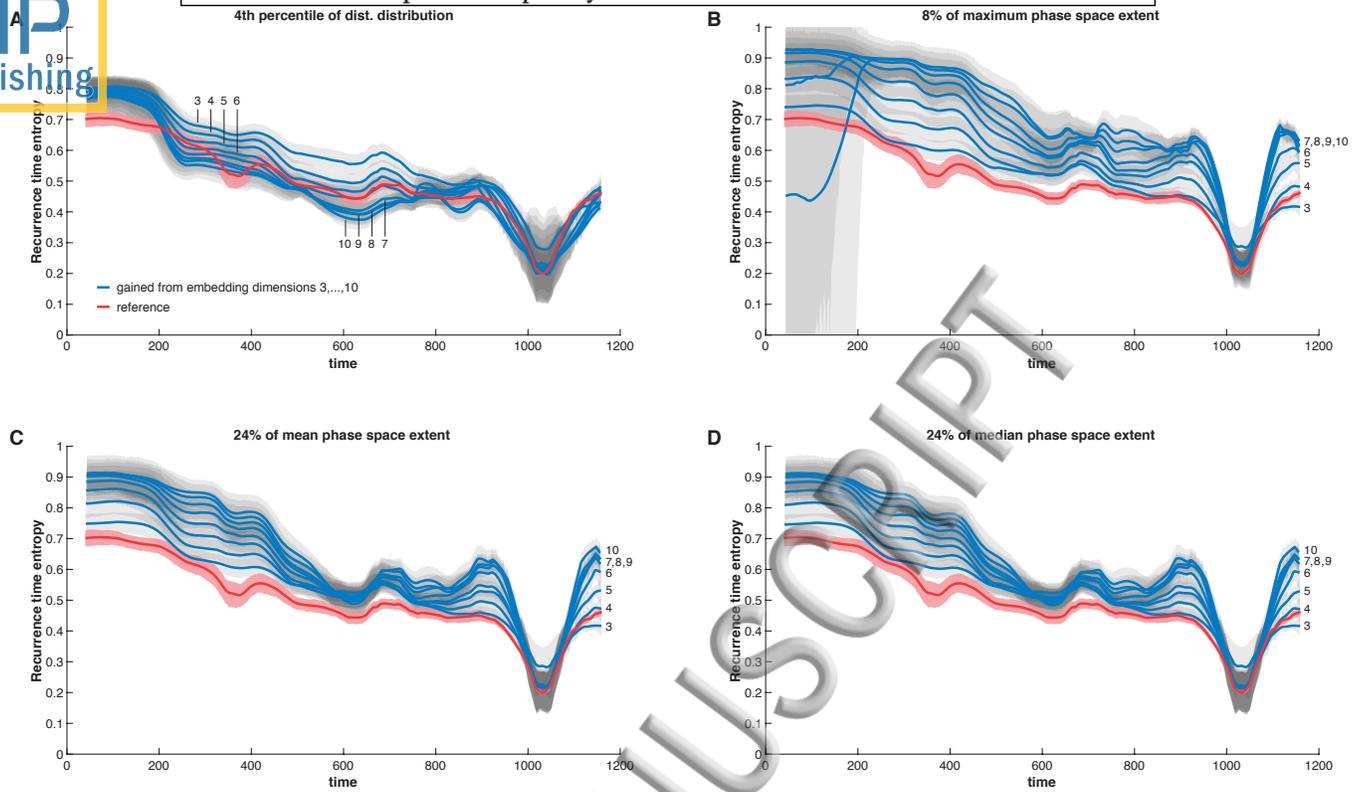


FIG. 2. Time-dependence of RTE (ensemble means and two-sided 90% confidence intervals from 1,000 independent realizations) based on the y component of the non-stationary Lorenz-63 system (see text for details) using the L_2 norm. The blue lines show the results for time-delay embedding with different embedding dimensions ($m = 3, \dots, 10$) and for four different methods to select the recurrence threshold according to (A) a certain percentile of the distance distribution and some percentage of the (B) maximum, (C) mean and (D) median distance between state vectors on the reconstructed attractor. The actual threshold values (4th percentile, 8%, 24% and 24%, respectively) have been chosen such that the global recurrence rate of approximately 4% is achieved for each method in the embedding scenario with $m = 3$. The red line shows the reference time series gained from 1,000 independent realizations of the non-stationary Lorenz-63 system by randomly choosing initial conditions and using all three components as state variables. Shaded areas (gray and red) indicate the two-sided 90% confidence intervals estimated from the respective ensembles.

is clearly lost (Fig. 2B,C,D), and only the limit cycle regime (plus some shorter sections before) are properly revealed by the estimates obtained in the reconstructed state space.

Considering the results of Section II, the reason for the failure of the methods based on individual location parameters (maximum, mean, median) of the pairwise distance distribution between state vectors for higher embedding dimensions is the change in the shape of that distribution beyond its characteristic location and range parameters. Appendix C demonstrates this effect on the RPs in some more detail. Hence, we argue that selecting the recurrence threshold at some percentile of the distance distribution is to be preferred if we aim to obtain stable results for a broad range of embedding dimensions, which is the case if we wish to automatically choose fixed recurrence thresholds for the analysis of arbitrary complex systems.

We note that the presented example has focused on a recurrence characteristic that is particularly well suited

for detecting transitions between chaotic and periodic dynamics and is linked to a dynamical invariant. Other recurrence characteristics, like classical RQA measures or recurrence network characteristics, have been found to exhibit less stable variations with changing embedding dimension (not shown) and are therefore not further discussed here. Clarifying the reasons for the different behaviors of different recurrence characteristics will be an important subject of future work.

IV. CONCLUSIONS

We have discussed the changing shape of the distribution of pairwise distances between state vectors obtained by time delay embedding with increasing embedding dimension and its implications for different methods for selecting temporally fixed recurrence thresholds. While specific values of the recurrence threshold should always be chosen based upon a multitude of criteria ranging

from time series length to different topological and/or geometric characteristics of the associated trajectory, we have provided both empirical arguments and numerical indications that selecting the recurrence threshold at a prescribed percentile of the distance distribution (i.e., conserving the global recurrence rate) results in quantitative recurrence characteristics that are more stable under changes of the embedding dimension than when using alternative approaches. In the latter context, we have demonstrated that measures from RQA and related frameworks may exhibit a crucial dependence on the embedding dimension when selecting the recurrence threshold according to a certain percentage of the mean or maximum state space diameter, as sometimes suggested in other works^{7,10}. This also indicates that some alternative approaches, such as normalizing the time series and applying a uniform threshold independent of the embedding dimension and the considered norm¹⁵, are not likely to perform well for any kind of data, when neglecting the effect on the distance distribution with increasing embedding dimension.

At the conceptual level, our general discussion of the changing shape of distance distributions with embedding dimension has led to some interesting follow-up questions associated with the convergence properties of these distributions at high embedding dimensions, which should be further addressed in future studies. Notably, the relationship between the distribution of L_∞ distances and extreme value statistics clearly deserves further investigations to fully understand the emerging shape of the distributions as the embedding dimension becomes large. As a cautionary note, we emphasize that the considerations presented in this work relate exclusively to the concept of time delay embedding as the most widely applied embedding technique, but not necessarily to methodological alternatives like derivative embedding³⁰, for which the metric properties of different components of the embedding vector cannot be easily related to each other.

Taken together, the results presented in this work are important for automatizing the problem of data-adaptive recurrence threshold selection, which is key for further widening the scope of applications of recurrence plots, recurrence quantification analysis and related techniques across scientific disciplines. Especially in the context of long time series originating from non-stationary systems, which frequently appear in many fields of science, a generally applicable approach is crucial for obtaining reliable and easily interpretable results.

ACKNOWLEDGMENTS

This work has been financially supported by the German Research Foundation (DFG projects no. MA4759/8 and MA4759/9), the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement no. 691037 (project QUEST), and the German Federal Ministry of Education

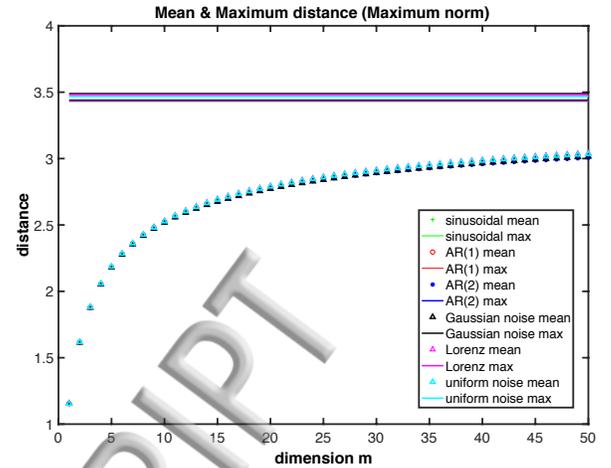


FIG. 3: Mean $d_{mean}^{(\infty)}$ and maximum $d_{max}^{(\infty)}$ L_∞ distance between all pairs of state vectors as a function of the embedding dimension m for different types of time series: polychromatic harmonic oscillation with periods 3, 50 and 500; auto-regressive processes of first and second order with parameters $\varphi_1 = 0.5, \varphi_2 = 0.3$; random numbers of standard Gaussian (zero mean and unit variance) and uniform (unit variance) distributions, and y component of the Lorenz-63 system (Eq. (11), see Section III) with control parameters $\sigma = 10, \beta = 8/3$ and r linearly increasing from 180 (chaotic regime) to 210 (periodic regime) as a function of the embedding dimension m .

and Research via the Young Investigators Group CoSyCC² (Complex Systems Approaches to Understanding Causes and Consequences of Past, Present and Future Climate Change, grant no. 01LN1306A).

Appendix A: Influence of embedding dimension on the variations in the maximum and mean pairwise distances

As discussed in Section II, we show some numerical results illustrating the general behavior of mean and maximum L_∞ and L_2 distances for different types of systems in Figs. 3 and 4, respectively. For a theoretical explanation of the observed changes with increasing embedding dimension, see Section II.

Appendix B: Empirical shape parameters of the distance distributions for different systems

In order to further characterize the shape of the empirically observed pairwise distance distributions shown in Fig. 1 in more detail, we consider two standard characteristics from descriptive statistics. On the one hand,

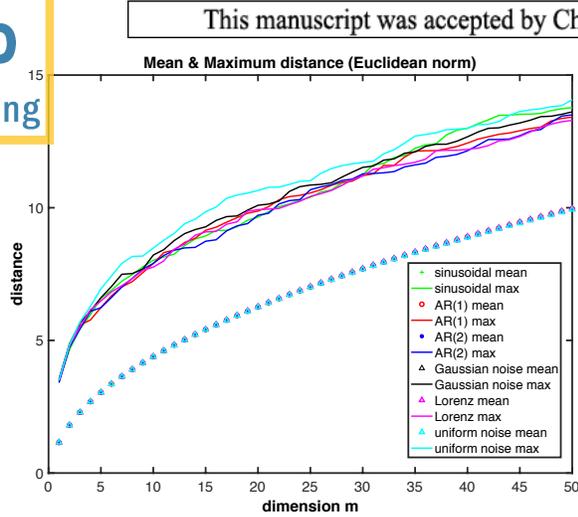


FIG. 4. Same as in Fig. 3 for L_2 distances.

the skewness

$$\hat{s} = \frac{\frac{1}{N_d} \sum_{i=1}^{N_d} (d_i - \bar{d})^3}{\left(\sqrt{\frac{1}{N_d} \sum_{i=1}^{N_d} (d_i - \bar{d})^2} \right)^3} \quad (\text{B1})$$

of the distribution measures its asymmetry around the sample mean distance \bar{d} . On the other hand, we study the associated Shannon entropy

$$\hat{h} = - \sum_{j=1}^{N_b} p_j \frac{\log(p_j)}{\log(N_b)} \quad (\text{B2})$$

providing an integral measure of the heterogeneity of the distribution of d . Here, j enumerates the bins of a histogram of the values of d with N_b bins and relative frequencies p_j , and N_d is the number of pairwise distances in the sample (i.e., the number of independent entries of the distance matrix \mathbf{d} , $N_d = N_{\text{eff}}(N_{\text{eff}} - 1)/2$). The bin width has been selected by first computing the optimum value according to the Freedman-Diaconis rule³¹ for each embedding dimension m and then averaging over all corresponding values and taking the resulting mean to keep N_b fixed for each considered setting. Specifically, for the time series drawn from the Gaussian and uniform distributions, $N_{b,L_2} = 355$ and $N_{b,L_\infty} = 286$, while for the Lorenz system, $N_{b,L_2} = 701$ and $N_{b,L_\infty} = 771$.

According to the corresponding normalization, \hat{h} assumes its maximum of one in case of a uniform distribution (since then, $p_j = 1/N_b \quad \forall j = 1, \dots, N_b$, i.e., for each (binned) distance within $[d_{\min}, d_{\max}]$). In turn, the more heterogeneous (e.g., spiky or generally asymmetric) the distribution of distances gets, the lower \hat{h} .

Figure 5 shows the resulting behavior of both characteristics for the L_2 (panels A,C,E) and L_∞ (panels B,D,F) distances obtained from uniform and Gaussian distributed noise as well as for the non-stationary Lorenz-63 system (Eq. (11), see Section III) in dependence on the

embedding dimension. The results complement the qualitative description based on a visual inspection of Fig. 1 as given in Section II. In case of the L_2 norm and time series drawn from uniform and Gaussian distributions (Fig. 5 A,C) we observe the skewness converging towards zero (symmetric Gaussian distribution) and the entropy reflecting this convergence towards a normal distribution by a downward trend until the skewness approaches zero as m further increases. Although the theoretically predicted Gaussian shape for high m is visually apparent in case of the time series from the Lorenz-63 system (see Fig. 1 E), the skewness takes clearly non-zero negative values while the entropy constantly decreases with increasing m , indicating an asymmetric shape (Fig. 5 E). In case of the L_∞ norm, the considered maximum embedding dimension appears not suited for observing convergence of both shape parameters.

Appendix C: RPs and RQA for one realization of the non-stationary Lorenz system

For further illustrating the RPs resulting from the time-dependent Lorenz-63 system discussed in Section III, we show here the results for just one example trajectory corresponding to a set of randomly chosen initial conditions $x(0) = 0.9649$, $y(0) = 0.1576$, $z(0) = 0.9706$. As before, we embed the y component time series and study the RP for each previously discussed threshold selection method. Then, we use a running window over each (global) RP with a window size of $w = 400$ and mutual shift of $ws = 40$ data points, i.e., 90% overlap between consecutive windows.

The RPs and the associated time-dependent recurrence characteristic RTE (Eq. (13)) for a “normal” three-dimensional embedding with time delay $\tau = 4$, consistent with the first local minimum of the mutual information²², are shown in Fig. 6, using the Euclidean norm. We compare the results for four different threshold selection methods but similar effective threshold values (corresponding to a global recurrence rate of $RR \approx 4\%$), which are thus expected to give comparable results. The left panel corresponds to the recommended method of taking a certain percentile of the distance distribution, while the other three panels are based on thresholds selected according to some percentage of the maximum, mean and median distance of state vectors on the attractor in the reconstructed state space. Comparing the different panels, as expected there are hardly any marked differences in the RPs or the temporal changes of RTE . The transition from a chaotic regime into a periodic one is well reflected by a constantly decreasing RTE , which takes its minimum for the limit cycle behavior between $t_1 \approx 1,000$ and $t_2 \approx 1,080$.

However, if choosing a higher-dimensional embedding (e.g., $m = 10$) motivated by the non-stationarity of the system, the RP becomes almost completely white if the recurrence threshold is chosen based upon the same per-

This manuscript was accepted by Chaos. Click [here](#) to see the version of record.

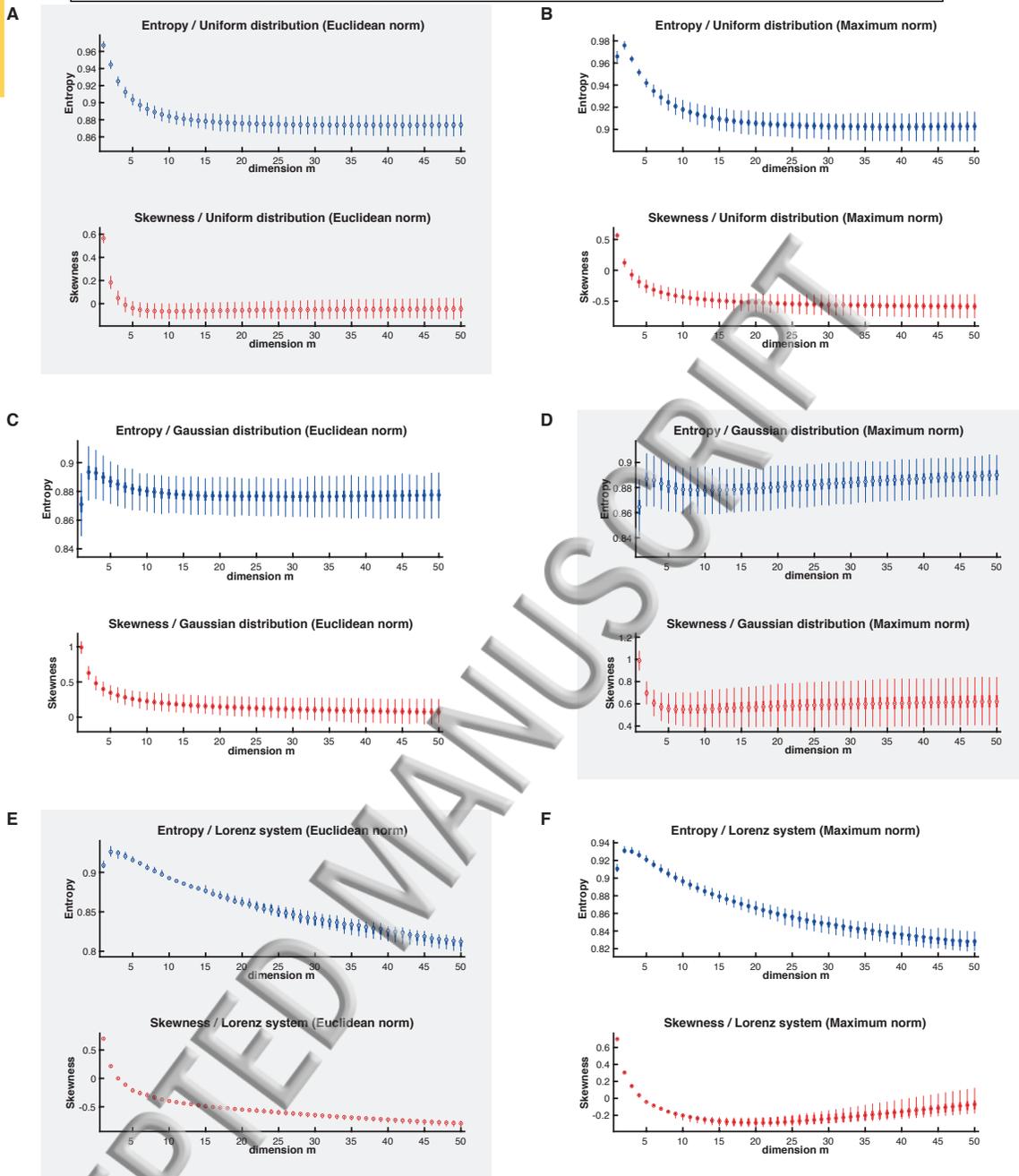


FIG. 5. Skewness (red) and Shannon entropy (blue) of the L_2 (A,C,E) and L_∞ (B,D,F) distances of $N = 1,500$ independent random numbers with uniform (A,B) and Gaussian (C,D) distribution and (E,F) the y component of the Lorenz-63 system (Eq. (11), $N = 6,000$, see Section III) with control parameters $\sigma = 10$, $\beta = 8/3$ and r linearly increasing from 180 (chaotic regime) to 210 (periodic regime) as a function of the embedding dimension m . For the two noise series, box plots show the variability estimated from 1,000 independent realizations for each data set, using a random number generator. In case of the Lorenz-63 system the variability is estimated from 10 independent realizations of the non-stationary Lorenz-63 equations with randomly chosen initial conditions.

centages of the maximum, mean or median state space distances as used before (Fig. 7). In this case the RTE is still able to detect the transitory limit cycle regime, but one loses information about the chaotic regime before. In contrast, we retain the same density of recurrences

and, hence, resolution of the RP as for $m = 3$ when fixing the threshold according to the whole distance distribution (left panel in Fig. 7). Here, the overall behavior of RTE from the lower-dimensional ($m = 3$) case is qualitatively retained, although the periodic regime is less well

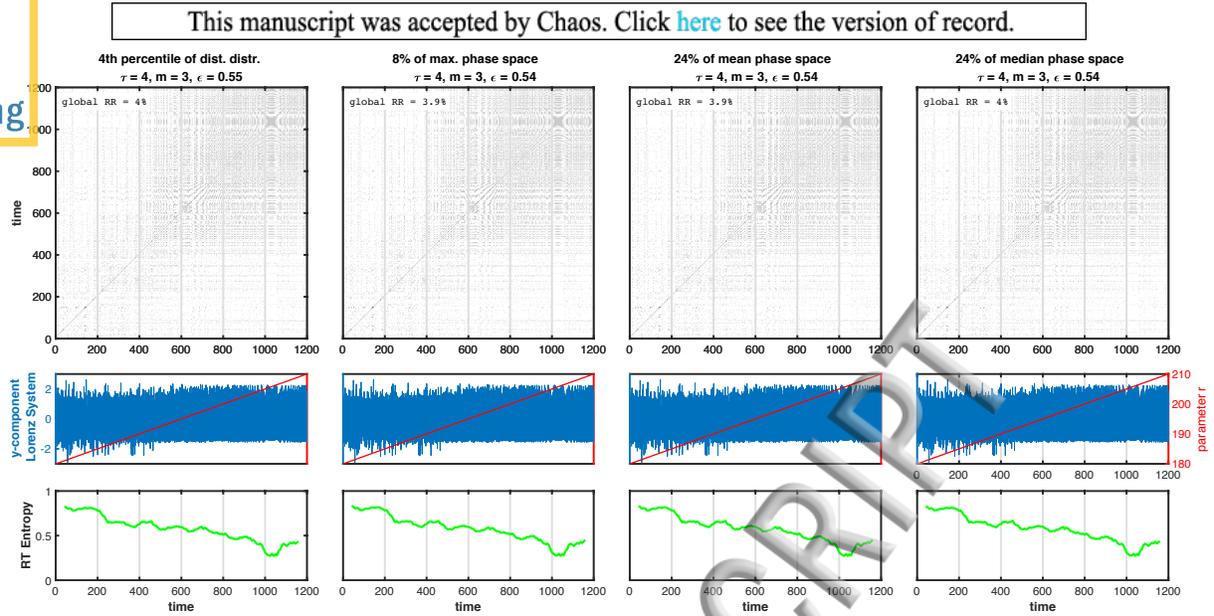


FIG. 6. RPs, according time series (blue), time-dependence of the control parameter r (red) and recurrence characteristic RTE (green) based on the y component of the non-stationary Lorenz-63 system (see text for details), using the L_2 norm. Shown are the results for low-dimensional embedding ($m = 3$) and for four different methods to select the recurrence threshold according to a certain percentile of the distance distribution and some percentage of the maximum, mean or median distances of state vectors on the reconstructed attractor (from left to right). The actual threshold values (4th percentile, 8%, 24% and 24%, respectively) have again been chosen such that the global recurrence rate for each method in this embedding scenario is $\approx 4\%$.

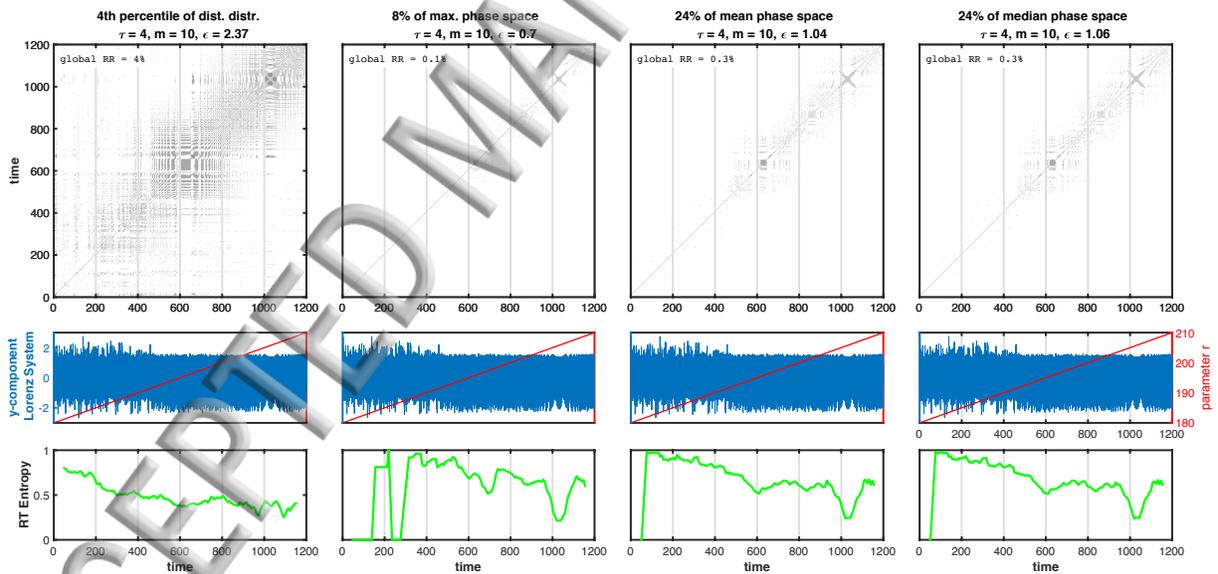
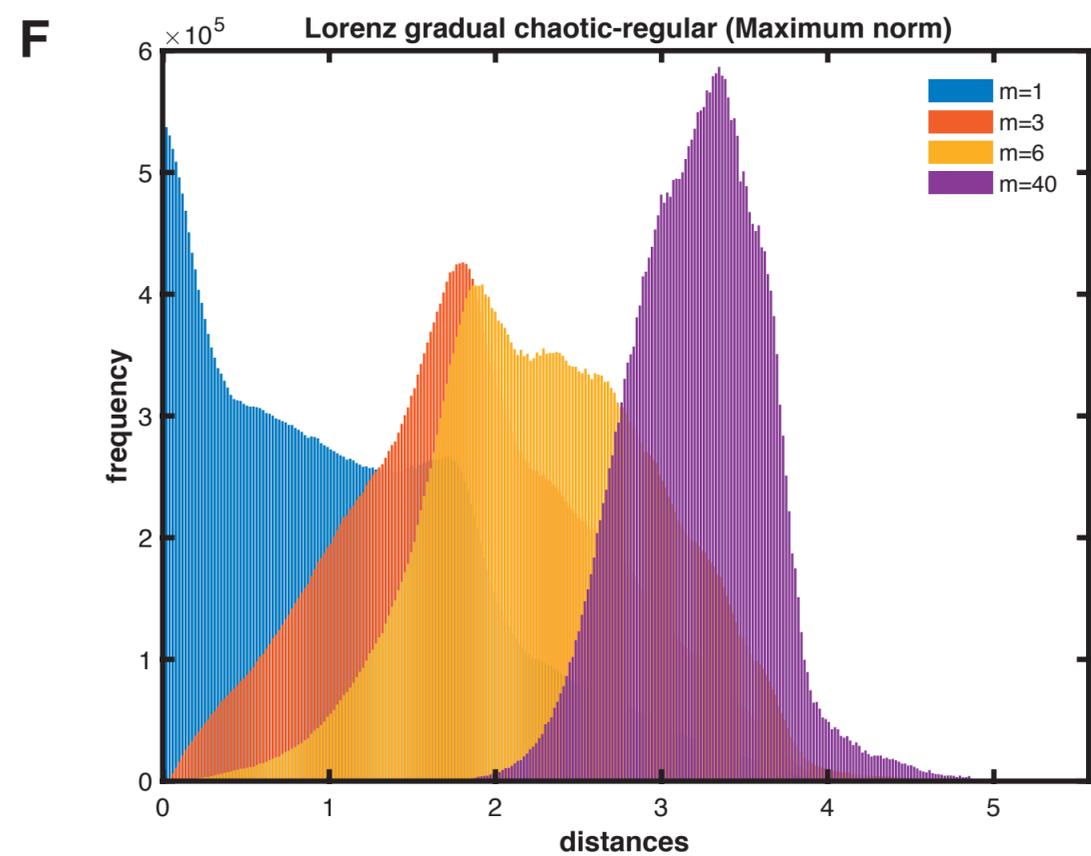
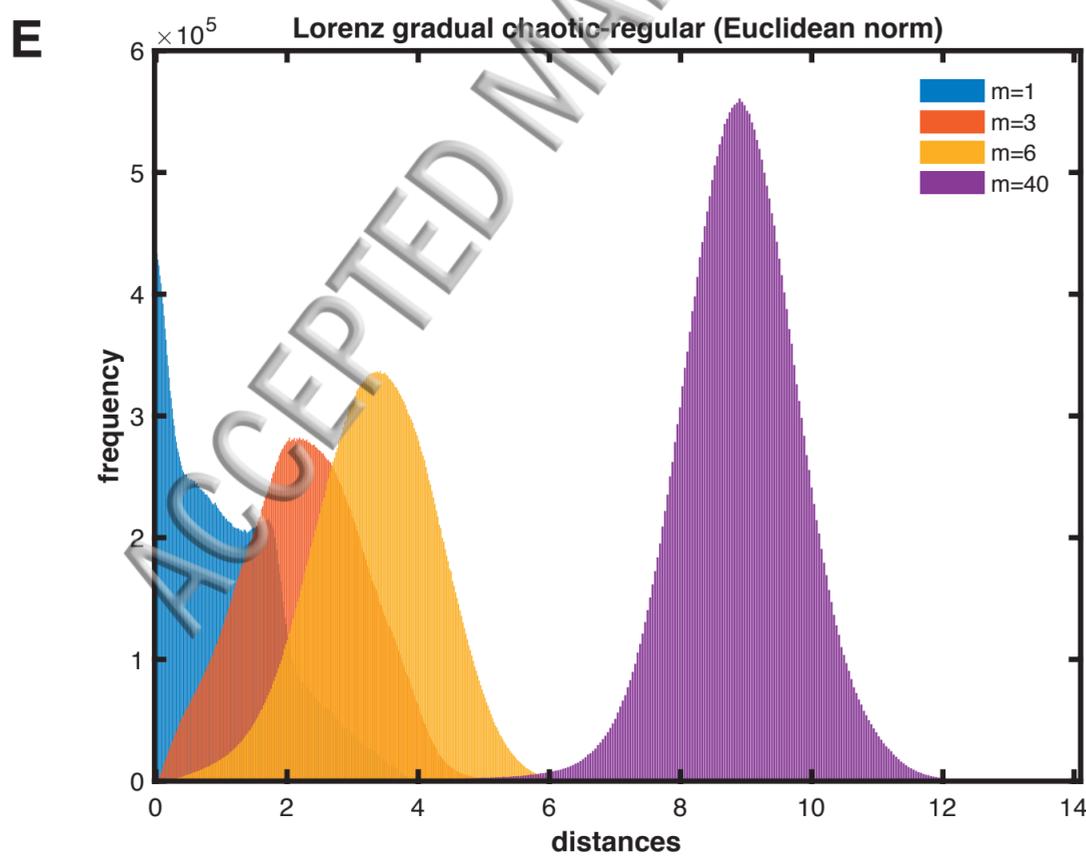
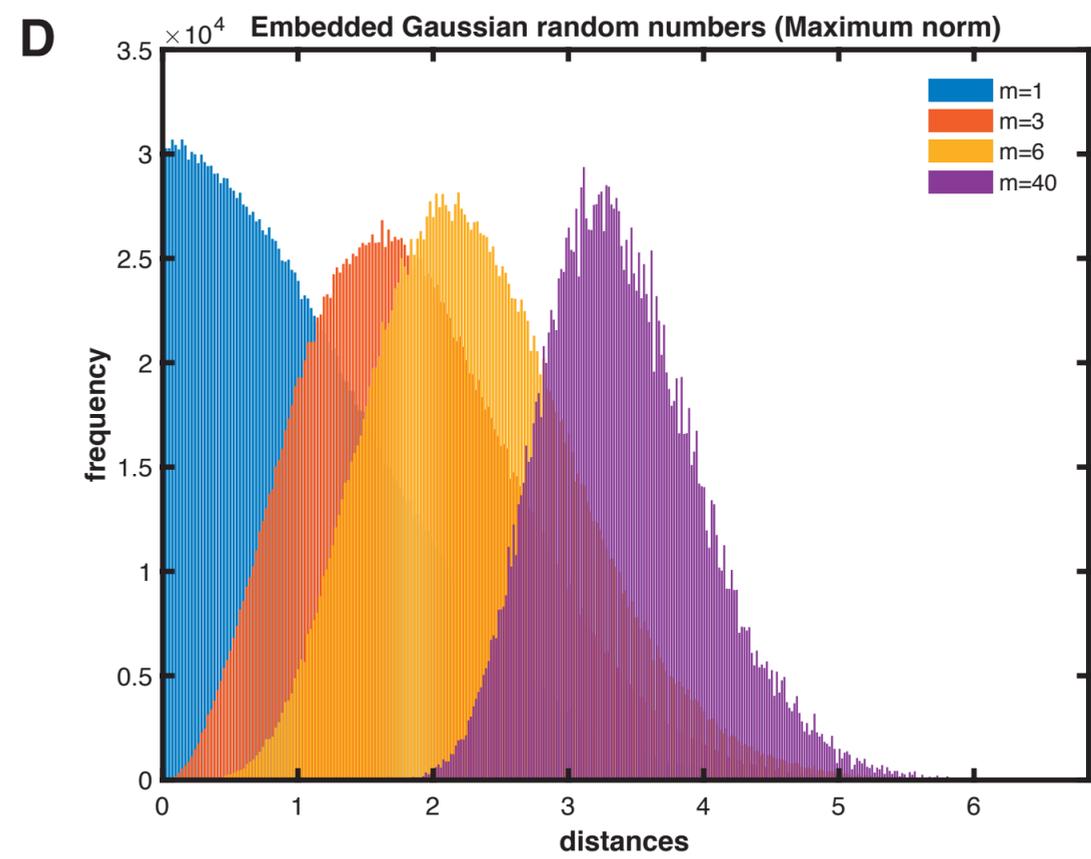
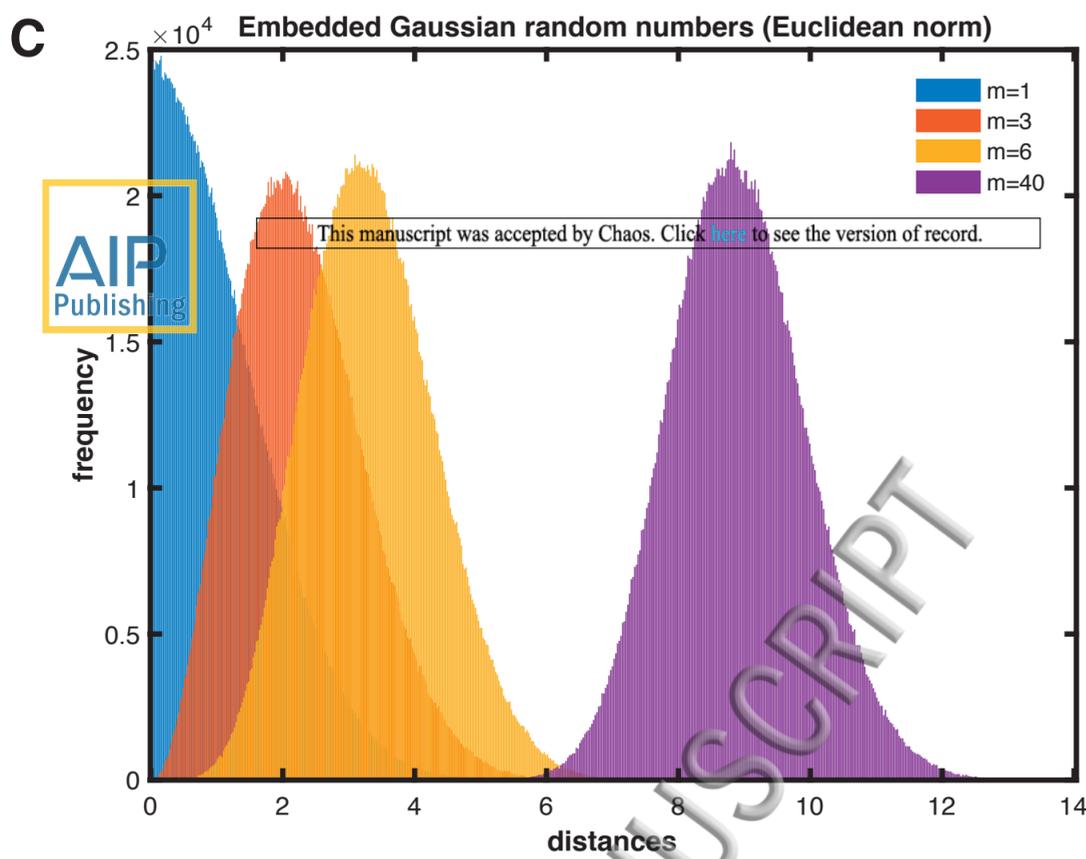
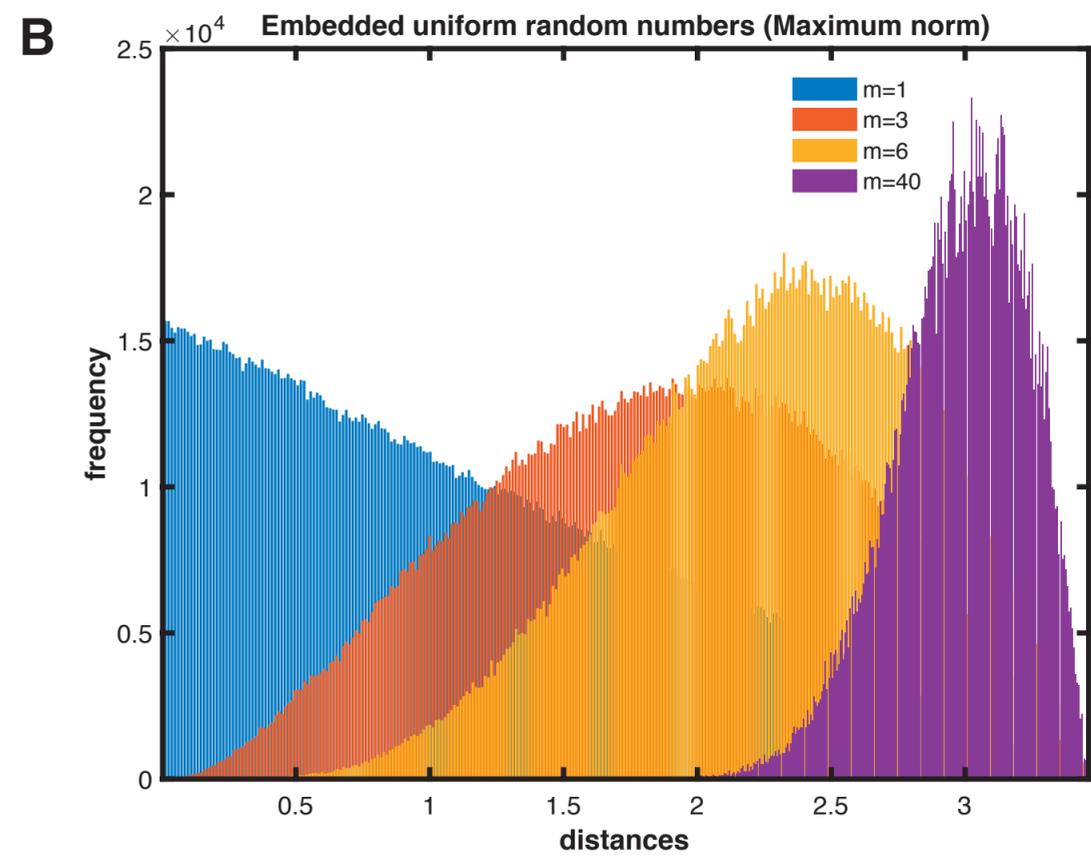
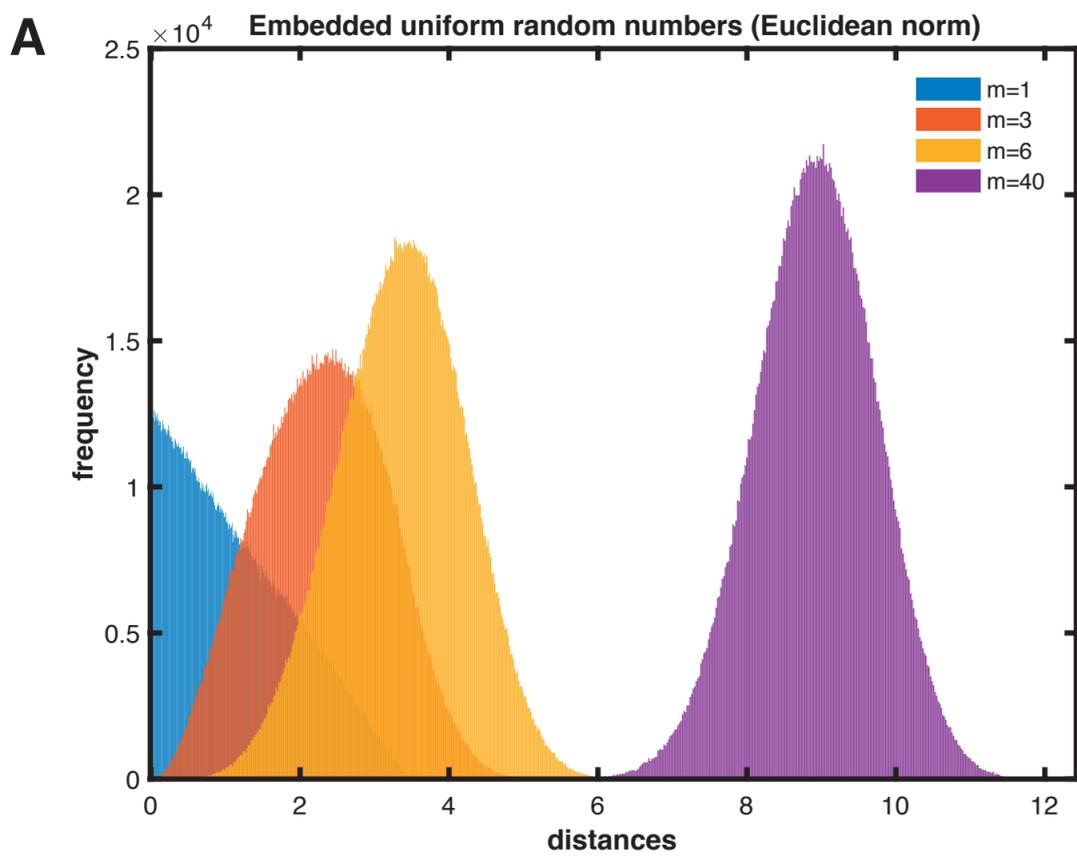
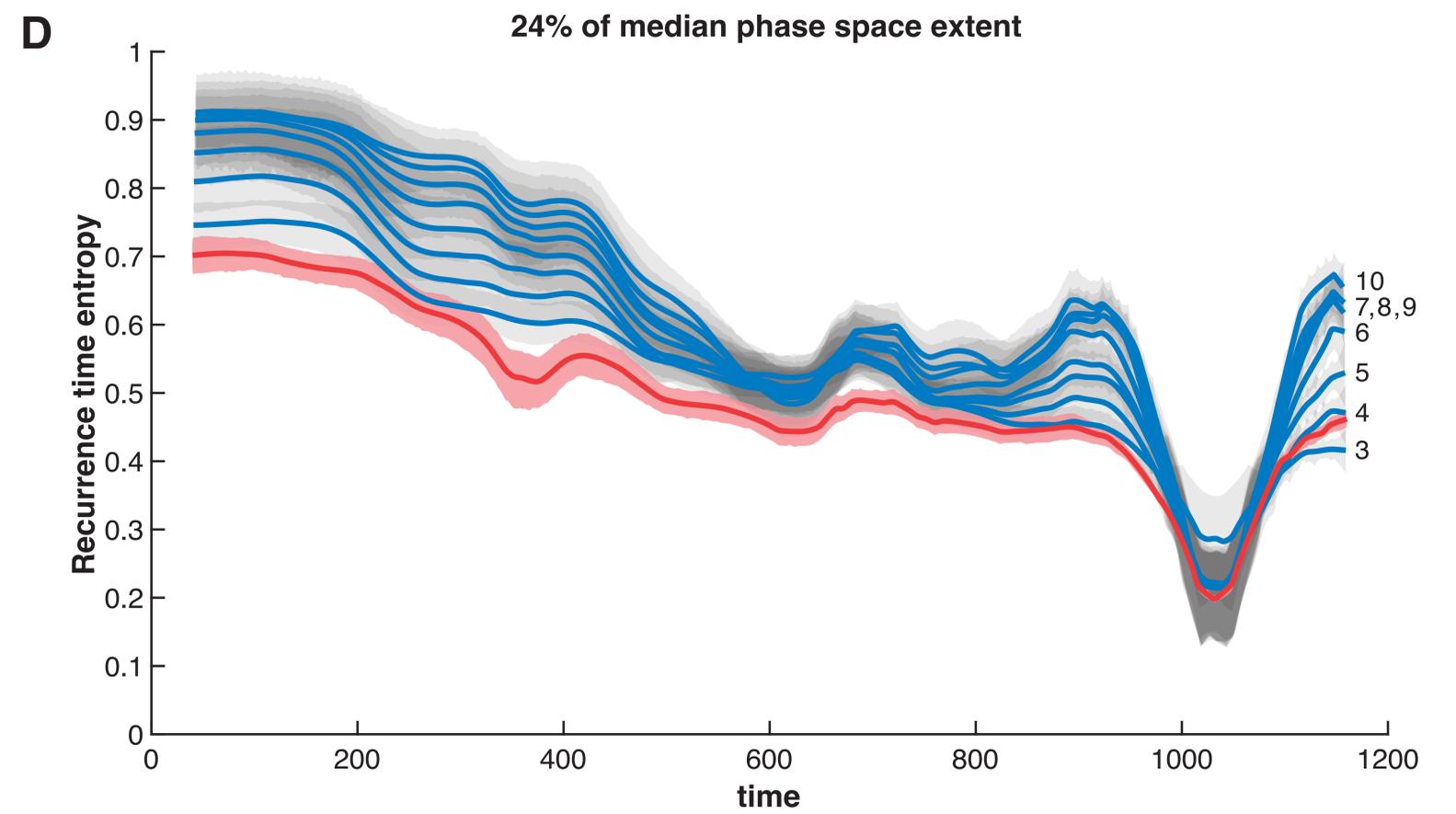
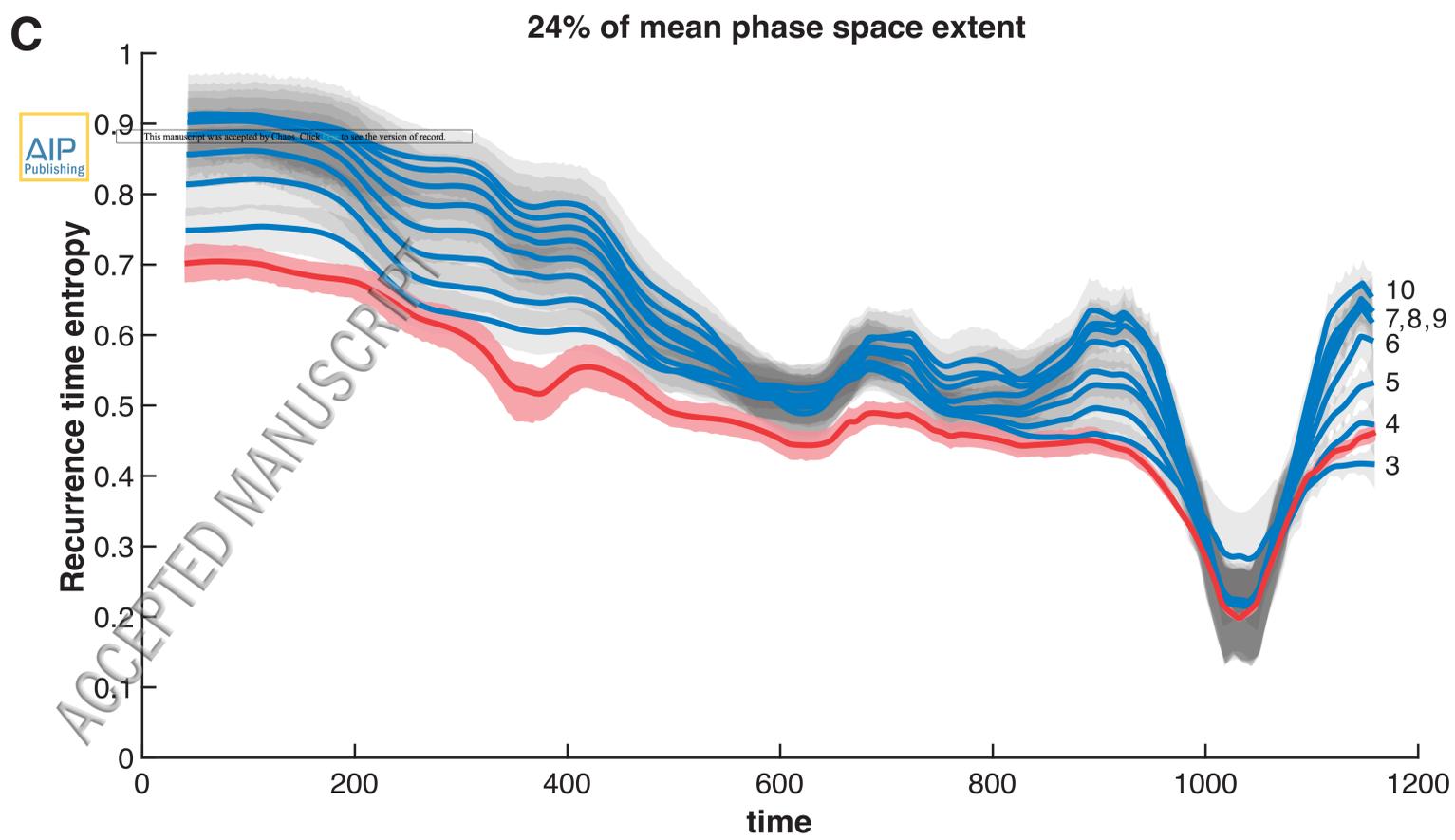
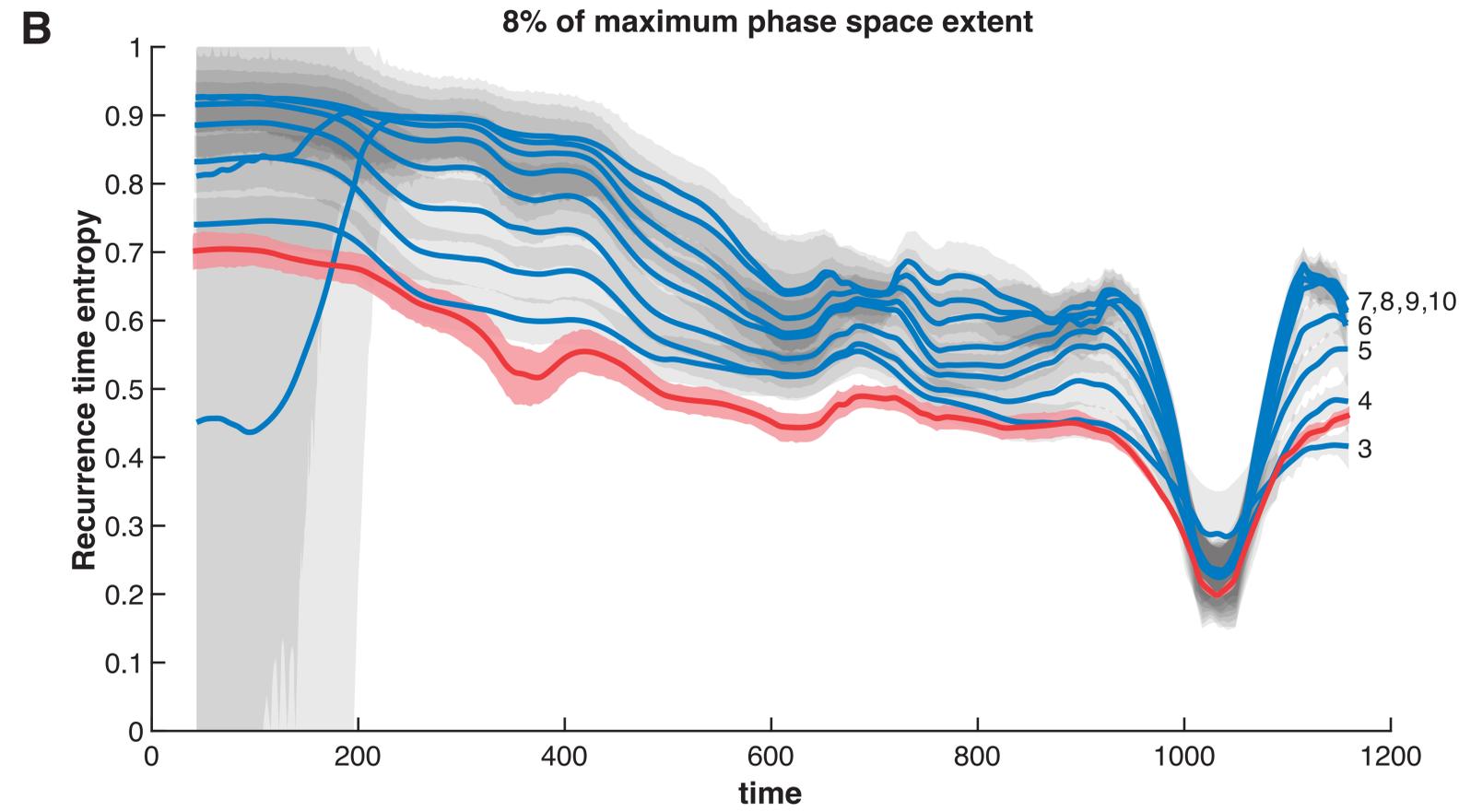
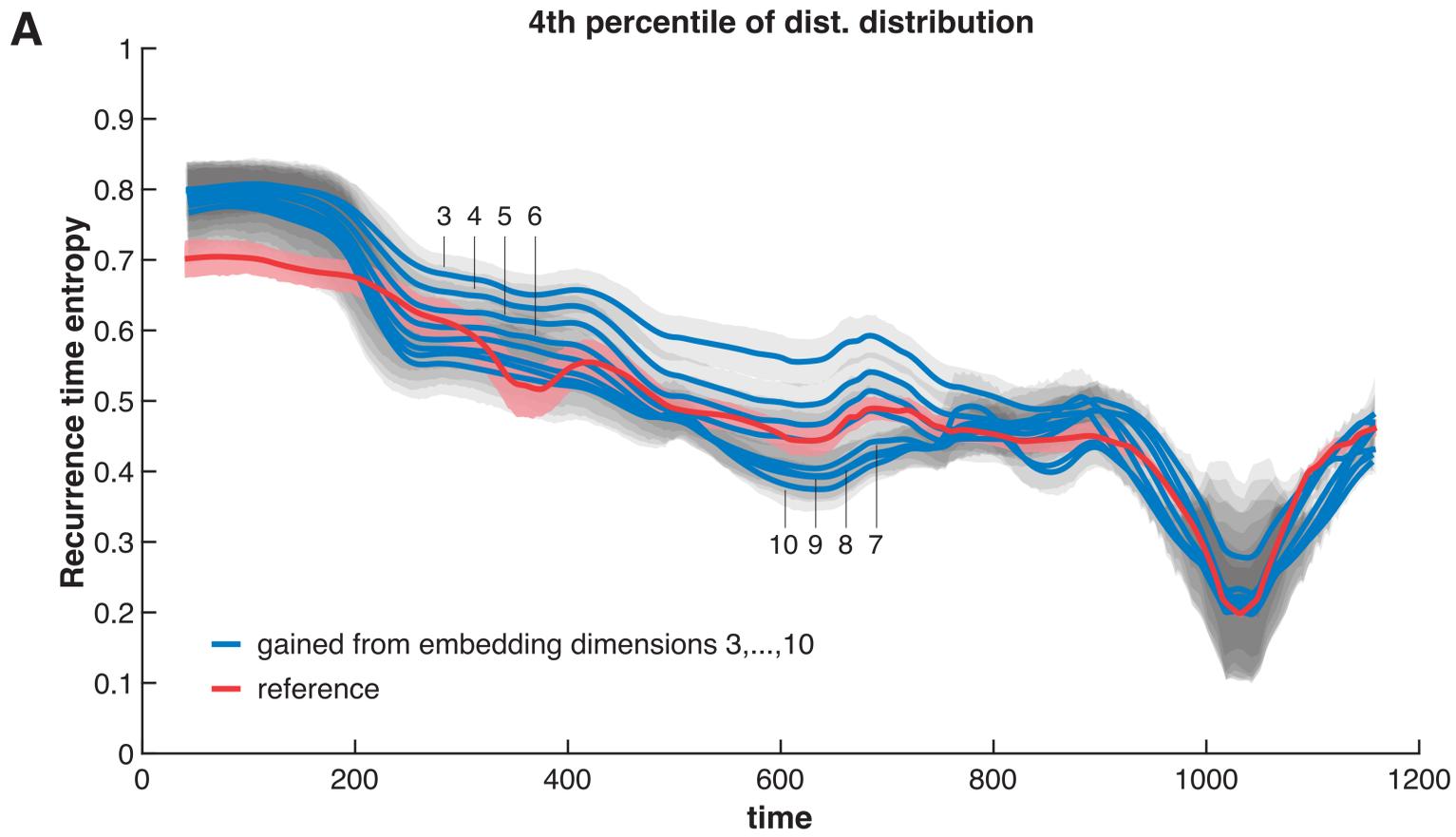


FIG. 7. Same as in Fig. 6, but for ten-dimensional embedding. In comparison to Fig. 6, three of the four methods lead to a marked drop in the global recurrence rate and a resulting change in the RTE values. Only for a recurrence threshold corresponding to the same percentile of the distance distribution, the results are qualitatively stable over the full considered time evolution.

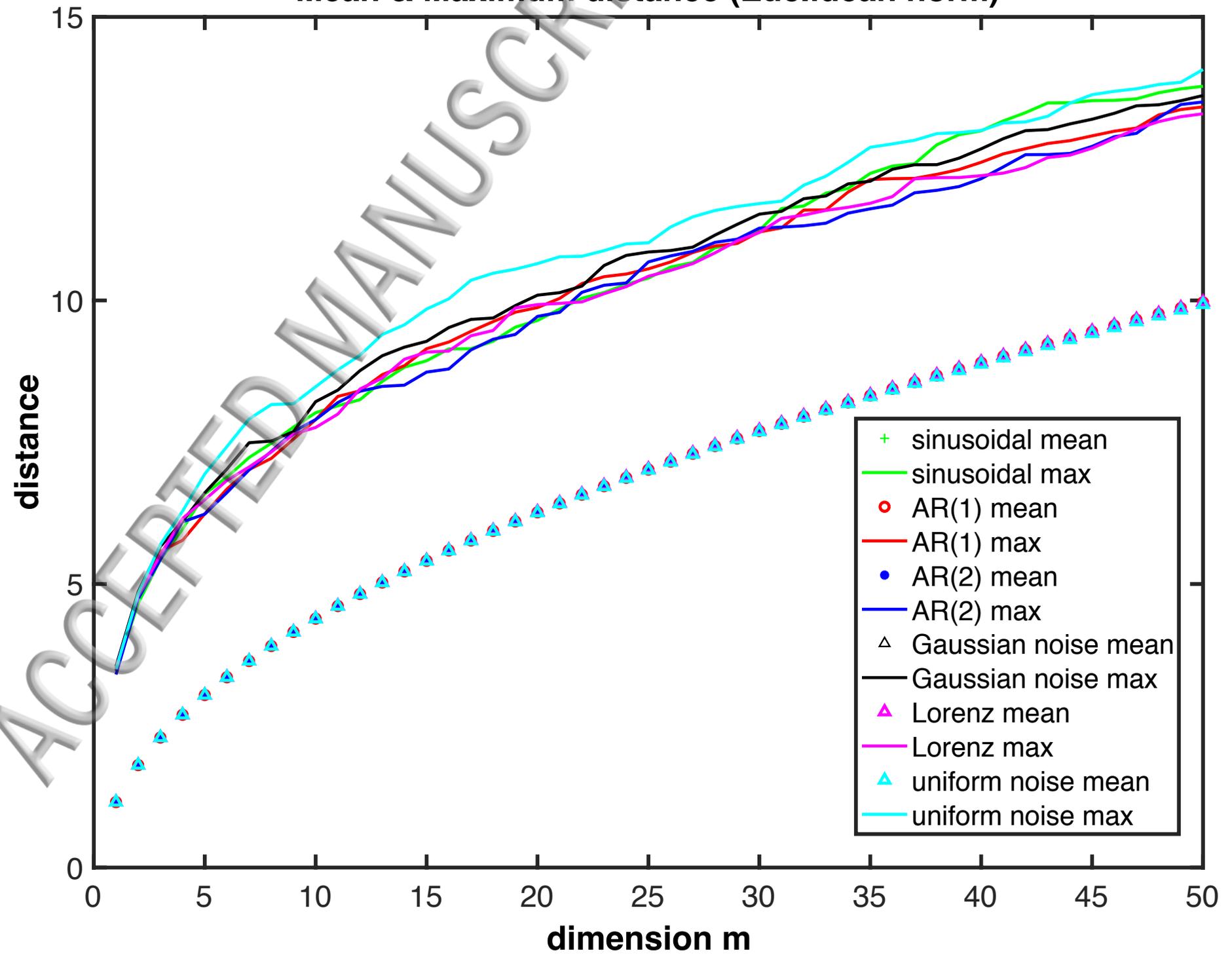
expressed than in the former case.

- ¹F. Takens, "Detecting strange attractors in turbulence," Lecture Notes in Mathematics **898**, 366–381 (1981).
- ²C. Geier and L. A. Aguirre, "Investigating nonlinear dynamics from time series: The influence of symmetries and the choice of observables," Chaos **12**, 549–558 (2002).
- ³N. Packard, J. Crutchfield, J. Farmer, and R. Shaw, "Geometry from a time series," Phys. Rev. Lett. **45**, 712–716 (1980).
- ⁴J.-P. Eckmann, S. Olliffson Kamphorst, and D. Ruelle, "Recurrence Plots of Dynamical Systems," Europhysics Letters **4**, 973–977 (1987).
- ⁵N. Marwan, M. C. Romano, M. Thiel, and J. Kurths, "Recurrence Plots for the Analysis of Complex Systems," Physics Reports **438**, 237–329 (2007).
- ⁶G. M. Mindlin and R. Gilmore, "Topological analysis and synthesis of chaotic time series," Physica D **58**, 229–242 (1992).
- ⁷J. P. Zbilut and C. L. Webber, Jr., "Embeddings and delays as derived from quantification of recurrence plots," Physics Letters A **171**, 199–203 (1992).
- ⁸J. P. Zbilut, J.-M. Zaldívar-Comenges, and F. Strozzi, "Recurrence quantification based Liapunov exponents for monitoring divergence in experimental data," Physics Letters A **297**, 173–181 (2002).
- ⁹M. Thiel, M. C. Romano, J. Kurths, R. Meucci, E. Allaria, and F. T. Arecchi, "Influence of observational noise on the recurrence quantification analysis," Physica D **171**, 138–152 (2002).
- ¹⁰S. Schinkel, O. Dimigen, and N. Marwan, "Selection of recurrence threshold for signal detection," European Physical Journal – Special Topics **164**, 45–53 (2008).
- ¹¹L. Matassini, H. Kantz, J. A. Holyst, and R. Hegger, "Optimizing of recurrence plots for noise reduction," Physical Review E **65**, 021102 (2002).
- ¹²N. Marwan, J. F. Donges, Y. Zou, R. V. Donner, and J. Kurths, "Complex network approach for recurrence analysis of time series," Physics Letters A **373**, 4246–4254 (2009).
- ¹³R. V. Donner, J. Heitzig, J. F. Donges, Y. Zou, N. Marwan, and J. Kurths, "The Geometry of Chaotic Dynamics – A Complex Network Perspective," European Physical Journal B **84**, 653–672 (2011).
- ¹⁴J. F. Donges, J. Heitzig, R. V. Donner, and J. Kurths, "Analytical framework for recurrence network analysis of time series," Physical Review E **85**, 046105 (2012).
- ¹⁵R. Jacob, K. P. Harikrishnan, R. Misra, and G. Ambika, "Uniform framework for the recurrence-network analysis of chaotic time series," Physical Review E **93**, 012202 (2016).
- ¹⁶D. Eroglu, N. Marwan, S. Prasad, and J. Kurths, "Finding recurrence networks' threshold adaptively for a specific time series," Nonlinear Processes in Geophysics **21**, 1085–1092 (2014).
- ¹⁷M. Wiedermann, J. F. Donges, J. Kurths, and R. V. Donner, "Mapping and discrimination of networks in the complexity-entropy plane," Physical Review E **96**, 042304 (2017).
- ¹⁸M. Koebbe and G. Mayer-Kress, "Use of Recurrence Plots in the Analysis of Time-Series Data," in *Proceedings of SFI Studies in the Science of Complexity*, Vol. XXI, edited by M. Casdagli and S. Eubank (Addison-Wesley, Redwood City, 1992) pp. 361–378.
- ¹⁹A. Zimek, E. Schubert, and H. P. Kriegel, "A survey on unsupervised outlier detection in high-dimensional numerical data," Statistical Analysis and Data Mining **5**, 363–387 (2012).
- ²⁰T. Sauer, J. A. Yorke, and M. Casdagli, "Embedology," J. Statist. Phys. **65**, 579–616 (1991).
- ²¹R. Hegger, H. Kantz, L. Matassini, and T. Schreiber, "Coping with Nonstationarity by Overembedding," Physical Review Letters **84**, 4092–4095 (2000).
- ²²A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," Physical Review A **33**, 1134–1140 (1986).
- ²³M. R. Leadbetter, G. Lindgren, and H. Rootzen, *Extremes and Related Properties of Random Sequences and Processes* (Springer, New York, 1983).
- ²⁴H. D. Abarbanel, *Analysis of Observed Chaotic Data* (Springer, New York, 1996).
- ²⁵M. Thiel, M. C. Romano, and J. Kurths, "Spurious structures in recurrence plots induced by embedding," Nonlinear Dynamics **44**, 299–305 (2006).
- ²⁶E. N. Lorenz, "Deterministic nonperiodic flow," Journal of the Atmospheric Sciences **20**, 130–141 (1963), [https://doi.org/10.1175/1520-0469\(1963\)020<0130:DNF;2.0.CO;2](https://doi.org/10.1175/1520-0469(1963)020<0130:DNF;2.0.CO;2).
- ²⁷E. J. Ngamga, S. Bialonski, N. Marwan, J. Kurths, C. Geier, and K. Lehnertz, "Evaluation of selected recurrence measures in discriminating pre-ictal and inter-ictal periods from epileptic EEG data," Physics Letters A **380**, 1419–1425 (2016).
- ²⁸M. A. Little, P. E. McSharry, S. J. Roberts, D. A. E. Costello, and I. M. Moroz, "Exploiting Nonlinear Recurrence and Fractal Scaling Properties for Voice Disorder Detection," BioMedical Engineering OnLine **6**, 1–19 (2007).
- ²⁹M. S. Baptista, E. J. Ngamga, P. R. F. Pinto, M. Brito, and J. Kurths, "Kolmogorov-Sinai entropy from recurrence times," Physics Letters A **374**, 1135–1140 (2010).
- ³⁰J. Lekscha and R. V. Donner, "Phase space reconstruction for non-uniformly sampled noisy time series," ArXiv e-prints (2018), arXiv:1801.09517 [physics.data-an].
- ³¹D. Freedman and P. Diaconis, "On the histogram as a density estimator: l_2 theory," Z. Wahrscheinlichkeitstheorie verw. Gebiete **57**, 453–476 (1981).

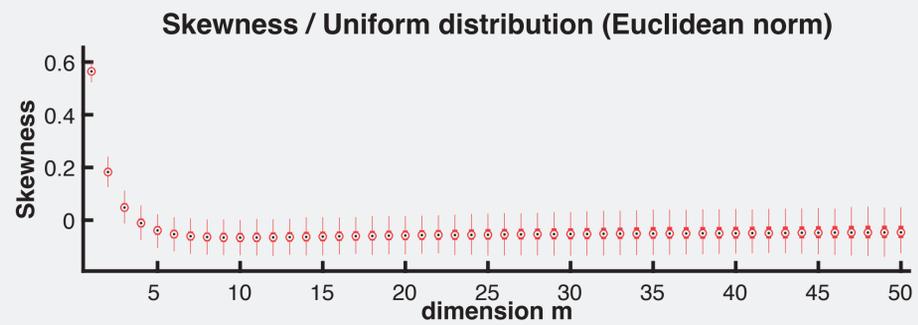
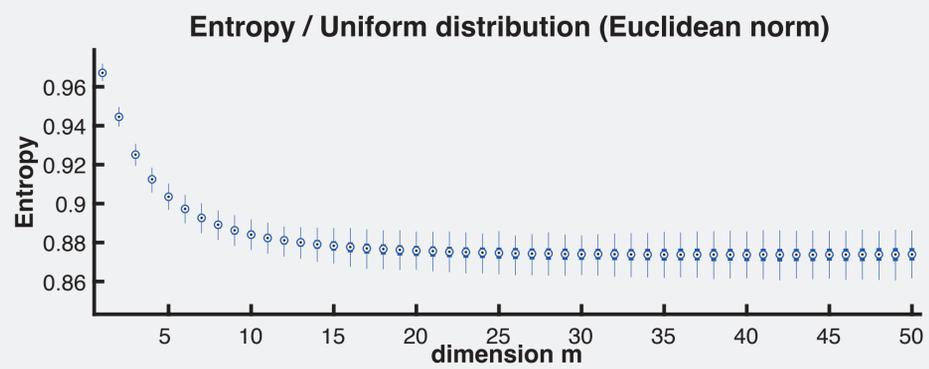




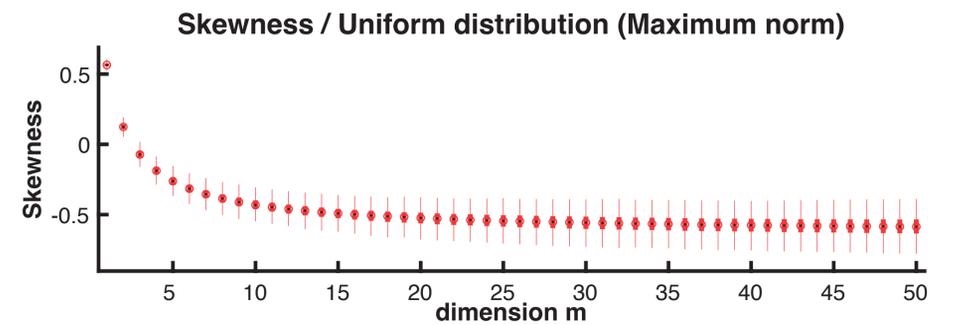
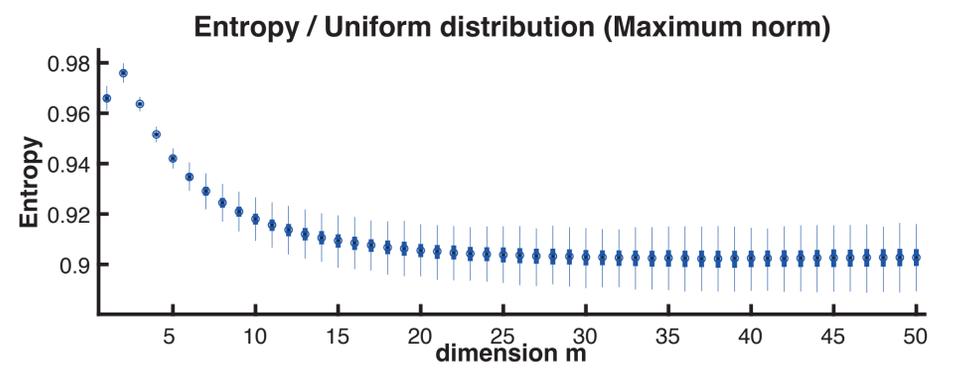
Mean & Maximum distance (Euclidean norm)



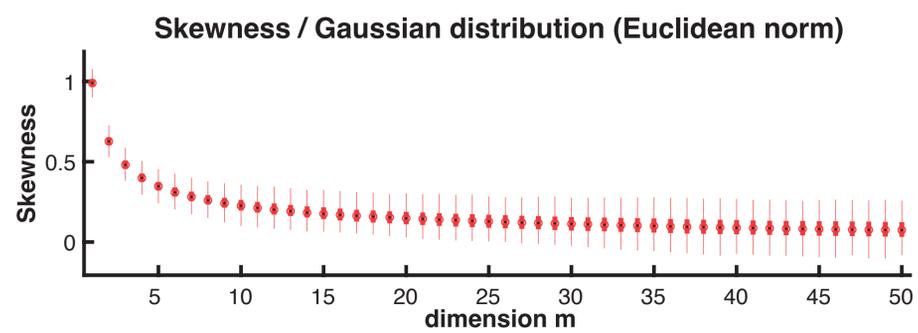
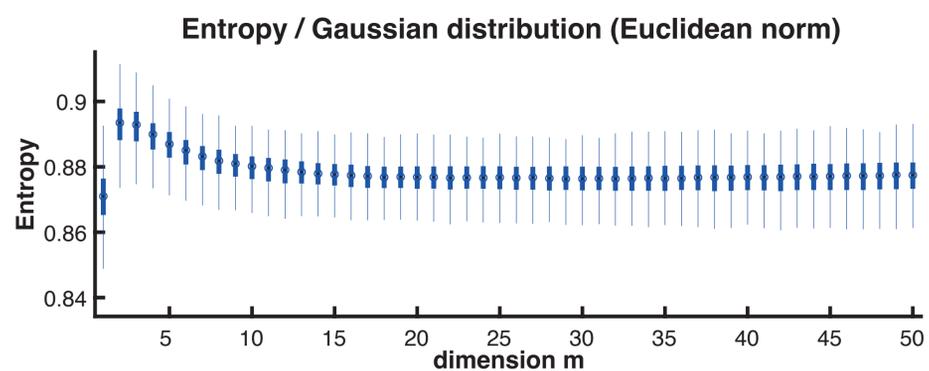
A



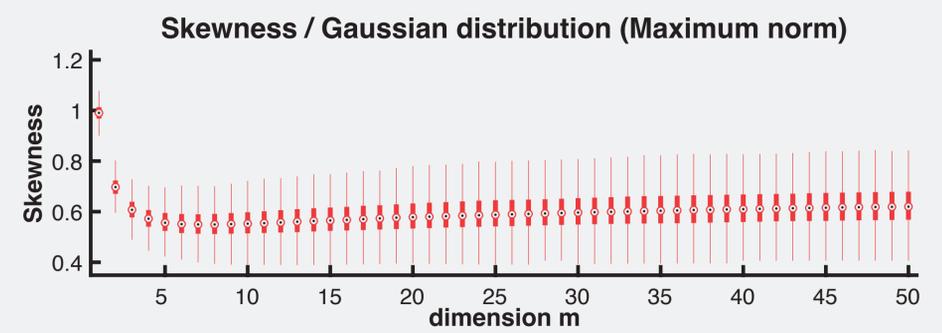
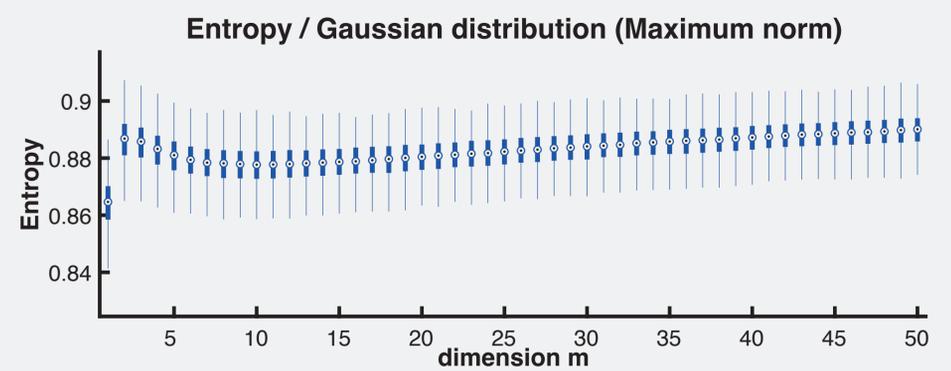
B



C

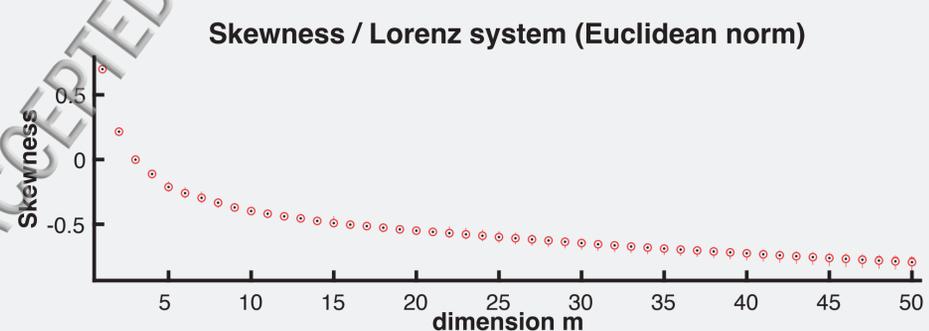
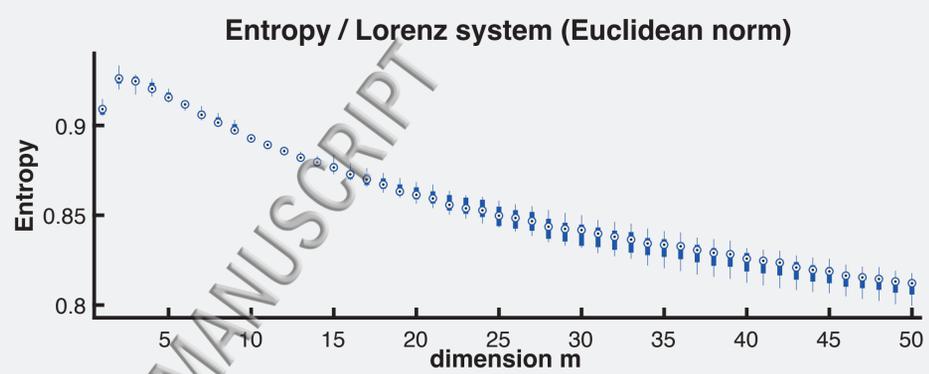


D

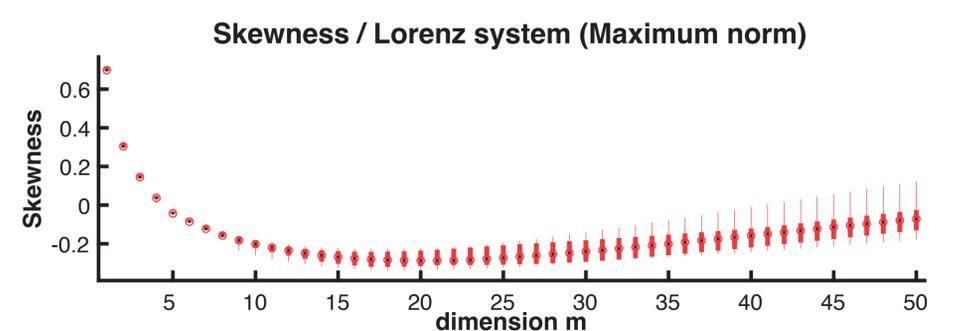
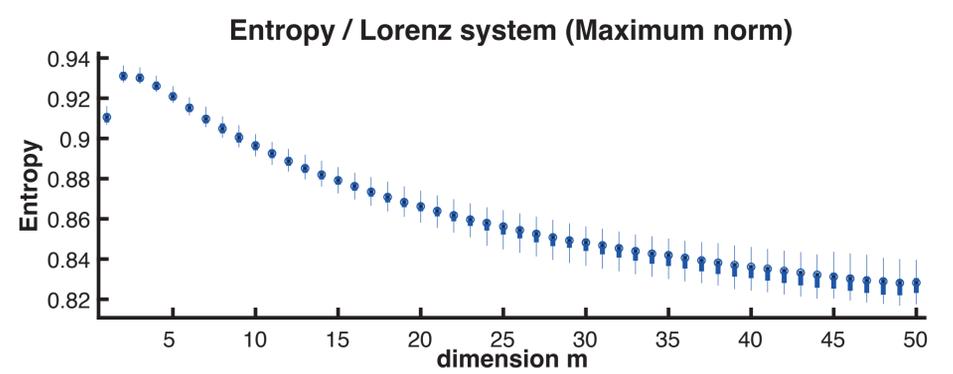


This manuscript was accepted by Chaos. Click [here](#) to see the version of record.

E

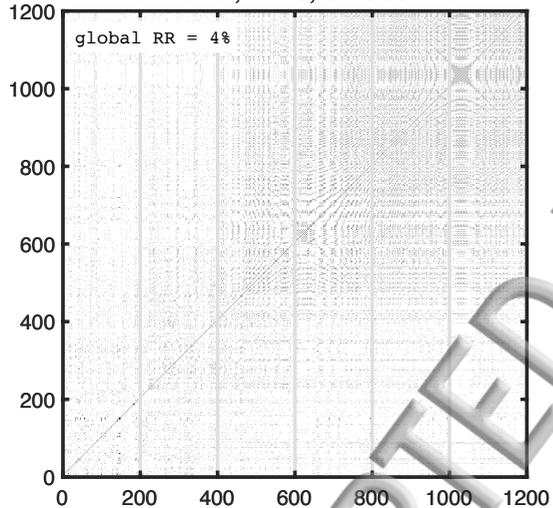


F



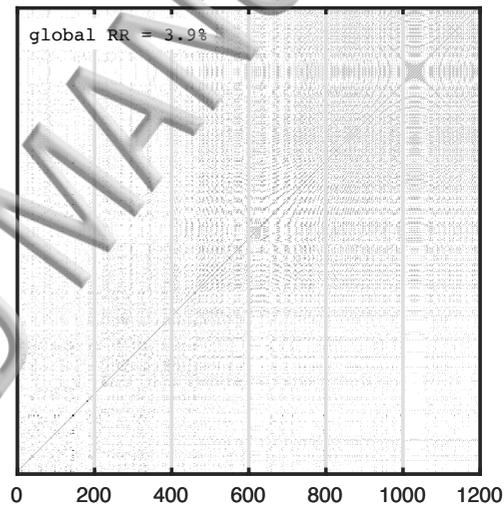
4th percentile of dist. distr.

$\tau = 4, m = 3, \epsilon = 0.55$



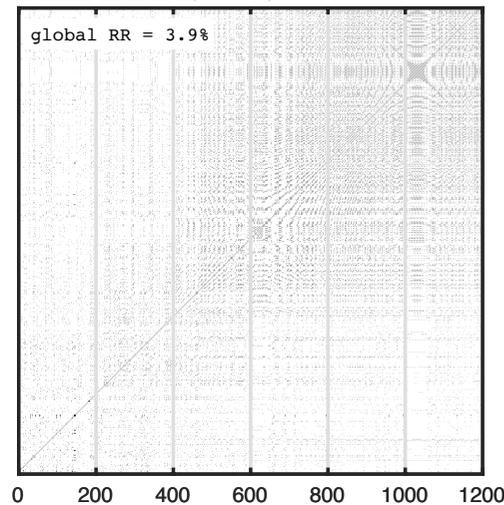
8% of max. phase space

$\tau = 4, m = 3, \epsilon = 0.54$



24% of mean phase space

$\tau = 4, m = 3, \epsilon = 0.54$



24% of median phase space

$\tau = 4, m = 3, \epsilon = 0.54$

