ORIGINAL PAPER

Optimal state space reconstruction via Monte Carlo decision tree search

K. Hauke Kraemer · **Maximilian Gelbrecht** ·
Induja Pavithran · **R. I. Sujith** ·
Norbert Marwan

Abstract A novel idea for an optimal time delay state space reconstruction from uni- and multivariate time series is presented. The entire embedding process is considered as a game, in which each move corresponds to an embedding cycle and is subject to an evaluation through an objective function. This way the embedding procedure can be modeled as a tree, in which each leaf holds a specific value of the objective function. By using a Monte Carlo ansatz, the proposed algorithm populates the tree with many leafs by computing different possible embedding paths and the final embedding is chosen as that particular path, which ends at the leaf with the lowest achieved value of the objective function. The method aims to prevent getting stuck in a local minimum of the objective function and can be used in a modular way, enabling practitioners to choose a statistic for possible delays in each embedding cycle as well as a suitable objective function themselves. The proposed method guarantees the optimization of the chosen objective function over the parameter space of the delay embedding as long as the tree is sampled sufficiently. As a proof of concept, we demonstrate the superiority of the proposed method over the classical time delay embedding methods using a variety of application examples. We compare recurrence plot-based statistics inferred from reconstructions of a Lorenz-96 system and highlight an improved forecast accuracy for map-like model data as well as for palaeoclimate isotope time series. Finally, we utilize state space reconstruction for the detection of causality and its strength between observables of a gas turbine type thermoacoustic combustor.

K. H. Kraemer · M. Gelbrecht · N. Marwan
Potsdam Institute for Climate Impact Research, Member of the Leibniz Association, 14473 Potsdam, Germany

K. H. Kraemer (✉)
Institute of Physics and Astronomy, University of Potsdam, 14476 Potsdam, Germany
e-mail: hkraemer@pik-potsdam.de

M. Gelbrecht
Department of Mathematics and Computer Science, Freie Universität Berlin, Arnimallee 14, 14195 Berlin, Germany

M. Gelbrecht
TUM School of Engineering and Design, Technische Universtiät München, 80333 Munich, Germany

I. Pavithran
Department of Physics, IIT Madras, Chennai 600036, India

R. I. Sujith
Department of Aerospace Engineering, IIT Madras, Chennai 600036, India

N. Marwan
Institute of Geosciences, University of Potsdam, 14476 Potsdam, Germany

1 Introduction

The famous embedding theorems of Whitney [1], Mañé [2], and Takens [3] together with their enhancement by Sauer et al. [4] allow a high dimensional state space reconstruction from (observed) uni- or multivariate time series. Computing dynamical invariants [5–9] from the observed system, making meaningful predictions even for chaotic or stochastic systems [10–16], detecting causal interactions [17–19] or nonlinear noise reduction algorithms [20,21] all rely explicitly or implicitly on (time delay) embedding [22] the data into a reconstructed state space. Other ideas rather than time delay embedding (TDE) are also possible [22–27], but due to its simple use and its proficient outcomes in a range of situations, TDE is by far the most common reconstruction technique. Suppose there is a multivariate dataset consisting of M time series $s_i(t)$, $i = 1, \ldots, M$. The basic idea is to use lagged values of the available time series as components of the reconstruction vector

$$\mathbf{v}(t) = \big(s_{i_1}(t - \tau_1), s_{i_2}(t - \tau_2), \ldots, s_{i_m}(t - \tau_m)\big).$$
(1)

Here, the delays τ_j are multiples of the sampling time Δt and the indices $i_1, i_2, \ldots, i_m$ each denote the time series index $i \in [1, \ldots, M]$, which has been chosen in the 1st, 2nd, $\ldots$, mth embedding cycle. The total number of delays τ_j, $j = [1, \ldots, m]$, i.e., the embedding dimension m, its values and the corresponding time series s_{i_j}, $i_j \in [1, \ldots, M]$ need to fulfill certain criteria to guarantee the equivalence to the unknown true attractor, e.g., the embedding dimension must suffice $m \geqslant 2D_B + 1$, with D_B being the unknown box-counting dimension (see Casdagli et al. [28], Gibson et al. [24], Uzal et al. [29] or Nichkawde [30] for a profound overview of the problem). Picking optimal embedding parameters τ_j and m comes down to make the resulting components of the reconstruction vectors $\mathbf{v}(t)$ as independent as possible [4,22], but at the same time not too independent, in order to keep sufficient information of the correlation structure of the data [28,29,31–33]. Besides some unified approaches [34–44], which tackle the estimation of the delays τ_j and the embedding dimension m simultaneously, most researchers use two different methods to perform the reconstruction.

(1) A statistic determines the delays τ_j, we call it Λ_τ throughout this paper. Usually, $\tau_1 = 0$, i.e., the first component of $\mathbf{v}(t)$ is the unlagged time series s_{i_1} in Eq. (1). For embedding a univariate time series, $s_{i_1} = \ldots = s_{i_m} = s(t)$, the approach to choose τ_2 from the first minimum of the auto-mutual information [45,46] is most common. All consecutive delays are then simply integer multiples of τ_2. Other ideas based on different statistics like the auto-correlation function of the time series have been suggested [23,32,33,40,47–50]. However, by setting τ_j, $j > 2$ to multiples of τ_2, one ignores the fact that this "measure" of independence strictly holds only for the first two components of reconstruction vectors ($m = 2$) [51,52], even though in practice it works fine for most cases. More sophisticated ideas, like high-dimensional conditional mutual information [53,54] and other statistics [54–59], some of which include non-uniform delays and the extension to multivariate input data [30,38,39,53,54,60,60–65,65], have been presented.

(2) A statistic, we call it Γ throughout this paper, which serves as an objective function and quantifies the goodness of a reconstruction, given that delays τ_j have been estimated. The embedding process is thought of as an iterative process, starting with an unlagged (given) time series s_{i_1}, i.e., $\tau_1 = 0$. In each embedding cycle $\mathscr{D}_d$, $[d = 1, \ldots, m]$, a time series s_{i_d} lagged by τ_d gets appended to obtain the actual reconstruction vectors $\mathbf{v}_d(t) \in \mathbb{R}^{d+1}$ and these are compared to the reconstruction vectors $\mathbf{v}_{d-1}(t)$ of the former embedding cycle (if $d = 1$, $\mathbf{v}_{d-1}(t)$ is simply the time series s_{i_1}). This "comparison" is usually achieved by the amount of false nearest neighbors (FNN) [66–70], some other neighborhood-preserving-idea [71,72], or more ambitious ideas [29,30].

We have recently proposed an algorithm [41], which minimizes the L-statistic [29] (the objective function) in each embedding cycle $\mathscr{D}_d$ over possible delay values in this embedding cycle determined by a continuity statistic [65]. Nichkawde [30] minimizes the FNN-statistic in each embedding cycle over time delays given by a statistic, which maximizes directional derivatives of the actual reconstruction vectors. However, it cannot be ruled out that these approaches result in achieving a local minimum of the corresponding objective function, rather than attaining the global minimum.

Here, we propose a Monte Carlo Decision Tree Search (MCDTS) idea to ensure the reach of a global minimum of a freely selectable objective function Γ, e.g., the L- or FNN-statistic or any other suitable statistic, which evaluates the goodness of the reconstruction with respect to the task. A statistic Λ_τ, which guides the pre-selection of potential delay values in each embedding cycle (such as the continuity statistic or conditional mutual information), is also freely selectable and can be tailored to the research task. This modular construction might be useful for practitioners, since it has been pointed out that optimal embedding parameters—thus also the used statistics to approximate them—depend on the research question, e.g., computing dynamical invariants or prediction [63,64,73–75]. Thus, the proposed method is neither restricted to the auto-mutual information, in order to measure the independence of consecutive reconstruction vector components, nor does it necessarily rely on the ubiquitous false nearest neighbor statistic. Independently from the chosen statistic for potential time delays and from the chosen objective function, the proposed method computes different embedding pathways in a randomized manner and structures these paths as a tree. Consequently, it is able to reveal paths through that tree— if there are any—which lead to a lower value of the objective function than paths, which strictly minimize the costs in each embedding cycle. Given a sufficiently high number of samplings, MCDTS guarantees to optimize the chosen objective function Γ over the (delay embedding-) parameter space. In Sect. 2, we describe this method before we apply it to paradigmatic examples in Sect. 3, which include Recurrence Analysis, nearest-neighbor-based time series prediction and causal analysis based on convergent cross mapping.

2 Method

When embedding a time series, in each embedding cycle a suitable delay, and for multivariate data a suitable time series, has to be chosen. While the final embedding vector is invariant to the order of chosen components, the embedding process, and the used statistics and methods to suggest suitable delays, generally depend on all the previous embedding cycles[1]. It
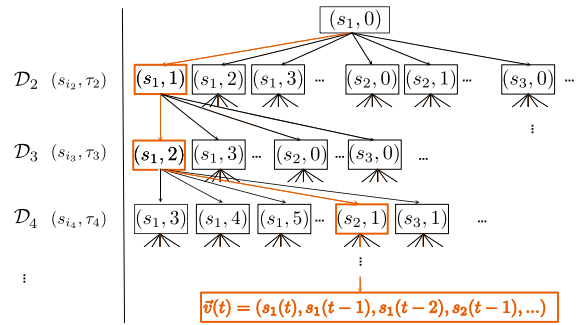

Fig. 1 All possible embeddings of a time series visualized by a tree. Each leaf of the tree symbolizes one embedding cycle $\mathscr{D}_d$ using one selected time series s_{i_d} from the multivariate data set and delay τ_d. Marked in orange is one chosen full embedding

seems therefore natural to visualize all possible embedding cycles in a tree-like hierarchical data structure as shown in Fig. 1. The initial time series s_{i_1} with delay $\tau_1 = 0$ forms the root of the tree, and each possible embedding cycle $\mathscr{D}_d$ is a leaf or node of the tree. With the large amount of possible delays and time series to choose from, this decision tree becomes too large to fully compute it. At the same time, aforementioned statistics like the continuity statistic or conditional mutual information can guide us in pre-selecting potentially suitable delay values and an objective function like the L- or FNN-statistic can pick the most suitable delay value of the pre-selection by quantifying the quality of the reconstruction in each embedding cycle. Throughout this paper, we denote a statistic, which pre-selects potential delay values as Λ_τ and the objective function as Γ. The task to embed a time series can then be interpreted as minimizing $\Gamma(i_1, i_2, .., i_m, \tau_1, \tau_2, ..., \tau_m)$. Visualizing this with a tree as in Fig. 1, we actually perform a tree search to minimize Γ. However, always choosing the leaf of the tree that decreases Γ the most might lead only to a local minimum.

As we strive to find a global minimum and cannot compute the full embedding tree, we proceed by sampling the tree. This approach is inspired by the Monte Carlo Tree Search algorithms that were originally envisioned to master the game of Go [76]. Ultimately computer programs based on these algorithms were able to beat a reigning world champion, a feat that was long

---

[1] The continuity statistic $\langle \varepsilon^\star \rangle(\tau)$ that is used later in this article is one example for such a statistic that depends on all previous embedding cycles.

thought to be impossible for computer programs [77]. Adapting this idea to the embedding problem, we proceed as follows. We randomly sample the full tree, for each embedding cycle we compute the change in the objective function Γ and pick for the next embedding cycle preferably those delays that decrease Γ further. Each node $\mathcal{N}_d$ of the tree encodes one possible embedding cycle and holds the time series used $[s_{i_1}, \ldots, s_{i_d}]$, the delays used until this node $[\tau_1, \ldots, \tau_d]$, i.e., the current *path* through the tree up to node $\mathcal{N}_d$, and a value of the objective function Γ_d. We sample the tree N_{trial}-times in a two-step procedure:

– *Expand:* Starting from the root, for each embedding cycle $\mathcal{D}_d$, possible next steps $(s_{i_j}, \tau_j, \Gamma_j)$ are either computed using suitable statistics Λ_τ and Γ or, if there were already previously computed ones, they are looked up from the tree. We consider the first embedding cycle $\mathcal{D}_2$ and use the continuity statistic $\langle \varepsilon^\star \rangle(\tau)$ for Λ_τ. Then, for each time series s_i the corresponding local maxima of all $\langle \varepsilon^\star \rangle(\tau)$ (for a univariate time series there will only be one $\langle \varepsilon^\star \rangle(\tau)$) that determines the set of possible delay values τ_2 (see the rows in Figs. 1, 2 corresponding to $\mathcal{D}_2$). Then, one of the possible τ_2's is randomly chosen with probabilities computed with a softmax of the corresponding values of Γ_j. Due to its normalization, the softmax function is able to convert all possible values of Γ_j to probabilities with $p_j = \exp(-\beta \Gamma_j)/\sum_k \exp(-\beta \Gamma_k)$. This procedure is repeated (consecutive rows for $\mathcal{D}_3 \ldots$, etc., in Figs. 1, 2) until the very last computed embedding cycle $\mathcal{D}_{m+1}$. This is, when the objective function Γ_{m+1} cannot be further decreased for any of the τ_{m+1}-candidates. Figure 2 visualizes this procedure.
– *Backpropagation:* After the tree is expanded, the final value Γ_m is backpropagated through the taken *path* of this trial, i.e., to all leafs (previous embedding cycles d), that were visited during this expand, updating their Γ_d values to that of the final embedding cycle.
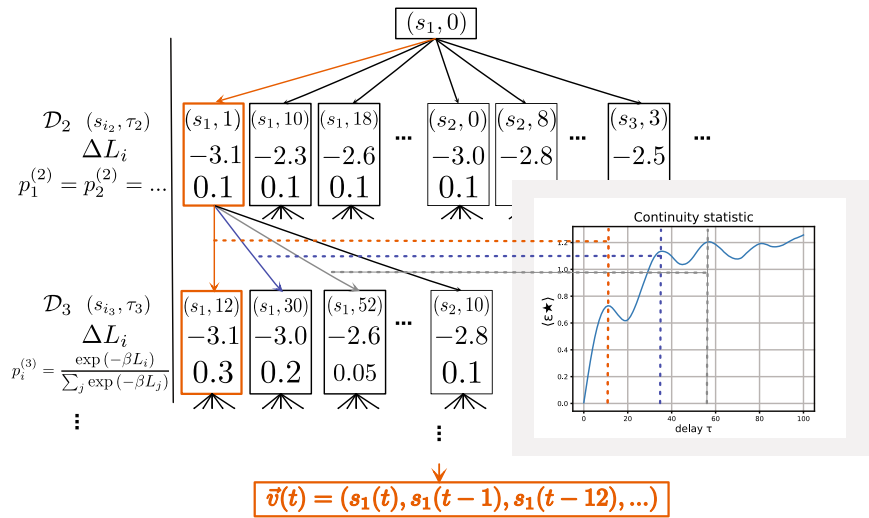
With this two-step procedure, we iteratively build up the part of the tree that leads to embedding with the smallest values for the objective function. The following two refinements are made to improve this general strategy: in case of multivariate time series input, the probabilities are chosen uniformly random in the zeroth embedding cycle $\mathcal{D}_1$. This ensures an even sampling over the given time series, which can all serve as a valid first component of the final reconstruction vectors. Additionally, as soon as a Γ_j is found that is smaller than the previous global minimum, this embedding cycle is directly chosen and not randomized via the softmax function. This also means that for the very first trial always the smallest value of Γ_j is chosen, resulting in a good starting point for the further Monte Carlo search of the tree. In case, the continuity statistic $\langle \varepsilon^\star \rangle(\tau)$ is used as the delay pre-selection statistic Λ_τ and the ΔL-statistic [29] as the objective function Γ, the first sample thus is identical to the PECUZAL algorithms [41] and every further sample improves upon this embedding further minimizing ΔL. Aside from the choice of Λ_τ and Γ, the two hyperparameters of the method are the number of trials N_{trials} and the β parameter of the probability distribution choosing the next delay value. The parameter β governs how likely it is that the minimum of all Γ_i is chosen, i.e., in the extreme cases for $\beta = 0$ the possible delay times are chosen uniformly random and for $\beta \to \infty$ always the smallest Γ_i is chosen. For the tree search algorithms, this means that β governs how "wide" the tree search is, larger β values search the tree more along the already found previously found minima, whereas for smaller values the tree search will stress previously unvisited paths through tree stronger. The default value for β which is used in all shown results is $\beta = 2$.

The computational complexity of this algorithm obviously scales with the number of trials N_{trials}, even though already computed embedding cycles are not computed again in later trials. When sampling the tree many times, the path through the tree of the first few embedding cycles will likely often be the same as that of previous trials. In these cases, computing the delay-preselection and objective function will be identical to that of previous trials. All the values of possible delays and values of the objective function that are computed in previous trials are saved during the tree search and are reused when the same embedding cycle needs to be computed again.

Otherwise, the complexity depends on the chosen delay pre-selection function Λ_τ and the objective function Γ. It has to be clear that the algorithm is computationally much more demanding than a classical TDE. However, once an embedding is computed for a specified system, it can be reused in later applications.

Fig. 2 Visualization of the expand step of the MCDTS algorithm. Here, we exemplary use the continuity statistic $\langle \varepsilon^\star \rangle (\tau)$ as the delay pre-selection statistic Λ_τ and the ΔL-statistic [29] as the objective function Γ, as it has been utilized in the recently proposed PECUZAL algorithm [41]

3 Applications

In this section, we present the potential of the proposed MCDTS method by various applications. Here, we aim to provide suggestions and show that there are a number of state-space based applications that directly benefit from our method or provide better results than with the state-of-the-art embedding techniques. A variety of applications are presented to support the fact that different research questions elicit different embedding behavior and that our proposed method is able to optimize the embedding with respect to different study objectives. In particular, we investigate the influence of the state space reconstruction parameters on a recurrence analysis of the chaotic Lorenz-96 system (Sect. 3.1), a nearest-neighbor time series prediction for the chaotic Hénon map and for a palaeoclimate dataset (Sects. 3.2, 3.3), and last but not least, a causal analysis of two physical observables of a combustion process (Sect. 3.4). The selected applications cover many areas of nonlinear time series analysis, and it is not our intention here to propose new techniques for prediction or causal analysis which are necessarily superior to other, alternative approaches. We rather chose well established state-space-based methods and use them to show how our proposed method optimizes results with respect to the chosen embedding.

3.1 Recurrence properties of the Lorenz-96 system

At first, we consider a potentially higher dimensional nonlinear dynamical system and compare the recurrence properties of its dynamics as derived from the original set of system variables with such by applying the different embedding approaches. We utilize the Lorenz-96 system [79], a set of N ordinary first-order differential equations

$$\frac{dx_i}{dt} = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F, \qquad (2)$$

with x_i being the state of the system of node $i = 1, \ldots, N$ and it is assumed that the total number of nodes is $N \geq 4$. We can think of this system as a ring-like structure of N coupled oscillators— each representing some atmospheric quantity—all connected to the same forcing. The forcing constant F serves as the control parameter. Here, we vary F from $F = 3.7$ to 4.0 in steps of 0.002 covering limit cycle dynamics as well as chaos. We set $N = 8$, randomly choose the initial condition to $u_0 = [0.590; 0.766; 0.566; 0.460; 0.794; 0.854; 0.200; 0.298]$, and use a sampling time of $\Delta t = 0.1$. By discarding the first 2500 points of the integration as transients, we get time series consisting of 5000 samples for each of the encountered values of F. We focus on two scenarios: (1) only the time series of the 2nd node (univariate embedding) and (2) three time series of nodes 2, 4, and 7 are used to mimic a uni- and a multivariate embedding case. For each of these time series, we perform an embedding, using three classic time delay approaches as proposed by Kennel et al. [69] (5%-

threshold), Cao [66] (slope threshold of 0.2), and Hegger and Kantz [67] (5%-threshold) with a uniform delay value estimated as the first minimum of the auto-mutual information (only applicable to the univariate case) and the recently proposed PECUZAL algorithm [41]. For our proposed MCDTS approach, we embed the data using the continuity statistic $\langle \varepsilon^\star \rangle(\tau)$ as the delay pre-selection statistic Λ_τ. For the objective function Γ, we try two different approaches, namely the ΔL-statistic [29] (MCDTS-C-L) as well as the FNN-statistic [67] (MCDTS-C-FNN). In all approaches, we discard serially correlated points from the nearest neighbor search by setting a Theiler window [80] to the first minimum of the mutual information. An overview over all MCDTS implementations and abbreviations is given in Table 1.

By varying the control parameter F, the system varies its dynamics which is well represented by a change in the recurrence behavior [81]. In previous work, we have demonstrated that recurrence quantification analysis (RQA) can be used to qualitatively characterize the typical dynamical properties of the Lorenz-96 system such as chaotic or periodic dynamics [82]. We, therefore, compare the recurrence properties of all reconstructed trajectories to recurrence properties of the true trajectory (obtained from the numerical integration) by using RQA. The neighborhood relations of a trajectory can be visualized in a recurrence plot (RP), a binary, square matrix $\mathbf{R}$ representing the recurrences of states $\mathbf{x}_i$ ($i = 1, \ldots, N$, with N the number of points forming the trajectory) in a d-dimensional (optionally reconstructed) state space [83,84]

$$R_{i,j}(\varepsilon) = \Theta\left(\varepsilon - \|\mathbf{x}_i - \mathbf{x}_j\|\right), \qquad \mathbf{x} \in \mathbb{R}^d, \qquad (3)$$

with $\|\cdot\|$ a norm, ε a recurrence threshold, and Θ the Heaviside function. There are numerous ideas of how to quantify a RP [84,85]. Some statistics are based on the distribution of recurrence points, some on the diagonal line structures, some on the vertical structures, and it is also possible to use complex-network measures, when interpreting $\mathbf{R}$ (subtracting the main diagonal) as an adjacency matrix $\mathbf{A} = \mathbf{R} - \mathbb{1}$ of a recurrence network (RN) [86]. Some of these quantifiers are related to dynamical invariants [87,88].

For our purpose of comparing different aspects of recurrence properties of original and reconstructed trajectories, we use the transitivity (*TRANS*) of the ε-RN, the determinism (*DET*), the mean diagonal line length (L_{mean}), the maximum diagonal line length ($L_{\max}$) and its reciprocal (*DIV*), the entropy of diagonal line lengths (*ENTR*), the *TREND*, the mean recurrence time (*MRT*), the recurrence time entropy (*RTE*), and finally, the joint recurrence rate fraction (*JRRF*). JRRF measures the accordance of the recurrence plot of the (true) reference system, $\mathbf{R}^{\mathrm{ref}}$ with the RP of the reconstruction, $\mathbf{R}^{\mathrm{rec}}$.

$$\mathrm{JRRF} = \frac{\sum_{i,j}^N JR_{i,j}}{\sum_{i,j}^N R_{i,j}^{\mathrm{ref}}}, \quad \mathrm{JRRF} \in [0, \ 1] \qquad (4)$$

$$\mathbf{JR} = \mathbf{R}^{\mathrm{ref}} \circ \mathbf{R}^{\mathrm{rec}}. \qquad (5)$$

We compute both, $\mathbf{R}^{\mathrm{ref}}$ and $\mathbf{R}^{\mathrm{rec}}$, by fixing the recurrence threshold corresponding to a global recurrence rate (RR) of 5% in order to ensure comparability [89]. Although the quantification measures depend crucially on the chosen recurrence threshold, the particular choice we make here is not so important, since we apply it to all RPs we compare. $RR = 5\%$ ensures a proper resolution of the inherent structures to be quantified by the ten aforementioned measures.

The described procedure is schematically illustrated in Fig. 3. For each reconstruction method and for each of the ten RQA-statistics, the mean squared error (MSE) with respect to the RQA-statistics of the true reference trajectory is computed (normalized to the reference RQA-values). The pairwise comparison of the MSEs is evaluated as the percentage of the ten RQA-MSEs, which take a lower MSE (Fig. 4). For instance, a value of 70% in the table indicates that for seven out of the ten considered RQA-quantifiers the normalized mean squared error for the reconstruction method displayed on the y-axis is *lower* than for the reconstruction method displayed on the x-axis. The m-notation indicates the multivariate embedding approach, where three instead of one time series have been passed to the reconstruction methods ($x_2(t)$, $x_4(t)$, and $x_7(t)$, see Fig. 3). Since the classic TDE algorithms from Cao, Kennel et al., and Hegger & Kantz are not able to handle multivariate input data, only PECUZAL and the proposed MCDTS-idea combined with the L-statistic and with the FNN-statistic are considered in the multivariate scenario. The superiority over the three classic TDE methods is discernible in values > 50% for PECUZAL and MCDTS in the first three columns. While we would expect a better reconstruction for the multivariate cases—because we simply provide more information—our proposed method also performs better in the univariate case when the FNN-statistic is

Table 1 The different implementations of the MCDTS algorithm used throughout the article and their choice of delay-preselection and objective function, which is minimized through the tree search

Notation	Delay pre-selection method Λ_τ	Objective function Γ
MCDTS-C-FNN	Maxima of continuity statistic $\langle \varepsilon^\star \rangle(\tau)$ [65,78]	FNN [67–69]
MCDTS-C-L	Maxima of continuity statistic $\langle \varepsilon^\star \rangle(\tau)$	$L/\Delta L$-statistic [29,41]
MCDTS-C-MSE	Maxima of continuity statistic $\langle \varepsilon^\star \rangle(\tau)$	Mean squared prediction error
MCDTS-C-MSE-KL	Maxima of continuity statistic $\langle \varepsilon^\star \rangle(\tau)$	Mean Kullback-Leibler divergence of true and predicted
MCDTS-C-CCM	Maxima of continuity statistic $\langle \varepsilon^\star \rangle(\tau)$	Negative CCM-correlation coefficient [18]
MCDTS-R-MSE-KL	Given range of delay values	Mean Kullback-Leibler divergence of true and predicted
MCDTS-R-MSE	Given range of delay values	Mean squared prediction error

Fig. 3 Schematic visualization of the data analysis for the Lorenz-96 system, Eq. (2) (see text for details). In case of the univariate approach, the $x_2(t)$-time series gets embedded by all considered reconstruction methods, for the multivariate approach, three time series ($x_2(t)$, $x_4(t)$ and $x_7(t)$) are passed to the reconstruction algorithms. From the reconstructed attractors, we obtain a recurrence plot and quantify it (RQA) by using ten different quantifiers. The same is done for the reference trajectory gained from all 8 time series from the numerical integration. Repeating the analysis for time series corresponding to varying values of the control parameter F of the system, we finally obtain time series of the RQA-quantifiers for each reconstruction method as well as for the *true* trajectory

used as an objective function. When using MCDTS with the L-statistic, there is hardly any improvement discernible, while the computational costs are magnitudes higher. Here, PECUZAL reveals better results, even though it uses the same statistics. However, combined with the FNN-statistic our proposed idea performs very well in the univariate case and reveals excellent results for the multivariate case.
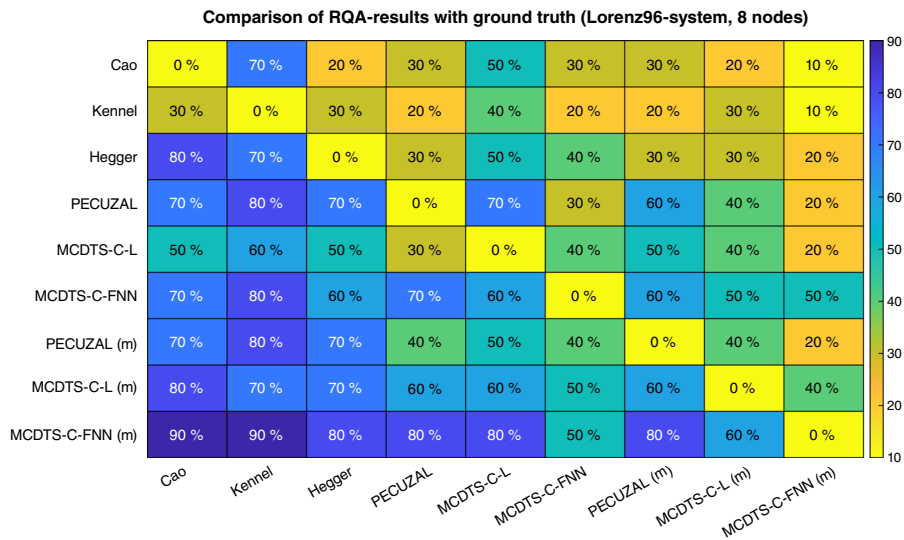
Fig. 4 Results of the analysis of the Lorenz-96 system with varying control parameter and for all considered reconstruction approaches (see Table 1 for notations). Shown is the pairwise comparison of the normalized mean squared error of all considered ten RQA-quantifiers with respect to the truth RQA-time series (see text for details). For instance, a value of 70% in the table indicates that for seven out of the ten considered RQA-quantifiers the normalized mean squared error for the reconstruction method displayed on the y-axis is *lower* than for the reconstruction method displayed on the x-axis

3.2 Short time prediction of the Hénon map time series

In the following, a state space reconstruction $\mathbf{v}(t)$ of a single time series $s(t)$ is used to further predict its course. Besides a very recent idea [90] to train neural ordinary differential equations on a reconstructed trajectory, which then allows prediction, several attempts have been published [10–16] which more or less rely on the same basic idea. For the last vector of the reconstructed trajectory, denoted with a time-index l, $\mathbf{v}(t_l)$, a nearest neighbor search is performed. Then, these neighbors are used to predict the future value of this point T time steps ahead, $\mathbf{v}(t_{l+T})$. Knowledge of the used embedding, which led to the reconstruction vectors $\mathbf{v}(t)$, then allows to read the prediction of the time series $s(t_l + T)$ from the predicted reconstruction vector $\mathbf{v}(t_{l+T})$. Usually, $T = 1$, i.e., the forecast is iteratively build by appending $\mathbf{v}(t_{l+T})$ to the trajectory $\mathbf{v}(t_i)$, $i = 1, \ldots, l$, and this procedure is repeated N times, in order to obtain an N-step prediction. The aforementioned approaches differ from the way they construct a local model of the dynamics based on the nearest neighbors. For instance, Farmer and Sidorowich [11] proposed a linear approximation, i.e., a linear polynomial is fitted to the pairs $(\mathbf{v}(t_{nn_i}), \mathbf{v}(t_{nn_i+T}))$, where nn_i

denotes the ith nearest neighbor time-index. Sugihara and May [16] used a simplex with minimum diameter to select the nearest neighbor indices nn_i and projected this simplex T steps into the future. The prediction is then being made by computing the location of the original predictee $\mathbf{v}(t_l)$ within the range of the projected simplex, "giving exponential weight to its original distances from the relevant neighbors." Here, a much simpler idea is considered: a *zeroth-order approximation* of the local dynamics. The prediction is simply the projection of the nearest neighbor of $\mathbf{v}(t_l)$, denoted by the index nn_1, $\mathbf{v}(t_{l+T}) = \mathbf{v}(t_{nn_1+T})$. It is clear that the performance of all prediction approaches based on an approximation of the local dynamics by making use of nearest neighbors will crucially depend on the length of the training set. By training set, we mean the time series $s(t)$, which has been used to construct the trajectory $\mathbf{v}(t)$. We hypothesize that the accuracy of such a prediction will also depend on the reconstruction method, especially when the training set is rather short (Small and Tse [43] and also Bradley and Kantz [73]). In particular, Garland and Bradley [74] have shown that accurate predictions can be achieved with the aforementioned *zeroth-order approximation* when using an incomplete embedding of the data, i.e., reconstructions

that do not satisfy the theoretical requirements on the embedding dimension in Takens' sense.

As a proof of concept, we now use the described nearest-neighbor prediction method to predict the x-time series of the Hénon map [91], even though other simple models like low order polynomial models might be superior for such noise-free and pure deterministic dynamics (we provide a more challenging example in Sect. 3.3). The time series $x_{i+1} = y_i + 1 - ax_i^2$ and $y_{i+1} = bx_i$, with standard parameters $a = 1.4$, $b = 0.3$ and 100 randomly chosen different initial conditions are used. For each of those 100 samples x- and y-time series of length $N = 10{,}030$ are obtained (transients removed). The first 10,000 points of the time series are used for state space reconstruction (both time series for the multivariate cases, only the x-time series in the univariate case), while the last 30 points are the prediction test set (only the x-time series is predicted). The same reconstruction methods as in Sect. 3.1 are used, but for MCDTS we try two different delay pre-selection statistics Λ_τ. Rather than only considering the continuity-statistic (denoted as C in the model description) we also look at a whole range of delay values $\tau = 0, \ldots, 50$ (denoted as R in the model description). For the objective function Γ, we try

- The ΔL-statistic (denoted as L in the model description),
- The FNN-statistic (denoted as *FNN* in the model description),
- The root mean squared in-sample one-step prediction error on the first component of the reconstruction vectors, i.e., the x-time series (denoted as *MSE* in the model description), and finally
- The mean Kullback–Leibler-distance of the in-sample one-step prediction and the "true" trajectory points (denoted as *MSE-KL* in the model description).

By "in-sample," we mean the training set, which is used for the reconstruction. For all MCDTS implementations and abbreviations, see again Table 1. The accuracy of the prediction is evaluated by the normalized root-mean-square forecast error (*RMS*),

$$e_{\text{rms}}(T) = \frac{\sqrt{\left\langle \left[x_{\text{pred}}(T) - x_{\text{true}}(T) \right]^2 \right\rangle}}{\sqrt{\left\langle \left[x_{\text{true}}(T) - \langle x_{\text{true}}(T) \rangle \right]^2 \right\rangle}}$$

with index *true* denoting the test set values. This way $e_{\text{rms}}(T) = 0$ indicates a perfect prediction, whereas

$e_{\text{rms}}(T) \approx 1$ means that the prediction is not better than a constant mean-predictor of the test set. Figure 5 shows the mean forecast accuracy for the traditional TDE methods (Cao, Kennel et al., Hegger & Kantz) and two selected MCDTS approaches as a function of the prediction time. The largest Lyapunov exponent is estimated to $\lambda_1 \approx 0.419$, and we display *Lyapunov times* on the x-axis, i.e., units of $1/\lambda_1$. As in Sect. 3.1, m indicates the multivariate case, in which both, x- and y-time series are fed into the reconstruction algorithms. The results for all discussed reconstruction methods can be found in Appendix A (Fig. 8). As expected, the forecast accuracy is worse in case of added white noise (Fig. 5B) and the predictions based on multivariate reconstructions perform slightly better. The MCDTS-based forecasts perform significantly better than the forecasts based on the traditional TDE methods. Even though the continuity statistic constitutes a reasonable delay pre-selection statistic with a clear physical meaning, when utilized in our MCDTS approach (MCDTS-C-), it performs not as good as if we would not pre-select delays on the basis of some statistic, but try delays in a whole range of values ($\tau \in [0, 50]$, MCDTS-R-). At least, this statement holds for this example of the Hénon map time series.

A Wilcoxon rank sum test is applied to underpin the better performance of the MCDTS-approaches in comparison with the classical time delay methods. Therefore, we define a threshold $\zeta = 0.1$ and compute the prediction times for which $e_{\text{rms}}(T)$ first exceeds ζ for all trials and for all considered reconstruction methods. These distributions of prediction times for each method are used for the statistical test with the null hypothesis that two considered distributions have equal medians. The tests complement the visual analysis of Figs. 5 and 8. A significantly better forecast performance ($\alpha = 0.01$) than the classic time delay embedding methods for PECUZAL and all considered MCDTS-based approaches, but the ones combined with the FNN-statistic (MCDTS-FNN) can be verified for the noise free case. In the case of the noise corrupted time series PECUZAL (m), all MCDTS-MSE-approaches and MCDTS-C-L (m) achieve a significantly better prediction performance than the classical time delay methods.

Some remarks: Together with PECUZAL (m) and MCDTS-R-MSE (m), MCDTS-C-L (m) achieves the overall best results (Fig. 8). The choice of the threshold ζ is obviously subjective, but a range of thresholds gave similar results and the "grouping" of the results accord-
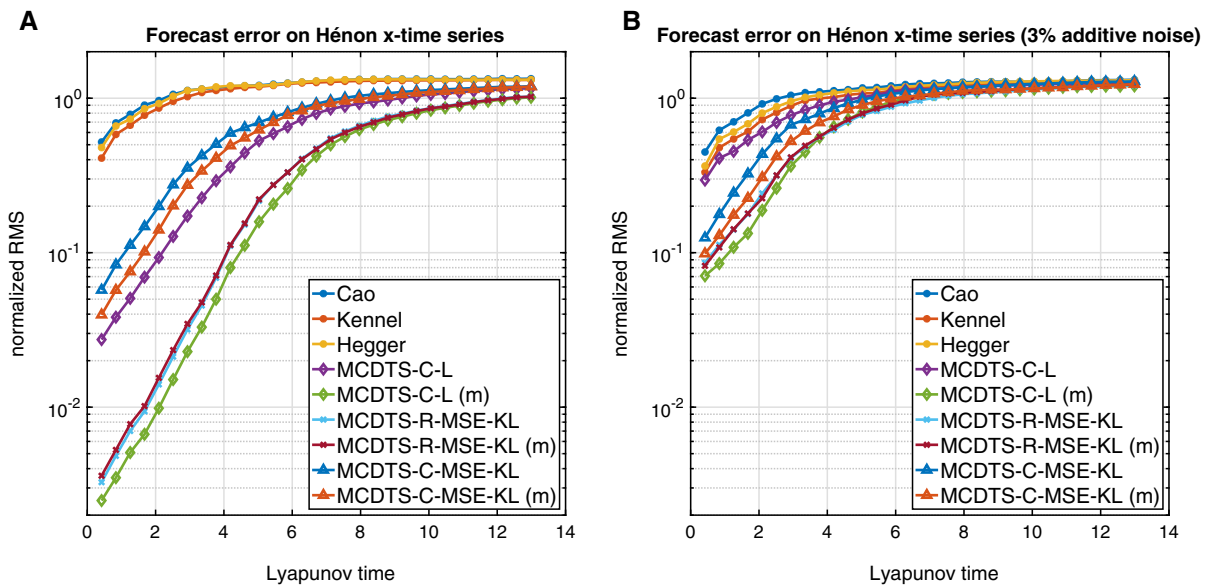
A

B

Fig. 5 **A** Normalized root-mean-square prediction errors (*RMS*) for the Hénon *x*-time series and for selected reconstruction methods (see Fig. 8 for all mentioned approaches and Table 1) as a function of the prediction time. Shown are mean values of a distribution of 100 trials with different initial conditions. For the prediction, we use a one step ahead zeroth-order approximation on the nearest neighbor of the last point of the reconstructed trajectory and iteratively repeated that procedure 30 times in order to obtain a prediction of 31 samples in total for each trial. **B** Same as in **A** but with 3% additive white noise

ing to the different techniques is clearly discernible already when looking at the mean (Figs. 5, 8). We have to mention that we could not achieve results as shown here for continuous systems like the Lorenz-63 or the Rössler model. In those cases, the difference in the prediction accuracy was not as clear as it is in the Hénon example and not significant, for both, noise-free and noise corrupted time series. We also investigated the influence of the time series length of the training sets, but the results did not change much. All reconstruction methods gave similar prediction results. We could, however, observe that simple and incomplete embeddings, i.e., a too low embedding dimension, often—but not always—led to similarly good prediction results, when compared to "full" embeddings. This was true for the continuous examples (not shown in this work), but this also holds for the Hénon example shown here, where the MCDTS-C-L approach does not yield the best results in the univariate case, although it targets the total minimum of the *L*-objective-function, which the authors consider to be a suitable cost-function for a good/full embedding. These observations are in line with the findings of Garland and Bradley [74] and the fact that our reconstruction methods tend to suggest higher dimensional embeddings with smaller delays

in the presence of noise support the findings of Small and Tse [43]. The FNN-statistic does not seem to be useful in the prediction application shown here, since all approaches which make use of it (including classic TDE) perform clearly worse compared to the other methods used.

3.3 Improved short-time predictions for CENOGRID

To demonstrate that the prediction procedure from the preceding section works for real, noisy data, we apply it to the recently published CENOzoic Global Reference benthic foraminifer carbon and oxygen Isotope Dataset (CENOGRID) [92]. The temperature-dependent fractionation of carbon and oxygen isotopes in benthic foraminifera is an important means to reconstruct past global temperatures and environmental conditions. Moreover, the Cenozoic is interesting, because it provides an analogue of future greenhouse climate and how and which regime shifts in large-scale atmospheric and ocean circulation can be expected in the future warming climate. Predicting these data may be unrealistic and not motivated by an actual research question. However, this task shall serve as a proof of concept. The non-stationarity and noise level of CENOGRID make prediction particularly difficult.

The dataset consists of a detrended $\delta^{18}O$ and a detrended $\delta^{13}C$ isotope record with a total length of $N = 13,421$ samples and a sampling period of $\Delta t = 5000$yrs (Fig. 9 in Appendix B). Here, we make predictions on the $\delta^{13}C$ isotope record. The first 13,311 samples are used as a training set, from which state space reconstructions are obtained. The remaining 110 samples of the $\delta^{13}C$ record act as the test set. For 100 different starting points in the test set, we make 10-step-ahead predictions for each reconstruction method by using the embedding parameters gained from the training and with the iterative *zeroth-order approximation* prediction procedure described in Sect. 3.2. This way we simulate different initial conditions for the prediction and obtain a distribution of forecasts for each reconstruction method. We again use a Wilcoxon rank sum test on these distributions in order to see whether predictions based on some reconstruction method are significantly better than the predictions obtained from classic TDE (Cao, Kennel et al., Hegger & Kantz). Only one of the applied reconstruction methods (listed in Table 1), MCDTS-R-MSE (m), score significantly better predictions (highly significant for prediction horizons up to $4\Delta t$ and significant for prediction horizon up to $5\Delta t$). Figure 6A shows the mean normalized root mean square prediction error gained from the 100 predictions for the classic TDE and the mentioned MCDTS-R-MSE (m). The distribution of all prediction trials for the best performing classic TDE method (Hegger & Kantz) and for MCDTS-R-MSE (m) is shown in panels B, C. Even though the multivariate approach MCDTS-R-MSE (m) could have been used both, the $\delta^{18}O$ and the $\delta^{13}C$ time series for the reconstruction, it only uses $\delta^{13}C$ lagged by 1 and 2 samples in a three-dimensional reconstruction. The classic TDE methods and all other reconstruction methods (listed in Table 1, not shown in Fig. 9) revealed higher dimensional embeddings (Table 2). Yet, all these higher dimensional reconstructions give poor prediction results, except for MCDTS-C-MSE-KL (m), which gives significant better predictions ($\alpha = 0.05$) than the classic TDE methods at least for the one-step-ahead prediction.

3.4 Estimating causal relationship of observables of a thermoacoustic system

As a final proof of concept, we utilize state space reconstruction for detecting causality between observables

X and Y in a turbulent combustion flow in a gas turbine. It is possible to infer a causal relationship between two (or more) time series $x(t)$ and $y(t)$ via *convergent cross mapping (CCM)* [18,19,93], which—in contrast to Granger causality [94]—also works for time series stemming from nonseparable systems, i.e., deterministic dynamical systems. The CCM method "tests for causation by measuring the extent to which the historical record of Y values can reliably estimate states of X. This happens only if X is causally influencing Y." [18] This also incorporates the embedding theorems [1–3] in a sense that a state space reconstruction based on $x(t)$ is diffeomorphic to a reconstruction of $y(t)$, if $x(t)$ and $y(t)$ describe the same dynamical system *and* the embedding parameters have been chosen correctly. To check for a causal relationship from $X \rightarrow Y$, a state space reconstruction of $y(t)$ yields a trajectory $\mathbf{v}_y(t) \in \mathbb{R}^m$, with m denoting the embedding dimension, which is then used for estimating values of $x(t)$, namely $\hat{x}(t)$. It is said that $\mathbf{v}_y(t)$ *cross-maps* $x(t)$, in order to get estimates $\hat{x}(t)$. Technically, this is done by first searching for $m + 1$ nearest neighbors of a point corresponding to a time index $t' \in t$, i.e., find the $m + 1$ time indices t'_{NN_i}, $i = 1, \ldots, m + 1$ of the nearest neighbors of $\mathbf{v}_y(t')$. Further, these time indices t'_{NN_i} are used to "identify points (neighbors) in X (a putative neighborhood) to estimate $x(t')$ from a locally weighted mean of the $m + 1$ $x(t'_{NN_i})$ values" [18]:

$$\hat{x}(t') = \sum w_i x(t'_{NN_i}), \qquad i = 1, \ldots, m + 1, \qquad (6)$$

with the weighting w_i based on the nearest neighbor distance to $\mathbf{v}_y(t')$.

$$w_i = u_i / \sum u_j, \qquad j = 1, \ldots, m + 1 \qquad (7)$$

$$u_i = \exp\left[-\|\mathbf{v}_y(t') - \mathbf{v}_y(t'_{NN_i})\| / \|\mathbf{v}_y(t') - \mathbf{v}_y(t'_{NN_1})\|\right] \qquad (8)$$

with $\|\cdot\|$ a norm (we used Euclidean distances). Finally, the agreement of the cross-mapped estimates $\hat{x}(t')$ with the true values $x(t')$ is quantified for all considered $t' \in t$, e.g., by computing a linear Pearson correlation ρ_{CCM}, which has been done in this study. The clou is that the estimation skill, here represented by ρ_{CCM}, increases with the considered amount of data used, if X indeed causally influences Y. This is because the attractor—represented by the reconstruction vectors $\mathbf{v}_y(t)$—gets resolved better with increasing time
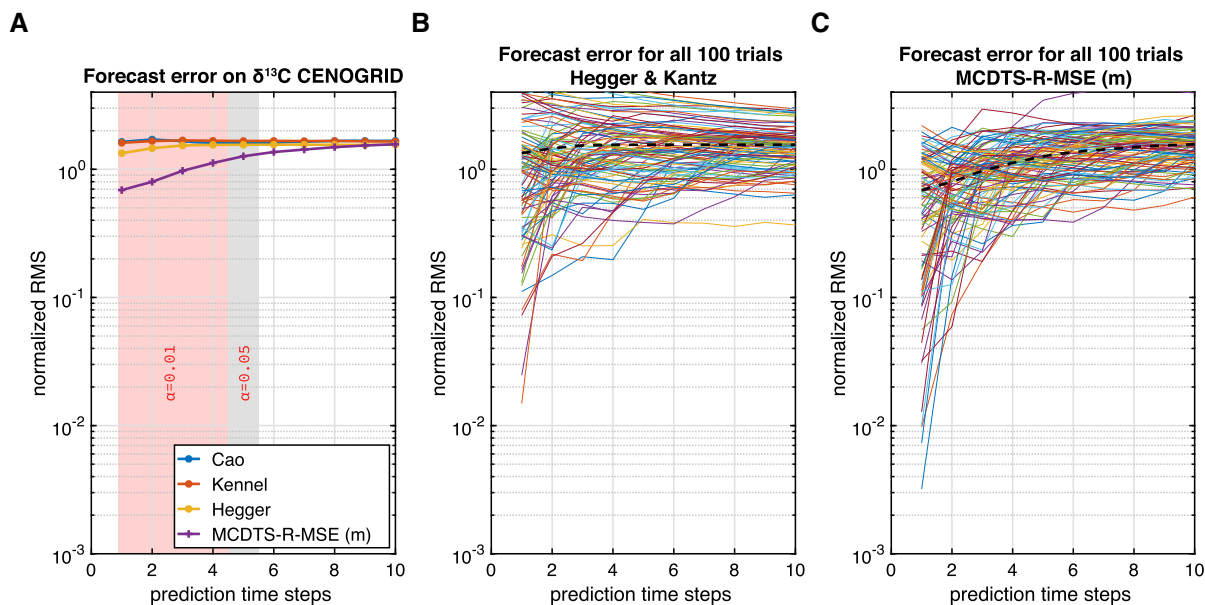
Fig. 6 A Mean normalized root mean square prediction error for four selected reconstruction methods on the δ^{13}C CENOGRID record. **B** Prediction error for all 100 trials for the classic TDE method of [67] (yellow line in panel A). **C** Prediction error for all 100 trials for the MCDTS-R-MSE (m) method (purple line in panel A). The forecasts based on this method are significantly better than for all three classic TDE methods (up to 4 prediction time steps under a significance level $\alpha = 0.01$ and up to 5 prediction time steps under a significance level $\alpha = 0.05$)

series length, resulting in closer nearest neighbors and therefore a better concordance of $\hat{x}(t)$ and $x(t)$, i.e., an increase in ρ_{CCM} with increasing time series length. This *convergence* of the estimation skill based on cross-mapping is a necessary condition for causation, not only a high value of ρ_{CCM} itself (Fig. 7A). Although the embedding process is key to a successful application of CCM to data, its influence has not been discussed by Sugihara et al. [18]. However, Schiecke et al. [95] discussed the impact of the embedding parameters on CCM briefly and we hypothesize that the embedding method can play a crucial role, when analyzing real-world data. Therefore, we utilize the MCDTS framework in the following way. As a delay pre-selection method Λ_τ, we use the reliable continuity statistic $\langle \varepsilon^\star \rangle (\tau)$ [65,78]. As a suitable objective function Γ, we use the negative of the corresponding ρ_{CCM}, i.e., MCDTS optimizes the embedding with respect to maximizing ρ_{CCM} of two given time series. According to our abbreviation-scheme given in Table 1, we will refer to this approach as MCDTS-C-CCM.

We apply the CCM-method to time series data that spans the different dynamical regimes of a thermoa-

coustic system. Here, we investigate the mutual causal influence of two recorded variables of the thermoacoustic system, namely the pressure and the heat release rate fluctuations (Fig. 10). The original experiments were performed on a turbulent combustor with a rectangular combustion chamber (length 700 mm, cross-section 90 mm × 90 mm, Fig. 11). In such a combustion experiment, a fixed vane swirler is used to stabilize the flame and a central shaft that supports the swirler injects the fuel through four radial injection holes. The fuel used is liquefied petroleum gas (60% butane and 40% propane). The airflow enters through the inlet to the combustion chamber. The partially premixed reactant mixture is ignited using a spark plug. Once the flame is established in the combustor, we continuously varied the control parameter (mass flow rate of air, which, in turn, varies the Reynolds number[2] and the equivalence ratio[3]) to observe the dynamical transitions in the

---

[2] Reynolds number is $\frac{\rho U D}{\mu}$, where ρ is the density, U is a characteristic velocity, D is a characteristic dimension (the diameter) and μ is the viscosity.

[3] Equivalence ratio is the ratio between the actual fuel-air ratio to the stoichiometric fuel-air ratio.
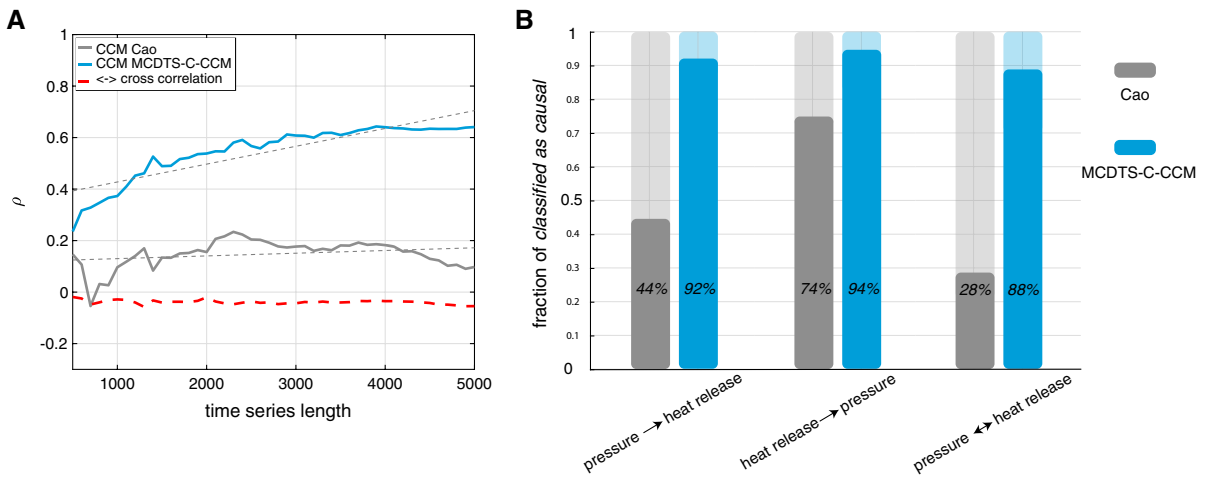
Fig. 7 **A** Linear correlation coefficient of convergent cross mapping (CCM) *heat release → pressure* as a function of the considered time series length for Cao's embedding method (gray) and the proposed MCDTS embedding (blue) exemplary shown for one out of 50 drawn sub-samples of length $N = 5000$ from the entire time series (Fig. 10, c.f. Table 1 for abbreviations). While the dashed black lines show the linear trend for both CCM correlations, the dashed red line shows the Pearson linear correlation between the heat release and the pressure time series, indicating no influence. We ensure convergence of the cross mapping,

and, thus, a true causal relationship, if there is a positive trend in the CCM-correlation over increasing time series length (slope of the dashed black lines) and when the last point of the CCM-correlation (i.e., longest considered time series length) exceeds a value of 0.2 (in the shown case Cao's method does not detect a causal influence of the heat release to the pressure). We test this on all 50 sub-samples for both causal directions. **B** True classified causal relationships as a fraction of all sub-samples based on the embedding of each time series using Cao's method and our proposed MCDTS method

system. Acoustic pressure fluctuations were measured using a piezoelectric transducer (PCB103B02) and heat release rate using a photomultiplier tube (Hamamatsu H10722-01) at a sampling rate of 4 kHz.

The interactions between the turbulent flow, the unsteady fluctuations of the flame due and the acoustic field of the chamber lead to different dynamical states. As the airflow rate increases, the system transitions from a state of stable operation (which comprises high dimensional chaos having low amplitude [96]) to intermittency, a state that comprises bursts of periodic oscillations amid epochs of aperiodicity [97], and then to limit cycle [98]. The self-sustained limit cycle oscillations represent a state of oscillatory instability, known as thermoacoustic instability [99]. When the flow rate of air is further increased, the flame loses its stability inside a combustor and blows out. The pressure and heat release rate data capture the transition through all these dynamical states in sequence. In the many different dynamical regimes recorded in the time series, we expect the strength of causal interference between the heat release and the pressure to vary. But in all dynamics, we expect a mutual causal interaction between heat release and pressure. Moreover, since a possible asym-

metric bi-directional coupling between heat release and pressure has been discovered in a stationary setup of a very similar experiment [100], we would also expect that the heat release rate has a slightly stronger causal influence on pressure than vice versa.

In short, the goal here is twofold:

1. Prove the expected mutual causal relationship between heat release rate and pressure as well as
2. the hypothesized asymmetry in its strengths by applying MCDTS-C-CCM on a range of time series, sampled from the entire record (Fig. 10).

We compare it to results obtained from using the CCM method with the classical embedding approach of Cao [66]. Specifically, we set up the following workflow for this analysis:

1. 50 time indices $t' \in t$ are drawn randomly, where t covers the entire record.
2. For each of these indices t', time series of length $N = 5000$ for pressure and heat release are obtained and standardized to zero mean and unit variance (Fig 10).
3. Both time series samples (of full length $N = 5000$) each are embedded using Cao's method

as a classical reference and our proposed framework MCDTS-C-CCM with 100 trials (Table 1). Based on the obtained reconstructions, $\rho_{\text{CCM-Cao}}$ and $\rho_{\text{CCM-MCDTS}}$ are computed for both directions as a function of increasing time series length as exemplary shown in Fig. 7A.

4. To ensure convergence in the CCM-sense, we fit a linear model to ρ_{CCM} (dashed black lines in Fig. 7A) and whenever that model gives a positive slope *and* the last value of ρ_{CCM} (i.e., for the longest considered time series of length $N = 5000$) exceeds a value of 0.2, we infer a true causal relationship.

5. When we can detect a causal relation simultaneously in both directions, we compute the average of the pointwise difference $\rho_{\text{CCM heat}\rightarrow\text{pressure}} - \rho_{\text{CCM pressure}\rightarrow\text{heat}}$

The minimum considered value of 0.2 for ρ_{CCM} is an arbitrary and subjective choice, and we could have made other choices. But since this procedure is applied to $\rho_{\text{CCM-Cao}}$ and $\rho_{\text{CCM-MCDTS}}$ at the same time, we think this is reasonable and it prevents samples to be accounted for as "true causal" when there is near-0 ρ_{CCM}, but a positive linear trend. Results do only change slightly when varying this value in some interval [0.2 0.3]. Figure 7B summarizes the results obtained for both considered embedding methods. Shown are the classification results for correctly deducing a causal influence of pressure on heat release (left panel) and of heat release on pressure (middle panel) based on our definition (item 4 in the list above). Thus, in this first step, we do not measure the strength of the causal relationship, but rather test whether such a relationship actually exists. While MCDTS-C-CCM maintains a correct classification in 92% of all cases considered (50 samples) for *pressure → heat release* and 94% for *heat release → pressure*, Cao's method is only able to correctly classify 44% and 74%, respectively. These results themselves already demonstrate a clear advantage of our proposed method, but recall that we expect a causal relationship between heat release *and* pressure simultaneously for each sample. The right panel of Fig. 7B reveals that in 88% of all cases considered, MCDTS-C-CCM is able to detect a mutual causal relationship, while Cao's method managed to do so in only 28% of the cases.

Furthermore, we try to validate a hypothesis made by Godavarthi et al. [100] that heat release has a stronger effect on pressure than vice versa for most of the consid-

ered dynamics. The problem of measuring the strength of a causal relationship is twofold: First, the experiment considered here exhibits a number of different dynamics due to the continuously changing control parameter. The hypothesis of an asymmetry in the strength of the interaction was made for stationary cases and four considered dynamics the authors investigated. Second, in describing the CCM method, Sugihara et al. [18] merely described that in the case of a stronger causal effect of X on Y, cross-mapping X with $\mathbf{v}_y$ converges *faster* than the other way around. Thus, we would have to define what *faster* means with respect to our experimental curves like the ones shown in Fig. 7A. That would mean introducing some parameters on which the results would depend too much. Here, we pursue a simpler idea in order to detect the strength of a causal interaction. For samples where a causal relation in *both* directions has been detected, we compute the average of the pointwise difference of the CCM-correlation coefficients, i.e., $\Delta\rho_{\text{CCM}} = \rho_{\text{CCM heat}\rightarrow\text{pressure}} - \rho_{\text{CCM pressure}\rightarrow\text{heat}}$. When this difference is positive, we claim that heat release stronger effects pressure in a causal sense than vice versa. Our analysis reveals that the proposed method is able to reflect the hypothesized stronger causal effect of the heat release on pressure data. Figure 12 shows that for 29 of the 50 samples ($\sim$ 58%) $\Delta\rho_{\text{CCM}}$ is indeed positive. Using the Cao method, we were able to derive such a result in only $\sim$ 26% of all samples. In this case, however, only $\sim$ 28% of the samples were found to be mutually causally related at all (cf. Fig. 7B). Within the group of mutually causally related samples, the assumed asymmetry is reflected very well (13 of 14 mutually causally related samples had a positive $\Delta\rho_{\text{CCM}}$).

The proposed MCDTS reconstruction approach shows a clear advantage when using it together with the CCM method. Not only is the general classification ability remarkable, but the MCDTS reconstructions also allow verification of an assumed asymmetric causal interaction, which would be limited by the classical time delay method.

4 Conclusions

A novel perspective of the embedding process has been proposed, in which the state space reconstruction from single time series can be treated as a game, in which each move corresponds to an embedding cycle and is subject to an evaluation through an objective function.

It is possible to model different embeddings, i.e., different choices of delay values and time series (if there are multivariate data at hand) in the embedding cycles, in a tree like structure. Consequently, our approach randomly *samples* this tree, in order to ensure the finding of a global minimum of the chosen objective function. We leave it to practitioners which state space evaluation statistic, i.e., objective function, they use, since different research questions require different reconstruction approaches. There is also a free choice of a delay pre-selection method for each embedding cycle, e.g., using the minima of the auto-mutual information statistic. We recommend the combination of the continuity statistic of Pecora et al. [65] as a delay pre-selection method together with the *L*-statistic of Uzal et al. [29] as an objective function as a very good "all-rounder" for many research questions in nonlinear time series analysis, as already shown by Kraemer et al. [41] (PECUZAL algorithm). Since the sampling of the tree is a random procedure, the proposed idea only yields converging embedding parameters for a sufficient sampling size N_{trial}. In our numerical investigations, $N_{\text{trial}} = 50$ usually led to satisfying results for univariate cases and $N_{\text{trial}} = 80$ for multivariate embedding scenarios. Our proposed method initializes in a local minimum of the objective function, which is achieved by minimizing the objective function in each embedding cycle to the maximum extend. So in practice, even setting N_{trial} too low would lead to similar—but never worse—results as the state-of-the-art methods. Moreover, the proposed method is not limited to delay pre-selection and objective functions that take into account certain physical constraints. It would also optimize the reconstruction vectors of the state space for research questions such as classification, where we could speak of a feature or latent space instead of the state or phase space notation associated with statistical physics. We exemplified the use of such a modular algorithm by combining different objective- and delay pre-selection functions. Its superiority to classical time delay embedding methods has been demonstrated for a recurrence analysis of the Lorenz-96 system, a prediction of the *x*-time series of the chaotic Hénon map and the δ^{13}C CENOGRID record as well as on studying causal interactions between variables in a combustion process.

With these applications, we showed the advantage MCDTS brings for any kind of method that utilizes an embedding such as recurrence analysis, embedding-based predictions of time series, or causal analysis with convergent cross-mapping. It, thus, has potential in many applications and disciplines, everywhere where such phase space-based approaches are used, but an automatic phase space reconstruction is required. The latter is of increasing interest, e.g., for big data analysis, analysis with highly reliable requirements (e.g., in medical applications), and also for deep learning-based frameworks.

Data and Code Availability The study that we present here is available as a fully reproducible code base [103]. The method itself will be available along with detailed documentation in the Julia language within the *DynamicalSystems.jl* [101] framework soon.

Declarations

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

Appendix

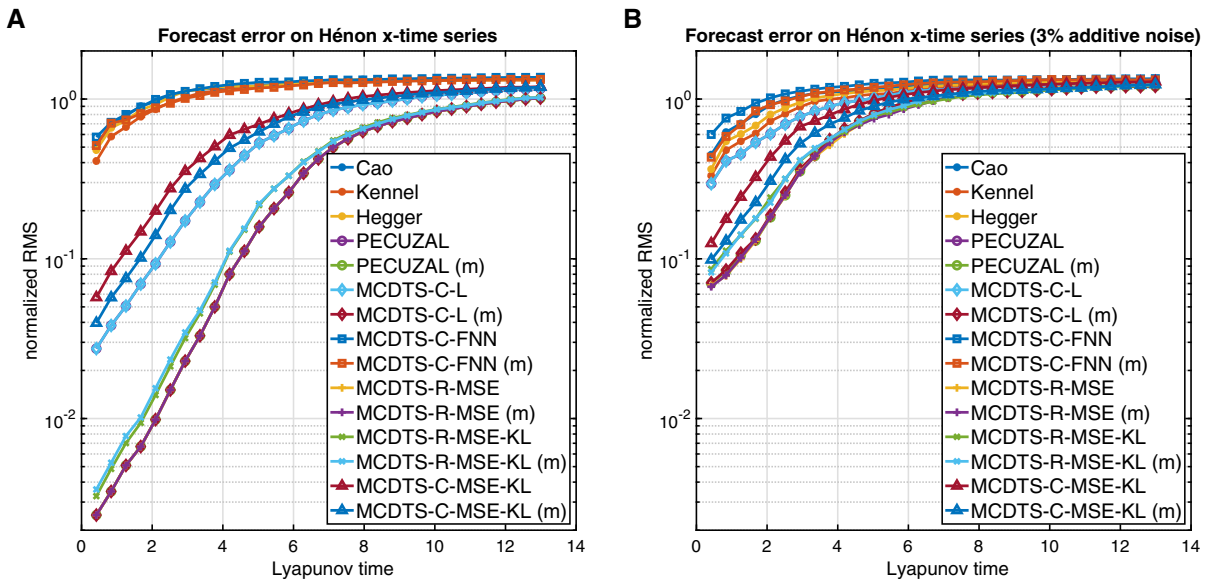A: Forecast of Hénon map time series

Fig. 8 A Normalized root-mean-square prediction errors (*RMS*) for the Hénon *x*-time series for all discussed reconstruction methods as a function of the prediction time. Shown are mean values of a distribution of 100 trials with different initial conditions. For the prediction, we used a one-step-ahead zeroth-order approximation on the nearest neighbor of the last point of the reconstructed trajectory and iteratively repeated that procedure 30 times in order to obtain a prediction of 31 samples in total for each trial. **B** Same as in **A** but with 3% additive white noise

B: CENOGRID prediction

Fig. 9 A Detrended δ^{13}C and **B** δ^{18}O time series of a total length of $N = 13,421$ samples and a sampling period of $\Delta t = 5000$yrs [92]

Table 2 Obtained embedding parameters for the different reconstruction methods. Time series index 1 in the third column corresponds to the detrended $\delta^{13}C$ and time series index 2 to the detrended $\delta^{18}O$ record shown in Fig. 9. For a description of the reconstruction methods see Table 1. The sequence of the delays (center column) and time series (right column) are a result of the embedding cycles which have been passed through in the corresponding reconstruction methods, which is why they are not necessarily ordered. For a reconstruction based on these embedding parameters, it would make no difference whether delays and corresponding time series were sorted beforehand

Reconstruction method	Chosen delays [in index values]	Chosen time series
Cao [66]	0,14,28,42,56,70,84	1,1,1,1,1,1,1
Kennel et al. [69]	0,14,28,42,56,70	1,1,1,1,1,1
Hegger and Kantz [67]	0,14,28,42,56,70,84,98	1,1,1,1,1,1,1,1
PECUZAL [41]	0,13,7,10,56,27,3,5,77,42,20,17,15	1,1,1,1,1,1,1,1,1,1,1,1,1
PECUZAL (m) [41]	0,13,7,10,56,27,3,5,77,42,20,17,15	1,1,1,1,1,1,1,1,1,1,1,1,1,1
MCDTS-C-FNN	0,69,84,54,37,48,63,26,12,20,16	1,1,1,1,1,1,1,1,1,1,1
MCDTS-C-FNN (m)	0,53,51,98,33,67,73,91,25,40,46	1,2,1,2,2,2,2,1,1,1,1
MCDTS-C-MSE-KL	0,54,117,93,126	1,1,1,1,1
MCDTS-C-MSE-KL (m)	0,109,78,198,37,155,53,95,133	1,2,1,2,2,1,1,1,1
MCDTS-R-MSE-KL	0,39,38,40,1,27	1,1,1,1,1,1
MCDTS-R-MSE-KL (m)	0,41,48,33,45,31	1,2,1,1,1,2
MCDTS-R-MSE	0,28,1,2	1,1,1,1
MCDTS-R-MSE (m)	0,1,2	1,1,1

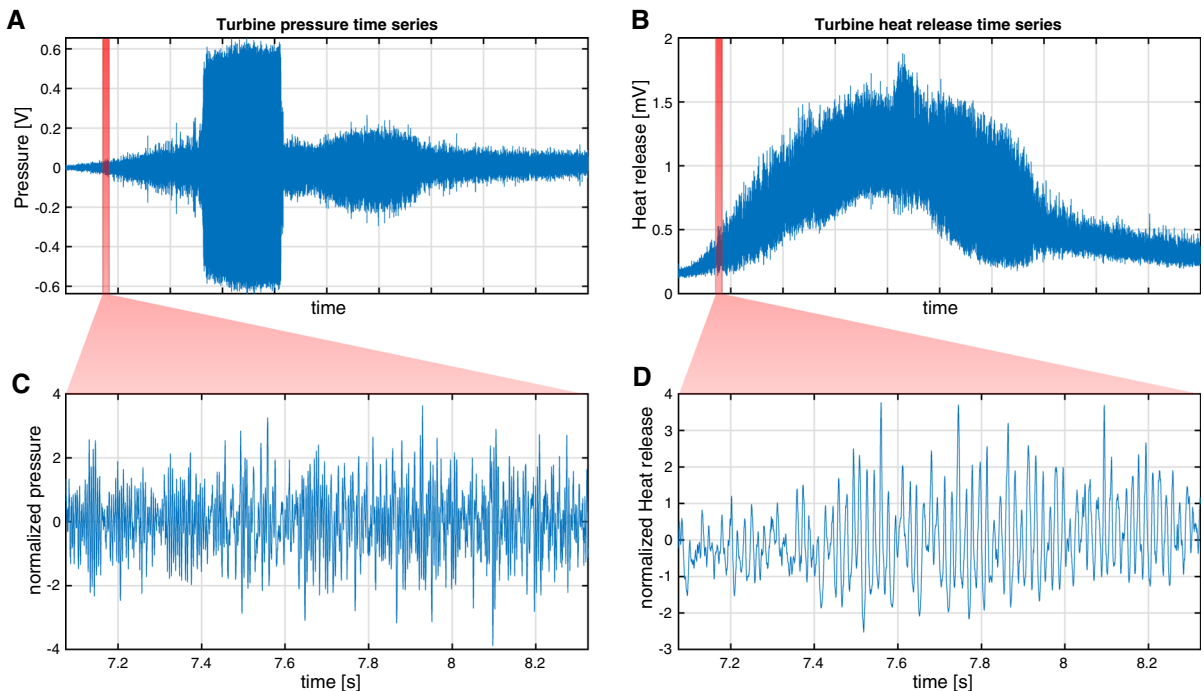C: Causal relationships in combustion process

Fig. 10 Entire time series of length $N' = 400,000$ of **A** turbine pressure (measured in V) and **B** combustion heat release (measured in mV). This is a non-stationary setup with a linearly varying control parameter (air flow rate) leading to bifurcations and different dynamics. For the calculations in Sect. 3.4, we sampled both time series 50 times in subsamples of length $N = 5000$. Panels **C, D** show one sample for both cases. Each sample has been normalized to zero mean and unit variance before we applied the embedding and the CCM

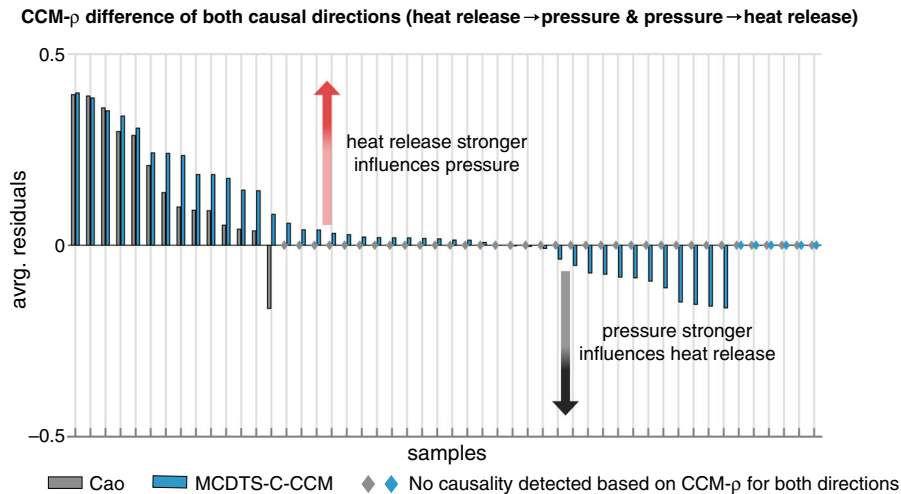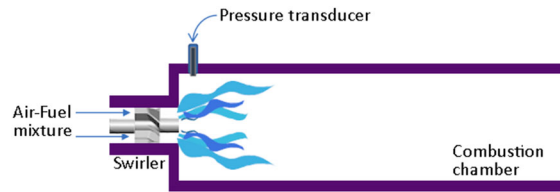**Fig. 11** Schematic experimental setup of the combustion experiment, see main text for details

CCM-ρ difference of both causal directions (heat release → pressure & pressure → heat release)

Fig. 12 Average pointwise difference of the CCM-correlation coefficients for the direction *heat release* → *pressure* and vice versa for both underlying reconstruction approaches. For a better visualization, we sorted these values here separately for both methods. A positive value indicates that the heat release has a stronger causal influence on pressure than vice versa, which is the expectation value. Diamonds indicate cases, where we could

not deduce a causal relationship for both directions in one sample. As also shown in the right panel of Fig. 7B, MCDTS-C-CCM was able to correctly detect a mutual causal relationship in 88% of all considered samples (only 12 % marked with blue diamonds in this Figure), whereas in the case of Cao's reconstruction approach, we could only detect this in 28% of all cases (72 % marked with gray diamonds in this Fig.)

References

1. Whitney, H.: Differentiable manifolds. Ann. Math. **37**(3), 645–680 (1936)
2. Mañé, R.: On the dimension of the compact invariant sets of certain non-linear maps. In: Rand, D., Young, L.S. (eds.) Dynamical Systems and Turbulence, Warwick 1980, pp. 230–242. Springer, Berlin (1981)
3. Takens, F.: Detecting strange attractors in turbulence. In: Rand, D., Young, L.S. (eds.) Dynamical Systems and Turbulence, Warwick 1980, pp. 366–381. Springer, Berlin (1981)
4. Sauer, T., Yorke, J.A., Casdagli, M.: Embedology. J. Statist. Phys. **65**(3), 579–616 (1991). https://doi.org/10.1007/BF01053745
5. Grassberger, P., Procaccia, I.: Estimation of the kolmogorov entropy from a chaotic signal. Phys. Rev. A **28**, 2591–2593 (1983). https://doi.org/10.1103/PhysRevA.28.2591
6. Grassberger, P., Procaccia, I.: Measuring the strangeness of strange attractors. Phys. D Nonlinear Phe-

nom. **9**(1), 189–208 (1983). https://doi.org/10.1016/0167-2789(83)90298-1
7. Hentschel, H., Procaccia, I.: The infinite number of generalized dimensions of fractals and strange attractors. Phys D Nonlinear Phenom. **8**(3), 435–444 (1983). https://doi.org/10.1016/0167-2789(83)90235-X
8. Kantz, H.: A robust method to estimate the maximal lyapunov exponent of a time series. Phys. Lett. A **185**(1), 77–87 (1994). https://doi.org/10.1016/0375-9601(94)90991-1
9. Kantz, H., Schürmann, T.: Enlarged scaling ranges for the ks-entropy and the information dimension. Chaos Interdiscip. J. Nonlinear Sci. **6**(2), 167–171 (1996). https://doi.org/10.1063/1.166161
10. Casdagli, M.: Nonlinear prediction of chaotic time series. Phys. D Nonlinear Phenom. **35**(3), 335–356 (1989). https://doi.org/10.1016/0167-2789(89)90074-2
11. Farmer, J.D., Sidorowich, J.J.: Predicting chaotic time series. Phys. Rev. Lett **59**, 845–848 (1987). https://doi.org/10.1103/PhysRevLett.59.845
12. Isensee, J., Datseris, G., Parlitz, U.: Predicting spatio-temporal time series using dimension reduced local states.

J. Nonlinear Sci. **30**(3), 713–735 (2020). https://doi.org/10.1007/s00332-019-09588-7

13. Kantz, H., Schreiber, T.: Nonlinear Time Series Analysis, 2nd edn. Cambridge University Press, Cambridge (2003)

14. Parlitz, U., Merkwirth, C.: Prediction of spatiotemporal time series based on reconstructed local states. Phys. Rev. Lett **84**, 1890–1893 (2000). https://doi.org/10.1103/PhysRevLett.84.1890

15. Ragwitz, M., Kantz, H.: Detecting non-linear structure and predicting turbulent gusts in surface wind velocities. Europhys. Lett. (EPL) **51**(6), 595–601 (2000). https://doi.org/10.1209/epl/i2000-00379-x

16. Sugihara, G., May, R.M.: Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. Nature **344**(6268), 734–741 (1990). https://doi.org/10.1038/344734a0

17. Feldhoff, J.H., Donner, R.V., Donges, J.F., Marwan, N., Kurths, J.: 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

18. Sugihara, G., May, R., Ye, H., Hsieh, Ch., Deyle, E., Fogarty, M., Munch, S.: Detecting causality in complex ecosystems. Science **338**(6106), 496–500 (2012). https://doi.org/10.1126/science.1227079

19. Ye, H., Deyle, E.R., Gilarranz, L.J., Sugihara, G.: Distinguishing time-delayed causal interactions using convergent cross mapping. Sci. Rep. **5**(1), 14750 (2015). https://doi.org/10.1038/srep14750

20. Kantz, H., Schreiber, T., Hoffmann, I., Buzug, T., Pfister, G., Flepp, L.G., Simonet, J., Badii, R., Brun, E.: Nonlinear noise reduction: a case study on experimental data. Phys. Rev. E **48**, 1529–1538 (1993). https://doi.org/10.1103/PhysRevE.48.1529

21. Matassini, L., Kantz, H., Hołyst, J., Hegger, R.: Optimizing of recurrence plots for noise reduction. Phys. Rev. E (2002). https://doi.org/10.1103/PhysRevE.65.021102

22. Packard, N.H., Crutchfield, J.P., Farmer, J.D., Shaw, R.S.: Geometry from a time series. Phys. Rev. Lett **45**, 712–716 (1980). https://doi.org/10.1103/PhysRevLett.45.712

23. Broomhead, D., King, G.P.: Extracting qualitative dynamics from experimental data. Phys. D Nonlinear Phenom. **20**(2), 217–236 (1986). https://doi.org/10.1016/0167-2789(86)90031-X

24. Gibson, J.F., Doyne Farmer, J., Casdagli, M., Eubank, S.: An analytic approach to practical state space reconstruction. Phys. D Nonlinear Phenom. **57**(1), 1–30 (1992)

25. Lu, Z., Hunt, B.R., Ott, E.: Attractor reconstruction by machine learning. Chaos Interdiscip. J. Nonlinear Sci. (6),(2018). https://doi.org/10.1063/1.5039508

26. Mann, B., Khasawneh, F., Fales, R.: Using information to generate derivative coordinates from noisy time series. Commun. Nonlinear Sci. Numer. Simul. **16**(8), 2999–3004 (2011). https://doi.org/10.1016/j.cnsns.2010.11.011

27. Parlitz, U.: Identification of true and spurious lyapunov exponents from time series. Int. J. Bifurcation Chaos **02**(01), 155–165 (1992). https://doi.org/10.1142/S0218127492000148

28. Casdagli, M., Eubank, S., Farmer, J., Gibson, J.: State space reconstruction in the presence of noise. Phys. D Nonlinear Phenom. **51**(1), 52–98 (1991). https://doi.org/10.1016/0167-2789(91)90222-U

29. Uzal, L.C., Grinblat, G.L., Verdes, P.F.: Optimal reconstruction of dynamical systems: a noise amplification approach. Phys. Rev. E (2011). https://doi.org/10.1103/PhysRevE.84.016223

30. Nichkawde, C.: Optimal state-space reconstruction using derivatives on projected manifold. Phys. Rev. E (2013). https://doi.org/10.1103/PhysRevE.87.022905

31. Eftekhari, A., Yap, H.L., Wakin, M.B., Rozell, C.J.: Stabilizing embedology: geometry-preserving delay-coordinate maps. Phys. Rev. E **97**(2), 022222 (2018)

32. Kugiumtzis, D.: State space reconstruction parameters in the analysis of chaotic time series –the role of the time window length. Phys. D Nonlinear Phenom. **95**(1), 13–28 (1996)

33. Rosenstein, M.T., Collins, J.J.: Luca] CJD reconstruction expansion as a geometry-based framework for choosing proper delay times. Phys. D Nonlinear Phenom. **73**(1), 82–98 (1994). https://doi.org/10.1016/0167-2789(94)90226-7

34. Buzug, T., Pfister, G.: Comparison of algorithms calculating optimal embedding parameters for delay time coordinates. Phys. Nonlinear Phenom. **58**(1), 127–137 (1992). https://doi.org/10.1016/0167-2789(92)90104-U

35. Buzug, T., Pfister, G.: Optimal delay time and embedding dimension for delay-time coordinates by analysis of the global static and local dynamical behavior of strange attractors. Phys. Rev. A **45**, 7073–7084 (1992). https://doi.org/10.1103/PhysRevA.45.7073

36. Buzug, T., Reimers, T., Pfister, G.: Optimal reconstruction of strange attractors from purely geometrical arguments. Europhys. Lett. (EPL) **13**(7), 605–610 (1990). https://doi.org/10.1209/0295-5075/13/7/006

37. Gao, J., Zheng, Z.: Local exponential divergence plot and optimal embedding of a chaotic time series. Phys. Lett. A **181**(2), 153–158 (1993). https://doi.org/10.1016/0375-9601(93)90913-K

38. Garcia, S.P., Almeida, J.S.: Multivariate phase space reconstruction by nearest neighbor embedding with different time delays. Phys. Rev. E (2005). https://doi.org/10.1103/PhysRevE.72.027205

39. Garcia, S.P., Almeida, J.S.: Nearest neighbor embedding with different time delays. Phys. Rev. E (2005). https://doi.org/10.1103/PhysRevE.71.037204

40. Kember, G., Fowler, A.: A correlation function for choosing time delays in phase portrait reconstructions. Phys. Lett. A **179**(2), 72–80 (1993). https://doi.org/10.1016/0375-9601(93)90653-H

41. Kraemer, K.H., Datseris, G., Kurths, J., Kiss, I.Z., Ocampo-Espindola, J.L., Marwan, N.: A unified and automated approach to attractor reconstruction. New J. Phys. (2021). https://doi.org/10.1088/1367-2630/abe336

42. Liebert, W., Pawelzik, K., Schuster, H.G.: Optimal embeddings of chaotic attractors from topological considerations. Europhys. Lett. (EPL) **14**(6), 521–526 (1991). https://doi.org/10.1209/0295-5075/14/6/004

43. Small, M., Tse, C.: Optimal embedding parameters: a modelling paradigm. Phys. D Nonlinear Phenom. **194**(3), 283–296 (2004). https://doi.org/10.1016/j.physd.2004.03.006

44. TSONIS AA,: Reconstructing dynamics from observables: The issue of the delay parameter revisited. Int. J. Bifurcation Chaos **17**(12), 4229–4243 (2007). https://doi.org/10.1142/S0218127407019913

45. Fraser, A.M., Swinney, H.L.: Independent coordinates for strange attractors from mutual information. Phys. Rev. A **33**, 1134–1140 (1986). https://doi.org/10.1103/PhysRevA.33.1134

46. Liebert, W., Schuster, H.: Proper choice of the time delay for the analysis of chaotic time series. Phys. Lett. A **142**(2), 107–111 (1989). https://doi.org/10.1016/0375-9601(89)90169-2

47. Aguirre, L.A.: A nonlinear correlation function for selecting the delay time in dynamical reconstructions. Phys. Lett. A **203**(2), 88–94 (1995). https://doi.org/10.1016/0375-9601(95)00392-G

48. Albano, A., Passamante, A., Farrell, M.E.: Using higher-order correlations to define an embedding window. Phys. D Nonlinear Phenom **54**(1), 85–97 (1991). https://doi.org/10.1016/0167-2789(91)90110-U

49. Albano, A.M., Muench, J., Schwartz, C., Mees, A.I., Rapp, P.E.: Singular-value decomposition and the grassberger-procaccia algorithm. Phys. Rev. A **38**, 3017–3026 (1988). https://doi.org/10.1103/PhysRevA.38.3017

50. Cao, L., Mees, A., Judd, K.: Dynamics from multivariate time series. Phys. D Nonlinear Phenom. **121**(1), 75–88 (1998). https://doi.org/10.1016/S0167-2789(98)00151-1

51. Fraser, A.M.: Reconstructing attractors from scalar time series: a comparison of singular system and redundancy criteria. Phys. D Nonlinear Phenom. **34**(3), 391–404 (1989). https://doi.org/10.1016/0167-2789(89)90263-7

52. Grassberger, P., Schreiber, T., Schaffrath, C.: Nonlinear time sequence analysis. Int. J. Bifurcation Chaos **01**(03), 521–547 (1991). https://doi.org/10.1142/S0218127491000403

53. Jia, Z., Lin, Y., Liu, Y., Jiao, Z., Wang, J.: Refined nonuniform embedding for coupling detection in multivariate time series. Phys. Rev. E (2020). https://doi.org/10.1103/PhysRevE.101.062113

54. Vlachos, I., Kugiumtzis, D.: Nonuniform state-space reconstruction and coupling detection. Phys. Rev. E (2010). https://doi.org/10.1103/PhysRevE.82.016207

55. Cai, W.D., Qin, Y.Q., Yang, B.R.: Determination of phase-space reconstruction parameters of chaotic time series. Kybernetika **44**(4), 557–570 (2008)

56. Gao, J., Zheng, Z.: Direct dynamical test for deterministic chaos and optimal embedding of a chaotic time series. Phys. Rev. E **49**, 3807–3814 (1994). https://doi.org/10.1103/PhysRevE.49.3807

57. Kim, H., Eykholt, R., Salas, J.: Nonlinear dynamics, delay times, and embedding windows. Phys. D Nonlinear Phenom. **127**(1), 48–60 (1999). https://doi.org/10.1016/S0167-2789(98)00240-1

58. Matilla-García, M., Morales, I., Rodríguez, J.M., Ruiz Marín, M.: Selection of embedding dimension and delay time in phase space reconstruction via symbolic dynamics. Entropy (2021). https://doi.org/10.3390/e23020221

59. Perinelli, A., Ricci, L.: Identification of suitable embedding dimensions and lags for time series generated by chaotic, finite-dimensional systems. Phys. Rev. E (2018). https://doi.org/10.1103/PhysRevE.98.052226

60. Han, M., Ren, W., Xu, M., Qiu, T.: Nonuniform state space reconstruction for multivariate chaotic time series. IEEE Trans. Cybern. **49**(5), 1885–1895 (2019)

61. Hirata, Y., Aihara, K.: Dimensionless embedding for nonlinear time series analysis. Phys. Rev. E (2017). https://doi.org/10.1103/PhysRevE.96.032219

62. Hirata, Y., Suzuki, H., Aihara, K.: Reconstructing state spaces from multivariate data using variable delays. Phys. Rev. E (2006). https://doi.org/10.1103/PhysRevE.74.026202

63. Holstein, D., Kantz, H.: Optimal markov approximations and generalized embeddings. Phys. Rev. E (2009). https://doi.org/10.1103/PhysRevE.79.056202

64. Judd, K., Mees, A.: Embedding as a modeling problem. Phys. D Nonlinear Phenom. **120**(3), 273–286 (1998). https://doi.org/10.1016/S0167-2789(98)00089-X

65. Pecora, L.M., Moniz, L., Nichols, J., Carroll, T.L.: A unified approach to attractor reconstruction. Chaos Interdiscip. J. Nonlinear Sci. (2007). https://doi.org/10.1063/1.2430294

66. Cao, L.: Practical method for determining the minimum embedding dimension of a scalar time series. Phys. D Nonlinear Phenom. **110**(1), 43–50 (1997). https://doi.org/10.1016/S0167-2789(97)00118-8

67. Hegger, R., Kantz, H.: Improved false nearest neighbor method to detect determinism in time series data. Phys. Rev. E **60**, 4970–4973 (1999). https://doi.org/10.1103/PhysRevE.60.4970

68. Kennel, M.B., Abarbanel, H.D.I.: False neighbors and false strands: a reliable minimum embedding dimension algorithm. Phys. Rev. E (2002). https://doi.org/10.1103/PhysRevE.66.026209

69. Kennel, M.B., Brown, R., Abarbanel, H.D.I.: Determining embedding dimension for phase-space reconstruction using a geometrical construction. Phys. Rev. A **45**, 3403–3411 (1992). https://doi.org/10.1103/PhysRevA.45.3403

70. Tang, Y., Krakovská, A., Mezeiová, K., Budáčová, H.: Use of false nearest neighbours for selecting variables and embedding parameters for state space reconstruction. J. Complex Syst. (2015). https://doi.org/10.1155/2015/932750

71. Aleksić, Z.: Estimating the embedding dimension. Phys. D Nonlinear Phenom. **52**(2), 362–368 (1991). https://doi.org/10.1016/0167-2789(91)90132-S

72. Čenys, A., Pyragas, K.: Estimation of the number of degrees of freedom from chaotic time series. Phys. Lett. A **129**(4), 227–230 (1988). https://doi.org/10.1016/0375-9601(88)90355-6

73. Bradley, E., Kantz, H.: Nonlinear time-series analysis revisited. Chaos Interdiscip. J. Nonlinear Sci. 25(9):097610, (2015). https://doi.org/10.1063/1.4917289

74. Garland, J., Bradley, E.: Prediction in projection. Chaos Interdiscip. J. Nonlinear Sci. (2015). https://doi.org/10.1063/1.4936242

75. Wendi, D., Marwan, N., Merz, B.: In search of determinism-sensitive region to avoid artefacts in recurrence plots. Int. J. Bifurcation Chaos **28**(1), 1850007 (2018). https://doi.org/10.1142/S0218127418500074

76. Coulom, R.: Efficient selectivity and backup operators in monte-carlo tree search. In: van den Herik HJ, Ciancarini

P, Donkers HHLMJ (eds.) Computers and Games, pp. 72–83. Springer, Berlin (2007)

77. Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., Hassabis, D.: Mastering the game of go with deep neural networks and tree search. Nature **529**(7587), 484–489 (2016). https://doi.org/10.1038/nature16961

78. Pecora, L.M., Carroll, T.L., Heagy, J.F.: Statistics for mathematical properties of maps between time series embeddings. Phys. Rev. E **52**, 3420–3439 (1995). https://doi.org/10.1103/PhysRevE.52.3420

79. Lorenz, E.: Predictability: a problem partly solved. In: Seminar on Predictability, 4-8 September 1995, ECMWF, ECMWF, Shinfield Park, Reading, vol **1**, pp 1–18, (1995). https://www.ecmwf.int/node/10829

80. Theiler, J.: Spurious dimension from correlation algorithms applied to limited time-series data. Phys. Rev. A **34**, 2427–2432 (1986). https://doi.org/10.1103/PhysRevA.34.2427

81. Karimi, A., Paul, M.R.: Extensive chaos in the Lorenz-96 model. Chaos **20**(4), 043105 (2010). https://doi.org/10.1063/1.3496397

82. Marwan, N., Foerster, S., Kurths, J.: Analysing spatially extended high-dimensional dynamics by recurrence plots. Phys. Lett. A **379**, 894–900 (2015). https://doi.org/10.1016/j.physleta.2015.01.013

83. Eckmann, J.P., Oliffson Kamphorst, S., Ruelle, D.: Recurrence plots of dynamical systems. Europhys. Lett. **4**(9), 973–977 (1987). https://doi.org/10.1209/0295-5075/4/9/004

84. Marwan, N., Romano, M.C., Thiel, M., Kurths, J.: Recurrence plots for the analysis of complex systems. Phys. Rep. **438**(5–6), 237–329 (2007). https://doi.org/10.1016/j.physrep.2006.11.001

85. Braun, T., Unni, V.R., Sujith, R.I., Kurths, J., Marwan, N.: Detection of dynamical regime transitions with lacunarity as a multiscale recurrence quantification measure. Nonlinear Dyn. (2021). https://doi.org/10.1007/s11071-021-06457-5

86. Zou, Y., Donner, R.V., Marwan, N., Donges, J.F., Kurths, J.: Complex network approaches to nonlinear time series analysis. Phys. Rep. **787**, 1–97 (2019). https://doi.org/10.1016/j.physrep.2018.10.005

87. Baptista, M.S., Ngamga, E.J., Pinto, P.R.F., Brito, M., Kurths, J.: Kolmogorov-Sinai entropy from recurrence times. Phys. Lett. A **374**(9), 1135–1140 (2010). https://doi.org/10.1016/j.physleta.2009.12.057

88. March, T.K., Chapman, S.C., Dendy, R.O.: Recurrence plot statistics and the effect of embedding. Phys. D **200**(1–2), 171–184 (2005). https://doi.org/10.1016/j.physd.2004.11.002

89. Kraemer, K.H., Donner, R.V., Heitzig, J., Marwan, N.: Recurrence threshold selection for obtaining robust recurrence characteristics in different embedding dimensions. Chaos Interdiscip. J. Nonlinear Sci. (2018)

90. Dhadphale, J., Unni, VR., Saha, A., Sujith, RI.: Neural ode to model and prognose thermoacoustic instability. 2106.12758, (2021)

91. Hénon, M.: A two-dimensional mapping with a strange attractor. Commun. Math. Phys. **50**(1), 69–77 (1976). https://doi.org/10.1007/BF01608556

92. Wendi, D., Marwan, N., Merz, B., Westerhold, T., Marwan, N., Drury, A.J., Liebrand, D., Agnini, C., Anagnostou, E., Barnet, J.S.K., Bohaty, S.M., De Vleeschouwer, D., Florindo, F., Frederichs, T., Hodell, D.A., Holbourn, A.E., Kroon, D., Lauretano, V., Littler, K., Lourens, L.J., Lyle, M., Pälike, H., Röhl, U., Tian, J., Wilkens, R.H., Wilson, P.A., Zachos, J.C.: An astronomically dated record of earth's climate and its predictability over the last 66 million years. Science **369**(6509), 1383–1387 (2020). https://doi.org/10.1126/science.aba6853, https://science.sciencemag.org/content/369/6509/1383

93. Clark, A.T., Ye, H., Isbell, F., Deyle, E.R., Cowles, J., Tilman, G.D., Sugihara, G.: Spatial convergent cross mapping to detect causal relationships from short time series. Ecology **96**(5), 1174–1181 (2015). https://doi.org/10.1890/14-1479.1

94. Granger, CWJ.: Investigating causal relations by econometric models and cross-spectral methods. Econometrica. **37**(3), 424–438 (1969)

95. Schiecke, K., Pester, B., Feucht, M., Leistritz, L., Witte, H.: Convergent cross mapping: basic concept, influence of estimation parameters and practical application. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp 7418–7421, (2015). https://doi.org/10.1109/EMBC.2015.7320106

96. Tony, J., Gopalakrishnan, E., Sreelekha, E., Sujith, R.: Detecting deterministic nature of pressure measurements from a turbulent combustor. Phys. Rev. E **92**(6), 062902 (2015)

97. Nair, V., Thampi, G., Sujith, R.: Intermittency route to thermoacoustic instability in turbulent combustors. J. Fluid Mech. **756**, 470–487 (2014)

98. Sujith, R., Unni, V.R.: Complex system approach to investigate and mitigate thermoacoustic instability in turbulent combustors. Phys. Fluid. **32**(6), 061401 (2020)

99. Juniper, M.P., Sujith, R.I.: Sensitivity and nonlinearity of thermoacoustic oscillations. Ann. Rev. Fluid Mech. **50**, 661–689 (2018)

100. Godavarthi, V., Pawar, S.A., Unni, V.R., Sujith, R.I., Marwan, N., Kurths, J.: Coupled interaction between unsteady flame dynamics and acoustic field in a turbulent combustor. Chaos Interdiscip. J. Nonlinear Sci. (11),(2018). https://doi.org/10.1063/1.5052210

101. Datseris, G.: Dynamicalsystems jl: A julia software library for chaos and nonlinear dynamics. J. Open Source Softw. 3(23):598, (2018). https://doi.org/10.21105/joss.00598,

102. Rackauckas, C., Nie, Q.: Differentialequations.jl–a performant and feature-rich ecosystem for solving differential equations in julia. J. Open Res. Softw. (2017). 10.5334/jors.151

103. Kraemer, KH., Gelbrecht, M.: hkraemer/MCDTS.jl: MCDTS Code Base for Reproducibility. (2022).https://doi.org/10.5281/zenodo.5877357