Testing for causality in reconstructed state spaces by an optimized mixed prediction method

Anna Krakovská* and Filip Hanzely
Institute of Measurement Science, Slovak Academy of Sciences, Dúbravská Cesta 9, 842 19 Bratislava, Slovakia
(Received 7 July 2016; published 3 November 2016)

In this study, a method of causality detection was designed to reveal coupling between dynamical systems represented by time series. The method is based on the predictions in reconstructed state spaces. The results of the proposed method were compared with outcomes of two other methods, the Granger VAR test of causality and the convergent cross-mapping. We used two types of test data. The first test example is a unidirectional connection of chaotic systems of Rössler and Lorenz type. The second one, the fishery model, is an example of two correlated observables without a causal relationship. The results showed that the proposed method of optimized mixed prediction was able to reveal the presence and the direction of coupling and distinguish causality from mere correlation as well.

DOI: 10.1103/PhysRevE.94.052203

I. INTRODUCTION

In recent years, the study of drive-response relationships between systems has attracted considerable attention. Relevant applications mainly concern areas such as economics, climatology, ecosystems, electrical activity of the brain, or cardiorespiratory relations.

The data subjected to causal analysis are very often in the form of time series. In such cases, the well-known Granger causality test, proposed in 1969, remains a popular method for analysis [1]. The Granger test focuses on determining whether one time series is useful in forecasting another. If yes, then the first system causally influences the second one. The conventional test of Granger's causality, based on autoregressive models, is particularly successful in the case of stochastic linearly interconnected systems.

However, in this study, we focus on another approach to causality analysis, namely, techniques operating in state spaces. With these methods in mind, the following type of coupling is usually considered:

$$\dot{X}(t) = F(X(t)),$$
$$\dot{Y}(t) = G(Y(t), X(t))$$

where X denotes the state vectors of the driving system and Y denotes the driven response. If there is an invertible functional relationship between X and Y, then there is said to be a generalized synchronization. The direction of coupling can only be uncovered when the coupling is weaker than the threshold for the emergence of synchronization [2]. Once the systems are synchronized, the future states of the driver X can be predicted from the response Y equally well and vice versa.

Let the systems X and Y be represented by a time series x and y, respectively. Then, the first step prior to the following types of causality analyses is to reconstruct a d-dimensional manifold M_X from lags of observable x so that the state of the system in time t is $X_t = (x_t, x_{t-\tau}, \ldots, x_{t-(d-1)\tau})$. The manifold M_Y is reconstructed analogously. From Taken's theorem [3], we know that given certain conditions, the reconstructed manifold is diffeomorphic to the original one. Suppose that X causally influences Y. Then it is highly likely that the neighboring states in M_Y are mapped to the neighboring states in M_X rather than in the case of uncoupled systems. On the other hand, the fact that X drives Y also results in the opposite effect, in that it is highly likely that the neighboring states in M_X are mapped to neighboring states in M_Y; however, the probability is higher in the opposite direction. Therefore, it is essential to examine both directions. Several methods have been proposed to infer causality from the asymmetry of cross-mappings between close neighbors in M_X and M_Y [4–10]. However, as Faes *et al.* [11] point out, it is often very difficult to obtain nontrivial directional information through these methods, as the asymmetry measures do not respond in a simple and stable manner with regard to varying the dynamical properties of the systems, the coupling degree, or the noise amount.

In this study, we are interested in methods working in state spaces, but with more direct reference to Granger's original evaluation of causality. In Ref. [12], for example, the authors sought a nonlinear extension of Granger's idea. They started with a standard delay embedding reconstruction. Then, locally, the dynamics were approximated using an autoregressive model. Granger's causality was assessed on all short pieces of the trajectory and the results were then averaged over the entire state portrait.

More global consideration of past dynamics on the attractor has resulted in a further class of methods [13,14], which attempt to determine whether a prediction made in a reconstructed state space of a time series from one system improves when data from another system are included in the state space reconstruction. The potential of such an approach, called mixed state analysis, has been identified in Refs. [11,13,15].

Monitoring the predictability improvement can be seen as an analogy to methods based on transfer entropy or conditional mutual information [2,16]. The information-theoretic measures and predictability improvement both measure the change of uncertainty of the future of a signal when estimates with and without additional knowledge of another system are compared. However, authors in Ref. [14] argue that, for short time series, assessment of the predictability is more appropriate.

In this study, we propose an optimized mixed prediction method of causality detection.

*krakovska@savba.sk

The remainder of this paper is organized as follows. We first introduce the new mixed prediction method and the methods used for comparison, which are Granger's test of causality [1] and the convergent cross-mapping (CCM) method of Sugihara *et al.* [10]. The third section presents the results obtained by testing coupled Rössler and Lorenz systems. This is followed by the results for a fishery model. Finally, the findings are summarized and discussed.

II. METHODS

A. Granger's VAR test

We say that variable x causes another variable y in Granger meaning [1], if better prediction of y is produced using information from both x and y rather than using only information from y. In recent decades, a number of modifications and extensions of the Granger causality test have been proposed. However, the original analysis has been performed by fitting an autoregressive model to the variable as follows:

$$y(t) = \sum_{j=1}^{p_y} c_{y,j}\, y(t-j) + \sum_{j=1}^{p_x} c_{x,j}\, x(t-j) + \epsilon(t),$$

where p_x, p_y are numbers of lagged observations from variables x, y; $c_{y,j}$, $c_{x,j}$ are coefficients of the model, and ϵ is the vector of residuals. The Granger causality is tested by performing the F test of the null hypothesis that all coefficients $c_{x,j}$ are zeros. Rejection of the hypothesis means that the coefficients corresponding to the data from variable x are statistically significantly different from zero. Then we conclude that the information from x increases the predictability of y, i.e., the process x Granger causes y.

For the Granger causality analysis, we used the Matlab code of Chandler Lutz [17]. The numbers p_x, p_y of samples for linear regressions were estimated with the help of the Bayesian information criterion. Then the test statistics S was calculated:

$$S = \frac{(R - R_U)/p_y}{R_U/[T - (p_x + p_y + 1)]} \sim F_{p_y, T-(p_x+p_y+1)},$$

where R and R_U are residual sums of squares when x is not considered in the model and when x is considered in the model, respectively. T is the size of the data, and F denotes the cumulative distribution function of the F distribution with p_y and $T - (p_x + p_y + 1)$ degrees of freedom.

The null hypothesis $H_0 : \forall j : c_{x,j} = 0$ of noncausality was tested by the F test.

B. Convergent cross-mapping

As a second testing method, we used one of the approaches that rely on evaluating distances of conditioned neighbors in reconstructed state spaces. Essentially, the methods determine whether the time indices of nearby points in the historical data of M_Y can be used to identify nearby points in M_X. This idea can be successfully applied through measures like M or L introduced in Refs. [8] and [9] or by a method called convergent cross-mapping (CCM) proposed by Sugihara *et al.* [10]. In the latter article, the value of similarity between points X_n and estimates $\hat{X}_n$ is evaluated by the correlation between them. A high value of the correlation exponent indicates that system X drives system Y. For the same reasons as mentioned above, the opposite direction needs to also be checked.

Sugihara *et al.* [10] emphasized that an important part of their method involves convergence, a key property that might distinguish causation from mere correlation, which is important, among other places, in a situation wherein correlation between two systems is produced by a shared third driving variable. Such a correlation can be falsely declared as causality. In these cases, we may ask whether cross-mapped estimates converge to correct values for an increasing number of used data. For causally coupled systems, the estimates improve with the length of the time series. Lack of convergence, on the other hand, indicates the absence of actual causality.

C. Optimized mixed prediction

In this section, we introduce a new method of causality detection. The method builds on the so-called mixed state analysis presented in Refs. [13] and [14]. Similarly, as in the case of Granger testing, a comparison of forecasts is used to search for causal links. However, the method proposed here, let us call it optimized mixed prediction (OMP), works in reconstructed multidimensional state spaces. The main idea of the method is as follows. The predictions in the spaces M_X and M_Y, reconstructed as described above, can be compared to predictions computed in M_{X+Y}, i.e., in a mixed state space built from delayed observables of both systems. If, for example, the predictability of Y in M_{X+Y} is better than its predictability in M_Y, we expect that X causally affects Y. The proposed method builds on this idea, with special attention attributed to the optimization of the reconstructed space based on the weighting of each of its coordinates. This is explained in the following.

With regard to the actual method of prediction, we find historical data similar to the current situation and assume that the system will continue in the same way as it has in the past. This forecasting technique is generally known as the method of analogues [18]. There are many ways to predict a follower of point Y_n, the simplest version being finding its nearest neighbor Y_i from past states on the reconstructed trajectory and declaring $Y_{n+1} = Y_{i+1}$. A modification, which was used in this study, improves the simplest version by considering several neighbors and utilizing the direction in which the images of the discovered neighbors moved in the past.

More precisely, the predictive method described above functions as follows. At first, k nearest neighbors of point Y_n are determined with K being a set of their time indices. In methods using nearest neighbors, it is fairly common to consider exponential weighting for the neighbors Y_i, $i \in K$ based on distances from Y_n:

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

Then the prediction $\hat{Y}_{n+1}$ is given by

$$\hat{Y}_{n+1} = \frac{\sum_{i \in K} w_i (Y_n + Y_{i+1} - Y_i)}{\sum_{i \in K} w_i}.$$

The smaller the variance of the errors $e = Y - \hat{Y}$, the better the prediction.

After selecting a predictive method, we build the mixed prediction method by comparing the residuals obtained from the prediction of Y using only the information from system Y and residuals from the prediction of Y using information from both X and Y.

The specific steps of the OMP method to detect causality from X to Y are as follows:

(1) State space reconstruction M_Y of Y without information from X, with states in time t represented as $[y_t, y_{t-\tau_Y}, \ldots, y_{t-(d_Y-1)\tau_Y}]$ is performed. In order to obtain the best possible predictions, some coordinates of M_Y are suppressed in a manner that is explained later. Then predictions $\hat{Y}_n$ of large enough statistical sample of N points spread over the entire state portrait are computed. To each predicted point, a sufficiently long piece of the previous trajectory has to be available, in order to find near neighbors from the past. The error vector is $e_n = Y_n - \hat{Y}_n$.

(2) State space reconstruction M_{X+Y} with the addition of information from X (mixed state space) is performed. Here the state corresponding to time t is $[y_t, y_{t-\tau_Y}, \ldots, y_{t-(d_Y-1)\tau_Y}, x_t, x_{t-\tau_X}, \ldots, x_{t-(d_X-1)\tau_X}]$. Again, to avoid problems pertaining to high dimensionality of the reconstructed spaces and to obtain the best possible predictions, impacts of the coordinates of M_{X+Y} are adjusted. Analogously as in step 1, one-point predictions of system Y are computed. The predictions are denoted as $\hat{Y}_n^X$, whereas the corresponding error vector is denoted as $e_n^X = Y_n - \hat{Y}_n^X$.

(3) To decide whether the addition of information from X improves the prediction of Y (X drives Y), the null hypothesis H_0: $\mathrm{Var}(e^X) \geqslant \mathrm{Var}(e)$ is tested. If the Kolmogorov-Smirnov test does not reject that e or e^X are normally distributed, then the F test is used. Otherwise, a nonparametric bootstrap method can be used [19]. If H_0 is rejected on a 5% significance level, than we accept that $\mathrm{Var}(e^X) < \mathrm{Var}(e)$, or equivalently, that inclusion of knowledge of X significantly improves the prediction of Y.

Causality in the opposite direction, i.e., from Y to X, is investigated analogously, after exchanging the roles of X and Y in the above instructions.

Now, let us identify some pitfalls of the OMP method.

First, making predictions in reconstructed state space requires a trajectory long enough to cover the state portrait sufficiently densely. This is important because to each predicted point we need its close neighbors from the past to capture and predict the underlying dynamics. Furthermore, the accuracy of the prediction also depends on the embedding dimension d and on the time delay τ used for reconstruction: values required for the first step of the OMP algorithm. Therefore, it is worthwhile to pay close attention to the selection of these parameters. The best combinations of dimension and time delay are given by some optimum time window $TW = (d-1)\tau$. As demonstrated in Ref. [20] for the same one-point prediction algorithm as used in the present study, the appropriate time window seems to be close to half of the mean orbital period, if it can be approximated by examining the oscillatory patterns in the data. Alternatively, we can calculate the prediction errors for several combinations of embedding parameters and choose one that leads to the lowest error.

Another aspect that requires caution relates to the size of the reconstructed spaces, especially the mixed state space. With a limited amount of samples, increasing the embedding dimension leads to a less dense occupancy of the space. Therefore, worsening of the prediction simply owing to a higher than necessary dimension of mixed state space M_{X+Y} compared to self-predictability in M_X or M_Y can be expected. To alleviate this problem, we optimized the reconstructed state portraits M_X, M_Y, and M_{X+Y}. More specifically, we suppressed those directions that proved unnecessary for successful prediction. Let d be the embedding dimension, i.e., the number of shifted copies of time series forming the reconstructed space. To each of the components, we intended to assign a weight, so that the result stretched the spaces in the directions of the elementary basis vectors. W needs to be a non-negative vector with the sum of all elements equal to 1. The initial weights of $W_j = \frac{1}{d}$ for $j = 1, \ldots, d$ were adjusted using a pattern search algorithm for constrained optimization (with implementation in Matlab). The final setting of W needs to reveal the basis vectors that allow for the lowest variance of residuals when using our predictor. Adjusting the weights improves predictions in M_X, M_Y, and M_{X+Y} and minimizes the dimensionality of the spaces.

The biggest advantage of the OMP method is that it allows very efficient decisions regarding causality even in complicated cases such as the relation between complex nonlinear systems, or apparent causality due to external forcing. This is demonstrated in the following examples.

III. RESULTS

A. Rössler → Lorenz

To test the methods described above, we chose two nontrivial examples. The first one was a coupling of two chaotic systems. Causality and synchronization in chaotic systems was discussed already in 1990 [21]. It has been shown that it is possible to "lock" one chaotic system to the other to get them to synchronize.

In this example, the Rössler system (x_1, x_2, x_3) drives the Lorenz system (y_1, y_2, y_3):

$$\dot{x}_1 = -6(x_2 + x_3), \quad \dot{x}_2 = 6(x_1 + 0.2x_2),$$
$$\dot{x}_3 = 6[0.2 + x_3(x_1 - 5.7)], \quad \dot{y}_1 = 10(-y_1 + y_2),$$
$$\dot{y}_2 = 28y_1 - y_2 - y_1 y_3 + Cx_2^2, \quad \dot{y}_3 = y_1 y_2 - \tfrac{8}{3} y_3. \qquad (1)$$

C represents the coupling strength. For $C = 0$, systems X and Y are not coupled. For higher coupling strengths, there is a considerable causal link from X to Y until generalized synchronization seems to occur just before $C = 3$. Two-dimensional plots of Fig. 1 illustrate how the originally independent Lorenz system ($C = 0$) is enslaved by the Rössler system for higher couplings.

A total of 100 000 data points of solutions of the ordinary differential equations (1) were obtained numerically by the fourth-order Runge-Kutta method. The solutions were computed for coupling strengths chosen from 0 to 4 with a step of 0.1. The starting point was [0,0,0.4,0.3,0.3,0.3]. The first 1000 data points were discarded. As regards the sampling of the resulting trajectory, one typical run around the attractor took about 10 points.

The same example of the Rössler → Lorenz system has been studied, among others, in Refs. [2,8,20,22–25].

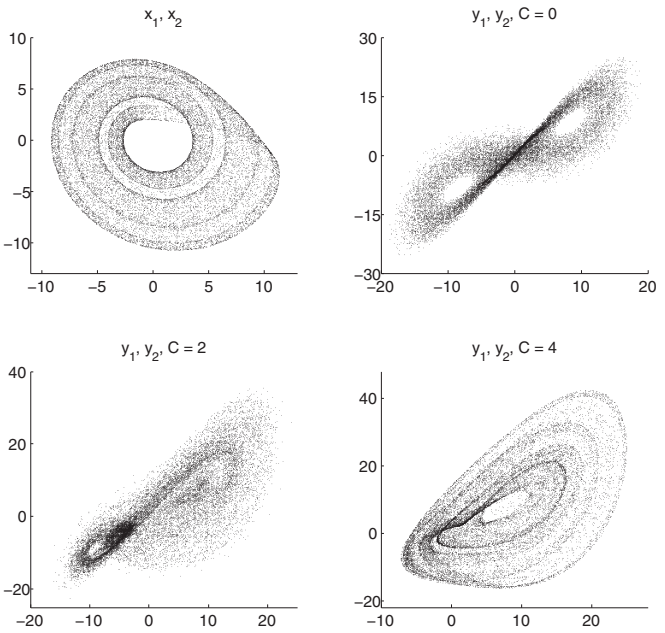

FIG. 1. Rössler system driving Lorenz system (1). 20 000 data points of x_1 and x_2 variables of the driving Rössler system (left, top); y_1 and y_2 variables of the Lorenz system itself (right, top); y_1 and y_2 variables of the driven Lorenz system with coupling strength $C = 2$ (left, bottom) and $C = 4$ (right, bottom).

The so-called interaction graph, easy to interpret from Eqs. (1) if defined as in Ref. [26], shows that the systems are coupled through one-way driving relationship between variables x_2 and y_2 (see Fig. 2). Since the Lorenz system Y does not drive the Rössler system X, we do not want to find the causal link in the direction from Y to X. On the other hand, we want to recover the outlined causal relations between variables of systems X and Y. As authors in Ref. [26] emphasized, both direct and indirect interactions can be recovered, although they cannot be reliably distinguished using comparisons of state space reconstructions.

In the following, we suppose that we know one variable of the driving Rössler system and one variable of the responsive Lorenz system, and we would like to find out whether there is a causal relationship between the two systems.

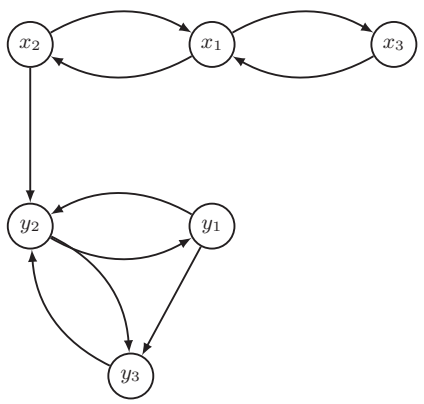

FIG. 2. Interaction graph for Rössler system driving Lorenz system (1).

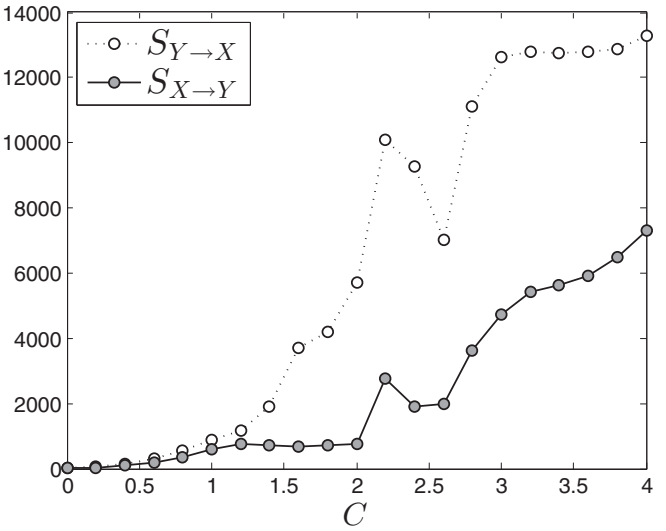

FIG. 3. Granger VAR test statistics S estimated in both directions using variables x_2, y_2 of the coupled Rössler and Lorenz systems.

1. Granger causality test

First, we searched for causality by the Granger test. For each of the 21 couplings, we used the entire datasets of 100 000 data points. A maximum lag of 15 was considered.

Figure 3 shows the results of the Granger test for causality between x_2 and y_2. However, other combinations of one variable from X and one from Y led to the same results. For $C = 0$, a causal link was not detected ($p = 0.11$ for direction $X \to Y$ and $p = 0.52$ for $Y \to X$). Except for zero coupling, causality was detected in both directions, although only detections in the direction $X \to Y$ are correct. Causality, whether correct or false, was determined at a high level of significance (in each case p value $< 10^{-15}$). The S statistics is even mostly much higher for the incorrect $Y \to X$ direction than for $X \to Y$. It means that, as expected, the Granger VAR method failed on the Rössler→Lorenz data. We used this example to describe how easily false results can be obtained if the Granger test of causality is inappropriately applied.

2. Convergent cross-mapping

As a second testing method, convergent cross-mapping was used. Since CCM is a state-space based method, we had to make reconstructions of state portraits first. We reconstructed M_X and M_Y from 10 000 points of observables x_2 and y_2, respectively, using delay $\tau = 1$ and embedding dimension $d = 7$. The number of nearest neighbors was set to eight. Testing by CCM led to the conclusion that X drives Y until the onset of synchronization between the couplings 2 and 3. Figure 4 shows that there is a small difference between the cross-mapping (CM) measures in direction $X \to Y$ and $Y \to X$ both for very weak couplings and for couplings higher than the threshold for the anticipated general synchronization. For the remaining couplings, CM($X \to Y$) is higher than CM($Y \to X$), which means that the direction of the causal influence was correctly detected, even though the differences in cases of strong couplings (above 1.5) did not seem to be significant.

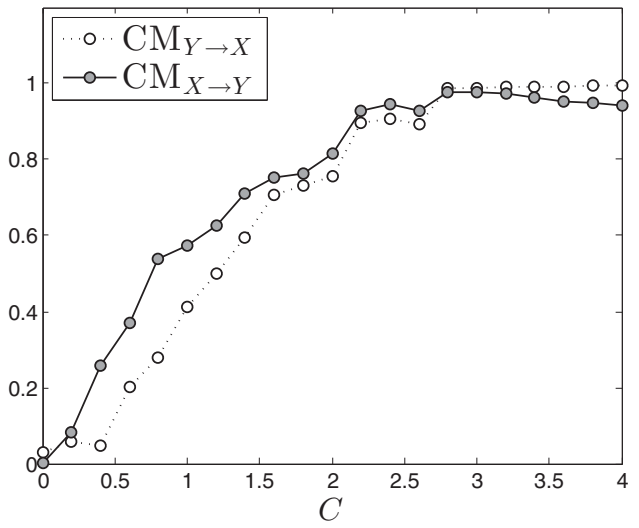

FIG. 4. Estimates of the CM measure in both directions using variables x_2, y_2 of the coupled Rössler and Lorenz systems.

We also considered the convergence of the results for an increasing number of data. Figure 5 shows the finding for the observables x_2 and y_2 under coupling $C = 1$. For segments of increasing length, the correlation between the two observables was computed. For each length, the average from correlations for 500 randomly chosen segments was considered. Besides that, for each length the cross-mappings were evaluated. Although, as we know, X drives Y, the CCM results indicate bidirectional influence: with increasing number of samples, the cross-mappings in both directions improved to a certain level. Although the values seem to be more convincing for $X \to Y$ than in the reverse, the link $Y \to X$ is difficult to exclude only based on this figure. Hence, the results are not very convincing, despite the fact that we worked with clean and quite long data,

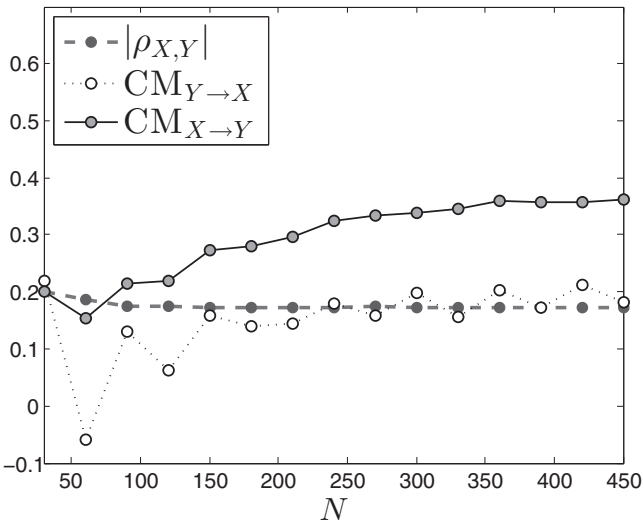

FIG. 5. Test for convergence of cross-mappings for Rössler → Lorenz system, with coupling $C = 1$, using variables x_2, y_2. For segments of increasing length, the correlation $\rho_{X,Y}$ between the two observables is compared to CM measures. Based on this figure, the correct link of $X \to Y$ cannot be unambiguously identified.

and we looked at a relatively clear example of $C = 1$ where the CM measures were well separated.

3. Optimized mixed prediction

In the newly proposed optimized mixed prediction method, the embedding parameters were individually selected for each investigated coupling strength. For this, we calculated the prediction errors for several combinations of parameters and selected one that leads to the lowest error. As a result, the space M_X was reconstructed with delay $\tau = 1$ and embedding dimension $d = 5$, while for M_Y delay $\tau = 1$ and embedding dimensions $d = 4$, 5, 6, or 7 (the stronger the coupling, the higher the value of d) were optimal. Then the corresponding mixed state spaces M_{X+Y} had dimensions from 9 to 12.

As explained earlier, we used weighting to minimize problems with high dimensions. In order to compute the optimal weights, we made predictions on 1000 randomly chosen data points. For each point, the nearest neighbors were found in a set of the preceding 50 000 points on the trajectory.

Let us recall at this point that we have to be careful when searching for the nearest neighbors. In order to capture the global dynamics, we are interested in previous visits of the trajectory to places near the current position of the predicted point. We have to ignore those points which are close to the predicted point just because they are close in time. Avoiding the adjacent data points is known as the Theiler correction [27]. For our sparsely sampled data, the correction was not required, but in the case of dense sampling, the Theiler correction must be taken into account.

Tables I and II show optimal weights found in case of six coupling values. Table I relates to reconstruction of the Lorenz system (Y) without the information from the Rössler system (X), while Table II shows optimal weights in the reconstruction of Y using the information from X. Obviously, for zero coupling, information from X does not help in the prediction of Y, and the corresponding coordinates were given zero weight. It is similar for the couplings above a synchronization level. Otherwise, the observable from X contributes significantly to the reconstruction created to predict Y.

After the weights had been adjusted, we made predictions of 5000 data points, each based on eight nearest neighbors from 50 000 past data points.

Figure 6 shows that predictions of y_2 without considering x_2 are clearly poorer than predictions made with the help of x_2. For couplings below the synchronization level, the link was

TABLE I. The optimal weights, rounded to two decimal places, for the reconstruction of Y (Lorenz system represented by y_2) for predicting Y without using the information from X (Rössler system represented by x_2). Results for six different couplings.

C	y_t	y_{t-1}	y_{t-2}	y_{t-3}	y_{t-4}	y_{t-5}	y_{t-6}
0	0.32	0.45	0.14	0.09	–	–	–
0.8	0.16	0.35	0.17	0.19	0.14	–	–
1.6	0.22	0.27	0.11	0.11	0.08	0.1	0.1
2.4	0.28	0.26	0.09	0.07	0	0.19	0.11
3.2	0.36	0.26	0.1	0.09	0	0.09	0.09
4	0.24	0.3	0.08	0.05	0.06	0.16	0.1

TABLE II. The optimal weights, rounded to two decimal places, for the reconstruction of Y (Lorenz system represented by y_2) for predicting Y using the information from X (Rössler system represented by x_2). Results for six different couplings.

C	y_t	y_{t-1}	y_{t-2}	y_{t-3}	y_{t-4}	y_{t-5}	y_{t-6}	x_t	x_{t-1}	x_{t-2}	x_{t-3}	x_{t-4}
0	0.31	0.45	0.14	0.09	–	–	–	0	0	0	0	0
0.8	0.19	0.27	0.01	0.04	0.07	–	–	0.11	0.09	0.05	0.03	0.14
1.6	0.14	0.24	0.07	0.03	0.02	0.05	0.01	0.16	0.01	0.01	0.06	0.2
2.4	0.14	0.17	0.04	0.02	0	0.04	0.01	0.29	0.02	0.07	0.1	0.1
3.2	0.28	0.31	0.13	0.03	0.02	0.17	0.04	0	0	0	0.02	0
4	0.32	0.24	0.04	0.01	0.06	0.09	0.1	0.01	0.03	0	0.1	0

detected with high statistical significance (p value $< 10^{-15}$), indicating that system X drives Y. As we can see when compared to CCM (Fig. 4), in the case of the OMP method the differences between the evaluated measures are distinctive even for strong couplings approaching the synchronization level.

When we focus on the opposite direction (shown in Fig. 7), we see that the self-predictability of x_2 is very good: the variances of residuals are small. Adding information from system Y did not improve the prediction, confirming that system Y does not drive X. The null hypothesis, that the prediction using only information from X is better, was never rejected.

B. Fishery model

In the case of a real-world time series, it may happen that there is a correlation between systems that are falsely declared as causality. The fishery model, selected as our second test data and used also in Ref. [10], illustrates such a situation very well. Mutually independent fish populations can have common peak recruitment seasons owing to favorable environmental conditions and be correlated.

We analyzed a standard logistic model of two noninteracting fish populations that share common environmental (weather) forcing:

$$\begin{aligned} R_X(n+1) &= X(n)\{3.1\,[1 - X(n)]\}\exp(0.5\,T), \\ R_Y(n+1) &= Y(n)\{2.9\,[1 - Y(n)]\}\exp(0.6\,T), \\ X(n+1) &= 0.4\,X(n) + \max(R_X(n-3),0), \\ Y(n+1) &= 0.35\,X(n) + \max(R_Y(n-3),0). \end{aligned} \tag{2}$$

The variables X and Y (see Fig. 9) denote sizes of two different fish populations, R_X and R_Y are recruitments of populations, and T is an environmental influence represented here by a red noise. The corresponding interaction graph can be seen in Fig. 8.

We generated 100 000 data with starting points $R_X(i) = R_Y(i) = X(i) = Y(i) = 0.5$ for $i \in \{1,2,3,4\}$.

The variable T was defined as follows:

$$T(i) = p \sum_{j=i-14}^{i} T'(j),$$

$$T' \sim N(0,1).$$

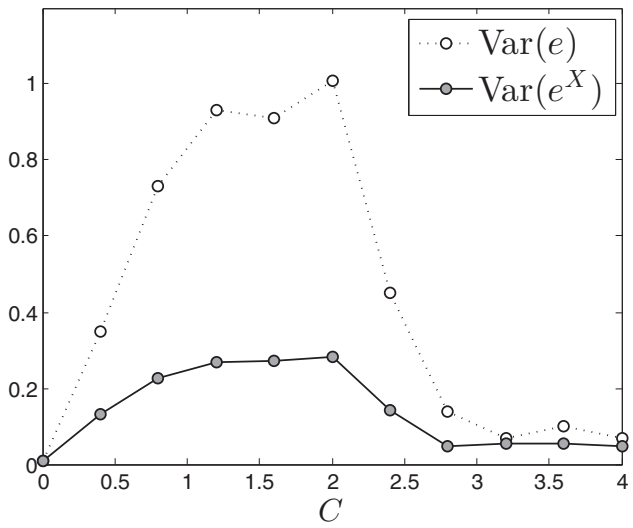

FIG. 6. Testing whether X (Rössler system) drives Y (Lorenz system) by the OMP method. Variances of residuals of predictions of y_2 with or without knowledge of x_2 are depicted. Adding information from X improved predictions, indicating that $X \to Y$.

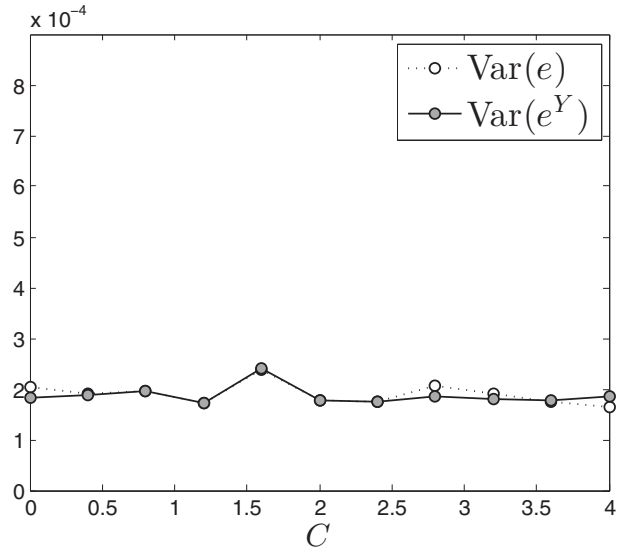

FIG. 7. Testing whether Y (Lorenz system) drives X (Rössler system) by the OMP method. Variances of residuals of predictions of x_2 with or without knowledge of y_2 are depicted. Adding information from Y did not improve predictions, indicating that Y does not drive X.

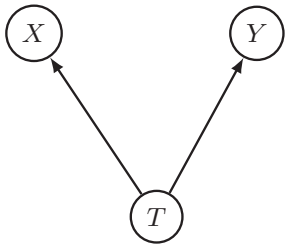

FIG. 8. Interaction graph for fishery model (2) of two populations driven by an environmental variable.

The variable p was chosen so that Var(T) is equal to 1.

1. Granger causality test

As in the previous case, again, we started with the Granger VAR test. We considered lags up to the value of 15. A bidirectional causal link was (incorrectly) detected between X and Y. In both directions p value $< 10^{-15}$. The fishery model is actually another improper example for application of the Granger test. The reason is that the test does not account for latent confounding effects represented here by the shared environmental forcing.

2. Convergent cross-mapping

The fishery model was used in Ref. [10] to illustrate the role of the convergence of cross-mapping results for an increasing number of data. As Fig. 10 shows, significant cross-correlation between species suggests that they might be coupled. However, cross-mappings do not converge for an increasing number of data in any of the two directions, indicating that X and Y are not coupled. We can conclude that the CCM method was able to visually distinguish true causal relation from a mere correlation produced by shared driving variable.

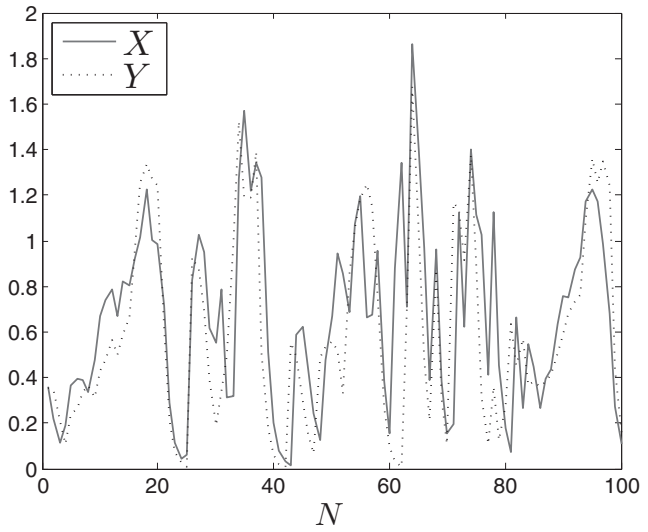

FIG. 9. Sample of 100 data points from the behavior of populations X and Y of the fishery model.

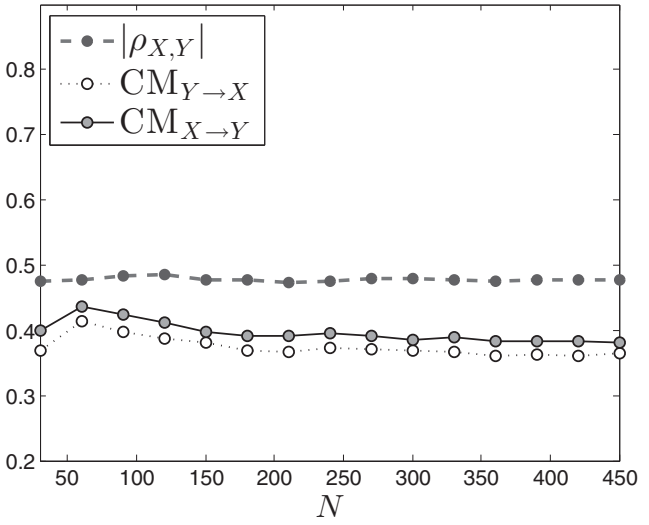

FIG. 10. Convergence of CM investigated in fishery model. The figure indicates that the X and Y are correlated but not causally related.

3. Optimized mixed prediction

Finally, the fishery model was tested by the proposed OMP method.

To estimate the proper embedding parameters, we evaluated the prediction errors for different combinations of parameters and looked for a combination leading to the lowest error. We found that for reconstruction of spaces M_X and M_Y delay $\tau = 1$ and embedding dimension $d = 6$ were optimal. Then the corresponding mixed state space M_{X+Y} had dimension 12.

Next, we looked for optimal weights to reduce the dimensionality of the reconstructed spaces. The optimal weights were estimated based on 1000 predictions, using 50 000 past data points to search for nearest neighbors. For an optimal prediction of y the next weights were found in M_Y:

y_t	y_{t-1}	y_{t-2}	y_{t-3}	y_{t-4}	y_{t-5}
0.128	0.172	0.178	0.183	0.177	0.162

Then, we added the delay coordinates of variable x to form a 12-dimensional M_{X+Y} and looked for the optimal weights for predicting y. However, the weights for the six added x coordinates were found to be negligibly small. Obviously, knowing X did not help improve the prediction of y. This, in the next step, was also confirmed statistically. In spaces M_Y and M_{X+Y} with adjusted weights, we made 5000 one-point predictions based on 50 000 historical data points. The variance of the residuals for prediction in M_Y was not significantly higher than the variance obtained in M_{X+Y} (p value > 0.64). This rules out causality in direction $X \to Y$.

In an analogous way, the causality in direction $Y \to X$ was excluded (p value > 0.39).

It was confirmed that the OMP method was not misled by external force influencing two independent time series. Causal link was correctly rejected in both directions.

IV. CONCLUSION

In this study, three methods were tested to detect causality between two systems represented by time series. The first method is the Granger VAR test of causality. The other two methods, the convergent cross-mapping [10] method and optimized mixed prediction method introduced in this paper, work in reconstructed state spaces.

The methods were compared on test examples of fishery model and unidirectional connection of chaotic systems of Rössler and Lorenz type. The goal was to compare the approaches in relation to their ability to detect unidirectional coupling and to distinguish causality from the correlation.

The results show that blind application of the Granger VAR test easily leads to incorrect conclusions.

The CCM method results in graphs from which the causal relations can mostly be correctly inferred.

The OMP method, introduced in this paper, is similar to the original Granger approach in that it evaluates the predictability. To decide whether X drives Y the state space of Y was reconstructed in two ways: with information from X and without it. After optimization described in the text above, predictabilities of Y made in the two reconstructions were compared. The opposite direction of Y driving X was examined analogously. Then the direction of coupling was inferred based on asymmetries emerging from the calculations along the two possible causal directions. In the test examples, the predictability improvements behaved according to theoretical expectations.

Causality was significantly detectable even for limited number of data. In our test cases, a few hundred data points were sufficient to satisfy the key requirement for several orbits around the reconstructed state portrait. Increasing the number of data by increasing the sampling frequency is not substantially helpful. However, demand for data would grow with increased degrees of freedom of systems or in the presence of noise.

Investigating the effects of noise was left for future research. However, there were indications that the asymmetry associated with predictability improvement can even be enhanced in the presence of noise [11].

In OMP, like in all state space-based techniques, the embedding dimension and time delay selected for reconstruction play an important role in the successful application of the method. The choice of the parameters lacks general rules. According to our experience, a safe way to find the optimum embedding parameters is to calculate the prediction errors for several combinations of dimension and delay and choose the one that leads to the lowest error.

Less useful coordinates of the reconstructed spaces can be further suppressed by assigning a lower weight. The optimization problem related to the adjustments of weights is a complex and computationally expensive task. Therefore, it might be worthwhile to consider other techniques as an alternative to the Matlab implementation of the pattern search algorithm used here. In this context, the search for neighbors may also deserve attention. Replacement of the Matlab version of exhaustive nearest neighbors searcher by an algorithm presented in [28], which has been declared to be useful when the fractal dimension of the data set is considerably smaller than the dimension of the embedding space, could be considered.

An important part of the OMP method is the statistical testing of the variances of residuals, thanks to which we obtain conclusion on causation at a certain level of significance.

Finally, we conclude that the OMP method introduced in this paper managed to reveal the presence and the direction of coupling and also to differentiate correlation and causality. In addition to detecting causality, the method can be useful also to those who are interested in the topic of predicting causally influenced time series.

ACKNOWLEDGMENTS

This work was supported by the Slovak Grant Agency for Science (Grant No. 2/0011/16) and by the Slovak Research and Development Agency (Grant APVV-15-0295).

[1] C. W. Granger, Econometrica **37**(3), 424 (1969).
[2] M. Paluš and M. Vejmelka, Phys. Rev. E **75**, 056211 (2007).
[3] F. Takens, in *Dynamical Systems and Turbulence*, edited by D. A. Rand and L. S. Young (Springer-Verlag, Berlin, 1981), pp. 366–381.
[4] N. F. Rulkov, M. M. Sushchik, L. S. Tsimring, and H. D. I. Abarbanel, Phys. Rev. E **51**, 980 (1995).
[5] S. J. Schiff, P. So, T. Chang, R. E. Burke, and T. Sauer, Phys. Rev. E **54**, 6708 (1996).
[6] J. Arnhold, P. Grassberger, K. Lehnertz, and C. E. Elger, Physica D **134**, 419 (1999).
[7] R. Q. Quiroga, A. Kraskov, T. Kreuz, and P. Grassberger, Phys. Rev. E **65**, 041903 (2002).
[8] R. G. Andrzejak, A. Kraskov, H. Stögbauer, F. Mormann, and T. Kreuz, Phys. Rev. E **68**, 066202 (2003).
[9] D. Chicharro and R. G. Andrzejak, Phys. Rev. E **80**, 026217 (2009).
[10] G. Sugihara, R. May, H. Ye, C.-H. Hsieh, E. Deyle, M. Fogarty, and S. Munch, Science **338**, 496 (2012).
[11] L. Faes, A. Porta, and G. Nollo, Phys. Rev. E **78**, 026201 (2008).
[12] Y. Chen, G. Rangarajan, J. Feng, and M. Ding, Phys. Lett. A **324**, 26 (2004).
[13] M. Wiesenfeldt, U. Parlitz, and W. Lauterborn, Int. J. Bifurcat. Chaos **11**, 2217 (2001).
[14] U. Feldmann and J. Bhattacharya, Int. J. Bifurcat. Chaos **14**, 505 (2004).
[15] M. Lungarella, K. Ishiguro, Y. Kuniyoshi, and N. Otsu, Int. J. Bifurcat. Chaos **17**, 903 (2007).
[16] T. Schreiber, Phys. Rev. Lett. **85**, 461 (2000).
[17] C. Lutz, Granger causality test: Matlab function, http://www.mathworks.com/matlabcentral/ (2009).
[18] E. N. Lorenz, J. Atmos. Sci. **26**, 636 (1969).
[19] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap* (CRC Press, Boca Raton, FL, 1994).

[20] A. Krakovská, K. Mezeiová, and H. Budáčová, J. Complex Syst. **2015**, 932750 (2015).
[21] L. M. Pecora and T. L. Carroll, Phys. Rev. Lett. **64**, 821 (1990).
[22] K. Pyragas, Phys. Rev. E **54**, R4508 (1996).
[23] M. Le Van Quyen, J. Martinerie, C. Adam, and F. J. Varela, Physica D **127**, 250 (1999).
[24] R. Q. Quiroga, J. Arnhold, and P. Grassberger, Phys. Rev. E **61**, 5142 (2000).
[25] M. Paluš, V. Komárek, Z. Hrnčíř, and K. Štěrbová, Phys. Rev. E **63**, 046211 (2001).
[26] B. Cummins, T. Gedeon, and K. Spendlove, SIAM J. Appl. Dyn. Syst. **14**, 335 (2015).
[27] J. Theiler, Phys. Rev. A **34**, 2427 (1986).
[28] C. Merkwirth, U. Parlitz, and W. Lauterborn, Phys. Rev. E **62**, 2089 (2000).