# Optimal State Space Reconstruction via Monte Carlo Decision Tree Search

**Kai Hauke Kraemer** ( hkraemer@pik-potsdam.de )
Potsdam Institute for Climate Impact Research (PIK) e V: Potsdam-Institut fur Klimafolgenforschung (PIK) e V https://orcid.org/0000-0002-9943-5391

**Maximilian Gelbrecht**
Potsdam Institute for Climate Impact Research (PIK) e V: Potsdam-Institut fur Klimafolgenforschung (PIK) e V

**Induja Pavithran**
Indian Institute of Technology Madras

**R.I. Sujith**
Indian Institute of Technology Madras

**Norbert Marwan**
Potsdam Institute for Climate Impact Research (PIK) e V: Potsdam-Institut fur Klimafolgenforschung (PIK) e V

**Research Article**

**Keywords:** State Space Reconstruction, Embedding, Optimization, Time Series Analysis, Causality, Prediction, Recurrence Analysis

**Posted Date:** September 20th, 2021

**DOI:** https://doi.org/10.21203/rs.3.rs-899760/v1



**Version of Record:** A version of this preprint was published at Nonlinear Dynamics on March 2nd, 2022.
See the published version at https://doi.org/10.1007/s11071-022-07280-2.

**Nonlinear Dynamics manuscript No.**
(will be inserted by the editor)

# Optimal state space reconstruction via Monte Carlo Decision Tree Search

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

Received: date / Accepted: date

**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. 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 paleoclimate 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.

**Keywords** State Space Reconstruction · Embedding · Optimization · Time Series Analysis · Causality · Prediction · Recurrence Analysis

**PACS** 05.45.Tp, 05.90.+m, 89.90.+n, 02.70.Uu, 05.10.Ln, 05.45.-a, 05.45.Ac

K. Hauke Kraemer · Maximilian Gelbrecht · Norbert Marwan
Potsdam Institute for Climate Impact Research, Member of the Leibniz Association, 14473 Potsdam, Germany
E-mail: hkraemer@pik-potsdam.de
ORCID: 0000-0002-9943-5391

K. Hauke Kraemer
Institute of Physics and Astronomy, University of Potsdam, 14476 Potsdam, Germany

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

Induja Pavithran
Department of Physics, IIT Madras, Chennai 600036, India

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

Norbert Marwan
Institute of Geosciences, University of Potsdam, 14476 Potsdam, Germany

## 1 Introduction

The famous embedding theorems of Whitney [97], Mañé [64], and Takens [88] together with their enhancement by Sauer et al. [81] allow a high dimensional state space reconstruction from (observed) uni- or multivariate time series. Computing dynamical invariants [35, 36, 41, 49, 51] from the observed system, making meaningful predictions even for chaotic or stochastic systems [23, 15, 85, 79, 50, 74, 45], detecting causal interactions [86, 24, 98] or non-linear noise reduction algorithms [52, 68] all rely explicitly or implicitly on (time delay) embedding [72] the data into a reconstructed state space. Other ideas rather than time delay embedding (TDE) are also possible [72, 8, 32, 73, 65, 63], 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

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

Here the delays $\tau_j$ are multiples of the sampling time $\Delta t$ 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 $1^{\text{st}}, 2^{\text{nd}}, \ldots, m^{\text{th}}$ embedding cycle. The total number of delays $\tau_j,\ j = [1, \ldots, m]$, 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. [16], Gibson et al. [32], Uzal et al. [93] or Nichkawde [71] for a profound overview of the problem). Picking optimal embedding parameters $\tau_j$ and $m$ comes down to make the resulting components of the reconstruction vectors $\vec{v}(t)$ as independent as possible [72, 81], but at the same time not too independent, in order to keep some information of the correlation structure of the data [16, 80, 59, 22, 93]. Besides some unified approaches [11, 61, 9, 10, 53, 27, 84, 30, 29, 92, 58], which tackle the estimation of the delays $\tau_j$ and the embedding dimension $m$ simultaneously, most researchers use two different methods to perform the reconstruction.

(1) A statistic determines the delays $\tau_j$, we call it $\Lambda_\tau$ throughout this paper. Usually $\tau_1 = 0$, i.e., the first component of $\vec{v}(t)$ 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 $\tau_2$ from the first minimum of the auto-mutual information [26, 60] is most common. All consecutive delays are then simply integer multiples of $\tau_2$. Other ideas based on different statistics like the auto-correlation function of the time series have been suggested [8, 3, 2, 53, 80, 1, 59, 14]. However, by setting $\tau_j, j > 2$ to multiples of $\tau_2$, one ignores the fact that this "measure" of independence strictly holds only for the first two components of reconstruction vectors ($m = 2$) [37, 25], even though in practice it works fine for most cases. More sophisticated ideas, like high-dimensional conditional mutual information [94, 46] and other statistics [28, 56, 12, 94, 77, 69], some of which include non-uniform delays and the extension to multivariate input data [47, 44, 94, 71, 30, 29, 76, 43, 38, 46, 76, 42, 38], have been presented.

(2) A statistic, we call it $\Gamma$ throughout this paper, which serves as an objective function and quantifies the goodness of a reconstruction, given that delays $\tau_j$ 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 $\tau_d$, gets appended to obtain the actual reconstruction vectors $\vec{v}_d(t) \in \mathbb{R}^{d+1}$ and these are compared to the reconstruction vectors $\vec{v}_{d-1}(t)$ of the former embedding cycle (if $d = 1$, $\vec{v}_{d-1}(t)$ is simply the time series $s_{i_1}$). This "comparison" is usually achieved by the amount of false nearest neighbors (FNN) [55, 54, 13, 39, 89], some other neighborhood-preserving-idea [100, 4], or more ambitious ideas [93, 71].

We have recently proposed an algorithm [58], which minimizes the $L$-statistic [93] (the objective function) in each embedding cycle $\mathscr{D}_d$ over possible delay values in this embedding cycle determined by a continuity statistic [76]. Nichkawde [71] 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, 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 $\Lambda_\tau$, 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 [47, 44, 6, 31, 95]. 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. In Section 2 we describe this method before we apply it to paradigmatic examples in Section 3.

## 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. Generally, each embedding cycle depends on all the previous embedding cycles. It seems therefore natural to visualize all possible embedding cycles in a tree-like hierarchical data structure as shown in Figure 1. The initial time series $s_{i_1}$ with delay $\tau_1 = 0$ forms the root

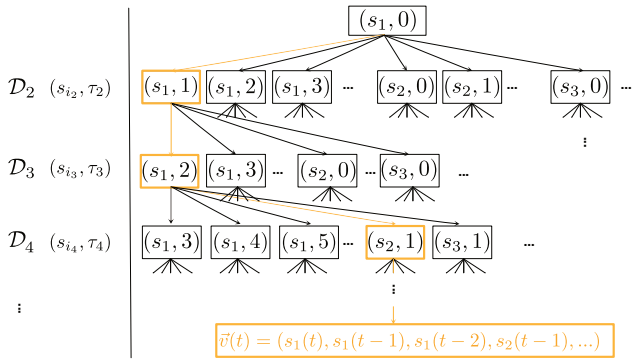


**Fig. 1** All possible embeddings of a time series visualized by a tree. Each leaf of the tree symbolizes one embedding cycle $\mathscr{D}_d$ using one selected time series $s_{i_d}$ from the multivariate data set and delay $\tau_d$. Marked in orange is one chosen full embedding.

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 $\Lambda_\tau$ and the objective function as $\Gamma$. 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 $\Gamma$. However, always choosing the leaf of the tree that decreases $\Gamma$ 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 [18]. Ultimately computer programs based on these algorithms were able to beat a reigning world champion, a feat that was long thought to be impossible for computer programs [83]. 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 $\Gamma$ and pick for the next embedding cycle preferably those delays that decrease $\Gamma$ further. Each node $\mathscr{N}_d$ 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 $\mathscr{N}_d$, and a value of the objective function $\Gamma_d$. We sample the tree $N_{\text{trial}}$-times in a two-step procedure:

- **Expand:** Starting from the root, for each embedding cycle $\mathscr{D}_d$, possible next steps $(s_{i_j},\tau_j,\Gamma_j)$ are either computed using suitable statistics $\Lambda_\tau$ and $\Gamma$ or, if there were already previously computed ones, they are looked up from the tree. We consider the first embedding cycle $\mathscr{D}_2$ and use the continuity statistic $\langle\varepsilon^\star\rangle(\tau)$ for $\Lambda_\tau$. 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 $\tau_2$ (see the rows in Figs. 1, 2 corresponding to $\mathscr{D}_2$). Then, one of the possible $\tau_2$'s is randomly chosen with probabilities computed with a softmax of the corresponding values of $\Gamma_j$. Due to its normalization, the softmax function is able to convert all possible values of $\Gamma_j$ to probabilities with $p_j = \exp(-\beta\Gamma_j)/\sum_k \exp(-\beta\Gamma_k)$. The parameter $\beta$ can be chosen to further stress smaller values of $\Gamma$. This procedure is repeated (consecutive rows for $\mathscr{D}_3\ldots$ etc. in Figs. 1, 2) until the very last computed embedding cycle $\mathscr{D}_{m+1}$. This is, when the objective function $\Gamma_{m+1}$ cannot be further decreased for any of the $\tau_{m+1}$-candidates. Figure 2 visualizes this procedure.
- **Backpropagation:** After the tree is expanded, the final value $\Gamma_m$ 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 $\Gamma_d$ 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 $\mathscr{D}_1$. This ensures an even sampling over the given time series, which can all serve as a valid 1st component of the final reconstruction vectors. Additionally, as soon as a $\Gamma_j$ is found that is smaller than the previous global minimum, this embedding cycle is directly chosen and not randomized via the softmax function.

The computational complexity of this algorithm obviously scales with the number of trials $N_{\text{trials}}$ even though already computed embedding cycles are not computed again in later trials. Otherwise the complexity depends on the chosen delay pre-selection function $\Lambda_\tau$ and the objective function $\Gamma$. 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.

## 3 Applications

### 3.1 Recurrence properties of the Lorenz-96 system

As a first, potentially higher dimensional example we utilized the Lorenz 96 system [62], a set of $N$ ordinary first-

**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 $\Lambda_\tau$ | objective function $\Gamma$ |
|---|---|---|
| MCDTS-C-FNN | maxima of contiunity statistic $\langle\varepsilon^\star\rangle(\tau)$ [75, 76] | FNN [55, 54, 39] |
| MCDTS-C-L | maxima of contiunity statistic $\langle\varepsilon^\star\rangle(\tau)$ | $L/\Delta L$-statistic [93, 58] |
| MCDTS-C-MSE | maxima of contiunity statistic $\langle\varepsilon^\star\rangle(\tau)$ | mean squared prediction error |
| MCDTS-C-MSE-KL | maxima of contiunity statistic $\langle\varepsilon^\star\rangle(\tau)$ | mean Kullback-Leibler divergence of true and predicted |
| MCDTS-C-CCM | maxima of contiunity statistic $\langle\varepsilon^\star\rangle(\tau)$ | negative CCM-correlation coefficient [86] |
| 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 |

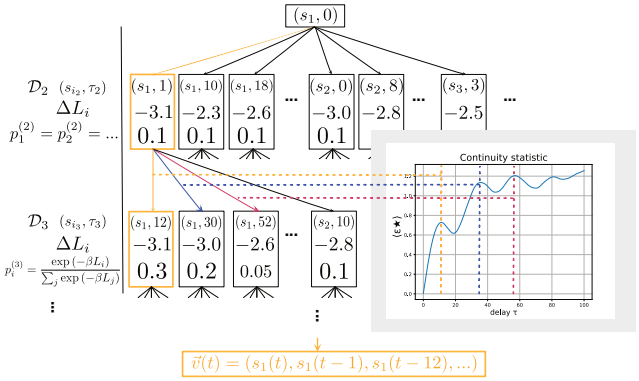


**Fig. 2** Visualization of the expand step of the MCDTS algorithm. Here we exemplary use the continuity statistic $\langle\varepsilon^\star\rangle(\tau)$ as the delay pre-selection statistic $\Lambda_\tau$ and the $\Delta L$-statistic [93] as the objective function $\Gamma$, as it has been utilized in the recently proposed PECUZAL algorithm [58].

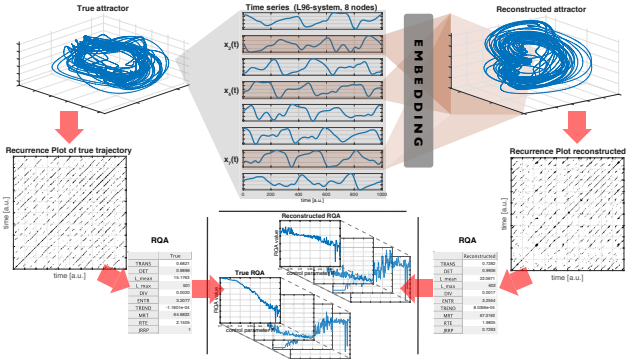


**Fig. 3** Schematic visualization of the data analysis for the Lorenz 96 system, Eq. (2) (see text for details). In case of the univariate approach the $x_2(t)$-time series gets embedded by all considered reconstruction methods, for the multivariate approach, three time series ($x_2(t)$, $x_4(t)$ and $x_7(t)$) 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.

order differential equations

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

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 $N \geq 4$. One 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 $N = 8$, the initial condition to $u_0 = [0.590; 0.766; 0.566; 0.460; 0.794; 0.854; 0.200; 0.298]$, used a sampling time of $\Delta t = 0.1$, discarded the first 2,500 points of the integration as transients and used time series consisting of 5,000 samples. We focused on two scenarios: (1) only the time series of the 2$^{\text{nd}}$ node (univariate embedding) and (2) three time series of nodes 2, 4, and 7 have been used to mimic a uni- and a multivariate embedding case. For each of these time series we performed an embedding, using three classic time delay approaches proposed by Kennel et al. [55] (5%-threshold), Cao [13] (slope threshold of 0.2) and Hegger and Kantz [39] (5%-threshold) with a uniform delay value estimated as the first minimum of the auto mutualinformation (only applicable to the univariate case) and the recently proposed PECUZAL algorithm [58]. For our proposed MCDTS approach we embedded the data using the continuity statistic $\langle\varepsilon^\star\rangle(\tau)$ as the delay pre-selection statistic $\Lambda_\tau$. For the objective function $\Gamma$ we tried two different approaches, namely the $\Delta L$-statistic [93] (MCDTS-C-L) as well as the FNN-statistic [39] (MCDTS-C-FNN). In all approaches we discarded serially correlated points from the nearest neighbor search, by setting a Theiler window [90] to the first minimum of the mutual information. An overview over all MCDTS implementations and abbreviations is given in Table 1.

We compared the recurrence properties of all reconstructed trajectories to recurrence properties of the true trajectory (obtained from the numerical integration) by using recurrence quantification analysis (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 $\vec{x}_i$ ($i = 1,\ldots,N$, with $N$ the number of points forming the trajectory) in a $d$-dimensional (optionally reconstructed) state space [21, 67]

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

with $\|\cdot\|$ a norm, $\varepsilon$ a recurrence threshold, and $\Theta$ the Heaviside function. There are numerous ideas of how to quantify a RP [67, 7]. 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 a an adjacency matrix $\mathbf{A} = \mathbf{R} - \mathbb{1}$ of a recurrence network (RN) [99]. Some of these quantifiers are related to dynamical invariants [66, 5]. Here we used the transitivity (*TRANS*) of the $\varepsilon$-RN, the *determinism* (*DET*), the *mean diagonal line length* ($L_{\text{mean}}$), the *maximal 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 the *joint recurrence rate fraction* (*JRRF*). JRRF measures the accordance of the recurrence plot of the (true) reference system, $\mathbf{R}^{\text{ref}}$ with the RP of the reconstruction, $\mathbf{R}^{\text{rec}}$.

$$\text{JRRF} = \frac{\sum_{i,j}^{N} JR_{i,j}}{\sum_{i,j}^{N} R^{\text{ref}}_{i,j}}, \quad \text{JRRF} \in [0,\ 1] \tag{4}$$

$$\mathbf{JR} = \mathbf{R}^{\text{ref}} \circ \mathbf{R}^{\text{rec}}. \tag{5}$$

We computed both, $\mathbf{R}^{\text{ref}}$ and $\mathbf{R}^{\text{rec}}$, by fixing the recurrence threshold corresponding to a global recurrence rate (*RR*) of 5% in order to ensure comparability [57]. Although the quantification measures depend crucially on the chosen recurrence threshold, the particular choice we made here is not so important, since we applied it to all RPs we compared. $RR = 5\%$ ensures a proper resolution of the inherent structures to be quantified by the ten aforementioned measures.

The described procedure is schematically illustrated in Figure 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 has been computed (normalized to the reference RQA-values). The pairwise comparison of the MSEs is shown in Figure 4 as the percentage of the ten RQA-MSEs, which take a lower MSE. 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 have expected a better reconstruction for the multivariate cases – because we simply provided more information – our proposed method also did better in the univariate case when the *FNN*-statistic has been used as an objective function. When using MCDTS with the *L*-statistic, there was hardly any improvement discernible, while the computational costs were magnitudes higher. Here PECUZAL gave better results, even though it uses the same statistics. However, combined with the *FNN*-statistic our proposed idea did very well in the univariate case and produced excellent results for the multivariate case.

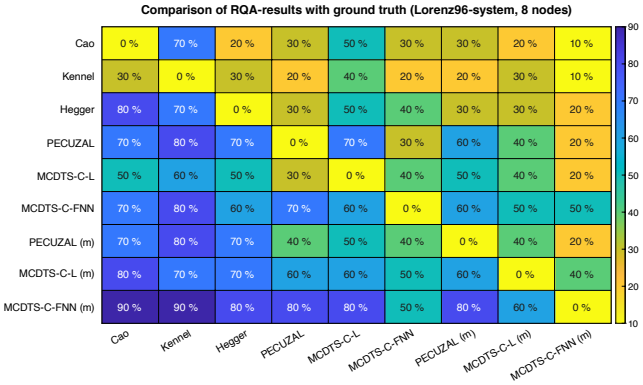


**Fig. 4** Results of the analysis of the Lorenz 96 system with varying control parameter and for all considered reconstruction approaches (see Table 1 for notations). Shown is the pairwise comparison of the normalized mean squared error of all considered ten RQA-quantifiers with respect to the truth 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 $\vec{v}(t)$ of a single time series $s(t)$ is used to further predict its course. Besides a very recent idea [20] to train neural ordinary differential equations on a reconstructed trajectory, which then allows prediction, several attempts have been published [23, 15, 85, 79, 50, 74, 45] which more or less rely on the same basic idea. For the last vector of the reconstructed trajectory, denoted with a time-index $l$, $\vec{v}(t_l)$, a nearest neighbor search is performed. Then these neighbors are used to predict the future value of this point $T$ time steps ahead, $\vec{v}(t_{l+T})$. Knowledge of the used embedding, which led to the reconstruction vectors $\vec{v}(t)$, then allows to read the prediction of the time series $s(t_l + T)$ from the predicted reconstruction vector $\vec{v}(t_{l+T})$. Usually $T = 1$, i.e., the forecast is iteratively build by appending $\vec{v}(t_{l+T})$ to the trajectory $\vec{v}(t_i)$, $i = 1,\ldots,l$, 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 near-

est neighbors. For instance, Farmer and Sidorowich [23] proposed a linear approximation, i.e., a linear polynomial is fitted to the pairs $(\vec{v}(t_{nn_i}), \vec{v}(t_{nn_i+T}))$, where $nn_i$ denotes the $i^{\text{th}}$ nearest neighbor time-index. Sugihara and May [85] 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 $\vec{v}(t_l)$ 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 $\vec{v}(t_l)$, denoted by the index $nn_1$, $\vec{v}(t_{l+T}) = \vec{v}(t_{nn_1+T})$. 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 $\vec{v}(t)$. We hypothesize that the accuracy of such a prediction will also depend on the reconstruction method, especially when the training set is rather short (c.f. Small and Tse [84] and also Bradley and Kantz [6]). In particular, Garland and Bradley [31] could show 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.

We used the time series of the Hénon map [40], $x_{i+1} = y_i + 1 - ax_i^2$ and $y_{i+1} = bx_i$, with standard parameters $a = 1.4$, $b = 0.3$ and 100 randomly chosen different initial conditions. For each of those 100 samples $x$- and $y$-time series of length $N = 10,030$ were obtained (transients removed). The first $10,000$ points of the time series were 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 were the prediction test set (only the $x$-time series has been predicted). The same reconstruction methods as in Section 3.1 were used, but for MCDTS we tried two different delay pre-selection statistics $\Lambda_\tau$. Rather then only considering the continuity-statistic (denoted as $C$ in the model description) we also looked at a whole range of delay values $\tau = 0, \ldots, 50$ (denoted as $R$ in the model description). For the objective function $\Gamma$ we tried

- the $\Delta L$-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*),

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

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 was 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 Section 3.1, $m$ indicates the multivariate case, in which both, $x$- and $y$-time series have been 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 performed significantly better than the forecasts based on the traditional TDE methods. Even though the continuity statistic constitutes a reasonable delay preselection 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 has been made to underpin the better performance of the MCDTS-approaches in comparison to the classical time delay methods. Therefore we defined a threshold $\zeta = 0.1$ and computed the prediction times for which $e_{\text{rms}}(T)$ first exceeds $\zeta$ for all trials and for all considered reconstruction methods. These distributions of prediction times for each method were 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

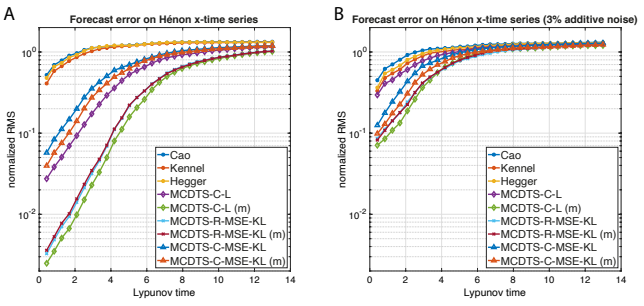


**Fig. 5** **A** Normalized root-mean-square prediction errors (*RMS*) for the Hénon $x$-time series and for selected reconstruction methods (see Fig. 8 for all mentioned approaches and Table 1) as a function of the prediction time. Shown are mean values of a distribution of 100 trials with different initial conditions. For the prediction we 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.

MCDTS-C-L (m) achieved 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) achieved the overall best results (Fig. 8). The choice of the threshold $\zeta$ is obviously subjective, but a range of thresholds gave similar results and the "grouping" of the results according 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 well 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 did 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 [31] 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 [84]. 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) performed 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 applied it to the recently published CENOzoic Global Reference benthic foraminifer carbon and oxygen Isotope Dataset (CENOGRID) [96]. 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.

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 $\Delta t = 5,000$yrs (Figure 9 in Appendix B). Here we made predictions on the $\delta^{13}$C isotope record. The first $13,311$ samples have been used as a training set, from which state space reconstructions were obtained. The remaining 110 samples of the $\delta^{13}$C record acted as the test set. For 100 different starting points in the test set we have made 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 Section 3.2. This way we have simulated different initial conditions for the prediction and obtained a distribution of forecasts for each reconstruction method. We again used 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), scored 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) are 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 used $\delta^{13}$C lagged by 1 and 2 samples in a 3-dimensional reconstruction. The classic TDE methods and all other reconstruction methods (listed in Table 1, not shown in Fig. 9) yielded higher dimensional embeddings (Table 2). Yet, all these higher dimensional reconstructions

gave poor prediction results, except for MCDTS-C-MSE-KL (m), which gave significant better predictions ($\alpha = 0.05$) than the classic TDE methods at least for the one-step-ahead prediction.

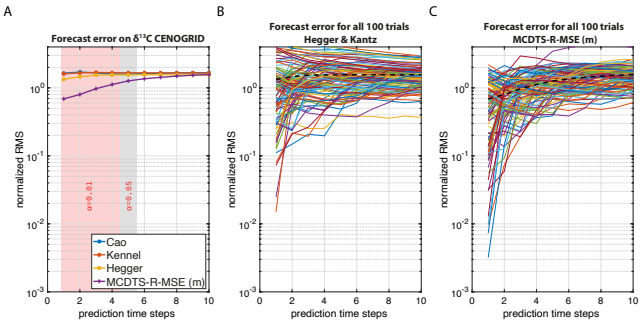


**Fig. 6** **A** Mean normalized root mean square prediction error for four selected reconstruction methods on the $\delta^{13}$C CENOGRID record. **B** Prediction error for all 100 trials for the classic TDE method of Hegger and Kantz [39] (yellow line in panel A). **C** Prediction error for all 100 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 $\alpha = 0.05$).

### 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)* [86, 98, 17], which – in contrast to Granger causality [34] – also works for time series stemming from non separable 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 $Y$."[86] This also incorporates the embedding theorems [97, 64, 88] in a sense that a state space reconstruction based on $x(t)$ is diffeomorph 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 $X \to Y$, a state space reconstruction of $y(t)$ yields a trajectory $\vec{v}_y(t) \in \mathbb{R}^m$, with $m$ denoting the embedding dimension, which is then used for estimating values of $x(t)$, namely $\hat{x}(t)$. It is said that $\vec{v}_y(t)$ *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 $\vec{v}_y(t')$. 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"[86]:

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

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

$$w_i = u_i / \sum u_j, \qquad j = 1, \ldots, m+1 \tag{7}$$

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

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 $\rho_{CCM}$, which has been done in this study. The clou is that the estimation skill, here represented by $\rho_{CCM}$, increases with the considered amount of data used, if $X$ indeed causally influences $Y$. This is because the attractor – represented by the reconstruction vectors $\vec{v}_y(t)$ – gets resolved better with increasing time series length, resulting in closer nearest neighbors and therefore a better concordance of $\hat{x}(t)$ and $x(t)$, i.e., an increase of $\rho_{CCM}$ 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 $\rho_{CCM}$ 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. [86]. However, Schiecke et al. [82] 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 utilized the MCDTS framework in the following way. As a delay pre-selection method $\Lambda_\tau$ we used the reliable continuity statistic $\langle \varepsilon^\star \rangle(\tau)$ [75, 76]. As a suitable objective function $\Gamma$ we used the negative of the corresponding $\rho_{CCM}$, i.e., MCDTS optimized the embedding with respect to maximizing $\rho_{CCM}$ 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 applied the CCM-method to time series data that spans the different dynamical regimes of a thermoacoustic system. Here we investigated the mutual causal influence of two recorded variables of the thermoacoustic system, namely the pressure and the heat release rate fluctuations (Fig. 10). The experiments are performed on a turbulent combustor with a rectangular combustion chamber (length = 700 mm, cross-section = 90 mm X 90 mm, Fig. 11). A fixed vane swirler is used to stabilize the flame. 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. We ignite the partially

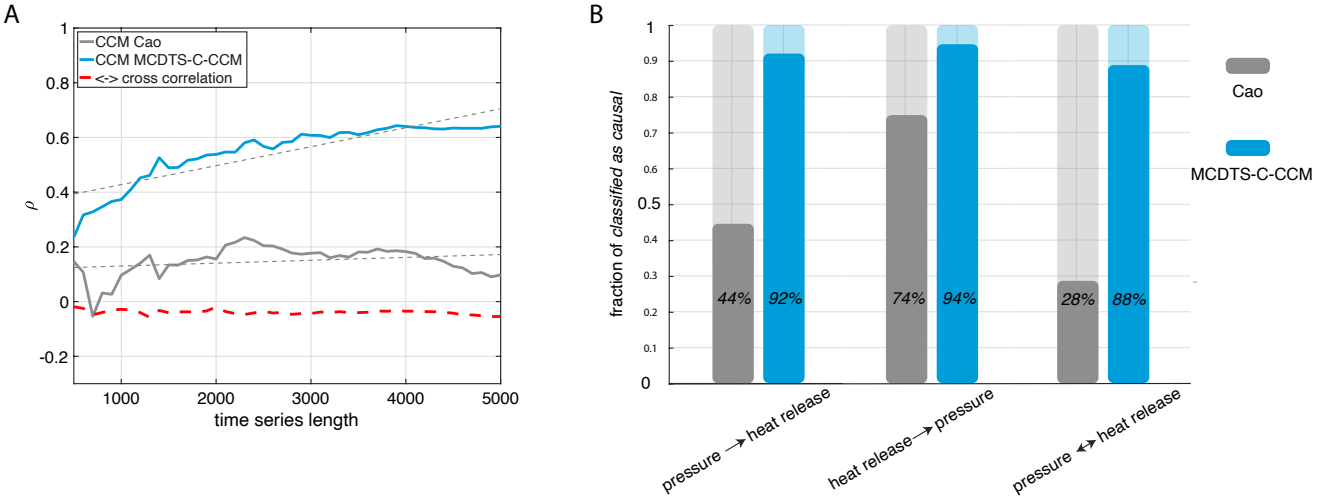


**Fig. 7** **A** Linear correlation coefficient of convergent cross mapping (CCM) *heat release* → *pressure* as a function of the considered time series length for Cao's embedding method (gray) and the proposed MCDTS embedding (blue) exemplary shown for on one out of 50 drawn sub-samples of length $N = 5,000$ 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 ensured convergence of the cross mapping, and, thus, a true causal relationship, if there was 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) exceeded a value of 0.2 (in the shown case Cao's method did not detect a causal influence of the heat release to the pressure). We tested 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.

premixed reactant mixture using a spark plug. Once the flame is established in the combustor, we continuously vary the control parameter (mass flow rate of air, which, in turn, varies the Reynolds number[1] and the equivalence ratio[2]) to observe the dynamical transitions in the system. We measure the acoustic pressure fluctuations 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 we increase the airflow rate, the system transitions from a state of stable operation (which comprises high dimensional chaos having low amplitude [91] to intermittency, a state that comprises bursts of periodic oscillations amid epochs of aperiodicity [70], and then to limit cycle [87]. The self-sustained limit cycle oscillations represent a state of oscillatory instability is known as thermoacoustic instability [48]. When we further increase the flow rate of air, 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 asymmetric bi-directional coupling between heat release and pressure has been discovered in a stationary setup of a very similar experiment [33] we would also expect that the heat release rate has a slightly stronger causal influence on pressure than vice versa.

The goal is to proof the expected mutual causal relationship and the asymmetry in its strengths by applying MCDTS-C-CCM on a range of time series, sampled from the entire record (Fig. 10). We compared it to results obtained from using the CCM method with the classical embedding approach of Cao [13]. Specifically, we set up the following workflow for this analysis:

1. 50 time indices $t' \in t$ were drawn randomly, where $t$ covers the entire record.
2. For each of these indices $t'$ time series of length $N = 5,000$ for pressure and heat release were obtained and standardized to zero mean and unit variance (c.f. Fig 10).
3. Both time series samples (of full length $N = 5,000$) each were 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 $\rho_{CCM-Cao}$ and $\rho_{CCM-MCDTS}$ were computed

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

[2] Equivalence ratio is the ratio between the actual fuel-air ratio to the stoichiometric fuel-air ratio.

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 fitted a linear model to $\rho_{CCM}$ (dashed black lines in Fig. 7A) and whenever that model gave a positive slope *and* the last value of $\rho_{CCM}$ (i.e. for the longest considered time series of length $N = 5{,}000$) exceeded a value of 0.2, we inferred a true causal relationship.
5. When we could detect a causal relation simultaneously in both directions, we computed the average of the pointwise difference $\rho_{CCM\ heat\to pressure} - \rho_{CCM\ pressure\to heat}$

The minimum considered value of 0.2 for $\rho_{CCM}$ is an arbitrary and subjective choice and we could have made other choices. But since this procedure had been applied to $\rho_{CCM-Cao}$ and $\rho_{CCM-MCDTS}$ 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 $\rho_{CCM}$, but a positive linear trend. Results did 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). While MCDTS-C-CMM maintains a correct classification in 92% of all cases considered (50 samples) for *pressure* $\to$ *heat release* and 94% for *heat release* $\to$ *pressure*, Cao's method is only able to correctly classify 44% and 74%, respectively. These results itself 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 Figure 7B reveals that in 88% of all cases considered, MCDTS-C-CCM was able to detect a mutual causal relationship, while Cao's method managed to do so in only 28% of the cases. Furthermore we tried to validate a hypothesis made by Godavarthi et al. [33] that heat release has a stronger effect on pressure than vice versa for most of the considered dynamics. The problem here is two-fold: 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 the description of the CCM method it is stated that in case of a stronger causal effect from $X$ to $Y$, cross-mapping $X$ using $\vec{v}_y$ converges *faster* then the other way round. Thus, we would have to define what *faster* means with respect to our experimental curves like the ones shown in Figure 7A. That would mean introducing some parameters on which the results would depend too much. Here we went for a way simpler idea, in order to detect the strength of a causal interaction. For samples where a causal relation in *both* direction has been detected we computed the average of the pointwise difference of the CCM-correlation coefficients, i.e. $\rho_{CCM\ Heat\to pressure} - \rho_{CCM\ pressure\to heat}$. When this difference is positive, we claim that heat release stronger effects pressure in a causal sense than vice versa. Figure 12 shows these results for all samples and for both considered reconstruction approaches. In conclusion, we were able to show a clear advantage in using the proposed MCDTS reconstruction approach together with the CCM method. Not only was the overall classification ability remarkable, but the MCDTS reconstructions also allowed verification of an assumed asymmetric causal interaction, which was not possible with 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* that 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-mutualinformation statistic. Since the sampling of the tree is a random procedure the proposed idea only yields converging embedding parameters for a sufficient sampling size $N_{\text{trial}}$. 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. 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 $\delta^{13}$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 embedding-based predictions of time series or causal analysis with convergent cross mapping.

**Acknowledgements** This work has been financially supported by the German Research Foundation (DFG projects MA4759/8 & MA4759/9). All computations have been carried out in the Julia language and made use of the packages *DynamicalSystems.jl* [19] and *DifferentialEquations.jl* [78].

## Declarations

### Conflict of interest

The authors declare that they have no conflict of interest.

### Code availability

The study that we present here is available as a fully reproducible code base **Repository will be published and cited here, when accepted**.

## A Forecast of Hénon map time series

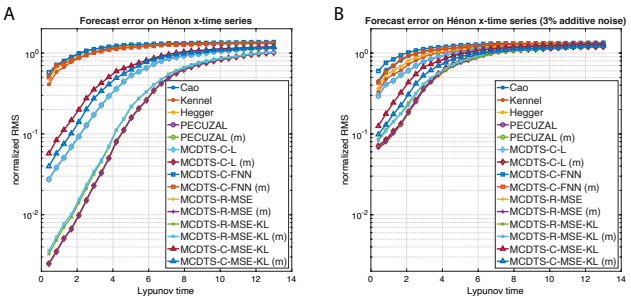


**Fig. 8** **A** Normalized root-mean-square prediction errors (*RMS*) for the Hénon $x$-time series for all discussed reconstruction methods as a function of the prediction time. Shown are mean values of a distribution of 100 trials with different initial conditions. For the prediction we used a one 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

## C Causal relationships in combustion process

## References


1. Aguirre LA (1995) A nonlinear correlation function for selecting the delay time in dynamical reconstructions. Physics Letters A 203(2):88 – 94, DOI https://doi.org/10.1016/0375-9601(95)00392-G, URL http://www.sciencedirect.com/science/article/pii/037596019500392G
2. Albano A, Passamante A, Farrell ME (1991) Using higher-order correlations to define an embedding window. Physica D: Nonlinear Phenomena 54(1):85 – 97, DOI https://doi.org/10.1016/0167-2789(91)90110-U, URL http://www.sciencedirect.com/science/article/pii/016727899190110U
3. Albano AM, Muench J, Schwartz C, Mees AI, Rapp PE (1988) Singular-value decomposition and the grassberger-procaccia algorithm. Phys Rev A 38:3017–3026, DOI 10.1103/PhysRevA.38.3017, URL https://link.aps.org/doi/10.1103/PhysRevA.38.3017
4. Aleksić Z (1991) Estimating the embedding dimension. Physica D: Nonlinear Phenomena 52(2):362 – 368, DOI https://doi.org/10.1016/0167-2789(91)90132-S, URL http://www.sciencedirect.com/science/article/pii/016727899190132S
5. Baptista MS, Ngamga EJ, Pinto PRF, Brito M, Kurths J (2010) Kolmogorov-Sinai entropy from recurrence times. Physics Letters A 374(9):1135–1140, DOI 10.1016/j.physleta.2009.12.057
6. Bradley E, Kantz H (2015) Nonlinear time-series analysis revisited. Chaos: An Interdisciplinary Journal of Nonlinear Science 25(9):097610, DOI 10.1063/1.4917289, URL https://doi.org/10.1063/1.4917289, https://doi.org/10.1063/1.4917289
7. Braun T, Unni VR, Sujith RI, Kurths J, Marwan N (2021) Detection of dynamical regime transitions with lacunarity as a multiscale recurrence quantification measure. Nonlinear Dynamics DOI 10.1007/s11071-021-06457-5
8. Broomhead D, King GP (1986) Extracting qualitative dynamics from experimental data. Physica D: Nonlinear Phenomena 20(2):217 – 236, DOI https://doi.org/10.1016/0167-2789(86)90031-X, URL http://www.sciencedirect.com/science/article/pii/016727898690031X
9. Buzug T, Pfister G (1992) Comparison of algorithms calculating optimal embedding parameters for delay time coordinates. Physica D: Nonlinear Phenomena 58(1):127 – 137, DOI https://doi.org/10.1016/0167-2789(92)90104-U, URL http://www.sciencedirect.com/science/article/pii/016727899290104U
10. Buzug T, Pfister G (1992) 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, DOI 10.1103/PhysRevA.45.7073, URL https://link.aps.org/doi/10.1103/PhysRevA.45.7073
11. Buzug T, Reimers T, Pfister G (1990) Optimal reconstruction of strange attractors from purely geometrical arguments. Europhysics Letters (EPL) 13(7):605–610, DOI 10.1209/0295-5075/13/7/006, URL https://doi.org/10.1209%2F0295-5075%2F13%2F7%2F006
12. Cai WD, Qin YQ, Yang BR (2008) Determination of phase-space reconstruction parameters of chaotic time series. Kybernetika 44(4):557–570
13. Cao L (1997) Practical method for determining the minimum embedding dimension of a scalar time series. Physica D: Nonlinear Phenomena 110(1):43 – 50, DOI https://doi.org/10.1016/S0167-2789(97)00118-8, URL http://www.sciencedirect.com/science/article/pii/S0167278997001188
14. Cao L, Mees A, Judd K (1998) Dynamics from multivariate time series. Physica D: Nonlinear Phenomena 121(1):75 – 88, DOI https://doi.org/10.1016/S0167-2789(98)00151-1, URL http://www.sciencedirect.com/science/article/pii/S0167278998001511
15. Casdagli M (1989) Nonlinear prediction of chaotic time series. Physica D: Nonlinear Phenomena 35(3):335–356, DOI https://doi.org/10.1016/0167-2789(89)90074-2, URL https://www.sciencedirect.com/science/article/pii/0167278989900742
16. Casdagli M, Eubank S, Farmer J, Gibson J (1991) State space reconstruction in the presence of noise. Physica D: Nonlinear Phenomena 51(1):52 – 98, DOI https://doi.org/10.1016/0167-2789(91)90222-U, URL http://www.sciencedirect.com/science/article/pii/016727899190222U
17. Clark AT, Ye H, Isbell F, Deyle ER, Cowles J, Tilman GD, Sugihara G (2015) Spatial convergent cross mapping to detect causal relationships from short time series. Ecology

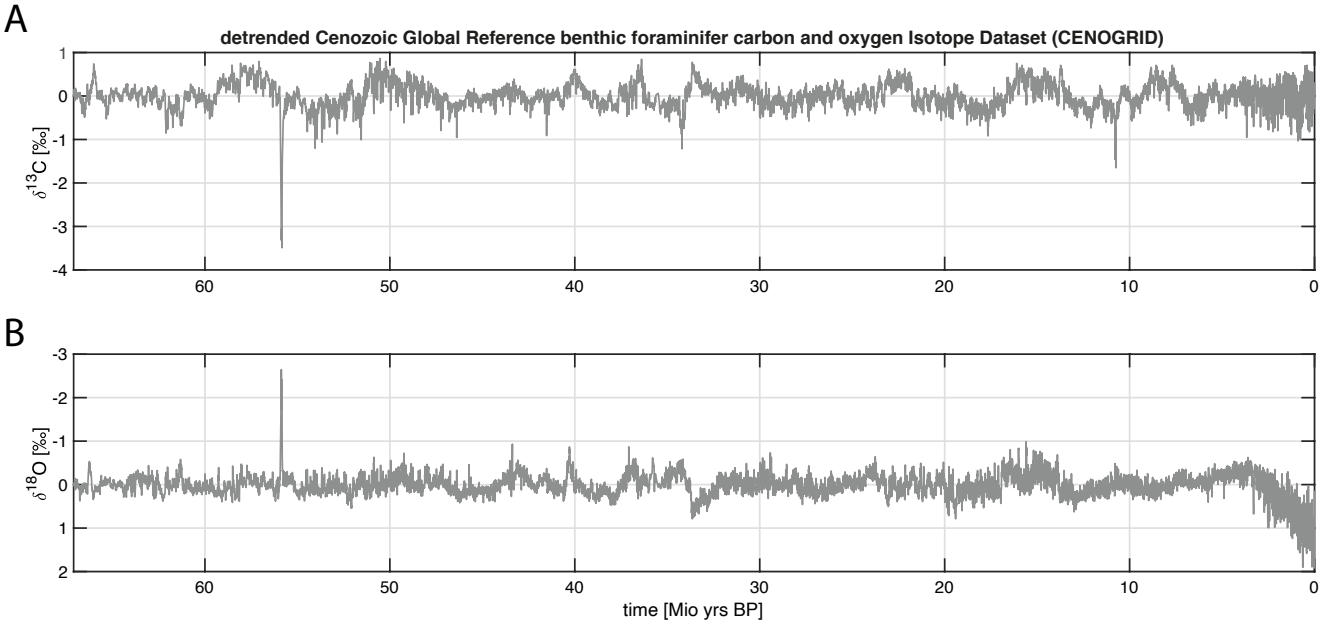


**Fig. 9** **A** Detrended $\delta^{13}$C and **B** $\delta^{18}$O time series of a total length of $N = 13,421$ samples and a sampling period of $\Delta t = 5,000$yrs [96].

**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 [13] | 0,14,28,42,56,70,84 | 1,1,1,1,1,1,1 |
| Kennel et al. [55] | 0,14,28,42,56,70 | 1,1,1,1,1,1 |
| Hegger and Kantz [39] | 0,14,28,42,56,70,84,98 | 1,1,1,1,1,1,1,1 |
| PECUZAL [58] | 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) [58] | 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 |
| 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 |

96(5):1174–1181, DOI https://doi.org/10.1890/14-1479.1, URL https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1890/14-1479.1, https://esajournals.onlinelibrary.wiley.com/doi/pdf/10.1890/14-1479.1

18. Coulom R (2007) Efficient selectivity and backup operators in monte-carlo tree search. In: van den Herik HJ, Ciancarini P, Donkers HHLMJ (eds) Computers and Games, Springer Berlin Heidelberg, Berlin, Heidelberg, pp 72–83
19. Datseris G (2018) Dynamicalsystems.jl: A julia software library for chaos and nonlinear dynamics. Journal of Open Source Software 3(23):598, DOI 10.21105/joss.00598, URL https://doi.org/10.21105/joss.00598
20. Dhadphale J, Unni VR, Saha A, Sujith RI (2021) Neural ode to model and prognose thermoacoustic instability. 2106.12758
21. Eckmann JP, Oliffson Kamphorst S, Ruelle D (1987) Recurrence Plots of Dynamical Systems. Europhysics Letters 4(9):973–977, DOI 10.1209/0295-5075/4/9/004
22. Eftekhari A, Yap HL, Wakin MB, Rozell CJ (2018) Stabilizing embedology: Geometry-preserving delay-coordinate maps. Phys Rev E 97:022222, DOI 10.1103/PhysRevE.97.022222, URL https://link.aps.org/doi/10.1103/PhysRevE.97.022222
23. Farmer JD, Sidorowich JJ (1987) Predicting chaotic time series. Phys Rev Lett 59:845–848, DOI 10.1103/PhysRevLett.59.845, URL https://link.aps.org/doi/10.1103/PhysRevLett.59.845
24. Feldhoff JH, Donner RV, Donges JF, Marwan N, Kurths J (2012) Geometric detection of coupling directions by means of inter-system recurrence networks. Physics Letters A 376(46):3504–3513, DOI 10.1016/j.physleta.2012.10.008
25. Fraser AM (1989) Reconstructing attractors from scalar time series: A comparison of singular system and redundancy criteria. Physica D: Nonlinear Phenomena 34(3):391 – 404, DOI https://doi.org/10.1016/0167-2789(89)90263-7, URL http://www.sciencedirect.com/science/article/

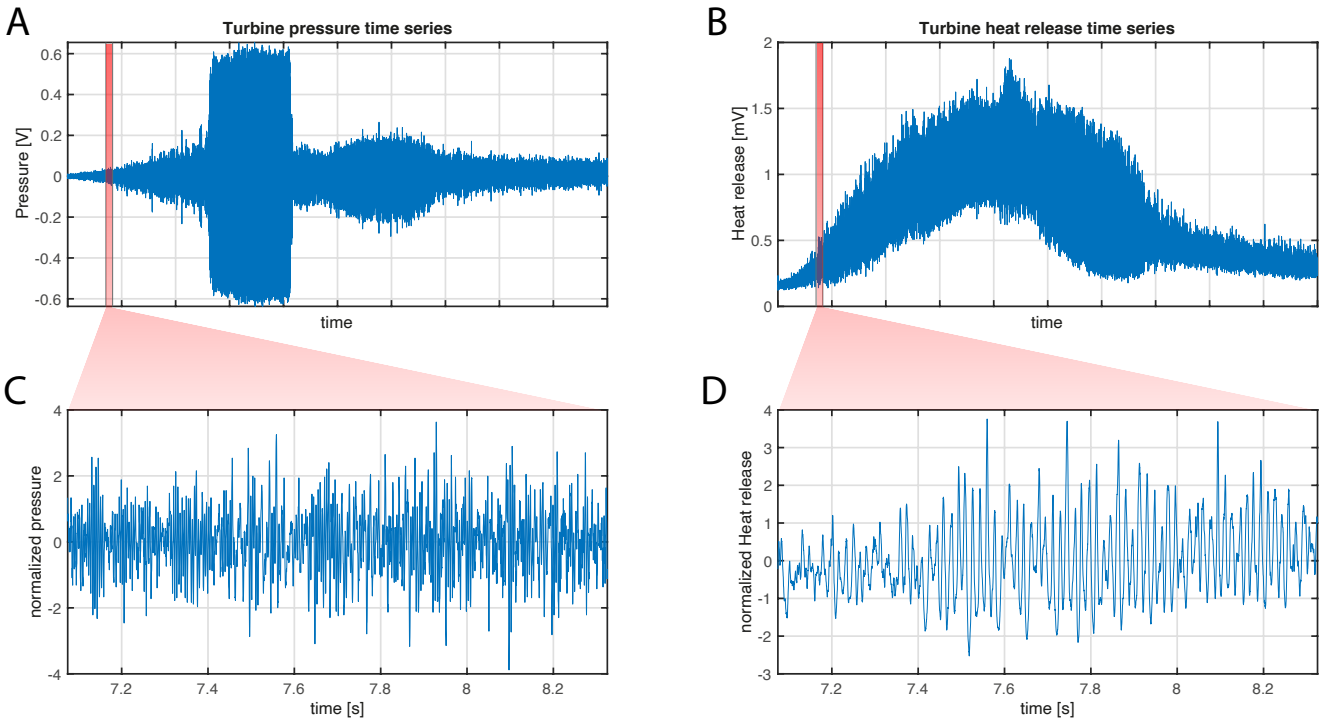


**Fig. 10** Entire time series of length $N' = 400,000$ of **A** turbine pressure (measured in V) and **B** combustion heat release (measured in mV). This is a non-stationary setup with a linearly varying control parameter (air flow rate) leading to bifurcations and different dynamics. For the calculations in Section 3.4 we sampled both time series 50 times in subsamples of length $N = 5,000$. 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.

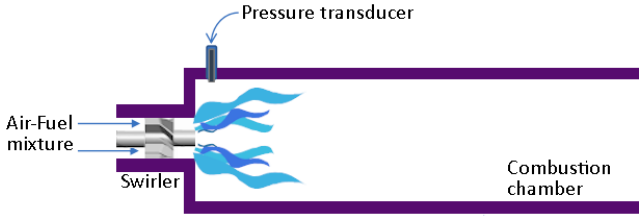


**Fig. 11** Schematic experimental setup of the combustion experiment, see main text for details.

pii/0167278989902637
26. Fraser AM, Swinney HL (1986) Independent coordinates for strange attractors from mutual information. Phys Rev A 33:1134–1140, DOI 10.1103/PhysRevA.33.1134, URL https://link.aps.org/doi/10.1103/PhysRevA.33.1134
27. Gao J, Zheng Z (1993) Local exponential divergence plot and optimal embedding of a chaotic time series. Physics Letters A 181(2):153 – 158, DOI https://doi.org/10.1016/0375-9601(93)90913-K, URL http://www.sciencedirect.com/science/article/pii/037596019390913K
28. Gao J, Zheng Z (1994) Direct dynamical test for deterministic chaos and optimal embedding of a chaotic time series. Phys Rev E 49:3807–3814, DOI 10.1103/PhysRevE.49.3807, URL https://link.aps.org/doi/10.1103/PhysRevE.49.3807
29. Garcia SP, Almeida JS (2005) Multivariate phase space reconstruction by nearest neighbor embedding with different time delays. Phys Rev E 72:027205, DOI 10.1103/PhysRevE.72.027205, URL https://link.aps.org/doi/10.1103/PhysRevE.72.027205
30. Garcia SP, Almeida JS (2005) Nearest neighbor embedding with different time delays. Phys Rev E 71:037204, DOI 10.1103/PhysRevE.71.037204, URL https://link.aps.org/doi/10.1103/PhysRevE.71.037204
31. Garland J, Bradley E (2015) Prediction in projection. Chaos: An Interdisciplinary Journal of Nonlinear Science 25(12):123108, DOI 10.1063/1.4936242, URL https://doi.org/10.1063/1.4936242, https://doi.org/10.1063/1.4936242
32. Gibson JF, Doyne Farmer J, Casdagli M, Eubank S (1992) An analytic approach to practical state space reconstruction. Physica D: Nonlinear Phenomena 57(1):1 – 30, DOI https://doi.org/10.1016/0167-2789(92)90085-2, URL http://www.sciencedirect.com/science/article/pii/0167278992900852
33. Godavarthi V, Pawar SA, Unni VR, Sujith RI, Marwan N, Kurths J (2018) Coupled interaction between unsteady flame dynamics and acoustic field in a turbulent combustor. Chaos: An Interdisciplinary Journal of Nonlinear Science 28(11):113111, DOI 10.1063/1.5052210, URL https://doi.org/10.1063/1.5052210, https://doi.org/10.1063/1.5052210
34. Granger CWJ (1969) Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37(3):424–438, URL http://www.jstor.org/stable/1912791
35. Grassberger P, Procaccia I (1983) Estimation of the kolmogorov entropy from a chaotic signal. Phys Rev A 28:2591–2593, DOI 10.1103/PhysRevA.28.2591, URL https://link.aps.org/doi/10.1103/PhysRevA.28.2591
36. Grassberger P, Procaccia I (1983) Measuring the strangeness of strange attractors. Physica D: Nonlinear Phenomena 9(1):189–208, DOI https://doi.org/10.1016/0167-2789(83)90298-1, URL https://www.sciencedirect.com/science/article/pii/0167278983902981

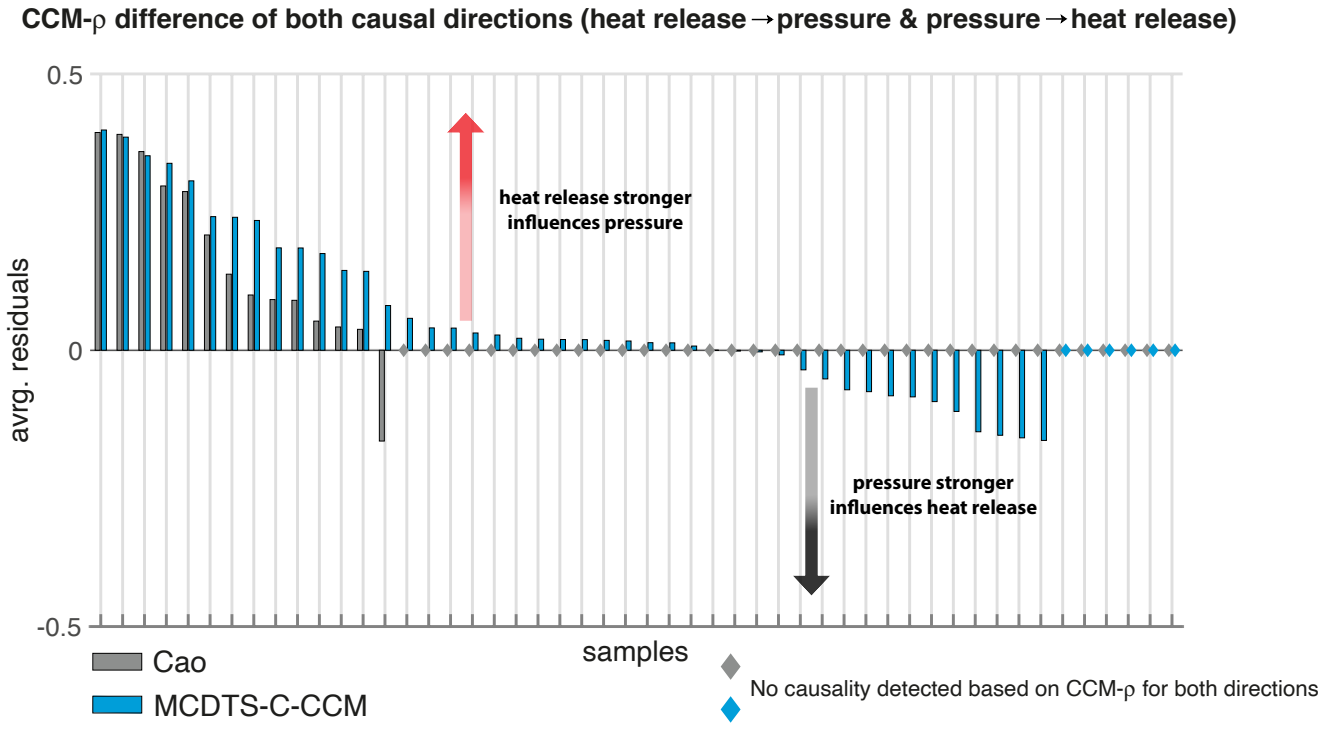


**Fig. 12** Average pointwise difference of the CCM-correlation coefficients for the direction *heat release* → *pressure* and vice versa for both underlying reconstruction approaches. For a better visualization we sorted these values here separately for both methods. A positive value indicates that 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 Figure).


37. Grassberger P, Schreiber T, Schaffrath C (1991) Nonlinear time sequence analysis. International Journal of Bifurcation and Chaos 01(03):521–547, DOI 10.1142/S0218127491000403, URL https://doi.org/10.1142/S0218127491000403, https://doi.org/10.1142/S0218127491000403
38. Han M, Ren W, Xu M, Qiu T (2019) Nonuniform state space reconstruction for multivariate chaotic time series. IEEE Transactions on Cybernetics 49(5):1885–1895
39. Hegger R, Kantz H (1999) Improved false nearest neighbor method to detect determinism in time series data. Phys Rev E 60:4970–4973, DOI 10.1103/PhysRevE.60.4970, URL https://link.aps.org/doi/10.1103/PhysRevE.60.4970
40. Hénon M (1976) A two-dimensional mapping with a strange attractor. Communications in Mathematical Physics 50(1):69–77, DOI 10.1007/BF01608556, URL https://doi.org/10.1007/BF01608556
41. Hentschel H, Procaccia I (1983) The infinite number of generalized dimensions of fractals and strange attractors. Physica D: Nonlinear Phenomena 8(3):435–444, DOI https://doi.org/10.1016/0167-2789(83)90235-X, URL https://www.sciencedirect.com/science/article/pii/016727898390235X
42. Hirata Y, Aihara K (2017) Dimensionless embedding for nonlinear time series analysis. Phys Rev E 96:032219, DOI 10.1103/PhysRevE.96.032219, URL https://link.aps.org/doi/10.1103/PhysRevE.96.032219
43. Hirata Y, Suzuki H, Aihara K (2006) Reconstructing state spaces from multivariate data using variable delays. Phys Rev E 74:026202, DOI 10.1103/PhysRevE.74.026202, URL https://link.aps.org/doi/10.1103/PhysRevE.74.026202
44. Holstein D, Kantz H (2009) Optimal markov approximations and generalized embeddings. Phys Rev E 79:056202, DOI 10.1103/PhysRevE.79.056202, URL https://link.aps.org/doi/10.1103/PhysRevE.79.056202
45. Isensee J, Datseris G, Parlitz U (2020) Predicting spatio-temporal time series using dimension reduced local states. Journal of Nonlinear Science 30(3):713–735, DOI 10.1007/s00332-019-09588-7, URL https://doi.org/10.1007/s00332-019-09588-7
46. Jia Z, Lin Y, Liu Y, Jiao Z, Wang J (2020) Refined nonuniform embedding for coupling detection in multivariate time series. Phys Rev E 101:062113, DOI 10.1103/PhysRevE.101.062113, URL https://link.aps.org/doi/10.1103/PhysRevE.101.062113
47. Judd K, Mees A (1998) Embedding as a modeling problem. Physica D: Nonlinear Phenomena 120(3):273 – 286, DOI https://doi.org/10.1016/S0167-2789(98)00089-X, URL http://www.sciencedirect.com/science/article/pii/S016727899800089X
48. Juniper MP, Sujith RI (2018) Sensitivity and nonlinearity of thermoacoustic oscillations. Annual Review of Fluid Mechanics 50:661–689
49. Kantz H (1994) A robust method to estimate the maximal lyapunov exponent of a time series. Physics Letters A 185(1):77–87, DOI https://doi.org/10.1016/0375-9601(94)90991-1, URL https://www.sciencedirect.com/science/article/pii/0375960194909911
50. Kantz H, Schreiber T (2003) Nonlinear Time Series Analysis, 2nd edn. Cambridge University Press, DOI 10.1017/CBO9780511755798

51. Kantz H, Schürmann T (1996) Enlarged scaling ranges for the ks-entropy and the information dimension. Chaos: An Interdisciplinary Journal of Nonlinear Science 6(2):167–171, DOI 10.1063/1.166161, URL https://doi.org/10.1063/1.166161, https://doi.org/10.1063/1.166161
52. Kantz H, Schreiber T, Hoffmann I, Buzug T, Pfister G, Flepp LG, Simonet J, Badii R, Brun E (1993) Nonlinear noise reduction: A case study on experimental data. Phys Rev E 48:1529–1538, DOI 10.1103/PhysRevE.48.1529, URL https://link.aps.org/doi/10.1103/PhysRevE.48.1529
53. Kember G, Fowler A (1993) A correlation function for choosing time delays in phase portrait reconstructions. Physics Letters A 179(2):72 – 80, DOI https://doi.org/10.1016/0375-9601(93)90653-H, URL http://www.sciencedirect.com/science/article/pii/037596019390653H
54. Kennel MB, Abarbanel HDI (2002) False neighbors and false strands: A reliable minimum embedding dimension algorithm. Phys Rev E 66:026209, DOI 10.1103/PhysRevE.66.026209, URL https://link.aps.org/doi/10.1103/PhysRevE.66.026209
55. Kennel MB, Brown R, Abarbanel HDI (1992) Determining embedding dimension for phase-space reconstruction using a geometrical construction. Phys Rev A 45:3403–3411, DOI 10.1103/PhysRevA.45.3403, URL https://link.aps.org/doi/10.1103/PhysRevA.45.3403
56. Kim H, Eykholt R, Salas J (1999) Nonlinear dynamics, delay times, and embedding windows. Physica D: Nonlinear Phenomena 127(1):48 – 60, DOI https://doi.org/10.1016/S0167-2789(98)00240-1, URL http://www.sciencedirect.com/science/article/pii/S0167278998002401
57. Kraemer KH, Donner RV, Heitzig J, Marwan N (2018) Recurrence threshold selection for obtaining robust recurrence characteristics in different embedding dimensions. Chaos: An Interdisciplinary Journal of Nonlinear Science 28(8):085720, DOI 10.1063/1.5024914, URL https://doi.org/10.1063/1.5024914, https://doi.org/10.1063/1.5024914
58. Kraemer KH, Datseris G, Kurths J, Kiss IZ, Ocampo-Espindola JL, Marwan N (2021) A unified and automated approach to attractor reconstruction. New Journal of Physics 23(3):033017, DOI 10.1088/1367-2630/abe336, URL https://doi.org/10.1088/1367-2630/abe336
59. Kugiumtzis D (1996) State space reconstruction parameters in the analysis of chaotic time series — the role of the time window length. Physica D: Nonlinear Phenomena 95(1):13 – 28, DOI https://doi.org/10.1016/0167-2789(96)00054-1, URL http://www.sciencedirect.com/science/article/pii/0167278996000541
60. Liebert W, Schuster H (1989) Proper choice of the time delay for the analysis of chaotic time series. Physics Letters A 142(2):107 – 111, DOI https://doi.org/10.1016/0375-9601(89)90169-2, URL http://www.sciencedirect.com/science/article/pii/0375960189901692
61. Liebert W, Pawelzik K, Schuster HG (1991) Optimal embeddings of chaotic attractors from topological considerations. Europhysics Letters (EPL) 14(6):521–526, DOI 10.1209/0295-5075/14/6/004, URL https://doi.org/10.1209%2F0295-5075%2F14%2F6%2F004
62. Lorenz E (1995) Predictability: A problem partly solved. In: Seminar on Predictability, 4-8 September 1995, ECMWF, ECMWF, Shinfield Park, Reading, vol 1, pp 1–18, URL https://www.ecmwf.int/node/10829
63. Lu Z, Hunt BR, Ott E (2018) Attractor reconstruction by machine learning. Chaos: An Interdisciplinary Journal of Nonlinear Science 28(6):061104, DOI 10.1063/1.5039508, URL https://doi.org/10.1063/1.5039508, https://doi.org/10.1063/1.5039508
64. Mañé R (1981) On the dimension of the compact invariant sets of certain non-linear maps. In: Rand D, Young LS (eds) Dynamical Systems and Turbulence, Warwick 1980, Springer Berlin Heidelberg, Berlin, Heidelberg, pp 230–242
65. Mann B, Khasawneh F, Fales R (2011) Using information to generate derivative coordinates from noisy time series. Communications in Nonlinear Science and Numerical Simulation 16(8):2999 – 3004, DOI https://doi.org/10.1016/j.cnsns.2010.11.011, URL http://www.sciencedirect.com/science/article/pii/S1007570410006003
66. March TK, Chapman SC, Dendy RO (2005) Recurrence plot statistics and the effect of embedding. Physica D 200(1–2):171–184, DOI 10.1016/j.physd.2004.11.002
67. Marwan N, Romano MC, Thiel M, Kurths J (2007) Recurrence Plots for the Analysis of Complex Systems. Physics Reports 438(5–6):237–329, DOI 10.1016/j.physrep.2006.11.001
68. Matassini L, Kantz H, Hołyst J, Hegger R (2002) Optimizing of recurrence plots for noise reduction. Phys Rev E 65:021102, DOI 10.1103/PhysRevE.65.021102, URL https://link.aps.org/doi/10.1103/PhysRevE.65.021102
69. Matilla-García M, Morales I, Rodríguez JM, Ruiz Marín M (2021) Selection of embedding dimension and delay time in phase space reconstruction via symbolic dynamics. Entropy 23(2), DOI 10.3390/e23020221, URL https://www.mdpi.com/1099-4300/23/2/221
70. Nair V, Thampi G, Sujith R (2014) Intermittency route to thermoacoustic instability in turbulent combustors. Journal of Fluid Mechanics 756:470–487
71. Nichkawde C (2013) Optimal state-space reconstruction using derivatives on projected manifold. Phys Rev E 87:022905, DOI 10.1103/PhysRevE.87.022905, URL https://link.aps.org/doi/10.1103/PhysRevE.87.022905
72. Packard NH, Crutchfield JP, Farmer JD, Shaw RS (1980) Geometry from a time series. Phys Rev Lett 45:712–716, DOI 10.1103/PhysRevLett.45.712, URL https://link.aps.org/doi/10.1103/PhysRevLett.45.712
73. Parlitz U (1992) Identification of true and spurious lyapunov exponents from time series. International Journal of Bifurcation and Chaos 02(01):155–165, DOI 10.1142/S0218127492000148, URL https://doi.org/10.1142/S0218127492000148, https://doi.org/10.1142/S0218127492000148
74. Parlitz U, Merkwirth C (2000) Prediction of spatiotemporal time series based on reconstructed local states. Phys Rev Lett 84:1890–1893, DOI 10.1103/PhysRevLett.84.1890, URL https://link.aps.org/doi/10.1103/PhysRevLett.84.1890
75. Pecora LM, Carroll TL, Heagy JF (1995) Statistics for mathematical properties of maps between time series embeddings. Phys Rev E 52:3420–3439, DOI 10.1103/PhysRevE.52.3420, URL https://link.aps.org/doi/10.1103/PhysRevE.52.3420
76. Pecora LM, Moniz L, Nichols J, Carroll TL (2007) A unified approach to attractor reconstruction. Chaos: An Interdisciplinary Journal of Nonlinear Science 17(1):013110, DOI 10.1063/1.2430294, URL https://doi.org/10.1063/1.2430294, https://doi.org/10.1063/1.2430294
77. Perinelli A, Ricci L (2018) Identification of suitable embedding dimensions and lags for time series generated by chaotic, finite-dimensional systems. Phys Rev E 98:052226, DOI 10.1103/PhysRevE.98.052226, URL https://link.aps.org/doi/10.1103/PhysRevE.98.052226
78. Rackauckas C, Nie Q (2017) Differentialequations.jl–a performant and feature-rich ecosystem for solving differential equations in julia. Journal of Open Research Software 5(1)

79. Ragwitz M, Kantz H (2000) Detecting non-linear structure and predicting turbulent gusts in surface wind velocities. Europhysics Letters (EPL) 51(6):595–601, DOI 10.1209/epl/i2000-00379-x, URL https://doi.org/10.1209/epl/i2000-00379-x
80. Rosenstein MT, Collins JJ, Luca] CJD (1994) Reconstruction expansion as a geometry-based framework for choosing proper delay times. Physica D: Nonlinear Phenomena 73(1):82 – 98, DOI https://doi.org/10.1016/0167-2789(94)90226-7, URL http://www.sciencedirect.com/science/article/pii/0167278994902267
81. Sauer T, Yorke JA, Casdagli M (1991) Embedology. Journal of Statistical Physics 65(3):579–616, DOI 10.1007/BF01053745, URL https://doi.org/10.1007/BF01053745
82. Schiecke K, Pester B, Feucht M, Leistritz L, Witte H (2015) 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, DOI 10.1109/EMBC.2015.7320106
83. Silver D, Huang A, Maddison CJ, 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 (2016) Mastering the game of go with deep neural networks and tree search. Nature 529(7587):484–489, DOI 10.1038/nature16961, URL https://doi.org/10.1038/nature16961
84. Small M, Tse C (2004) Optimal embedding parameters: a modelling paradigm. Physica D: Nonlinear Phenomena 194(3):283 – 296, DOI https://doi.org/10.1016/j.physd.2004.03.006, URL http://www.sciencedirect.com/science/article/pii/S0167278904001186
85. Sugihara G, May RM (1990) Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. Nature 344(6268):734–741, DOI 10.1038/344734a0, URL https://doi.org/10.1038/344734a0
86. Sugihara G, May R, Ye H, Hsieh Ch, Deyle E, Fogarty M, Munch S (2012) Detecting causality in complex ecosystems. Science 338(6106):496–500, DOI 10.1126/science.1227079, URL https://science.sciencemag.org/content/338/6106/496, https://science.sciencemag.org/content/338/6106/496.full.pdf
87. Sujith R, Unni VR (2020) Complex system approach to investigate and mitigate thermoacoustic instability in turbulent combustors. Physics of Fluids 32(6):061401
88. Takens F (1981) Detecting strange attractors in turbulence. In: Rand D, Young LS (eds) Dynamical Systems and Turbulence, Warwick 1980, Springer Berlin Heidelberg, Berlin, Heidelberg, pp 366–381
89. Tang Y, Krakovská A, Mezeiová K, Budáčová H (2015) Use of false nearest neighbours for selecting variables and embedding parameters for state space reconstruction. Journal of Complex Systems 2015:932750, DOI 10.1155/2015/932750, URL https://doi.org/10.1155/2015/932750
90. Theiler J (1986) Spurious dimension from correlation algorithms applied to limited time-series data. Phys Rev A 34:2427–2432, DOI 10.1103/PhysRevA.34.2427, URL https://link.aps.org/doi/10.1103/PhysRevA.34.2427
91. Tony J, Gopalakrishnan E, Sreelekha E, Sujith R (2015) Detecting deterministic nature of pressure measurements from a turbulent combustor. Physical Review E 92(6):062902
92. TSONIS AA (2007) Reconstructing dynamics from observables: The issue of the delay parameter revisited. International Journal of Bifurcation and Chaos 17(12):4229–4243, DOI 10.1142/S0218127407019913, URL https://doi.org/10.1142/S0218127407019913, https://doi.org/10.1142/S0218127407019913
93. Uzal LC, Grinblat GL, Verdes PF (2011) Optimal reconstruction of dynamical systems: A noise amplification approach. Phys Rev E 84:016223, DOI 10.1103/PhysRevE.84.016223, URL https://link.aps.org/doi/10.1103/PhysRevE.84.016223
94. Vlachos I, Kugiumtzis D (2010) Nonuniform state-space reconstruction and coupling detection. Phys Rev E 82:016207, DOI 10.1103/PhysRevE.82.016207, URL https://link.aps.org/doi/10.1103/PhysRevE.82.016207
95. Wendi D, Marwan N, Merz B (2018) In search of determinism-sensitive region to avoid artefacts in recurrence plots. International Journal of Bifurcation and Chaos 28(1):1850007, DOI 10.1142/S0218127418500074
96. Westerhold T, Marwan N, Drury AJ, Liebrand D, Agnini C, Anagnostou E, Barnet JSK, Bohaty SM, De Vleeschouwer D, Florindo F, Frederichs T, Hodell DA, Holbourn AE, Kroon D, Lauretano V, Littler K, Lourens LJ, Lyle M, Pälike H, Röhl U, Tian J, Wilkens RH, Wilson PA, Zachos JC (2020) An astronomically dated record of earth's climate and its predictability over the last 66 million years. Science 369(6509):1383–1387, DOI 10.1126/science.aba6853, URL https://science.sciencemag.org/content/369/6509/1383, https://science.sciencemag.org/content/369/6509/1383.full.pdf
97. Whitney H (1936) Differentiable manifolds. Annals of Mathematics 37(3):645–680, URL http://www.jstor.org/stable/1968482
98. Ye H, Deyle ER, Gilarranz LJ, Sugihara G (2015) Distinguishing time-delayed causal interactions using convergent cross mapping. Scientific Reports 5(1):14750, DOI 10.1038/srep14750, URL https://doi.org/10.1038/srep14750
99. Zou Y, Donner RV, Marwan N, Donges JF, Kurths J (2019) Complex network approaches to nonlinear time series analysis. Physics Reports 787:1–97, DOI 10.1016/j.physrep.2018.10.005
100. Čenys A, Pyragas K (1988) Estimation of the number of degrees of freedom from chaotic time series. Physics Letters A 129(4):227 – 230, DOI https://doi.org/10.1016/0375-9601(88)90355-6, URL http://www.sciencedirect.com/science/article/pii/0375960188903556

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- fighenonappendix.pdf