Implementation of two causal methods based on predictions in reconstructed state spaces

Anna Krakovská* and Jozef Jakubík
Institute of Measurement Science, Slovak Academy of Sciences, 841 04 Bratislava, Slovakia

(Received 17 April 2020; accepted 13 July 2020; published 7 August 2020)

If deterministic dynamics is dominant in the data, then methods based on predictions in reconstructed state spaces can serve to detect causal relationships between and within the systems. Here we introduce two algorithms for such causal analysis. They are designed to detect causality from two time series but are potentially also applicable in a multivariate context. The first method is based on cross-predictions, and the second one on the so-called mixed predictions. In terms of performance, the cross-prediction method is considerably faster and less prone to false negatives. The predictability improvement method is slower, but in addition to causal detection, in a multivariate scenario, it also reveals which specific observables can help the most if we want to improve prediction. The study also highlights cases where our methods and state-space approaches generally seem to lose reliability. We propose a new perspective on these situations, namely that the variables under investigation have weak observability due to the complex nonlinear information flow in the system. Thus, in such cases, the failure of causality detection cannot be attributed to the methods themselves but to the use of data that do not allow reliable reconstruction of the underlying dynamics.

DOI: 10.1103/PhysRevE.102.022203

I. INTRODUCTION

The best-known method for the detection of coupling between time series is the Granger's test, which evaluates the causal relations by a study of predictability in autoregressive models [1]. However, due to the linear concept, the Granger's method is not suitable for the analysis of complex nonlinear processes. Therefore, there is a growing body of work in the area of causality detection in nonlinear dynamical systems.

Some of the new methods that are generalizations of the Granger's prediction-based approach have been studied by Faes *et al.* in Ref. [2]. Although the results have been partially promising, the authors have pointed out that inferring directionality from asymmetries observed from nonlinear interdependence measures is often burdensome. It may be intricately related to differences between the dynamical properties of the systems and the strength of the coupling.

In 2018, a selection from the most promising causal methods, namely the extended Granger's test [3], the kernel version of the Granger's test [4], the conditional mutual information [5] or transfer entropy [6], two variants of the evaluation of cross mappings between state spaces [7,8], and an assessment of predictability improvement due to the use of mixed predictions [9], has been compared [10]. For data from complex dynamical systems, the results have been mixed. The method based on predictability improvement (PI) in reconstructed state spaces has been the only one to produce false-negative results on the tested data (16%). On the other hand, PI was also the only one that did not give any false positives, unlike some other methods, which saw causality in 68–94% of cases where none existed.

As it becomes increasingly clear that prediction evaluation has excellent potential in causal analysis, we will focus on these types of methods here. Specifically, we will provide detailed instructions for two implementations, both of which will make use of predictions in reconstructed state spaces.

The first method will use cross-predictions—predictions of the system based only on historical data from the other system. As far as we know, this idea was first used in the context of interconnection and synchronization in a 1995 article by Rulkov *et al.* [11] and a 1996 article by Schiff *et al.* [12].

The second method will evaluate whether the prediction of one series improves after using the so-called mixed prediction, combining past information from both time series. The ability of mixed forecasts to detect weak interactions below the threshold of synchronization was demonstrated in the example of two coupled Hénon maps by Wiesenfeldt *et al.* in 2002 [13] and Feldman and Bhattacharya in 2004 [14]. We will use the mixed predictions similarly as we proposed in 2016 [9].

In the following, we first explain the methodologies through an example of interconnected Rössler oscillators [15]. Then detailed implementation guides for the two causality detection methods are provided, and the results obtained by testing the coupled Rössler systems are presented. A discussion about observability related issues follows the results. Finally, in Appendix A, pseudocodes are included, allowing easy implementation of the methods in Matlab or other software environments.

II. DATA

Rössler 1.015 → Rössler 0.985

As the test example, we consider the unidirectional coupling of two Rössler systems, studied in detail by Paluš and

*krakovska@savba.sk

Vejmelka [5]. The equations for the driving system X and the driven system Y are as follows:

$$\begin{aligned}
\dot{x}_1 &= -\omega_1 x_2 - x_3 \\
\dot{x}_2 &= \omega_1 x_1 + 0.15 x_2 \\
\dot{x}_3 &= 0.2 + x_3(x_1 - 10) \\
\dot{y}_1 &= -\omega_2 y_2 - y_3 + C(x_1 - y_1) \\
\dot{y}_2 &= \omega_2 y_1 + 0.15 y_2 \\
\dot{y}_3 &= 0.2 + y_3(y_1 - 10).
\end{aligned} \tag{1}$$

In our example, the two Rössler systems are coupled through a one-way driving relationship between variables x_1 and y_1. X drives Y through the coupling term $C(x_1 - y_1)$ in the right-hand side of the equation for the y_1 variable. The parameter C states the coupling strength.

We use $\omega_1 = 1.015$ and $\omega_2 = 0.985$ in our test example. With this setting, the driver X is chaotic, while the driven system Y does not seem so, as its largest Lyapunov exponent is about 0 in the uncoupled case [5]. However, once the coupling with the chaotic driver X is active, also Y becomes a part of the overall chaotic dynamics.

With increasing coupling, the positive Lyapunov exponent of the response system (also known as the conditional Lyapunov exponent) can decrease and become negative at the C value at which synchronization takes place. The plots of the conditional Lyapunov exponents for the system given by Eq. (1) can be found in Ref. [5]. The plots indicate, similarly as the graph of the correlation dimensions presented in Ref. [16], that synchronization takes place for a coupling about $C = 0.12$.

As for the synchronization of coupled chaotic oscillators, they are often studied in terms of phase synchronizations. They occur when the coupled systems keep their phase difference bounded while their amplitudes are uncorrelated. If the studied coupling can be reflected in the dynamics of instantaneous phases, then the use of the phase synchronization perspective can decrease the dimensionality of the causality detection problem [5]. However, in this study, the term synchronization refers to the so-called generalized synchronization. This means that there is a one-to-one functional relation between the states of the systems, so the dynamical state of one system is completely determined by the state of the other. If the function is an identity, then the synchronization is called identical. It is essential to the causal methods that the direction of coupling can be inferred from data only when the underlying systems are coupled but not yet synchronized. Once the systems are synchronized, the future states of the driver X can be predicted from the past of X equally well as from the response Y and vice versa, and it is not possible to decide whether we are dealing with synchronized systems, highly correlated systems, or strongly bidirectionally interconnected systems [17].

Through our example of coupled Rössler systems, we can test different types of causal relations: causally independent time series ($C = 0$), unidirectional driving with increasing coupling strengths (C between zero and 0.12), synchronized states ($C > 0.12$), and bidirectional coupling (if investigating

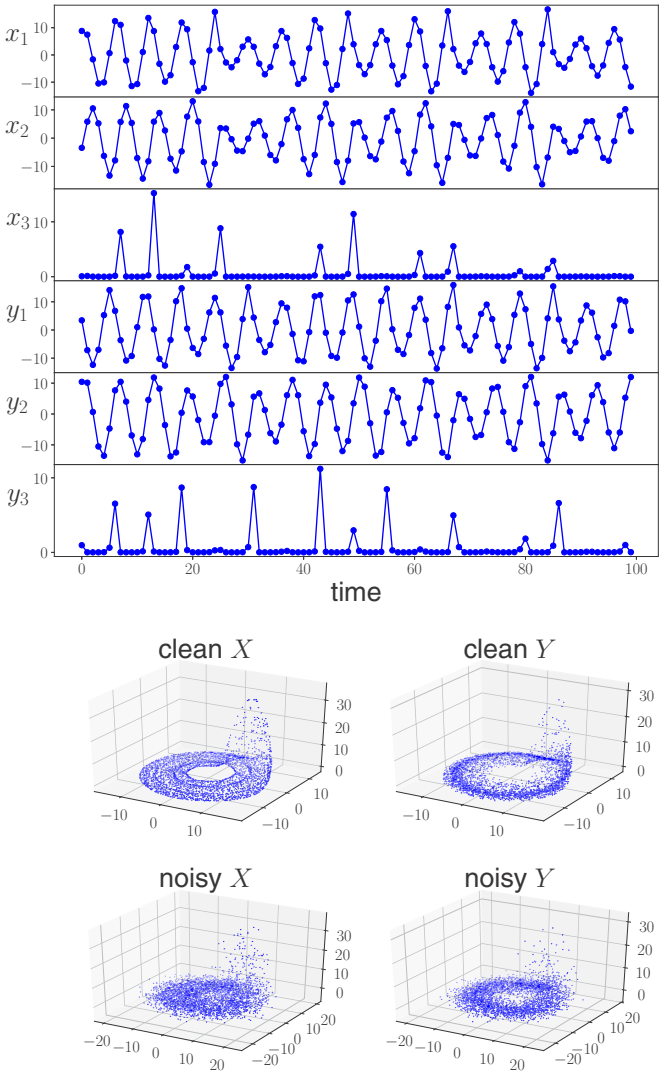

FIG. 1. Presentation of the test data. One hundred point samples of variables of Rössler systems in Eq. (1) coupled with the strength of $C = 0.05$. Followed by three-dimensional plots of 3600 points given by variables from the driver X and variables from the driven Y. The plots at the bottom are for the data with added noise (signal-to-noise ratio of 10).

two variables from the driver X or two from the responding subsystem Y).

We got our test time series as solutions of ordinary differential Eqs. (1) obtained by numerical integration using the adaptive Runge-Kutta method in Matlab. An integration step size of 0.1 and starting point [0,0,0.4,0,0,0.4] was used. We dropped the first 1000 points and saved every 10th value of the remaining data. This gave us about 6 points per one average orbit around the attractor. We repeated the numerical integration for coupling strengths C ranging from 0 to 0.18 with a step of 0.01. Figure 1 shows 100 points of each variable of Eq. (1), for coupling $C = 0.05$.

To characterize the relations between variables and subsystems in Eq. (1) we can also use a so-called interaction or fluence graph. We construct the fluence graph of our system in Fig. 2 by drawing a directed link from variable x to y whenever

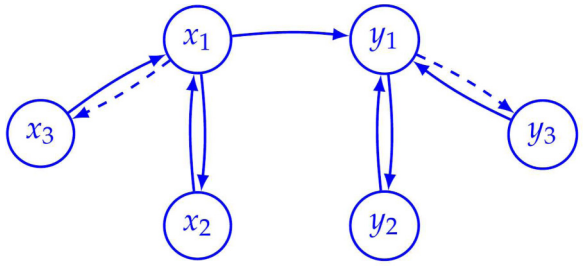

FIG. 2. Fluence graph of the coupled Rössler systems given by Eq. (1). The situation for interconnected systems with coupling C below the synchronization threshold. Dashed arrows symbolize nonlinearity in the connections.

x appears in the differential equation of y, that is, when the corresponding element of the Jacobian matrix of Eq. (1) is nonzero.

If we assume knowledge of one variable from X and one from Y, then our test example imitates a common practical problem of having two time series, representing two dynamical systems, with the task of finding out if there is a causal relationship between the systems.

$X \to Y$ denotes driving Y by X, and that is the causal link we would like to recover here if C is between zero and 0.12. Can we do this if we only know measurements of x_3 and y_3, for example, or some other pair of variables?

Cummins *et al.* have introduced a comprehensive theory showing that if we use causal methods requiring delay reconstruction techniques on deterministic systems, then the methods have limited capability [17]. The best we can hope for is finding the strongly connected components of the graph (sets of mutually reachable vertices) which represent distinct subsystems coupled through one-way driving links. We cannot recover self-loops, and we cannot distinguish the direct driving, indirect driving or correlate of a direct or indirect driving.

Suppose for a moment that we have time-series data for all six variables of our test example of coupled Rössler systems, but we do not know the relationships between them. In that case, complete pairwise causal testing should reveal all nine links in Fig. 2. However, in addition to these, we will also find 12 more links, since we cannot distinguish between direct and indirect driving. For example, we will find a bidirectional connection between x_2 and x_3 and between y_2 and y_3. We also find any variable from X driving any variable from Y. So, in summary, we would correctly determine that x_1, x_2, and x_3 form one subsystem X, while y_1, y_2, and y_3 form the second subsystem Y, and $X \to Y$. However, the exact position of the direct driving link ($x_1 \to y_1$) cannot be determined.

Another limitation concerns the possible presence of hidden variables. Most causal inference methods for time series are grounded on the assumptions that all common drivers are observed. If there are unmeasured common causes, then they may be responsible, for example, for the presence of correlated but causally unconnected variables. In the comparative article [10] we have used data from a fishery model as an example of two systems influenced by a common third, hidden driver. Most of the test methods have falsely detected bidirectional relation between the time series from the fishery model. The method based on the predictability improvement, on the other hand, (correctly) did not see any link. This was actually expected since the fact that one time series is correlated and causally unrelated to the other means that they share some common information, but it does not mean that there is any extra information needed to help predict the other series. However, in the presence of noise in the data the correlated observables might be able to help with each other's prediction and indicate a bidirectional link.

Although the main goal in this study is to explain how to use state-space predictions for causal analysis, we must emphasize that the danger of the above limitations must be taken into account when interpreting the results.

III. METHODS

The typical problem we are interested in here concerns two potentially coupled dynamical systems X and Y, of which we have no *a priori* knowledge of either their dynamics or their interdependence. Let each system, X and Y, be represented by a single time series, x and y, respectively.

Both of the presented causal methods will be based on predictions of the evolution of the systems in state spaces. To make such predictions when a system is represented by only one time series, the multidimensional state portrait of the dynamics has to be reconstructed first. Specifically, a d_X-dimensional manifold M_X is built from lags of observable x so that the state of the system in time t is

$$(x(t), x(t-\tau_X), x(t-2\tau_X), \ldots, x[t-(d_X-1)\tau_X]).$$

Using appropriate τ_Y and d_Y, also the manifold M_Y is reconstructed. According to Takens's theorem, under certain conditions, the reconstruction is equivalent, in the sense of diffeomorphism, to the original manifold [18]. Consequently, the reconstruction and the original share the same features in many ways. The most important thing for us is that the reconstruction is useful for predicting the system's evolution.

Theoretically, for noise-free data of unlimited length, the existence of a diffeomorphism between the original attractor and the reconstructed image is guaranteed for a sufficiently high embedding dimension m and almost any choice of delay τ. In practice, however, the data are of finite length and limited accuracy, and it is worthwhile to pay some attention to the selection of the embedding parameters d and τ. The essential requirement is to unfold the reconstruction of the trajectories sufficiently to avoid self-crossings and extreme closeness of distinct points. The best combinations of dimension and time delay are given by some optimum time window $(d-1)\tau$ [19]. As demonstrated in Ref. [20] for the same one-point prediction technique as will be used here, the size of the time window should not exceed the mean orbital period. Cyclicity in data can thus help with estimating the appropriate window. However, data can be broadband and lacking indication of any oscillatory pattern. Then, it is better to follow some proper invariant that is expected to reach an extremum for the appropriate embedding parameters. For example, we can calculate the prediction errors for several combinations of τ and m and choose the one that leads to the lowest error. Such a search for the embedding parameters will be included as a first step in the design of our causal algorithms.

Then, in the reconstructed state spaces, the self-predictions and cross-predictions or mixed predictions will be computed. To make predictions, we will use the method of analogs [21]. The method finds historical data similar to the current state and assumes that the system will continue analogously to what it did in the past. For example, to predict the follower of point Y_t in M_Y, we can find the time index i of its nearest neighbor from past states on the reconstructed trajectory and declare the follower's estimate $\hat{Y}_{t+1} = Y_{i+1}$. This simple procedure can be modified in various ways. In this study, we will average the followers of several neighbors while considering exponential weighting based on their distances from Y_t. More precisely, k nearest neighbors of point Y_t will be determined with K being a set of their time indices. For the neighbors Y_i, $i \in K$, exponential weighting w_i based on Euclidean distances from Y_t is considered:

$$w_i = e^{-\frac{\|Y_t - Y_i\|}{\min_{j\in K}\|Y_t - Y_j\|}}.$$

The prediction is then given by

$$\hat{Y}_{t+1} = \frac{\sum_{i\in K} w_i Y_{i+1}}{\sum_{i\in K} w_i}.$$

In this way, we obtain the self-prediction of y in M_Y and, analogously, the self-prediction of x in M_X, the mixed prediction of x in M_{XY}, and the mixed prediction of y in M_{YX}.

The way of cross-predicting (predicting the first system based on the history of the second system) is, of course, a bit different. To get the cross-prediction of some variable y for example, the indices of the neighbors of Y_t are taken from the neighbors of X_t found in M'_X. Note that M'_X is not the same as the manifold M_X, which was created to be optimal for self-predictions. Instead, in order to ensure better performance, we propose to find and use new reconstruction M'_X, guaranteeing minimal cross-prediction errors.

After all predictions are completed, the causality detection itself, based on the prediction errors $|y_{t+1} - \hat{y}_{t+1}|$, will proceed in one of the two following ways.

A. Cross-prediction method (CP method)

In the case of the CP method, we evaluate the mean absolute errors of the self-predictions of x in M_X and y in M_Y and the cross-predictions of x and y.

With this method, we use the next notation: $\mathcal{E}xX$ and $\mathcal{E}yY$ for mean absolute errors of self-predictions of x in M_X and y in M_Y, respectively, and $\mathcal{E}xY$ and $\mathcal{E}yX$ for mean absolute errors of cross-predictions of x and y, respectively.

To allow comparison of the predictability of two different time series we also use normalized errors with the corresponding notation being $n\mathcal{E}xX$, $n\mathcal{E}yY$, $n\mathcal{E}xY$, and $n\mathcal{E}yX$. Our choice for normalization is the mean absolute deviation.

Overall, we get the next four possible causal relations between X and Y with corresponding options for the prediction errors of the two self-predictions and two cross-predictions.

1. $X \perp\!\!\!\perp Y$ (X and Y are causally independent)

In the case of uncoupled and uncorrelated systems, X cannot be predicted from Y and vice versa. The normalized errors $n\mathcal{E}xY$ and $n\mathcal{E}yX$ are expected to be close to 1 as the cross-predictions are not better than taking the mean of the time series as the trivial estimate of the following value. This can be subjected to statistical testing of the null hypothesis H_0 that the cross-prediction does not overcome the trivial prediction against the alternative H_1 that it achieves lower errors.

2. $X \leftrightarrow Y$

For fully synchronized systems, as well as for bidirectionally coupled systems, the driver and the response are indistinguishable from each other. In such cases, due to cyclic flow of information, the causes and effects are entangled, which means that we can make predictions of x and y equally well in M_X as in M_Y. To decide whether this is the case, we can use a statistical test of the null hypothesis H_0 of the errors of cross-predictions being not significantly lower than the errors of the self-predictions against the alternative hypothesis H_1 of $\mathcal{E}xY < \mathcal{E}xX$ or $\mathcal{E}yX < \mathcal{E}yY$, respectively.

3. $X \rightarrow Y$

Provided that we ruled out the validity of either of the previous two cases, the remaining option is a one-way relationship. $n\mathcal{E}xY < n\mathcal{E}yX$ means that Y contains more information about X than vice versa, which, under the given conditions, is a sign of a unidirectional link from X to Y.

4. $Y \rightarrow X$

$n\mathcal{E}yX < n\mathcal{E}xY$ is the last option, corresponding to the unidirectional coupling, opposite to the previous case.

These rules are presented in a schematic form in Table I. Pseudocode of the CP algorithm can be found in Appendix A.

B. Predictability improvement method (PI method)

The second method is based on improving predictability with mixed predictions. This time, to detect the type of causality, we evaluate two self-predictions and two mixed predictions in a similar way as in Ref. [9]. The basic idea goes back to the so-called mixed-state analysis introduced in Refs. [13,14].

Again, suppose that we have two systems, X and Y, and each one is represented by one time series, x, and y, respectively. Similarly, as in the case of Granger's meaning, we say that variable x causes variable y if a better prediction of y can be obtained by using the information from both x and y rather than only the information from y. However, unlike the Granger's method, this one operates in multidimensional state spaces.

The original version of our PI method, introduced in Ref. [9], has attributed special attention to the optimization of the reconstructed manifolds based on the weighting of all their time-delayed coordinates. Less useful coordinates of the reconstructions were suppressed by assigning a lower weight. However, the optimization problem related to the adjustments of weights is a challenging and computationally expensive task. To avoid this complication, we used a simpler variant in this study.

Here the mixed prediction of y, i.e., the prediction using information from both Y and X, will be the prediction

TABLE I. Summary of the rules for deriving causal relationships between systems X and Y based on errors and normalized errors of self-predictions and cross-predictions ($\mathcal{E}xX$, $\mathcal{E}yY$, $\mathcal{E}xY$, $\mathcal{E}yX$, $n\mathcal{E}xY$, $n\mathcal{E}yX$) in the CP method and errors of self-predictions and mixed predictions ($\mathcal{E}xX$, $\mathcal{E}yY$, $\mathcal{E}xXY$, $\mathcal{E}yYX$) in the PI method.

Causal relation	CP method	PI method
$X \perp\!\!\!\perp Y$	$n\mathcal{E}xY \approx 1$ $n\mathcal{E}yX \approx 1$	$\mathcal{E}xXY \geqslant \mathcal{E}xX$ $\mathcal{E}yYX \geqslant \mathcal{E}yY$
$X \leftrightarrow Y$	$\mathcal{E}xY \leqslant \mathcal{E}xX \vee \mathcal{E}yX \leqslant \mathcal{E}yY$	$\mathcal{E}xXY < \mathcal{E}xX$ $\mathcal{E}yYX < \mathcal{E}yY$
$X \rightarrow Y$	$n\mathcal{E}xY < n\mathcal{E}yX$	$\mathcal{E}yYX < \mathcal{E}yY$ $\mathcal{E}xXY \geqslant \mathcal{E}xX$
$Y \rightarrow X$	$n\mathcal{E}yX < n\mathcal{E}xY$	$\mathcal{E}xXY < \mathcal{E}xX$ $\mathcal{E}yYX \geqslant \mathcal{E}yY$

in M_{YX}, where M_{YX} is made up of M_Y extended by the weighted M_X: $M_{YX} = [y(t), y(t-\tau_Y), \ldots, y(t-(d_Y-1)\tau_Y), w_X x(t), w_X x(t-\tau_X), \ldots, w_X x(t-(d_X-1)\tau_X)]$, where the weight w_X represents the impact of M_X. The task is to find out whether for some weight w_X the predictions of y in the mixed space M_{YX} are more accurate than the self-predictions in M_Y. Essential for the PI method is that an improvement in predictability can only occur if X affects Y. Otherwise, x coordinates will not help to improve the reconstruction or the prediction of the underlying dynamics.

After exchanging the roles of X and Y above we get instructions to compute the mixed predictions of x needed for the testing of causality in the opposite direction, i.e., from Y to X.

For the prediction errors in the PI method, we will use the following notation: $\mathcal{E}xX$ and $\mathcal{E}yY$ for mean absolute errors of self-predictions of x in M_X and y in M_Y, respectively, and $\mathcal{E}xXY$ and $\mathcal{E}yYX$ for mean absolute errors of mixed predictions of x in M_{XY} and y in M_{YX}, respectively.

The essence of the PI method is to determine if the mixed prediction errors can be lower than the self-prediction errors, meaning that one variable helps the other with the prediction. For example, to assess whether adding information from X improves the prediction of Y, the null hypothesis of no predictability improvement, i.e., $H_0 : \mathcal{E}yYX \geqslant \mathcal{E}yY$, is statistically tested against the alternative hypothesis $H_1 : \mathcal{E}yYX < \mathcal{E}yY$. If H_0 is rejected on a 5% significance level, then we accept that the inclusion of the knowledge of X significantly improves the prediction of Y. Analogously, we test the opposite direction (causal connection from Y to X).

The test results lead us to the next four possible causal connections between X and Y with corresponding relations between the prediction errors of the self-predictions and mixed predictions.

1. $X \perp\!\!\!\perp Y$

The first option is that prediction of neither observable can be improved using information from the other system. Therefore, we declare systems X and Y causally independent when the null hypothesis H_0 of no predictability improvement is not rejected for either of them.

2. $X \leftrightarrow Y$

Once the systems are synchronized or bidirectionally connected, we can understand the observables as part of the same dynamic system with cyclic information flow. In such a case, the mixed delayed reconstruction M_{XY} may contain more information about the underlying dynamics than M_X alone, resulting in a more accurate prediction of the x observable compared to its self-prediction. The same goes for improving the predictability of y. In practice, however, despite the bidirectional connection, we may see a predictability improvement in only one direction. Such cases are then confused with unidirectional links. This asymmetry is related to the phenomenon of the so-called observability of the variable, which will be discussed later.

3. $X \rightarrow Y$

If we ruled out the previous two cases, then a unidirectional causal link is the remaining option. If we can accept the alternative hypothesis that $\mathcal{E}yYX < \mathcal{E}yY$, then we conclude that there is a unidirectional link from X to Y.

4. $Y \rightarrow X$

After excluding all of the previous options, only $\mathcal{E}xXY < \mathcal{E}xX$ remains, corresponding to the unidirectional coupling between the driver Y and response X.

The pseudocode of the PI algorithm can also be found in Appendix A.

For both methods, the rules for inferring causal relations from the prediction errors are summarized in Table I.

IV. RESULTS

Let us now demonstrate how the methods just presented handle the detection of causal relationships in our test example of interconnected Rössler oscillators.

For example, suppose that we have a few thousand data points of variable $x = x_1$ of the driving Rössler system X and variable $y = y_1$ of the driven system Y and we would like to know whether there is a causal relationship between the two systems. In fact, we tested all pairs of time series selected from the six variables of the coupled systems, each pair for 19 different coupling values C. We also tested the noisy versions of all these time series, with a signal-to-noise ratio of 10. White Gaussian noise was added to mimic the effect of random processes typically present in real-world data. Since the noise is additive, it is not dependent on the state of the system.

As the first step, we found optimal embedding parameters for the tested time series x and y to build the reconstructions of the manifolds M_X and M_Y. To do so, we made predictions in manifolds given by several different combinations of parameters m and τ and selected the one that led to the lowest errors. For example, for variable x_1, the reconstruction M_X with time

delay $\tau_X = 1$ and dimension $d_X = 4$ was optimal. For variable y_1 from the driven subsystem, manifold M_Y given by delay $\tau_Y = 1$ and dimension $d_Y = 4, 5, 6,$ or 7 (the stronger the coupling, the higher the embedding dimension) proved to be optimal. In the CP method, we set the embedding parameters separately for self-predictions and cross-predictions. With the PI method, in addition to m and τ, we also optimized the impact w_X of the system X, and the impact w_Y of the system Y when creating the mixed manifolds M_{YX} and M_{XY}.

After reconstructing all the manifolds, we computed the self-predictions and cross-predictions for the CP method and the self-predictions and mixed predictions for the PI method. In each case, 600 points were predicted, with the help of 3000 historical points of the reconstructed trajectory.

Figures 3, 4 and Table II demonstrate the performance of the two methods. Let us review the results in more detail.

A. Results of the CP method in bivariate analysis

1. $X \perp\!\!\!\perp Y$

At zero coupling, we can take any variable from X and any one from Y to get an example of a pair of mutually causally independent time series. The errors for $C = 0$ in the top four pictures in Fig. 3 show the results of the CP method for the pair of x_1 and y_1 variables and the pair of x_3 and y_3. High cross-prediction errors both for clean and noisy data unmistakably show that there is no relationship between the systems.

2. $X \leftrightarrow Y$

As mentioned above, the systems become synchronized for coupling around the value $C = 0.12$. Above this threshold, the driven system is no longer distinguishable from the driver, and the observable from Y can be predicted just as well as the one from X in any of the reconstructed spaces M_X, M_Y. See the comparatively small self-prediction and cross-prediction errors for $C > 0.12$ in Figs. 3(a)–3(d). These enabled the CP method to detect the threshold of synchronization reliably.

As mentioned, in our context, the situation with synchronized systems is essentially the same as cases of bidirectional connections. Figure 3 presents two examples of bidirectional coupling. In the first case, we pretended to know only the historical data from x_1 and x_2. These variables form part of the driving Rössler system with a cyclical flow of information. As a result, the equal success of all types of predictions allowed the CP method easy detection of bidirectional causality in this case. See Figs. 3(e) and 3(f).

The second example of a two-way causality looks a bit more complicated. We used variables y_2 and y_3 that interact bidirectionally within the driven subsystem Y. The prediction errors are significantly higher than in the case of x_1 and x_2. This is partly due to the higher complexity of the driven system Y compared to the driving system X alone and, in the case of y_3, the predictability is further reduced due to poor observability of this particular variable (to be discussed later). Nevertheless, the CP method was able to detect bidirectional causality even in this case. In Figs. 3(g) and 3(h) we can see that the errors of cross-predictions of y_3 are lower than the errors of its self-predictions, which is only possible with bidirectional coupling.

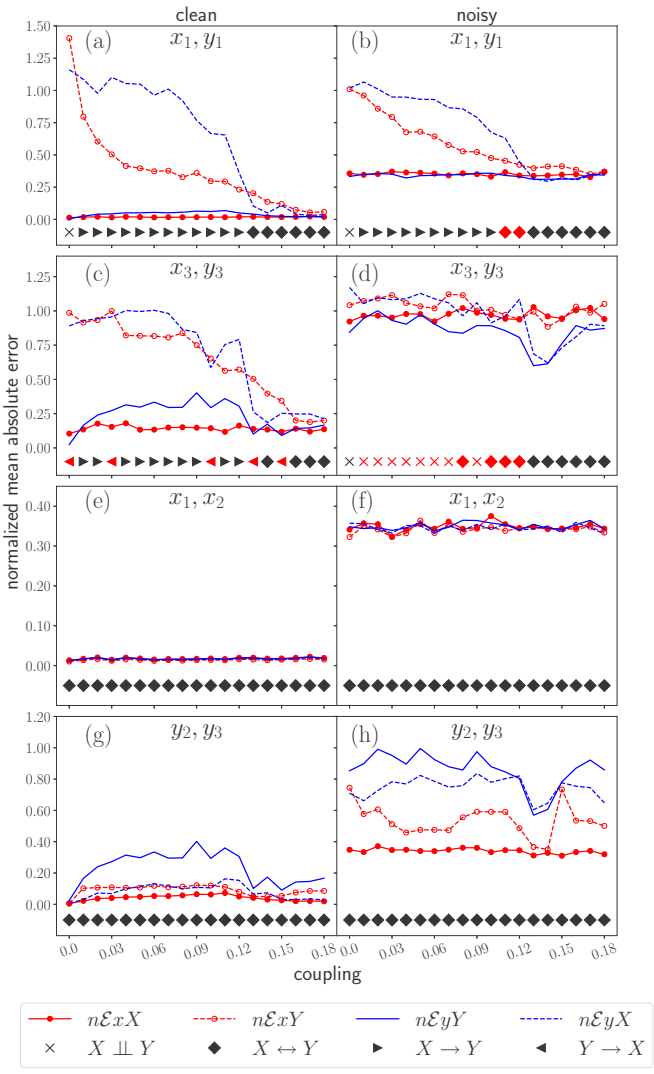

FIG. 3. Results of the CP method. Normalized mean absolute errors of self-predictions and cross-predictions for four pairs of variables of coupled Rössler systems given by Eq. (1). In each case, the results are given for 19 increasing values of the coupling strength C. Pictures on the left are for clean data, while pictures on the right are for data with a signal-to-noise ratio of 10. The figures also show concrete conclusions of the method about the type of causality, detected in individual cases. Red is used for symbols of incorrect detection results. Black is for correct detections.

3. $X \rightarrow Y$

In Figs. 3(a)–3(d), for coupling values higher than zero but lower than the synchronization threshold, we can see that $n\mathcal{E}xY < n\mathcal{E}yX$. Based on this, and after excluding $X \perp\!\!\!\perp Y$ and $X \leftrightarrow Y$, the CP method correctly concludes that system X causes system Y. We expected the same results when subjecting any one observable from the driving system X and one from the driven system Y to a causal test. However, the pictures for x_3 and y_3 in Figs. 3(c) and 3(d) show that with these problematic variables, causality detection was very unreliable.

To summarize, the CP method revealed the presence and direction of causal links in the case of x_1 and y_1 flawlessly. It

TABLE II. Summary of false-negative and false-positive results in the CP method and the PI method when applied to clean and noisy (signal-to-noise ratio of 10) data from coupled Rössler oscillators given by Eq. (1). In both clear and noisy cases, 570 causal links were tested. Good observability means that variables with high observability index (x_1, x_2, y_1, y_2) were used. Weak observability means that variable with low observability index (x_3 or y_3) was involved in the analysis.

		CP clean data	CP noisy data	PI clean data	PI noisy data
Good observability	False negatives	0%	0%	14%	5%
	False positives	0%	10%	3%	19%
Weak observability	False negatives	4%	6%	39%	15%
	False positives	7%	37%	0%	31%

was equally successful when any of the pairs (x_1, y_2), (x_2, y_1), or (x_2, y_2) were used for testing. However, if one of the time series tested was x_3 or y_3, or if we added noise to the data, revealing the one-way causality was no longer error free. On the other hand, the detection of bidirectional causality was without problems for any pair of variables and even at the signal-to-noise ratio of 10.

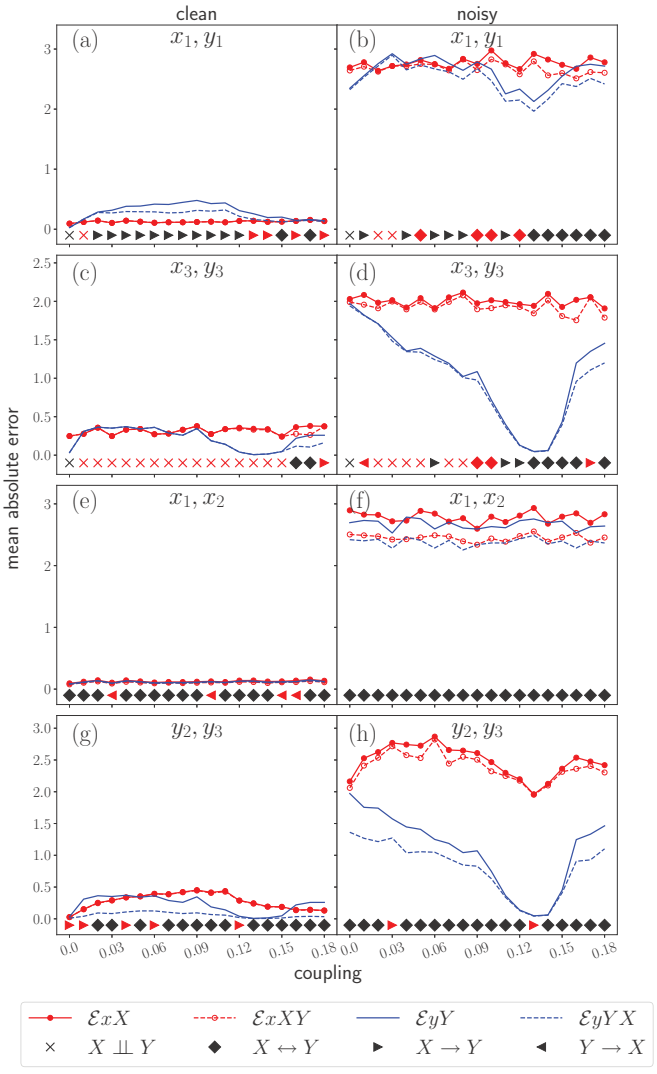

FIG. 4. Results of the PI method. Mean absolute errors of self-predictions and mixed predictions for four pairs of variables of coupled Rössler systems given by Eq. (1). As with the previous figure, red is used for symbols of incorrect detection results and black for correct detections.

B. Results of the PI method in bivariate analysis

1. $X \perp\!\!\!\perp Y$

The errors for $C = 0$ in Figs. 4(a)–4(d) show the situation for the pair of x_1 and y_1 variables and the pair of x_3 and y_3. In most cases, the inability to improve the predictions using the information from the other variable correctly indicated mutual independence.

2. $X \leftrightarrow Y$

The PI method relatively reliably revealed the onset of synchronization after the coupling strength C rose above 0.12. See the self-prediction and mixed prediction errors for high coupling values in the top four pictures of Fig. 4. For most cases, x could help significantly with the prediction of the observable y and vice versa.

The same was expected for a pair of variables from the same system with a cyclic flow of information. Figure 4 presents two such examples of bidirectional coupling, namely the pair of x_1 and x_2 and the pair of y_2 and y_3. For clean x_1 and x_2 variables, the self-prediction errors are tiny, and yet the mixed-prediction errors are often even smaller, indicating a bidirectional connection. In the case of the second pair, we can see that prediction of the problematic third variable y_3 can be greatly improved using the mixed state space, indicating the causal link $y_2 \to y_3$. However, as Figs. 4(g) and 4(h) show, the opposite direction of $y_3 \to y_2$ is more difficult to prove for some couplings.

3. $X \to Y$

When testing x_1 and y_1 for $0 < C < 0.12$ [Fig. 4(a)], we mostly observed a significant improvement in the predictability of the y_1 variable using mixed predictions, so that the PI method correctly identified the unidirectional causality. On the other hand, pictures for x_3 and y_3 in Figs. 4(c) and 4(d) show that when testing these problematic variables, causality detection was not reliable. The PI method failed even more often than the CP method. For most coupling values, x_3 was unable to improve the predictability of y_3, and therefore the presence of a causal link could not be identified.

For interested readers, in Appendix B we also present the results for two earlier methods working in state spaces [7,8].

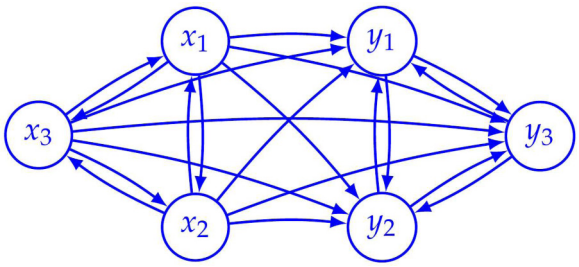

FIG. 5. Connections detected by the CP method in coupled Rössler systems given by Eq. (1) for coupling value of $C = 0.05$. The method found all the links it should have found.

C. Results in a multivariate scenario

Although the methods have so far been presented in the context of bivariate analysis, they can also be useful for large networks where we are interested in how many nodes we need to measure in order to characterize relationships across the studied network. In this sense, we can look at our test example of the coupled Rössler oscillators as a network of six nodes. Suppose that we know the measurements of all six variables (nodes), and not just two as before, and assume we do not know if there is any relationship between the nodes at all. What will our pairwise causal analysis reveal? The results, computed for one selected coupling value of $C = 0.05$, are visualized by the fluence graphs in Figs. 5 and 6.

Figure 5 shows the results of the CP method. Bivariate causal analyzes revealed all nine links of the original fluence graph presented already in Fig. 2. In addition to these, we correctly found 12 more links, which is consistent with the fact that it is impossible to distinguish between direct and indirect driving.

Figure 6 shows that the PI method did not find some of the links. The higher incidence of false negatives is not surprising since the CP method only determines whether, for example, M_Y allows at least some prediction of x, while in the PI method we ask if the mixed manifold M_{XY} contains enough extra information to improve the self-prediction of x significantly. The second question is usually harder to answer. However, in the presented case, the results of both methods made it possible to correctly conclude that we are dealing with a network that can be decomposed into two unidirectionally connected blocks X and Y, each with a cyclic flow of information. The results also allow us to declare that

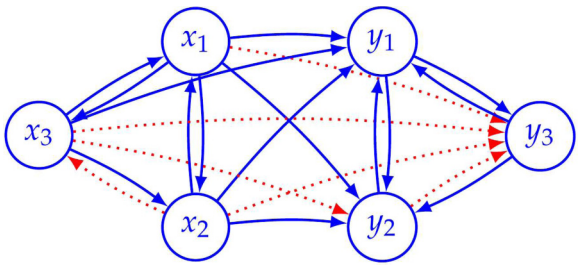

FIG. 6. Connections detected by the PI method in coupled Rössler systems given by Eq. (1) for coupling value of $C = 0.05$. The dotted links remained undetected.

measuring any one variable of the X subsystem together with any one variable of the Y subsystem should be sufficient to characterize and model the dynamics behind the behavior of all six nodes.

In the particular example shown in Figs. 5 and 6 (clean data and coupling 0.05), the CP method did not produce any false results and the PI method led to some false negatives with x_3 and y_3 variables. However, Table II shows that for different couplings, and especially after the addition of noise, false-positive results can also occur.

At this point, we just wanted to indicate that having a network of systems or dynamical observables, pairwise causal analysis can help to identify the significantly reduced set of nodes, providing sufficient information about the entire network dynamics.

V. DISCUSSION

As demonstrated so far, our test data not only revealed the strengths and weaknesses of the presented methods but also drew attention to situations where the methods can run into problems. It turns out that such situations occur when a variable with weak observability is used for reconstruction of the studied dynamics. To explain what we mean, let us take a closer look at the notion of observability in this section.

In control theory, the observability of a linear system means that the behavior of the entire system can be determined from its outputs. The linear system is either fully observable or not through a given set of measurements. The concepts of observability extended to nonlinear systems have been studied for more than two decades by Letellier *et al.* (see Refs. [22,23] and references to earlier work therein). An important step in the field was to recognize that the observability matrix that has to be of full rank is, in fact, the Jacobian matrix of the coordinate transformation between the original state space and the reconstructed space. So there is an obvious connection with the ability to reconstruct the dynamics of the system from one measured variable declared by the Takens theorem [18]. If the map is a global diffeomorphism, then there is full observability. If there is a global singularity in the observability matrix, then the system is nonobservable [23].

When considering the three-dimensional Rössler system, we expect that a proper delay-time reconstruction from any variable constitutes a diffeomorphism between the original manifold and the reconstructed image. Moreover, the variables are to be equally useful for the state space reconstruction. However, it has been shown that in the case of the variable x_2, derivative reconstruction is already globally diffeomorphic to the original state portrait in three dimensions $[x_2, \dot{x}_2, \ddot{x}_2]$ [22]. Even the global model can be obtained quite easily from the x_2 observable. On the other hand, for x_1 and x_3 at least a four-dimensional reconstruction space is needed. Other than derivative reconstruction, such as delay-time embedding, might even require an additional dimension to unfold the attractor sufficiently. Ultimately, however, increasing the dimension by adding additional coordinates should eliminate the problem of singularities for observable systems. Despite all these arguments, the third variable of the Rössler system is known to be an almost unusable basis for reconstruction and modeling [22]. Even if we used a large number of clean

data and high embedding dimensions, the effort to predict the reconstructed trajectories or to estimate the attractor's fractal complexity turned out poorly [20]. Similarly, in this study, once x_3 was among the variables examined, the causal analysis was more susceptible to failure. The same was true for the third variable of the driven Rössler system.

Obviously, even with global diffeomorphisms, one variable could be preferable to the other. So, besides the question of whether the system is observable or not, another exciting task is to sort the variables of observable systems in terms of their effectiveness in the reconstruction process and subsequent analysis [23]. To contribute to solving this problem, several measures (observability coefficients) have been introduced. Some of them even do not involve knowing the governing equations. We are not going into detail; the interested reader is referred to Refs. [22,23] and the references therein for additional information. What is important to recall for this study is that whichever observability coefficient we use for the Rössler system, the variables are ranked according to observability as $x_2 > x_1 > x_3$.

A closer look at the dynamical systems has revealed that the problematic variables are those that receive information about the rest of the system through a nonlinear function. Assuming that nonlinear links are generally responsible for the lack of local observability, Letellier *et al.* have proposed constructing a so-called pruned fluence graph considering only linear links—corresponding to the constant nonzero terms of the Jacobian matrix [24]. From this graph, we identify the largest connected subgraphs in which there is a directed path from each node to every other node and no outgoing links. In each one of those subgraphs, at least one node should be measured. The authors argue that such a selection is not only minimal but also provides good observability, although not necessarily the best one [23]. The technique was validated by investigating large-dimensional reaction networks for which the determinant of the observability matrix can be rigorously computed.

Let us see what all this means for our test example. The coupled Rössler oscillators form a six-dimensional system which is theoretically fully observable from two variables—one from X and one from Y. However, let us leave out the two nonlinear links in the interaction graph in Fig. 2 to get the prune fluence graph. In this new graph, the strongly connected components are only composed of variables x_1 and x_2 or y_1 and y_2, respectively. Variables x_3 and y_3 become disconnected from the rest of the system. The pruned fluence graph thus correctly separates those variables that have the worst observability indices. The same variables were problematic also in the causal analysis. To get closer to the complete picture of system dynamics, we should also measure the excluded variables. So in our test example, to get the most out of state-space-based causal methods, more variables are recommended to be measured than we would expect based on the Takens's theorem. The recommended combinations include (x_1, x_3, y_1, y_3), (x_1, x_3, y_2, y_3), (x_2, x_3, y_1, y_3), and (x_2, x_3, y_2, y_3).

VI. CONCLUSIONS

In this study, we were facing an interesting problem of causality detection, if the valid working hypothesis is that the investigated time series x and y are manifestations of dynamical systems X and Y, respectively.

Although it is not articulated often enough in the literature, reconstruction of the state space from a variable with weak observability can undermine the whole subsequent analysis. This applies to any analysis of the reconstructed dynamics, and as we demonstrated in this study, causal analysis is no exception.

However, we point out that even with weak observables, difficulties mainly occurred with detections of unidirectional causal links. Bidirectional coupling, as well as the absence of any interconnection, were detected quite reliably, especially with the CP method. This could be true in general, but it needs to be confirmed by more thorough research.

We also touched on the aspect of the noise level in data, although it was not systematically studied here. As regards the impact of additive noise on the results, we tested cases with a signal-to-noise ratio of 10. We found that, in several cases, this level of noise made it impossible to detect causality correctly. The role of noise and possibilities of its reduction are among the topics we want to address in more detail in the near future.

The main aim of this paper, however, was to present a detailed guide to the implementation of two methods of causality detection. Both methods work with signals, or time series, from which they reconstruct the underlying dynamics in multidimensional state spaces. In the reconstructed spaces, one-point predictions are calculated, and based on the prediction errors, the causal relationships are inferred.

If the measured data are good representatives of the governing dynamics, then both the CP method and the PI method work reliably, with the first method being significantly faster and therefore preferable for large data sets. Compared to cross-mapping techniques (see results in Appendix B), the prediction-based methods are much less prone to false-positive results.

The way we proposed to implement the idea of cross-predictions leads to a robust methodology that can be successfully used for routine causal analysis of real-world time-series data. The PI method, although computationally more demanding and more susceptible to false negatives, can be beneficial if we are not only interested in causality detection but also in the possibility of improving the prediction. Moreover, if multiple system measurements are available, then the PI method can help sort out those variables that are most suitable for predicting and modeling the underlying dynamics.

Of course, as we demonstrated in this study, poor observability of the variables tested, or too much noise in data can reduce the credibility of the results. However, it is not the methods themselves that are responsible for such a failure, but the properties of data that do not allow a proper reconstruction of dynamics. The more problematic the data, the more careful we must be with interpreting the results.

In addition to difficulties with observability, we must also mention once again the problems with an unobserved confounder (a hidden common cause of two measured variables). A basic assumption of most causal discovery methods from observational data is that there exist no latent variables that influence any pair of measured variables. The term causal is then understood relative to the given set. As a first step, we expect the methods to work in an ideal situation without hidden drivers. This is the main ambition of our prediction-based methods as well. However, in the end, the causal methods will also have to deal with unmeasured confounders, because they are ubiquitous in real-world applications. In the presence of confounders, the results may have more than one interpretation. A bidirected edge, for example, can indicate a causal link or just a correlation due to the effect of a common hidden driver or both. The problem of determining which of these is the case is far from solved. Even popular methods do not work, or they can handle only some special cases. The issue of hidden drivers was beyond the scope of this article, but it is a priority topic for future research. Readers who are interested in the ongoing research and current progress in understanding causal structures in the presence of confounders may refer to Ref. [25] and references therein.

Despite some open questions, we believe that predictions are extremely useful in finding causality in complex dynamic systems and networks. We hope that the detailed implementation instructions given here will contribute to the broader use of these state-space-based causal methods. They can be useful in bivariate causal detection and also for studying large dynamical networks whose nodes are characterized by time series.

ACKNOWLEDGMENTS

Supported by the Slovak Grant Agency for Science (Grant No. 2/0081/19) and by the Slovak Research and Development Agency (Grant No. APVV-15-0295).

APPENDIX A

In this section we present the pseudocodes of the proposed causal methods. Codes of Matlab implementation are available Ref. [26].

Algorithm 1: Nearest-neighbors prediction

Input:
y - time series for prediction
M - reconstructed manifold from historical values of y
Output:
$prediction\_y$

$m = \text{dimension}(M)$
for $i \leftarrow m, \text{length}(y)$ **do**
$point = [y(i-m+1), \ldots, y(i)]$
$[indices\_knn, distances\_knn] = \text{knnSearch}(point, M)$
$u = \exp(-distances\_knn/\min(distances\_knn))$
$weights = u/sum(u)$
$prediction\_y(i+1) = weights * y(indices\_knn+1)$
end for

Algorithm 2: Cross Prediction

Input:
x, y - time series for causality detection
$values_{(d,\tau,k)}$ - values for selecting the embedding dimension, time delay and number of nearest neighbors
Output:
$(d, \tau, k)_{(xX,xY,yY,yX)}$ - optimal values of d, τ, and k for predicting time series a based on information from manifold B
$n\mathcal{E}(xX, xY, yY, yX)$ - normalized mean absolute error of prediction of a based on B
$result$ - the type of causal relation between x and y detected

for a, B **in** $[xX, xY, yY, yX]$ **do**
$a\_hist, a\_predict = \text{split}(a)$
for d, τ, k **in** $values_d \times values_\tau \times values_k$ **do**
for $i \leftarrow d * \tau + 1, \text{length}(a\_hist)$ **do**
M_B_append$([B(i - d * \tau), B(i - (d-1) * \tau), \ldots, B(i)])$
end for
$z = \text{nn\_predict}(a\_predict, M_B)$ (ALGORITHM: 1)
$\mathcal{E} = \text{mae}(a\_predict, z)$
if $\mathcal{E} < \mathcal{E}aB$ **then**
$(d, \tau, k)_{aB} = d, \tau, k$
$\mathcal{E}aB = \mathcal{E}$
end if
end for
end for
$\mathcal{E}x\_baseline, \mathcal{E}y\_baseline$ - errors of trivial baseline prediction (taking the mean of the time series as the estimate of the following value)
if $\mathcal{E}xY > \mathcal{E}x\_baseline$ bf and $\mathcal{E}yX > \mathcal{E}y\_baseline$ **then**
$result = X \perp\!\!\!\perp Y$
else if $\mathcal{E}xY \leqslant \mathcal{E}xX$ **or** $\mathcal{E}yX \leqslant \mathcal{E}yY$ **then**
$result = X \leftrightarrow Y$
else if $n\mathcal{E}xY < n\mathcal{E}yX$ **then**
$result = X \rightarrow Y$
else if $n\mathcal{E}xY > n\mathcal{E}yX$ **then**
$result = Y \rightarrow X$
end if

Algorithm 3: Predictability Improvement

Input:
x, y - time series for causality detection
$values_{(d,\tau,k)}$ - values for selecting the embedding dimension, time delay and number of nearest neighbors
$values_w$ - values for selecting the weight of the helping system
Output:
$(d, \tau, k)_{(X,Y)}$ - optimal values of d, τ, and k for predicting time series a based on information from manifold B
$w_{(xXY,yYX)}$ - optimal weight of helping system
$\mathcal{E}(xX, xXY, yY, yYX)$ - mean absolute error of prediction of a based on information from manifold A or mixed manifold AB
$result$ - the type of causal relation between x and y detected

for a, A **in** $[xX, yY]$ **do**
$a\_hist, a\_predict = \text{split}(a)$
for d, τ, k **in** $values_d \times values_\tau \times values_k$ **do**
for $i \leftarrow d * \tau + 1, \text{length}(a\_hist)$ **do**
M_A_append$([A(i - d * \tau), A(i - (d-1) * \tau), \ldots, A(i)])$
end for
$z = \text{nn\_predict}(a\_predict, M_A)$(ALGORITHM: 1)
$\mathcal{E} = \text{mae}(a\_predict, z)$
if $\mathcal{E} < \mathcal{E}aA$ **then**
$(d, \tau, k)_A = d, \tau, k$
$\mathcal{E}aA = \mathcal{E}$
end if
end for
end for
$help_\tau = \min(\tau_X, \tau_Y)$
for a, A, B **in** $[xXY, yYX]$ **do**
for w **in** $values_w$ **do**
for $i \leftarrow d * \tau + 1, \text{length}(a\_hist)$ **do**
M_{AB}_append$([A(i - d_A * \tau_A), A(i - (d_A - 1) * \tau_A), \ldots, A(i),$
$w * B(i - d_B * help_\tau), w * B(i - (d_B - 1) * help_\tau), \ldots,$
$w * B(i)])$
end for
$z = \text{nn\_predict}([a\_predict, b\_predict], M_{AB})$ (ALGORITHM: 1)
$\mathcal{E} = \text{mae}(a\_predict, z)$
if $\mathcal{E} < \mathcal{E}aAB$ **then**
$w_{aAB} = w$
$\mathcal{E}aAB = \mathcal{E}$
end if
end for
end for
if $\mathcal{E}xXY \geqslant \mathcal{E}xX$ **and** $\mathcal{E}yYX \geqslant \mathcal{E}yY$ **then**
$result = X \perp\!\!\!\perp Y$
else if $\mathcal{E}xXY < \mathcal{E}xX$ **and** $\mathcal{E}yYX < \mathcal{E}yY$ **then**
$result = X \leftrightarrow Y$
else if $\mathcal{E}yYX < \mathcal{E}yY$ **then**
$result = X \rightarrow Y$
else if $\mathcal{E}xXY < \mathcal{E}xX$ **then**
$result = Y \rightarrow X$
end if

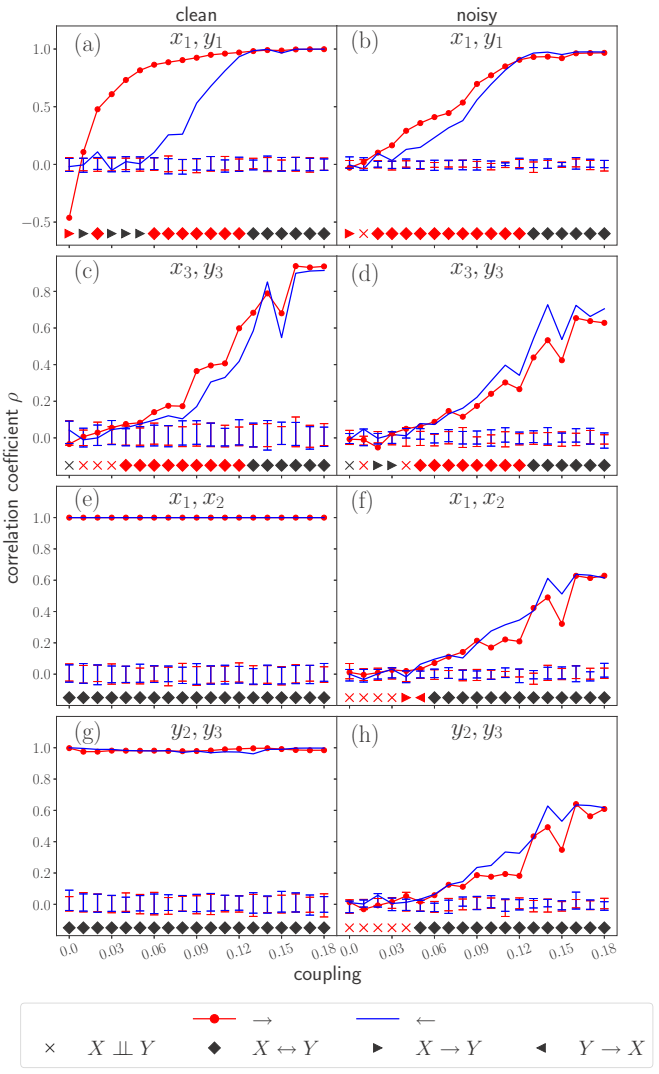

FIG. 7. Cross-map skill ρ for four pairs of variables of coupled Rössler systems given by Eq. (1) for increasing coupling values. The amount of 3700 data points was used for the computations. Embedding dimension within the range of 4 to 11 and time delay 1 was used for the reconstruction. For each coupling, a boxplot provides results for 100 surrogate time series. The line with dots (red) shows the result for the causal effect of the first variable from the title to the second. The line without markers represents the opposite direction. Pictures on the left are for clean data, while pictures on the right are for data with a signal-to-noise ratio of 10. The symbols at the bottom of each figure indicate concrete conclusions of the method about the type of causality. Red is used for symbols of incorrect detection results. Black is for correct detections.

APPENDIX B

This Appendix is intended for readers interested in comparing the prediction-based methods with other related causal methods operating in state spaces. The methods for comparison are based on the asymmetry of cross-mappings between close neighbors in the reconstructed manifolds M_X and M_Y.

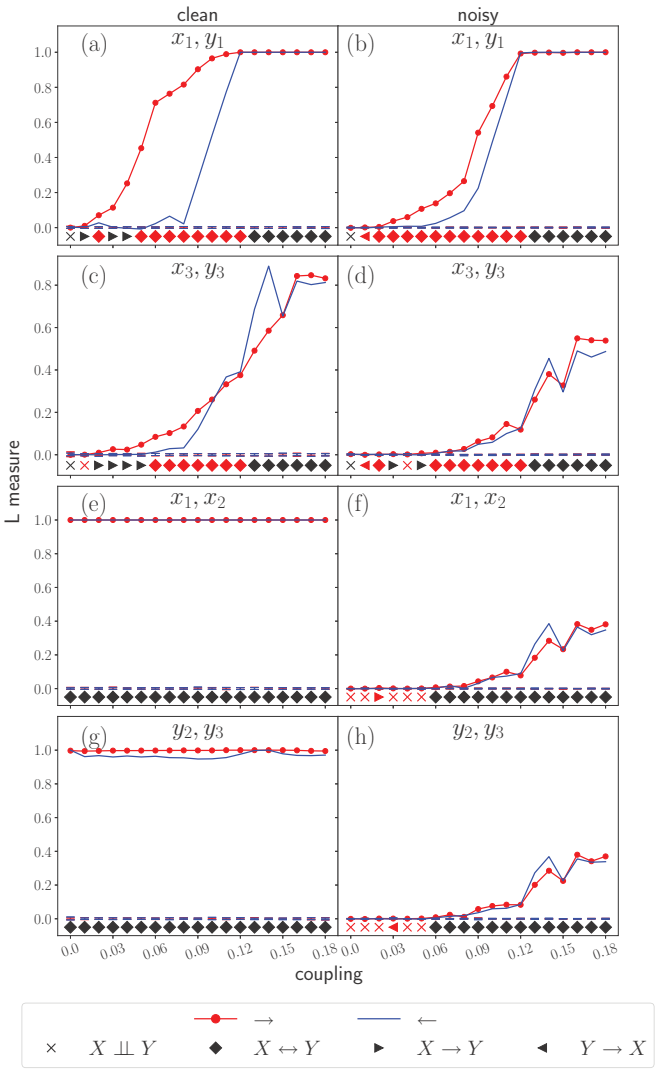

FIG. 8. Causal analysis for four pairs of variables of coupled Rössler systems given by Eq. (1). The difference compared to the previous figure is in the use of the L-measure [7] for the evaluation of the cross-mapping.

To infer causality from this asymmetry, the first method evaluates the rank-based measure L, proposed by Chicharro and Andrzejak in 2009 [7], and the second method, called convergent cross-mapping (CCM) and introduced by Sugihara *et al.* in 2012, uses the Pearson correlation coefficient between values estimated by cross-mapping and the actual values [8].

The CCM method, as originally proposed, determines causality based on the value of the correlation coefficient compared to some heuristically selected threshold. For the reliability of the results, we consider it more appropriate to add some form of testing, such as the use of surrogates, as we did in Ref. [10]. The skill ρ of cross-mapping higher then the threshold given by the surrogates indicates that one system drives the other.

Details on the implementation of the CCM and L-measure can be found in Ref. [10] and the Matlab codes for the methods are available at Ref. [27].

Figures 7 and 8 demonstrate the performance of the two methods. The cross-mapping techniques CCM and L produced a large number of false-positive results. For example, when we used variables of the system (1) having a high observability index, the CP method made no error and the PI method ended up with 14% false negatives and 3% false positives (see Table II). In contrast, although the CCM and the L-measure do not tend to give false-negative results, they did lead to 64% and 89% false positives, respectively. With additive noise or using the problematic variables x_3 and y_3, the results were even worse. This is also evident from the specific cases of bivariate causal analysis shown in Figs. 7 and 8).

[1] C. W. Granger, Investigating causal relations by econometric models and cross-spectral methods, Econometrica **37**, 424 (1969).

[2] L. Faes, A. Porta, and G. Nollo, Mutual nonlinear prediction as a tool to evaluate coupling strength and directionality in bivariate time series: Comparison among different strategies based on k nearest neighbors, Phys. Rev. E **78**, 026201 (2008).

[3] Y. Chen, G. Rangarajan, J. Feng, and M. Ding, Analyzing multiple nonlinear time series with extended Granger causality, Phys. Lett. A **324**, 26 (2004).

[4] D. Marinazzo, M. Pellicoro, and S. Stramaglia, Kernel method for nonlinear Granger causality, Phys. Rev. Lett. **100**, 144103 (2008).

[5] M. Paluš and M. Vejmelka, Directionality of coupling from bivariate time series: How to avoid false causalities and missed connections, Phys. Rev. E **75**, 056211 (2007).

[6] T. Schreiber, Measuring Information Transfer, Phys. Rev. Lett. **85**, 461 (2000).

[7] D. Chicharro and R. G. Andrzejak, Reliable detection of directional couplings using rank statistics, Phys. Rev. E **80**, 026217 (2009).

[8] G. Sugihara, R. May, H. Ye, C. H. Hsieh, E. Deyle, M. Fogarty, and S. Munch, Detecting causality in complex ecosystem, Science **338**, 496 (2012).

[9] A. Krakovská and F. Hanzely, Testing for causality in reconstructed state spaces by an optimized mixed prediction method, Phys. Rev. E **94**, 052203 (2016).

[10] A. Krakovská, J. Jakubík, M. Chvosteková, D. Coufal, N. Jajcay, and M. Paluš, Comparison of six methods for the detection of causality in a bivariate time series, Phys. Rev. E **97**, 042207 (2018).

[11] N. F. Rulkov, M. M. Sushchik, L. S. Tsimring, and H. D. Abarbanel, Generalized synchronization of chaos in directionally coupled chaotic systems, Phys. Rev. E **51**, 980 (1995).

[12] S. J. Schiff, P. So, T. Chang, R. E. Burke, and T. Sauer, Detecting dynamical interdependence and generalized synchrony through mutual prediction in a neural ensemble, Phys. Rev. E **54**, 6708 (1996).

[13] M. Wiesenfeldt, U. Parlitz, and W. Lauterborn, Mixed state analysis of multivariate time series, Int. J. Bifurcat. Chaos **11**, 2217 (2001).

[14] U. Feldmann and J. Bhattacharya, Predictability improvement as an asymmetrical measure of interdependence in bivariate time series, Int. J. Bifurcat. Chaos **14**, 505 (2004).

[15] O. E. Rössler, An equation for continuous chaos, Phys. Lett. A **57**, 397 (1976).

[16] A. Krakovská, J. Jakubík, H. Budáčová, and M. Holecyová, Causality studied in reconstructed state space. Examples of uni-directionally connected chaotic systems, arXiv:1511.00505 (2015).

[17] B. Cummins, T. Gedeon, and K. Spendlove, On the efficacy of state space reconstruction methods in determining causality, SIAM J. Appl. Dynam. Syst. **14**, 335 (2015).

[18] F. Takens, Detecting strange attractors in turbulence, in *Dynamical Systems and Turbulence*, edited by D. A. Rand and L. S. Young (Springer-Verlag, Berlin, 2002), p. 366.

[19] D. Kugiumtzis, State space reconstruction parameters in the analysis of chaotic time series—The role of the time window length, Physica D **95**, 13 (1996).

[20] A. Krakovská, K. Mezeiová, and H. Budáčová, Use of false nearest neighbours for selecting variables and embedding parameters for state space reconstruction, J. Complex Syst. **2015**, 932750 (2015).

[21] E. N. Lorenz, Atmospheric predictability as revealed by naturally occurring analogues, J. Atmospheric Sci. **26**, 636 (1969).

[22] L. A. Aguirre and C. Letellier, Observability of multivariate differential embeddings, J. Phys. A: Math. Gen. **38**, 6311 (2005).

[23] L. A. Aguirre, L. L. Portes, and C. Letellier, Structural, dynamical and symbolic observability: From dynamical systems to networks, PLoS ONE **13**, e0206180 (2018).

[24] C. Letellier, I. Sendina-Nadal, and L. A. Aguirre, Nonlinear graph-based theory for dynamical network observability, Phys. Rev. E **98**, 020303 (2018).

[25] C. Glymour, K. Zhang, and P. Spirtes, Review of causal discovery methods based on graphical models, Front. Genet. **10**, 524 (2019).

[26] J. Jakubík and A. Krakovská, Cross prediction and predictability improvement, https://www.mathworks.com/matlabcentral/fileexchange/75057-cross-prediction-and-predictability-improvement, MATLAB Central File Exchange, Retrieved July 29, 2020.

[27] J. Jakubík, A. Krakovská, L. CCM, and M causality, Convergent cross mapping, https://www.mathworks.com/matlabcentral/fileexchange/52964-convergent-cross-mapping, 2016.