Regular Article

Recurrence flow measure of nonlinear dependence

Tobias Braun[1,a] , K. Hauke Kraemer[1], and Norbert Marwan[1,2]

[1] Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, 14473 Potsdam, Germany
[2] Institute of Geoscience, University of Potsdam, Karl-Liebknecht-Straße 32, 14476 Potsdam, Germany

Abstract Couplings in complex real-world systems are often nonlinear and scale dependent. In many cases, it is crucial to consider a multitude of interlinked variables and the strengths of their correlations to adequately fathom the dynamics of a high-dimensional nonlinear system. We propose a recurrence-based dependence measure that quantifies the relationship between multiple time series based on the predictability of their joint evolution. The statistical analysis of recurrence plots (RPs) is a powerful framework in nonlinear time series analysis that has proven to be effective in addressing many fundamental problems, e.g., regime shift detection and identification of couplings. The *recurrence flow* through an RP exploits artifacts in the formation of diagonal lines, a structure in RPs that reflects periods of predictable dynamics. Using time-delayed variables of a deterministic uni-/multivariate system, lagged dependencies with potentially many time scales can be captured by the recurrence flow measure. Given an RP, no parameters are required for its computation. We showcase the scope of the method for quantifying lagged nonlinear correlations and put a focus on the delay selection problem in time-delay embedding which is often used for attractor reconstruction. The recurrence flow measure of dependence helps to identify non-uniform delays and appears as a promising foundation for a recurrence-based state space reconstruction algorithm.

1 Introduction

Measures of statistical dependence represent one of the cornerstones in the analysis of empirical data. The study of time series measured from complex real-world systems poses a broad variety of challenges in quantifying uni- and multivariate data sets, e.g., lagged dependencies, non-stationarity, noise contamination, uncertainties and the limited length of time series. The set of tools to detect and quantify statistical dependencies ranges from standard correlation analysis techniques [1] over graph theoretical approaches, such as complex networks [2, 3], towards causal discovery algorithms [4–6], and detecting critical transitions [7–9]. A notoriously challenging problem is to adequately quantify non-monotonous and nonlinear relationships in stochastic and deterministic systems. In this context, information theoretic measures have established as an effective framework [10–12]. However, popular methods, such as the mutual information (MI), are not designed to treat higher-dimensional data appropriately (even though extensions have been suggested [13, 14]). The nature and strength of links in complex systems additionally

often exhibits scale-dependence, i.e., multi-scale behavior [15]. This motivates the need of methods that are capable of unravelling dependencies at a broad range of scales, e.g., wavelet-based methods [16–18]. Few methods succeed to combine both capabilities of capturing nonlinear dependencies at multiple time scales [19, 20].

A powerful method that captures both nonlinear and multi-scale properties of a high-dimensional dynamical system is the recurrence plot (RP) [21]. An RP is a mathematically simple yet effective tool that encodes the tendency of a time series to recur to formerly visited states [22]. An RP is based on a binary recurrence matrix in which recurrences are marked by value one, giving rise to intriguing and well-interpretable structures in the RP. Various quantification measures can be applied to a recurrence matrix and prove powerful in classifying differing systems [23–26], identifying dynamical regime transitions [27, 28], and detecting non-linear correlations as well as synchronization [6, 29, 30]. Recurrence quantification analysis (RQA) based on diagonal lines in the RP not only allows identification of periodic behavior [18, 31], but also helps to identify unstable periodic orbits in high-dimensional chaotic systems [32]. The conceptual simplicity of RPs allows for a broad range of real-world applications, also for

[a] e-mail: tobraun@pik-potsdam.de (corresponding author)

challenging data that are event-like or unevenly sampled in time [33, 34]. Recurrence measures of dependence have not only facilitated the study of synchronization in dynamical systems [30, 35–37] but also have been extended to account for lagged and conditional dependencies [6, 29, 38, 39]. Further concepts, including symbolic analyses of relationships, have been conceived more recently [40]. Recurrence-based quantification of statistical dependencies, thus, bares high potential to meet the combination of the above-mentioned challenges. Here, we propose a novel recurrence-based measure of dependence that uses delay coordinates from a given observational time series. Since the measure is based on RPs, nonlinear dependencies with multiple time lags can be quantified which makes the measure applicable to the problem of non-uniform delay selection [19, 20]. The proposed dependence measure, thus, contributes to the challenge of characterizing complex real-world interactions using RPs.

This work is structured as follows: in Sect. 2, we introduce the recurrence flow as a measure of dependence along with a brief review of the RP method. We showcase its scope in different numerical experiments in Sect. 3, covering the characterization of lagged nonlinear dependence and delay selection for uniform and non-uniform TDE. We conclude our findings in Sect. 4.

2 Recurrence flow

We are interested in characterizing nonlinear dependencies in a deterministic, high-dimensional system that is represented by M observational time series $\{s_n(t) \,|\, n = 1, \ldots, M\}$. In general, the relationships between the different time series $s_n(t)$ and their coordinates do not need to be instantaneous but are often associated with time delays $\tau_1, \tau_2, \ldots, \tau_m$. Consequently, we define the recurrence flow measure of redundance to capture such lagged dependencies.

The key idea of the proposed measure is based on the existence of diagonal lines in RPs. An RP is a two-dimensional matrix that encodes how a system recurs to formerly visited states $\vec{v}_i$, $i = 1, \ldots, N$. In general, this representation can be computed for systems of any dimension d and is formally given by

$$R_{i,j}(\tau) = \Theta(\varepsilon - \|\vec{v}_i(\tau) - \vec{v}_j(\tau)\|)$$
$$= \begin{cases} 1 & \text{if } \|\vec{v}_i(\tau) - \vec{v}_j(\tau)\| \leq \varepsilon \\ 0 & \text{if } \|\vec{v}_i(\tau) - \vec{v}_j(\tau)\| > \varepsilon, \end{cases} \quad (1)$$

with two arbitrary times i and j, the vicinity threshold ε and a suitable norm $\|\cdot\|$. The states denoted by $\vec{v}$ are either given by the available components (state variables) of the system or, in case of only limited access to the state variables, by delayed copies of the one (or multiple) observational time series $s_n(t)$ of the studied system. In particular, $\vec{v}$ is then obtained by stacking these copies on top of each other as it is common practise in time delay embedding (TDE).

Diagonal lines of length L_d in an RP resemble periods of enhanced predictability as two trajectory segments at times i and j evolve in parallel in an ε-tube for L_d time instances. For a given system, this may reveal time periods of continuously high determinism or uncover abrupt regime shifts [41]. Properties of the diagonal line length distribution of an RP are linked to dynamical invariants of paradigmatic dynamical systems [42]. However, spurious artifacts are known to disrupt, lengthen or thicken diagonal lines due to erroneous computation of an RP [43]. An inadequate choice of the vicinity threshold ε will disrupt diagonal lines, thus underestimating the system's predictability. On the other hand, too high values will artificially merge fundamentally distinct regions of phase space. Sampling can also alter diagonal line structures: if the system is undersampled, diagonal lines might not emerge continuously as deterministic time intervals are not resolved sufficiently. On the other hand, over-sampling results in artificially thickened diagonal lines (*tangential motion*) [44]. Erroneous time delay embedding of uni-/multivariate time series can have several undesired effects on the formation of diagonal lines; if the embedding dimension is chosen too high, diagonal lines are artificially lengthened and can even emerge in absence of determinism for an uncorrelated stochastic process due to correlations in the underlying distance matrix [45]. A non-optimal choice for the embedding delay will result in diagonal lines that are perpendicular to the line of identity ($i = j$, LOI). This indicates inclusion of erroneous time scales [43]. The formation of perpendicular lines is caused by the ambiguity in the reconstruction that is introduced by not eliminating the full serial dependence; this results in close evolution of states both forward and backward in time: $\|\vec{v}_i - \vec{v}_j\| < \varepsilon$, $\|\vec{v}_{i+1} - \vec{v}_{j-1}\| < \varepsilon$, i.e., the trajectory segments closely evolve in parallel, but with opposite time directions. On top of that, additional deformations to a diagonal line can occur, e.g., in the form of bowed diagonal lines indicating that the evolution of states at different time intervals is similar but occurs with different velocity or temporal resolution [46].

We utilize the formation of such diagonal line artifacts (DLA) to identify time scales of the system that result in well-expressed diagonal lines. The proposed method is, thus, based on the assumption that the studied system exhibits (at least to some degree) deterministic dynamics which will result in meaningful diagonal lines in an RP. An effective way of retrieving information on the formation of DLA is given by scanning an RP diagonal-wise. We can identify an index for each diagonal at which the first recurrence pixel is located. It appears intuitive to regard these pixels as 'obstacles' to an imaginary fluid that flows along each diagonal into the RP and is not allowed to turn (Fig. 1C/E). The formation of DLA blocks the flow. As a basic example, we consider a noisy sinusoidal time series (Fig. 1A) with $n = 5000$ and a period of $T = 100$. Formation of perpendicular diagonal lines for $\tau = \mu T$, $\mu \in \mathbb{N}$ (Fig. 1B) reduces the flow through the RP compared to $\tau = T/4$ (Fig. 1D). We use the symmetry of the RP by only

flooding the upper triangular matrix to save computation time. The recurrence flow $\Phi(\tau)$ can be computed for varying delays τ and encodes similar information as an inverse autocorrelation function, yielding a continuous representation of the redundancy between the time series and its delayed version (Fig. 1F). For continuous variations of τ from $\tau = 0$ to $\tau = T/4$, the perpendicular diagonal lines are progressively eliminated. This reproduces the well-known result that a sinusoidal signal needs to be shifted by (odd multiples of) a quarter of its period against itself to minimize redundancy.

To quantify the flow through the RP, we define the recurrence flow matrix ϕ (Fig. 1C/E)

$$\phi_{i,j}(\tau) = \Big(1 - \Theta(\varepsilon - \|\vec{v}_i(\tau) - \vec{v}_j(\tau)\|)\Big)\,\Theta(\ell_j - i),$$
$$i, j = 1, \ldots, N, \tag{2}$$

with the time delay τ and the length ℓ_j of the jth flooded diagonal, i.e., the number of subsequent zeros up to the first one. A flooded diagonal ℓ_j has to be distinguished from a diagonal line L_d: it denotes a diagonal of the RP (parallel to the LOI) starting at time instance j that is flooded with a fictive fluid, regardless of whether any diagonal lines exist on this diagonal. Accordingly, the factor $\Theta(\ell_j - i)$ in Eq. (2) ensures that the flooding of the jth diagonal stops at the first recurrence on this diagonal. $\phi_{i,j}$ depends on the vicinity threshold ε, i.e., the fraction of recurrences. We fix ε at some reasonable value that corresponds to a fixed recurrence rate (RR).

We study the dependence on the time delay τ contained in the vector $\vec{v}(\tau)$ similar as it is done in TDE where the delays $\tau_1, \tau_2, \ldots, \tau_m$ for the different coordinates are free parameters and need to be chosen with respect to some notion of optimality [47]. The recurrence flow $\Phi(\tau)$ is computed by summing over the recurrence flow matrix $\phi_{i,j}$ at given τ and dividing by the number of non-recurrences (i.e., zeros in the RP):

$$\Phi(\tau) = \frac{\sum_{i,j=1}^{N} \phi_{i,j}(\tau)}{\sum_{i,j=1}^{N}(1 - R_{i,j}(\tau))}. \tag{3}$$

In a multivariate application, $\vec{v}$ can encompass time series from different systems to study their cross-dependencies. In such a scenario, it is more instructive to define the recurrence flow as a direct measure of correlation/redundance. We, thus, define the recurrence flow measure of redundance (RFMR) $\theta(\tau)$ as

$$\theta(\tau) = 1 - \Phi(\tau). \tag{4}$$

The significance of recurrence flow values can be tested against a random null model based on uncorrelated white noise (Appendix B). Finally, it needs to be noted that the idea of using RPs to identify optimal embedding parameters has been considered before, but to our best knowledge has not been performed systematically [48, 49].

3 Application to model examples

We now demonstrate the scope of the proposed method by highlighting two different potential applications: the quantification of nonlinear correlations (Sect. 3.1) and the identification of uniform embedding delays for TDE of nonlinear signals (Sect. 3.2).

3.1 Nonlinear dependence

We exemplify the efficacy of $\theta(\tau)$ as a nonlinear dependence measure for deterministic systems with a simple bivariate system:

$$x(t) = \sin(2\pi t/\omega) + \eta(t, \sigma_1),$$
$$y(t) = a x(t - \tilde{\tau})^2 + \eta(t, \sigma_2), \tag{5}$$

with frequency $\omega = 2\pi/T$, period T, time lag $\tilde{\tau} = 20$, and normal-distributed white noise processes $\eta(t, \sigma)$ with standard deviations σ_1 and σ_2. This system exhibits a sinusoidal cycle with frequency ω in its x-component. The y-component is nonlinearly coupled to $x(t)$ and exhibits a cycle with half of the period of $x(t)$. $y(t)$ follows $x(t)$ with a fixed time lag $\tilde{\tau}$. Both components are superimposed by measurement noise $\eta(t)$. We consider time series with $n = 5,000$ samples (Fig. 2A). Due to the specific coupling, the relationship between x and y is nonlinear (Fig. 2B).

We test whether we can detect the coupling and the corresponding time lag $\tilde{\tau}$ by computing $\theta(\tau)$ for delays in the range $\tau \in [-200, 200]$ (Fig. 2C). In fact, we find that $\theta(\tau)$ reaches local maxima at integer multiples of $\tilde{\tau}$, including $\tau = \tilde{\tau}$ (red dashed line). Similar results are obtained if the mutual information is used (Fig. 2C, black curve), confirming that nonlinear relationships between deterministic time series can be captured by recurrence flow in presence of measurement noise. In this case, the MI yields sharper peaks than the recurrence flow. However, the suggested recurrence flow measure provides advantages for high-dimensional data as discussed below.

3.2 Uniform time delay embedding

The ability of the recurrence flow to detect delayed dependencies between multiple variables motivates its use in the delay selection problem faced in TDE. Proficient delay selection must be based on a measure that captures the redundancy in a (potentially large) set of correlated time series. A popular solution is to use mutual information (MI). However, characterizing the required joint probability density function $p(s_1, s_2, \ldots, s_m)$ becomes cumbersome for a large number of variables m and is rendered infeasible for many high-dimensional real-world systems. Compared to nonlinear correlation measures like standard MI that are based on binning, recurrence flow offers the advantage that computation times increase less rapidly: given k_m bins, an increase in dimensionality $m \to m + 1$ results

Fig. 1 Recurrence flow for a noisy sinusoidal. **A** Time series of the noisy sinusoidal and **B/C** RPs for two different embedding vectors $\vec{v}(t)$: $\vec{v}(t) = [y(t), y(t+T)]$ and $\vec{v}(t) = [y(t), y(t+\frac{T}{4})]$ (left) and flooded RPs/recurrence flow matrices (right). **D** Recurrence flow $\Phi(\tau)$ through the RP for varying delay τ. The red dot marks the period T. $\Phi(\tau)$ is maximized for delays that are odd multiples of $T/4$
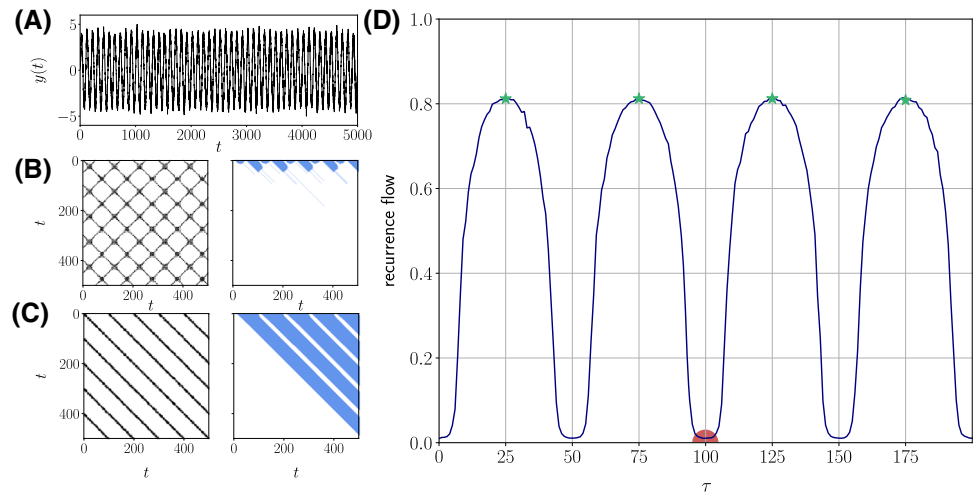

Fig. 2 Detection of lagged, nonlinear dependence between nonlinearly coupled sinusoidals. **A** The $y(t)$ time series (black) results from squaring $x(t)$ (blue) and shifting it by a fixed time delay. **B** Nonlinear relationship between $x(t)$ and $y(t)$. **C** Recurrence flow measure of redundance $\theta(\tau)$ (blue) detecting the time delay between $x(t)$ and $y(t)$ (red dashed line), confirmed by the cross-mutual information (black)
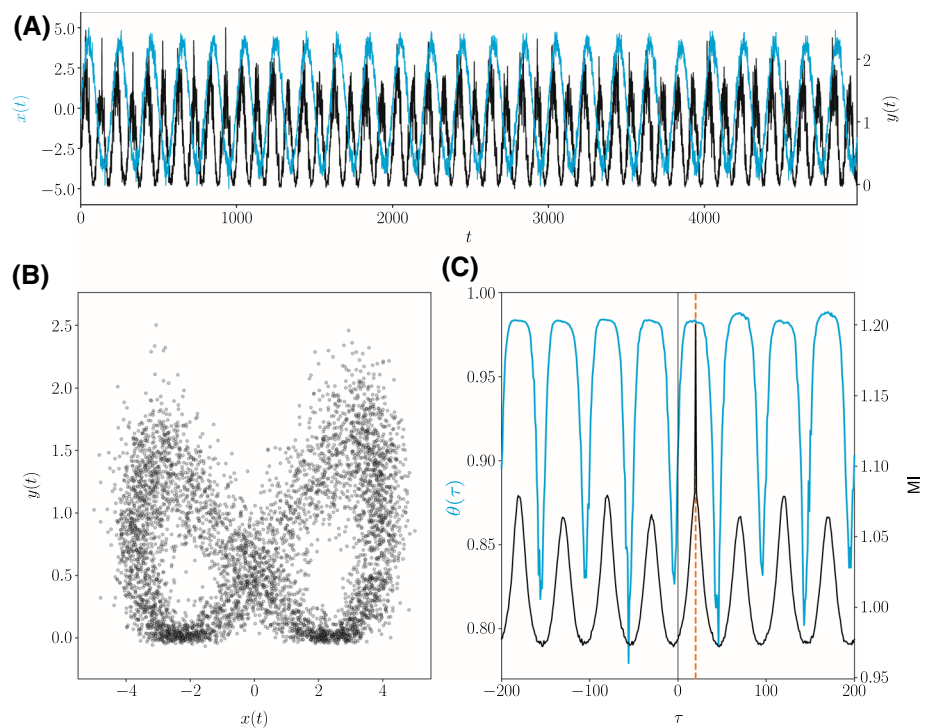
in $(k_m - 1)k_m$ additional bins while in the computation of Φ, a k-d-tree nearest neighbor search-based RP computation increases only linearly with the dimensionality of the system. It has yet to be noted that more sophisticated nearest neighbor-based approaches for MI computation do not suffer from this drawback [14].

Due to the popularity of this problem, other nonlinear correlation measures that do not suffer from the curse of dimensionality have been conceived [20, 50]. In order to validate the effectiveness of the measure proposed here for selecting embedding delays, we do not only compare it to the linear autocorrelation function (ACF) and the auto-mutual information (MI), but also to the delay selection method proposed in [20]: the *continuity statistic* is based on a hypothesis test of whether a component, added to an existing m-dimensional reconstruction vector, is functionally independent of all existing m components. This is tested by first defining a set of fiducial points and their k-nearest neighbors from the m-dimensional reconstruction vector. These are mapped onto the one-dimensional number line. The continuity statistic $\langle \varepsilon^* \rangle(\tau)$ denotes the smallest scale at which the number of observed neighbors mapped onto this line is larger than the number expected from a binomial distribution, implying a functional relationship. For further details, the reader is referred to [20, 51]. For the estimation of the optimal embedding dimension, we employ Cao's method with a threshold of $\Delta_{\mathrm{afn}} = 0.2$ for the change of the number of averaged false neighbors from $m \to m+1$ [52]. For the selection of embedding delays using the proposed
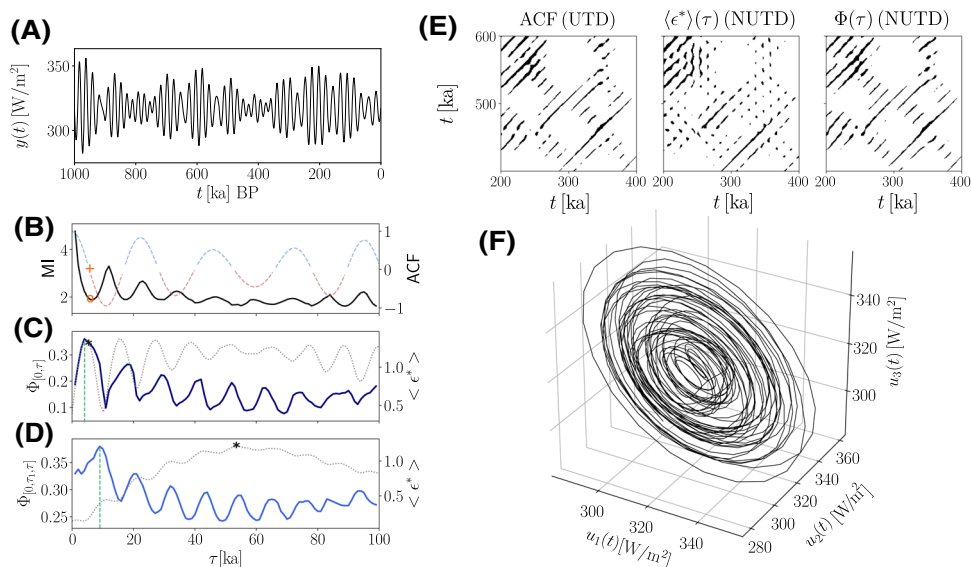
Fig. 3 Uniform delay selection for numerical insolation model. **A** Insolation time series $y(t)$ in W m^{-2}. **B** Serial dependence of $y(t)$, measured in terms of ACF (red/blue) and auto-MI (black) for the univariate time series, **C/D** as well as by the continuity statistic $\langle \varepsilon^* \rangle(\tau)$ (gray) and the recurrence flow $\Phi(\tau)$ (dark blue) for **C** two- and **D** three-dimensional embedding vectors $\vec{v}(t)$. The optimal delay is marked by a circle(/cross/star/vertical green line) for the ACF(/MI/continuity statistic/recurrence flow), respectively. **E** Zoomed RPs for the ACF, $\langle \varepsilon^* \rangle(\tau)$ and $\Phi(\tau)$ (from left to right). **F** Three-dimensional phase space reconstruction based on recurrence flow shows a spiral-like trajectory

recurrence flow measure, we identify all peaks of $\Phi(\tau)$ at different delays τ and choose the highest of these to obtain the corresponding embedding delay $\tau^{(\Phi)}$. In case there is several peaks of identical height, we choose the one for the smallest τ (i.e., the first one).

To illustrate the procedure, we consider a time series of the past 1 million years of insolation on Earth at 52.39° latitude (see Appendix A) [53]. The insolation depends on the Earth orbit and the Earth axis tilt and precession, thus, varies in specific cycles (Milankovich cycles). Cao's method suggests an embedding dimension $m = 4$. We present results for the second and third component of the embedding vector. Using first-crossing of ACF and first minimum of MI, both measures suggest a delay of $\tau_1^{(ACF)} = \tau_1^{(MI)} = 6$ka (Fig. 3A). Since none of both measures is capable of selecting different embedding delays for higher components of the embedding vector, this yields the three-dimensional embedding vector $\vec{v}(t) = [y(t), y(t - 6\text{ka}), y(t - 12\text{ka})]$. Next, we compute $\Phi(\tau)$ to check if the estimate of $\vec{v}(t)$ based on the traditional TDE metrics is confirmed: we find $\tau_1^{(\Phi)} = 5$ka in close agreement with the ACF and MI-criteria (Fig. 3C). Moreover, the continuity statistic $\langle \varepsilon^* \rangle(\tau)$ suggests the same embedding delay for the first component as the auto-correlation and MI (first local maximum). It does not indicate a global maximum at this delay, yielding a more ambiguous choice of the optimal embedding delay than $\Phi(\tau)$. Finally, we examine if both multi-dimensional measures suggest $\tau_2 = 2\tau_1$ for the second embedding delay as expected for a traditional uniform time delay embedding (UTDE). Interestingly, $\langle \varepsilon^* \rangle(\tau)$ offers only limited information

on an optimal embedding delay for the third component of $\vec{v}(t)$ (Fig. 3D). Multiple local maxima offer a variety of choices with no clear optimal value. We choose the global maximum (marked by star). Conversely, $\Phi(\tau)$ once more provides a clear choice for the second embedding delay with globally maximized flow for $\tau_2^{(\Phi)} = 2\tau_1^{(\Phi)}$.

We visually evaluate the quality of the resulting embeddings by comparing the line structures in the corresponding RPs (Fig. 3E). The enlarged details of the RPs illustrate how well deterministic intervals in the evolution of insolation are resolved based on the phase space reconstructions yielded by the ACF, continuity statistic, and recurrence flow. While the uniform embedding vectors obtained from the ACF and $\Phi(\tau)$ result in well-separated, undisturbed diagonal lines, multiple diagonal lines and the related cycles are poorly expressed in the phase space suggested by $\langle \varepsilon^* \rangle(\tau)$. The reconstructed phase space based on the embedding vector obtained from the recurrence flow criterion reveals several unstable periodic orbits (Fig. 3F), constituting a concentric spiral-like phase space trajectory in three dimensions.

Many real-world systems allow taking only a relatively short series of measurements for a single variable with high levels of superimposed measurement noise. We study how well a known phase space of a paradigmatic system can be reconstructed based on the four different measures considered above with increasing noise strength. In particular, we generate $n = 2000$ samples of a Rössler system (see Appendix A) such that the resulting trajectory only covers relatively few unstable periodic orbits. We reconstruct the

62

Eur. Phys. J. Spec. Top. (2023) 232:57–67

known three-dimensional phase space from the $y(t)$-component (Fig. 4A) with superimposed uncorrelated white noise realizations (Fig. 4B).

The noise strength (standard deviation of the noise) is varied in multiples of the standard deviation σ_{Roe} of the undisturbed $y(t)$. Even with only 10% measurement noise, the original attractor is already significantly less smooth (Fig. 4B). We compare the ACF, MI, continuity statistic, and recurrence flow as delay selection measures while only uniform time delays are considered, i.e., the optimal embedding delay is selected only once for the step from a one- to a two-dimensional embedding. This ensures that the two high-dimensional measures ($\Phi(\tau)$ and $\langle\varepsilon^*\rangle(\tau)$) can be compared adequately to the traditional measures.

To quantitatively evaluate the dependence of the reconstruction on noise strength, we generate an RP for each reconstruction and for each of the four delay selection methods. For each RP, we compute the joint recurrence rate fraction (JRRF):

$$\mathrm{JRRF} = \frac{\sum_{i,j}^{N} JR_{i,j}}{\sum_{i,j=1}^{N} R_{i,j}^{\mathrm{ref}}}, \qquad \mathrm{JRRF} \in [0,1]$$
$$\text{with}$$
$$\mathbf{JR} = \mathbf{R}^{\mathrm{ref}} \circ \mathbf{R}^{\mathrm{rec}},$$

from the RP of the (real) reference system $\mathbf{R}^{\mathrm{ref}}$ and the RP of the respective reconstruction $\mathbf{R}^{\mathrm{rec}}$. We use it to quantify the accordance of the real RP of the corresponding noisy Rössler system to the reconstructions with respect to its recurrence structure (the higher JRRF, the better the reconstruction).

We generate 50 individual noise realizations for each noise strength between 0 and 100% of the original standard deviation σ_{Roe} and average the corresponding JRRF values (Fig. 4C). As expected, with increasing noise level, the quality of the reconstruction decreases.

For noise strengths between 10 and 30%, the linear ACF performs worst in terms of JRRF while the MI and continuity statistic perform equally well. The almost perfect alignment of both might seem surprising but is due to the discreteness immanent in the delay selection. While their agreement might be interpreted in the way that this has to be the optimal delay (i.e., the reconstructed system preserves most of the recurrence structure), the recurrence flow $\Phi(\tau)$ shows superior performance with noise strengths up to 30%, i.e., three times the noise level illustrated in Fig. 4B. Beyond 30%, all four measures yield approximately the same performance (Fig. 4C inset), as for JRRF< 0.5, the alignment could be explained by random joint recurrences.

3.3 Non-uniform time delay embedding

For many real-world dynamical systems, it is not sufficient to consider only a single characteristic time scale. Instead, multi-scale systems are governed by a multitude of processes that imprint (quasi-)periodic cycles of various lengths onto the measured time series. The

selection of delays must account for this complexity by considering non-uniform embedding delays (non-uniform time delay embedding, NUTD). One of the most studied systems that exhibits multi-scale dynamics is the El Niño-Southern Oscillation (ENSO). ENSO represents a quasi-periodic climate pattern that is associated with spatio-temporal variations of sea surface temperatures in the central and eastern Pacific Ocean, oscillating between El Niño and La Niña events. We use the delay differential ENSO model proposed in [54] to examine if the recurrence flow can unveil distinct delays for a three-dimensional state space reconstruction of model time series. The model is based on a nonlinear delay differential equation and reproduces an abundance of key features of ENSO (see Appendix A). We study two different solution types that are associated with distinct dynamical regimes, i.e., a seasonal oscillation with superimposed faster low-amplitude oscillations prior to a period-doubling and irregular oscillations that are reminiscent of El Niño and La Niña events of random magnitudes. A more detailed discussion of these solution types can be found in [54].

The first solution type exhibits well-pronounced seasonal cycle and fast, amplitude-modulated wiggles on top (Fig. 5A). Cao's method suggests that this solution type can be embedded in a $m = 3$-dimensional embedding space. We find that MI does not yield an unambiguous choice for an embedding delay while the ACF suggests $\tau_1^{(\mathrm{ACF})} = 0.25$years, i.e., the expected value of a quarter of the seasonal cycle (Fig. 5B). The same delay is identified with $\Phi(\tau)$ whereby $\langle\varepsilon^*\rangle(\tau)$ yields a slightly higher optimal embedding delay (Fig. 5C). While traditional UTDE now suggests $\tau_2^{(\mathrm{ACF})} = 2\tau_1^{(\mathrm{ACF})}$ for a three-dimensional embedding, both $\Phi(\tau)$ and $\langle\varepsilon^*\rangle(\tau)$ instead show that a different choice yields a superior phase space reconstruction in terms of minimized redundancy (Fig. 5C/D). Both $\langle\varepsilon^*\rangle(\tau)$ and $\Phi(\tau)$ effectively uncover the faster cycle by means of local maxima. However, both also detect a delay that is the sum of the seasonal and the fast cycle as a promising candidate. The fact that the estimate of this conjoint cycle differs for both measures can be explained by the different estimates on τ_1. Since for $\Phi(\tau)$ both local maxima have the same height, we pick the first. The zoomed RPs clearly express that both NUTD selection methods entail more coherent diagonal lines with less perpendicular distortions (Fig. 5E). Despite the different embedding vectors $\vec{v}^{(\Phi)}$ and $\vec{v}^{(\varepsilon^*)}$, both reconstructions give a convincing representation of the seasonal cycle in the RP, respectively. The reconstructed phase space based on the delays selected from the recurrence flow yields a clear visualization of the system's periodic oscillations (Fig. 5F).

In the same manner, we study the second solution type of irregular ENSO-like oscillations (Fig. 6). Cao's method yields a 4-dimensional TDE. Both the ACF and MI identify different but similar delays slightly larger than $\tau_1 = 0.35$ years (Fig. 6B). $\langle\varepsilon^*\rangle(\tau)$ and $\Phi(\tau)$ agree on an embedding delay of $\tau_1^{(\Phi)} = \tau_1^{(\varepsilon^*)} = 0.33$ years

Fig. 4 Sensitivity of recurrence flow against measurement noise, compared to other dependence measures. **A** $y(t)$-component and reference phase space of Rössler system without noise contamination; **B** the same, but with 10% measurement noise. Time axis ticks refer to consecutive integration steps. **C** Performance of ACF, auto-MI, continuity statistic $\langle\varepsilon^*\rangle(\tau)$ and recurrence flow $\Phi(\tau)$ in terms of JRRF for noise strengths from 0% to 30% (and 100% inset)
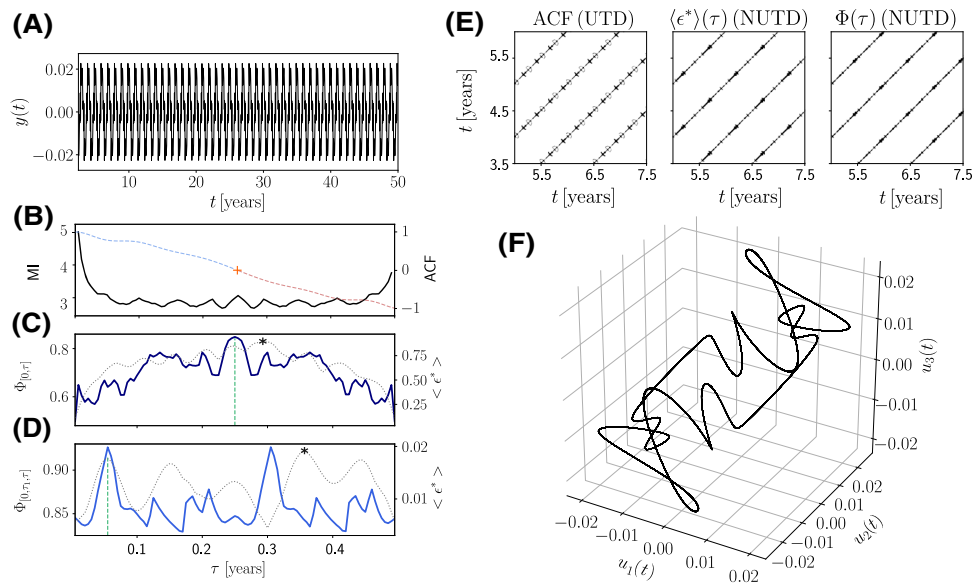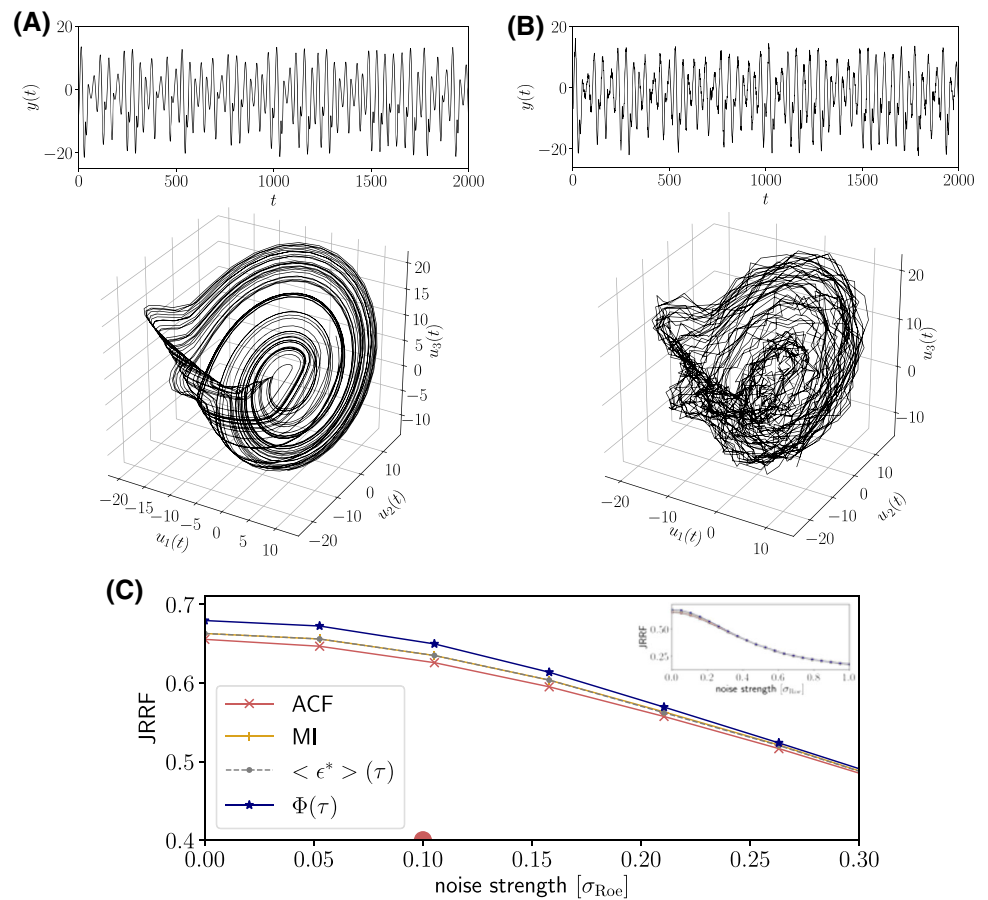

Fig. 5 Non-uniform delay selection for delay-differential ENSO model with **A** periodic dynamics, analogously to Fig. 3. Serial dependence of $y(t)$, measured in terms of **B** ACF (red/blue) and auto-MI (black) for the univariate time series, as well as by **C/D** the continuity statistic $\langle\varepsilon^*\rangle(\tau)$ (gray) and the recurrence flow $\Phi(\tau)$ (dark blue) for **C** two- and **D** three-dimensional embedding vectors $\vec{v}(t)$. The optimal delay is marked by a circle(/cross/star/vertical green line) for the ACF(/MI/continuity statistic/recurrence flow), respectively. **E** Zoomed RPs for the ACF, $\langle\varepsilon^*\rangle(\tau)$ and $\Phi(\tau)$ (from left to right). **F** Three-dimensional phase space reconstruction based on recurrence flow
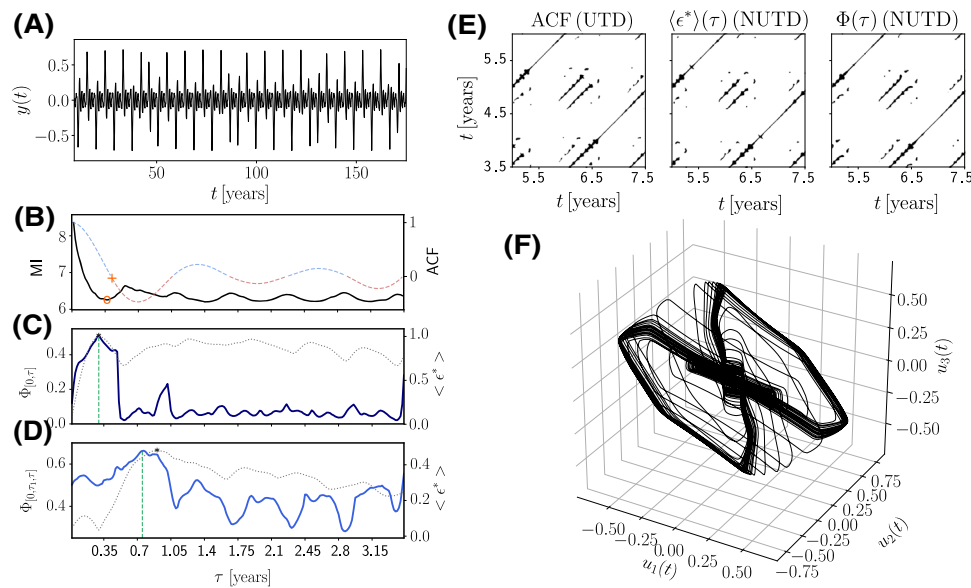
Fig. 6 Non-uniform delay selection for delay-differential ENSO model with **A** irregular oscillations, analogously to Fig. 3. Serial dependence of $y(t)$, measured in terms of **B** ACF (red/blue) and auto-MI (black) for the univariate time series, as well as by **C/D** the continuity statistic $\langle \varepsilon^* \rangle(\tau)$ (gray) and the recurrence flow $\Phi(\tau)$ (dark blue) for **C** two- and **D** three-dimensional embedding vectors $\vec{v}(t)$. The optimal delay is marked by a circle(/cross/star/vertical green line) for the ACF(/MI/continuity statistic/recurrence flow), respectively. **E** Zoomed RPs for the ACF, $\langle \varepsilon^* \rangle(\tau)$ and $\Phi(\tau)$ (from left to right). **F** Three-dimensional phase space reconstruction based on recurrence flow

which is slightly smaller than both $\tau_1^{(ACF)}$ and $\tau_1^{(MI)}$ (Fig. 6C). Again, the estimates $\tau_2^{(\Phi)}$ and $\tau_2^{(\varepsilon^*)}$ differ from the choice that would result from UTDE and indicate that describing the dominant variability in the observed irregular oscillations requires a multi-scale approach (Fig. 6D). Slight deviations in both estimates give rise to minor discrepancies between the emerging diagonal lines in the respective RPs (Fig. 6E). For none of the three reconstructions, DLA are entirely removed which hints at an optimal embedding dimension $m > 3$ as identified by Cao's method. This is supported by the displayed attractor reconstruction that has a two-winged structure reminiscent of the famous Lorenz attractor but potentially be unfolded further (Fig. 6F). However, both NUTD selection measures once more provide a more convincing result in terms of sparse, continuous diagonal lines than the ACF and, thus, capture the system's predictability more adequately.

4 Conclusion

The nonlinearity and scale-dependence of relationships observed in high-dimensional empirical data calls for appropriate and easily applicable methods. Recurrence plots offer a mathematically simple yet effective framework for the study of dependencies in high-dimensional dynamical systems and are an established tool in applied nonlinear time series analysis. For deterministic systems, diagonal lines in an RP yield valuable information on the evolution of a system's trajectory.

We propose a novel recurrence-based dependence measure, called recurrence flow. It builds on the fact that an RP can be computed from delayed copies of one (or multiple) time series that are stacked on top of each other as it is common practise in time delay embedding. The absence of spurious structures perpendicular to diagonal lines hints at a correct choice of (embedding) delays between the embedding vector's coordinates. We exploit the emergence of these structures to characterize the serial dependence in uni- and cross-dependence of multi-variate, high-dimensional systems. We demonstrated that the recurrence flow $\Phi(\tau)$ captures nonlinear, lagged dependencies in the presence of observational noise. Due to its conceptual proximity to time delay embedding, we put a focus on the delay selection problem that underlies attractor reconstruction. The recurrence flow effectively pinpoints uniform and non-uniform embedding delays in multi-dimensional nonlinear systems. It, thus, bares high potential to be used in a recurrence-based embedding algorithm which will be the focus of future work. Compared to existing nonlinear dependence measures, it offers several advantages: (i) it is not based on a binning procedure and does not suffer from the curse of dimensionality, (ii) it is able to detect non-uniform delays in high-dimensional multi-scale data, (iii) it performs relatively well in the presence of observational noise, and, (iv) for the considered examples, it yields well interpretable, unambiguous maxima in the delay selection procedure (unlike, e.g., the continuity statistic). The recurrence flow formalism appears promising for the conceptualization of

an automated recurrence-based time delay embedding algorithm.

Code availability Scripts and data used to perform this study are available via Zenodo (https://doi.org/10.5281/zenodo.6631137).

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Appendix A: Example systems

To assess the performance of $\Phi(\tau)$ as a measure to select embedding delays, the following systems are considered:

Numerical insolation model

The complex superposition of gravitational forces between the earth and the other planetary bodies in the solar system perturbs the earth's rotation on an elliptic orbit around the sun and its axial rotation. The cycles that manifest due to these variations control the earth's climate at time scales of millions of years and are called Milankovich cycles. The nature of the underlying perturbation renders the variations in insolation chaotic. The model proposed in [53] considers all nine planets in the solar system and describes the orbit of the moon separately. A Hamiltonian, consisting of an integrable and perturbation component, is numerically integrated with a symplectic integrator scheme (SABAC^4). Several dissipative effects (tides, core-mantle friction, climate friction) are included. The model returns time series for the earth's orbit's eccentricity, climatic precession, obliquity, and insolation, from which we study only the latter with $n = 1000$ samples. For the computation of RPs, we fix the recurrence rate to 5%.

Rössler system

The Rössler system is a three-dimensional, continuous dynamical system that generates a strange attractor:

$$\begin{aligned}
\dot{x} &= -y - z, \\
\dot{y} &= x + ay, \\
\dot{z} &= b + z(x - c).
\end{aligned}$$

The time series only covers a few unstable periodic orbits with 3000 samples from which we discard 1000 as transients, resulting in $N = 2000$. Uniform sampling intervals are fixed as $\Delta t = 0.10$. We set $a = 0.02925$, $b = 0.1$ and $c = 8.5$, ensuring chaotic dynamics. Independent realizations of uncorrelated white noise with different noise strengths between 0% and 100% are superimposed on the $y(t)$-component to mimic measurement noise. For the computation of RPs, we fix the recurrence rate to 8% regardless of the noise strength.

Delay differential ENSO model

As an example for a system with multiple characteristic time scales, we examine a delay differential model of ENSO. On top of the seasonal mode, ENSO represents the predominant mode of sea surface temperature (SST) variability in the tropical Pacific. The SSTs exhibit recurring variations with time scales between 2 and 7 years between two regimes of well-distinguishable SST anomalies: El Niño (warming phase) and La Niña (cooling phase). These variations disturb large-scale air transport in the tropics and induce a multitude of global climatic impacts, e.g., droughts and floods in Australia or South America. Several conceptual models have reproduced key features of this oscillation by including hypotheses on the mechanistic origins of ENSO, including negative and positive feedbacks of temperature anomalies and atmospheric circulation and potential resonance phenomena with the seasonal forcing. The model studied here is taken from [54] and mimics ENSO dynamics based on two key mechanisms, i.e., delayed negative feedback and seasonal forcing

$$\frac{\mathrm{d}y(t)}{\mathrm{d}t} = -\tanh(\kappa y(t - \zeta)) + b\cos(2\pi\omega t).$$

We set the frequency of the periodic forcing to seasonal forcing ($\omega = 1$) and fix $b = 1$. Variations in the delay ζ and the parameter κ give rise to dynamically

66

Eur. Phys. J. Spec. Top. (2023) 232:57–67

distinct time series. We study two solution types: a regular solution with a seasonal cycle and fast, amplitude-modulated wiggles ($\kappa = 100$, $\zeta = 0.025$) and a solution of irregular ENSO-like oscillations with stochastic amplitude variations ($\kappa = 50$, $\zeta = 0.42$). For both solution types, we generate $n = 10{,}000$ values. For the computation of RPs, we fix the recurrence rate to 5%.

Appendix B: Statistical significance

For an uncorrelated white noise time series of infinite length, the presence of recurrences along a diagonal can be described by a binomial distribution. The flow along each diagonal can consequently be regarded as an idealized sequence of Bernoulli trials for which a success is equivalent to a recurrence, i.e., a black pixel. The probability of having X unsuccessful Bernoulli trials (no recurrences) until a trial succeeds is given by the geometric distribution

$$P(X = n) = (1 - p)^{k-1} p \qquad (6)$$

with expectation value $1/p$. The probability of success p for each Bernoulli trial is given by the recurrence rate and depends on ε. It follows that we can derive the recurrence flow for an idealized uncorrelated white noise time series of length n as

$$\theta(\varepsilon) = 1 - \frac{n}{p(\varepsilon)}. \qquad (7)$$

Given an observational time series of length n and a suitable choice for the vicinity threshold ε, $p(\varepsilon)$ can be identified with the recurrence rate and Eq. (7) can be used to test whether the recurrence flow of the real signal can be distinguished from an uncorrelated random process.

References

1. J. Kwapień, P. Oświęcimka, S. Drożdż, Detrended fluctuation analysis made flexible to detect range of cross-correlated fluctuations. Phys. Rev. E **92**(5), 052815 (2015)
2. M. Newman, *Networks* (Oxford University Press, Oxford, 2018)
3. J.H. Feldhoff, R.V. Donner, J.F. Donges, N. Marwan, J. Kurths, Geometric detection of coupling directions by means of inter-system recurrence networks. Phys. Lett. A **376**(46), 3504–3513 (2012). https://doi.org/10.1016/j.physleta.2012.10.008
4. J. Runge, V. Petoukhov, J.F. Donges, J. Hlinka, N. Jajcay, M. Vejmelka, D. Hartman, N. Marwan, M. Paluš, J. Kurths, Identifying causal gateways and mediators in complex spatio-temporal systems. Nat. Commun. **6**(1), 1–10 (2015)
5. J. Runge, P. Nowack, M. Kretschmer, S. Flaxman, D. Sejdinovic, Detecting and quantifying causal associations in large nonlinear time series datasets. Sci. Adv. **5**(11), eaau4996 (2019)
6. A.M.T. Ramos, A. Builes-Jaramillo, G. Poveda, B. Goswami, E.E.N. Macau, J. Kurths, N. Marwan, Recurrence measure of conditional dependence and applications. Phys. Rev. E **95**, 052206 (2017). https://doi.org/10.1103/PhysRevE.95.052206
7. V. Dakos, M. Scheffer, E.H. van Nes, V. Brovkin, V. Petoukhov, H. Held, Slowing down as an early warning signal for abrupt climate change. Proc. Natl. Acad. Sci. **105**(38), 14308–14312 (2008)
8. C. Boettner, G. Klinghammer, N. Boers, T. Westerhold, N. Marwan, Early-warning signals for cenozoic climate transitions. Quat. Sci. Rev. **270**, 107177 (2021)
9. N. Boers, Early-warning signals for Dansgaard–Oeschger events in a high-resolution ice core record. Nat. Commun. **9**(1), 1–8 (2018)
10. G. Balasis, R.V. Donner, S.M. Potirakis, J. Runge, C. Papadimitriou, I.A. Daglis, K. Eftaxias, J. Kurths, Statistical mechanics and information-theoretic perspectives on complexity in the earth system. Entropy **15**(11), 4844–4888 (2013)
11. J. Sun, C. Cafaro, E.M. Bollt, Identifying the coupling structure in complex systems through the optimal causation entropy principle. Entropy **16**(6), 3416–3433 (2014)
12. M. Porfiri, R.R. Sattanapalle, S. Nakayama, J. Macinko, R. Sipahi, Media coverage and firearm acquisition in the aftermath of a mass shooting. Nat. Human Behav. **3**(9), 913–921 (2019)
13. B. Pompe, J. Runge, Momentary information transfer as a coupling measure of time series. Phys. Rev. E **83**(5), 051122 (2011)
14. A. Kraskov, H. Stögbauer, P. Grassberger, Estimating mutual information. Phys. Rev. E **69**(6), 066138 (2004)
15. A.P. Nawroth, J. Peinke, Multiscale reconstruction of time series. Phys. Lett. A **360**(2), 234–237 (2006)
16. D. Maraun, J. Kurths, Cross wavelet analysis: significance testing and pitfalls. Nonlinear Process. Geophys. **11**(4), 505–514 (2004)
17. A. Agarwal, R. Maheswaran, N. Marwan, L. Caesar, J. Kurths, Wavelet-based multiscale similarity measure for complex networks. Eur. Phys. J. B **91**(11), 1–12 (2018)
18. T. Braun, V.R. Unni, R.I. Sujith, J. Kurths, N. Marwan, Detection of dynamical regime transitions with lacunarity as a multiscale recurrence quantification measure. Nonlinear Dyn. **104**(4), 3955–3973 (2021)
19. I. Vlachos, D. Kugiumtzis, Nonuniform state-space reconstruction and coupling detection. Phys. Rev. E **82**(1), 016207 (2010)
20. L.M. Pecora, L. Moniz, J. Nichols, T.L. Carroll, A unified approach to attractor reconstruction. Chaos Interdiscip. J. Nonlinear Sci. **17**(1), 013110 (2007)
21. J.-P. Eckmann, S. Oliffson Kamphorst, D. Ruelle, Recurrence plots of dynamical systems. Europhys. Lett. **4**(9), 973–977 (1987). https://doi.org/10.1209/0295-5075/4/9/004
22. N. Marwan, M. Carmen Romano, M. Thiel, J. Kurths, Recurrence plots for the analysis of complex systems. Phys. Rep. **438**(5–6), 237–329 (2007)

23. M. Thiel, M.C. Romano, P.L. Read, J. Kurths, Estimation of dynamical invariants without embedding by recurrence plots. Chaos **14**(2), 234–243 (2004). https://doi.org/10.1063/1.1667633

24. S. Schinkel, O. Dimigen, N. Marwan, Selection of recurrence threshold for signal detection. Eur. Phys. J. Spec. Top. **164**(1), 45–53 (2008). https://doi.org/10.1140/epjst/e2008-00833-5

25. K. Klimaszewska, J.J. Żebrowski, Detection of the type of intermittency using characteristic patterns in recurrence plots. Phys. Rev. E **80**, 026214 (2009). https://doi.org/10.1103/PhysRevE.80.026214

26. Y. Hirata, K. Aihara, Devaney's chaos on recurrence plots. Phys. Rev. E **82**, 036209 (2010). https://doi.org/10.1103/PhysRevE.82.036209

27. N. Marwan, N. Wessel, U. Meyerfeldt, A. Schirdewan, J. Kurths, Recurrence plot based measures of complexity and its application to heart rate variability data. Phys. Rev. E **66**(2), 026702 (2002). https://doi.org/10.1103/PhysRevE.66.026702

28. N. Marwan, J.F. Donges, R.V. Donner, D. Eroglu, Nonlinear time series analysis of palaeoclimate proxy records. Quat. Sci. Rev. **274**, 107245 (2021)

29. M. Carmen Romano, M. Thiel, J. Kurths, C. Grebogi, Estimation of the direction of the coupling by conditional probabilities of recurrence. Phys. Rev. E **76**(3), 036211 (2007)

30. A.M. Nkomidio, E.K. Ngamga, B.R.N. Nbendjo, J. Kurths, N. Marwan, Recurrence-based synchronization analysis of weakly coupled bursting neurons under external ELF fields. Entropy **24**(2), 235 (2022). https://doi.org/10.3390/e24020235

31. J.P. Zbilut, N. Marwan, The Wiener–Khinchin theorem and recurrence quantification. Phys. Lett. A **372**(44), 6622–6626 (2008)

32. E. Bradley, R. Mantilla, Recurrence plots and unstable periodic orbits. Chaos Interdisc. J. Nonlinear Sci. **12**(3), 596–600 (2002)

33. A. Banerjee, B. Goswami, Y. Hirata, D. Eroglu, B. Merz, J. Kurths, N. Marwan, Recurrence analysis of extreme event-like data. Nonlinear Process. Geophys. **28**(2), 213–229 (2021)

34. C. Ozdes, D. Eroglu, Transformation cost spectrum for irregularly sampled time series. Eur. Phys. J. Spec. Top. (2022). https://doi.org/10.1140/epjs/s11734-022-00512-x

35. M.C. Romano, M. Thiel, J. Kurths, Generalized synchronization indices based on recurrence in phase space. AIP Conf. Proc. **742**, 330–336 (2004). https://doi.org/10.1063/1.1846494

36. M.C. Romano, M. Thiel, J. Kurths, I.Z. Kiss, J.L. Hudson, Detection of synchronization for non-phase-coherent and non-stationary data. Europhys. Lett. **71**(3), 466–472 (2005). https://doi.org/10.1209/epl/i2005-10095-1

37. D.V. Senthilkumar, M. Lakshmanan, J. Kurths, Transition from phase to generalized synchronization in time-delay systems. Chaos **18**(2), 023118 (2008). https://doi.org/10.1063/1.2911541

38. Y. Zou, M.C. Romano, M. Thiel, N. Marwan, J. Kurths, Inferring indirect coupling by means of recurrences. Int. J. Bifurc. Chaos **21**(04), 1099–1111 (2011)

39. B. Goswami, N. Marwan, G. Feulner, J. Kurths, How do global temperature drivers influence each other? Eur. Phys. J. Spec. Top. **222**(3), 861–873 (2013)

40. M. Porfiri, M.R. Marín, Transfer entropy on symbolic recurrences. Interdiscip. J. Nonlinear Sci. **29**(6), 063123 (2019)

41. T. Westerhold, N. Marwan, A.J. Drury, D. Liebrand, C. Agnini, E. Anagnostou, J.S.K. Barnet, S.M. Bohaty, D. De Vleeschouwer, F. Florindo et al., An astronomically dated record of earth's climate and its predictability over the last 66 million years. Science **369**(6509), 1383–1387 (2020)

42. M. Thiel, M. Carmen Romano, J. Kurths, Analytical description of recurrence plots of white noise and chaotic processes. Appl. Nonlinear Dyn. **11**(3), 20–30 (2003)

43. N. Marwan, How to avoid potential pitfalls in recurrence plot based data analysis. Int. J. Bifurc. Chaos **21**(04), 1003–1017 (2011)

44. K. Hauke Kraemer, N. Marwan, Border effect corrections for diagonal line based recurrence quantification analysis measures. Phys. Lett. A **383**(34), 125977 (2019)

45. M. Thiel, M. Carmen Romano, J. Kurths, Spurious structures in recurrence plots induced by embedding. Nonlinear Dyn. **44**(1), 299–305 (2006)

46. N. Marwan, J. Kurths, Line structures in recurrence plots. Phys. Lett. A **336**(4–5), 349–357 (2005). https://doi.org/10.1016/j.physleta.2004.12.056

47. K. Hauke Kraemer, M. Gelbrecht, I. Pavithran, R.I. Sujith, N. Marwan, Optimal state space reconstruction via Monte Carlo decision tree search. Nonlinear Dyn. **108**(2), 1525–1545 (2022)

48. J.P. Zbilut, C.L. Webber Jr., Embeddings and delays as derived from quantification of recurrence plots. Phys. Lett. A **171**(3–4), 199–203 (1992)

49. F.M. Atay, Y. Altıntaş, Recovering smooth dynamics from time series with the aid of recurrence plots. Phys. Rev. E **59**(6), 6593 (1999)

50. R.A.A. Ince, B.L. Giordano, C. Kayser, G.A. Rousselet, J. Gross, P.G. Schyns, A statistical framework for neuroimaging data analysis based on mutual information estimated via a Gaussian copula. Hum. Brain Mapp. **38**(3), 1541–1573 (2017)

51. K.H. Kraemer, G. Datseris, J. Kurths, I.Z. Kiss, J.L. Ocampo-Espindola, N. Marwan, A unified and automated approach to attractor reconstruction. New J. Phys. **23**(3), 033017 (2021)

52. L. Cao, Practical method for determining the minimum embedding dimension of a scalar time series. Phys. D Nonlinear Phenom. **110**(1–2), 43–50 (1997)

53. J. Laskar, P. Robutel, F. Joutel, M. Gastineau, A.C.M. Correia, B. Levrard, A long-term numerical solution for the insolation quantities of the earth. Astron. Astrophys. **428**(1), 261–285 (2004)

54. M. Ghil, I. Zaliapin, S. Thompson, A delay differential model of ENSO variability: parametric instability and the distribution of extremes. Nonlinear Process. Geophys. **15**(3), 417–433 (2008)