sciendo
ISSN 1335-8871
MEASUREMENT SCIENCE REVIEW
Journal homepage: https://content.sciendo.com

Against the Flow of Time with Multi-Output Models

Jozef Jakubík[1]*, Mary Phuong[2], Martina Chvosteková[1], Anna Krakovská[1]

[1]*Institute of Measurement Science, Slovak Academy of Sciences, Dúbravská cesta 9, Bratislava, Slovakia*
[2]*IST Austria, Am Campus 1, Klosterneuburg, Austria*

Abstract: Recent work has paid close attention to the first principle of Granger causality, according to which cause precedes effect. In this context, the question may arise whether the detected direction of causality also reverses after the time reversal of unidirectionally coupled data. Recently, it has been shown that for unidirectionally causally connected autoregressive (AR) processes $X \to Y$, after time reversal of data, the opposite causal direction $Y \to X$ is indeed detected, although typically as part of the bidirectional $X \leftrightarrow Y$ link. As we argue here, the answer is different when the measured data are not from AR processes but from linked deterministic systems. When the goal is the usual forward data analysis, cross-mapping-like approaches correctly detect $X \to Y$, while Granger causality-like approaches, which should not be used for deterministic time series, detect causal independence $X \perp\!\!\!\perp Y$. The results of backward causal analysis depend on the predictability of the reversed data. Unlike AR processes, observables from deterministic dynamical systems, even complex nonlinear ones, can be predicted well forward, while backward predictions can be difficult (notably when the time reversal of a function leads to one-to-many relations). To address this problem, we propose an approach based on models that provide multiple candidate predictions for the target, combined with a loss function that consideres only the best candidate. The resulting good forward and backward predictability supports the view that unidirectionally causally linked deterministic dynamical systems $X \to Y$ can be expected to detect the same link both before and after time reversal.

Keywords: Causality, deterministic dynamical systems, reversibility, multi-output prediction.

1. Introduction

Detecting causality between two deterministic dynamical systems or autoregressive (AR) processes is an open problem with an active community. Probably the best-known concept in this field is Granger causality. We say that process X *Granger causes* process Y if it is possible to obtain a better prediction of Y using both past values of X and Y than using only Y [1]. Considering the first principle of Granger causality that cause precedes effect, one might expect a change of the direction of causality from $X \to Y$ to $Y \to X$ when unidirectionally coupled systems and their time-reversals are analysed. It might even seem like a good idea to try to reverse the time to confirm the conclusion about the direction of the causal link. Since this issue is closely related to important open physical questions about the arrow of time and the mentioned change of causality direction would have a number of interesting applications in the analysis of real data, time-reversing approaches have received much attention recently [2–7]. Results have shown that there is no consensus on basic questions, such as whether detected causality should actually change for time-reversed data or whether the directions of the links may differ depending on the time series and the method of detection.

In articles [3, 4] the properties of Granger causality for reversed time were introduced and studied. Then in [6] the direction of information flow was studied for stochastic data defined by vector AR processes. It was shown that for AR processes X and Y connected in the $X \to Y$ way, and for the Granger causality test, we cannot expect to detect the opposite unidirectional $Y \to X$ link when we analyze time-reversed data. In fact, what we usually detect is a bi-directional $X \leftrightarrow Y$ link. Contrary to some expectations, a solely unidirectional $Y \to X$ is detected only when the product of the connection strength and the ratio of the predictive errors of the driver relative to the recipient is below a certain level [6]. The much more frequent $X \leftrightarrow Y$ result is quite surprising when it comes to the arrow of time associated with causality.

When we begin to investigate other processes than AR, further complications arise. In article [2], examples of different types of systems were used to illustrate the behaviour of the causal methods for time-reversed data. The paper has drawn attention to the fact that in the case of deterministic dynamical systems, the result of comparing forward and backward causal analysis is different than in the case of AR processes, which is interesting in the context of the first principle of Granger causality about the cause preceding the effect.

The question, then, is what to expect from a causal analysis of forward and time-reversed data when it comes to the interconnection of deterministic dynamical systems. In addressing this question, we will focus on unidirectionally coupled discrete dynamical systems, but the findings will have more

DOI: 10.2478/msr-2023-0023

*Corresponding author: jozef.jakubik.jefo@gmail.com (J. Jakubík)

general validity.

In causality detection methods, we can identify two basic approaches:

- Granger causality-like (GC-like) approaches – based on the ability to improve the fit of the original data with information from other variables (time series). These include, for example, Granger causality [1], kernel nonlinear Granger causality [8], and the predictability improvement method [9].
- Cross-mapping-like (CM-like) approaches – based on the ability to express one variable (time series) from other variables (time series). These include, for example, measures L and M [10], convergent cross-mapping [11], and the cross-prediction method [12].

In the first part of this study (Section 2), we discuss the theoretical expectations for forward and backward causal analysis of deterministic dynamical systems. We point out the inconsistency in causality detection after reversing the time for different systems. The problem we identify is the inability to learn one-to-many relations arising from the reversal of many-to-one functions, such as a Logistic map. In Section 2, we show how this problem can affect causality detection in reversed time. Then, in Section 3, we propose an approach to deal with one-to-many relations. The proposed approach is based on a multi-output model and a newly defined adaptive mean squared error (AMSE) loss (Section A). It illustrates the ability of the multi-output model and AMSE loss to overcome the problem of learning one-to-many relations that occur in causality detection in reversed time.

Consequently, based on the numerically achieved good forward and backward predictability, we conclude that for unidirectionally coupled deterministic dynamical systems $X \to Y$, the same $X \to Y$ link will be detected both for the direct and reversed data, using CM-like causal methods.

2. Forward and backward causality in dynamical systems

Consider a dynamical system (DS), which is a unidirectional coupling $X \to Y$ of two discrete systems:

$$\begin{aligned} x(t+1) &= f(x(t),\dots,x(t-\tau_x)),\\ y(t+1) &= g(y(t),\dots,y(t-\tau_y),x(t)), \end{aligned} \tag{1}$$

where τ_x and τ_y are integers.

The following definitions are useful mainly for backward analysis. We say that variable x is *expressible* from variable y if there exists a function h such that

$$x(t+1) = h(y(t),\dots,y(t-\tau_k)).$$

We call variable x *self-expressible* if it can be expressed as a function of its own previous values (e.g., variable x from (1)):

$$x(t+1) = f(x(t),\dots,x(t-\tau_x)).$$

The variable x, given by the function f:

$$x(t+1) = f(x(t),\dots,x(t-\tau_a),y(t),\dots,y(t-\tau_b)),$$

is called *reversible* if there exists a function g such that

$$x(t+1) = g(x(t+2),\dots,x(t+\tau_c),y(t+2),\dots,y(t+\tau_d)).$$

In the following sections, we will use examples of three coupled chaotic systems of Hénon and Logistic maps.

Hénon map $\to$ Hénon map:

$$\begin{aligned} x(t+1) &= 1.4 - x^2(t) + 0.3x(t-1),\\ y(t+1) &= 1.4 - (Cx(t)y(t) + (1-C)y^2(t)) + 0.3y(t-1), \end{aligned} \tag{2}$$

where C is the coupling strength and is equal to 0.4.

Hénon map $\to$ Logistic map with linear coupling coupling:

$$\begin{aligned} x(t+1) &= 1.4 - x^2(t) + 0.3x(t-1),\\ y(t+1) &= 3.7y(t)\,(1-y(t)) - 0.3x(t). \end{aligned} \tag{3}$$

Hénon map $\to$ Logistic map with square coupling:

$$\begin{aligned} x(t+1) &= 1.4 - x^2(t) + 0.3x(t-1),\\ y(t+1) &= 3.7y(t)\,(1-y(t)) - 0.3x^2(t). \end{aligned} \tag{4}$$

Let us break down the equations to determine which variables are expressible, self-expressible, and reversible. These properties play a role in explaining the diverse results of the causal analysis of dynamical systems.

First, we focus on DS given by (2), and show that not only x but also y is self-expressible. Each $y(t)$ can be expressed as a function of previous values $y(t-1),\cdots$:

$$y(t+1) = 1.4 - (Cx(t)y(t) + (1-C)y^2(t)) + 0.3y(t-1),$$

where

$$\begin{aligned} x(t) &= 1.4 - x^2(t-1) + 0.3x(t-2),\\ x(t-1) &= \frac{1.4 - y(t) - (1-C)y^2(t-1) + 0.3y(t-2)}{Cy(t-1)},\\ x(t-2) &= \frac{1.4 - y(t-1) - (1-C)y^2(t-2) + 0.3y(t-3)}{Cy(t-2)}. \end{aligned}$$

Now we show that the variable x is expressible from the variable y:

$$x(t+1) = 1.4 - x^2(t) + 0.3x(t-1),$$

where

$$\begin{aligned} x(t) &= 1.4 - x^2(t-1) + 0.3x(t-2),\\ x(t-1) &= \frac{1.4 - y(t) - (1-C)y^2(t-1) + 0.3y(t-2)}{Cy(t-1)},\\ x(t-2) &= \frac{1.4 - y(t-1) - (1-C)y^2(t-2) + 0.3y(t-3)}{Cy(t-2)}. \end{aligned}$$

Table 1. Forward causality analysis of unidirectional coupling $X \rightarrow Y$ of deterministic dynamical systems. Expected outcomes of GC-like and CM-like approaches when using single-output prediction methods.

GC-like	CM-like
$X \perp\!\!\!\perp Y$	$X \rightarrow Y$

For DS given by (3) and (4), we can derive the same properties: variable y (Logistic map) is self-expressible and variable x is expressible from variable y in both cases.

Both variables are reversible:

$$x(t-1) = \frac{-1.4 + x^2(t) + x(t+1)}{0.3},$$
$$y(t-1) = \frac{-1.4 + (Cx(t)y(t) + (1-C)y^2(t)) + y(t+1)}{0.3}$$

and it is easy to see that both time-reverted variables are self-expressible (x by definition) and the variable x is expressible from the variable y:

$$y(t-1) = \frac{-1.4 + (Cx(t)y(t) + (1-C)y^2(t)) + y(t+1)}{0.3},$$

where

$$x(t) = \frac{-1.4 + x^2(t+1) + x(t+2)}{0.3},$$
$$x(t+1) = \frac{0.3y(t) + 1.4 - y(t+2) - (1-C)y^2(t+1)}{Cy(t+1)},$$
$$x(t+2) = \frac{0.3y(t+1) + 1.4 - y(t+3) - (1-C)y^2(t+2)}{Cy(t+2)}.$$

The systems given by (3) and (4) include the non-reversible Logistic map.

In this paper we consider the unidirectional coupling $X \rightarrow Y$ (1) of two deterministic systems, where the representative variable x of system X is expressible from variable y and both x and y are self-expressible in the forward analysis. We would like to point out how the outputs of individual causal methods are influenced by these assumptions. We will apply causality detection to both forward and time-reversed data and deduce what may happen under the given conditions.

Table 1 summarises the obvious expectations in forward causal analysis of deterministic systems by the GC-like and the CM-like methods, which are as follows:

Forward 1 (GC-like) Since variable y is self-expressible, GC-like methods theoretically fail to detect $X \rightarrow Y$ causality, since there is no reason to expect an improvement in the prediction of y after adding information from variable x. The expected result would be the detection of independence $X \perp\!\!\!\perp Y$.

Forward 2 (CM-like) Since variable x is expressible from variable y, we can correctly detect the $X \rightarrow Y$ causal link with CM-like methods. The essence of the method is that the driving variable x can be expressed as a function of previous observations of the driven variable y, and therefore, it is possible to find a function that transforms Y to X.

The next step is to discuss the expectations of the reversed-time causal analysis. Recall that our ultimate goal is to determine whether, given an $X \rightarrow Y$ coupling of deterministic systems, the detected causality changes to $Y \rightarrow X$ after the time reversal of the observables.

Backwards 1 (GC-like) If the reversed variables x and y are self-expressible, then GC-like methods, analogous to the forward direction, theoretically find no improvement in prediction and thus no causal link. Thus, the theoretical result of GC-like methods for the reversed data should be $X \perp\!\!\!\perp Y$. As an example, see DS in (2) (Hénon map $\rightarrow$ Hénon map).

Backwards 2 (GC-like) If variable y is not reversible, but variable x is reversible, then the GC-like methods can detect causality $Y \rightarrow X$ because they can find an improvement in the backward prediction of y after adding information from x. The example DS can be found in (3) (Hénon map $\rightarrow$ Logistic map).

Backwards 3 (CM-like) If the reversed x is expressible from the reversed y, we can detect causality $X \rightarrow Y$ with CM-like methods. Thus, using this method, the observed causality does not change for the reversed data. DS in (2) (Hénon map $\rightarrow$ Hénon map) can serve as an example.

Backwards 4 (CM-like) If the reversed variable x is not expressible from the reversed variable y, we detect no causality with CM-like methods. Let us take (4) as an example (Hénon map $\rightarrow$ Logistic map with square coupling).

Backwards 5 (GC-like) It can be deduced that blind application of GC-like methods for irreversible x - depending on various conditions - can lead to any causal detection. For completeness, they are all listed in Table 2, but we will not discuss them in more detail because, as we have already emphasised, GC-like methods are actually inappropriate for application to data from dynamical systems.

In Table 2 we summarise the possible outcomes of causality detection for $X \rightarrow Y$ coupling of deterministic dynamical systems in the backward direction. Theoretically, as argued above, the conclusion of GC-like methods should be that the systems are detected as causally independent if both variables are reversible. If one or both variables are not reversible, the result may be different from the forward analysis. As summarised in Table 2, for the case of $X \rightarrow Y$ (in the forward sense), the inappropriate backward GC-like analysis can lead to any outcome.

Table 2. Backward causality analysis of unidirectional coupling $X \to Y$ of deterministic dynamical systems. Outcomes of GC-like and CM-like approaches when using single-output prediction methods.

	CM-like
x is expressible from reversed y	$X \to Y$
x is not expressible from reversed y	$X \perp\!\!\!\perp Y$
	GC-like
reversed x and y are self-expressible	$X \perp\!\!\!\perp Y$
not reversible y and reversible x	$X \to Y$
other options	$X \perp\!\!\!\perp Y, X \to Y, X \leftarrow Y, X \leftrightarrow Y$

Proper use of CM-like methods, on the other hand, will detect $X \to Y$ in the backward analysis if variable x is expressible from reversed variable y. Otherwise, CM-like methods will conclude that the systems are causally independent.

The theoretically expected results of the backward causal analysis, shown in Table 2, follow directly from the concepts behind CM-like and GC-like methods applied to data from deterministic systems. For clarity, let us illustrate them with our examples of coupled Hénon and Logistic maps - (2), (3) and (4).

A. Single-output model

Both GC-like and CM-like bivariate causal methods rely on the ability to predict observables using their own past (self-predictions) and the past of the second observable (cross-predictions) or to improve the self-prediction using the past of the other observable. Therefore, for our three coupled test systems given by (2), (3) and (4), we evaluated the predictions for direct and time-reversed variables in a standard manner.

As a single-output prediction model, we used a 3-layer neural network with 10 nodes on each layer and the hyperbolic tangent activation function $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ (there is no nonlinearity in the output layer). We estimated $x(t)$ based on three preceding (for forward prediction), or three succeeding (for backward prediction) observations. Therefore, there were three input nodes for self-predictions and six input nodes for mixed predictions. We trained the model with the MSE objective function on the first 10000 data points.

The numerical results for DS given by (2), (3) and (4) are presented in Table 3.

Since variable y is self-expressible, in the connection $X \to Y$ of deterministic dynamical systems, the prediction of the observable from Y cannot be improved by adding information from X, and GC-like methods are not theoretically appropriate to analyse these data. This is also evidenced by the numerical results in Table 3, where the forward self-predictions of the variables are not improved after adding information from the other time series - in all cases the prediction errors are smaller than 10^{-4}. This was already stated in Table 1, saying that the forward Granger causal analysis of the $X \to Y$ interconnection of the tested dynamical systems theoretically leads to an incorrect $X \perp\!\!\!\perp Y$ conclusion [6]. The numerical results in Table 3 are consistent with the theoretical expectations given in Table 1.

As for the CM-like methods, in the forward analysis, the numerical results in Table 3 show the successful cross-prediction of the observable from X and thus the detection of $X \to Y$ in all tested examples (prediction error is smaller than 10^{-4}). In the backward analysis, Table 2 shows that a CM-like method for different systems with the same coupling arrow $X \to Y$ can theoretically lead to the conclusions $X \to Y$ or $X \perp\!\!\!\perp Y$ for time-reversed data. For example, in testing systems with Logistic map involved, the cross-prediction was not successful and the conclusion of the CM-like method was $X \perp\!\!\!\perp Y$ (prediction error is significant - 0.018). Thus, with the CM-like methods used for causal analysis, the numerical results in Table 3 were again consistent with the expectations presented in Tables 1 and 2.

The ambiguity of numerical results is related to the difficult backward predictability of irreversible time series. In what follows, we propose a procedure for modifying the backward prediction process, that eliminates the irreversibility problem and the associated inconsistency in the conclusions of backward causal analysis.

3. Multi-output prediction

The inconsistent results of the backward causal analysis in Table 2 are related to the problems in predicting time-reversed data. Forecasting models (such as various types of regression) usually provide single-output results, although for processes that are not reversible in time (many-to-one function), predicting in the reverse direction should consider more different possible outcomes. The function $x \mapsto x^2$, for example, yields the value 1 for both $x = 1$ and $x = -1$, but in the backward direction we cannot unambiguously determine from which value (1 or -1) we obtained the outcome 1. In this case, the single-output approaches that minimise the MSE expect 0 to be the best prediction. That is why, even in the case of Logistic map, it is not possible to express $x(t)$ simply as a function of $x(t+1)$ from (3) and (4).

As we can see in Table 3, the prediction improvement varies by system type. The reversibility of the variable is crucial for backward causality detection.

Let us now return to the example of x^2. If it were possible to predict both the value 1 and the value -1, the prediction error in one of these cases would be 0. If we take only the value

Table 3. Mean squared error (MSE) of forward and backward prediction of 2000 points of the studied observables and conclusions derived for causal detection. The size of the error ~ 0 (in our case $< 10^{-4}$) depends on the number of data and the number of optimisation iterations.

	Hénon → Hénon		Hénon → Logistic		Hénon → Logistic with square coupling	
	forward	backward	forward	backward	forward	backward
X from X	~ 0	~ 0	~ 0	~ 0	~ 0	~ 0
X from X and Y	~ 0	~ 0	~ 0	~ 0	~ 0	~ 0
Y from Y	~ 0	~ 0	~ 0	0.018	~ 0	0.018
Y from X and Y	~ 0	~ 0	~ 0	0.013	~ 0	0.014
GC-like result	$X \perp\!\!\!\perp Y$	$X \perp\!\!\!\perp Y$	$X \perp\!\!\!\perp Y$	$X \rightarrow Y$	$X \perp\!\!\!\perp Y$	$X \rightarrow Y$
X from Y	~ 0	~ 0	~ 0	~ 0	~ 0	0.018
Y from X	0.02	0.02	0.018	0.018	0.018	0.018
CM-like result	$X \rightarrow Y$	$X \rightarrow Y$	$X \rightarrow Y$	$X \rightarrow Y$	$X \rightarrow Y$	$X \perp\!\!\!\perp Y$

with the smaller prediction error, the backward prediction result should be close to the true value. This is the main idea for the adaptive mean squared error objective function presented here.

A. Adaptive mean squared error

Let $\mathbf{X} \in \mathbb{R}^{n \times Q}$ be the input matrix and $\mathbf{Y} \in \mathbb{R}^n$ be the n-dimensional observation vector. Our goal is to estimate the one-to-many relationship between $\mathbf{X}$ and $\mathbf{Y}$. We denote by $\hat{\mathbf{Y}} \in \mathbb{R}^{n \times P}$ the estimates of $\mathbf{Y}$, where each row of $\hat{\mathbf{Y}}$ contains the P different candidate predictions of $\mathbf{Y}$ for a given $\mathbf{X}$.

We use an L-layer multi-output neural network as our model:

$$\hat{\mathbf{Y}}[i,:]^\intercal = \mathbf{b}_L + \mathbf{W}_L^\intercal \sigma\left(\dots \sigma\left(\mathbf{b}_1 + \mathbf{W}_1^\intercal \mathbf{X}[i,:]^\intercal\right)\right), \qquad (5)$$

where σ is an activation function, $\mathbf{W}_j$ is the weight matrix for the j-th layer, and $\mathbf{b}_j$ is the bias vector for the j-th layer.

We propose to train the model with the following objective, which we refer to as *adaptive mean squared error (AMSE)*:

$$\sum_{p=1}^{P} \frac{1}{|S_p|} \sum_{s \in S_p} s, \qquad (6)$$

where for $p = 1, \dots, P$, we define the sets S_p as

$$S_p = \big\{(\hat{\mathbf{Y}}[i,p] - \mathbf{Y}[i])^2 : p = \underset{p' \in \{1,\dots,P\}}{\arg\min}\, (\hat{\mathbf{Y}}[i,p'] - \mathbf{Y}[i])^2\big\}.$$

Intuitively, AMSE splits the training data into p subsets, each consisting of data points best predicted by the p-th model output. The AMSE loss is then calculated as the sum of the mean squared errors for the p subsets.

This multi-output design combined with the AMSE loss allows for better fitting of one-to-many data. For example, if we want to "invert" the quadratic function, as in the Logistic map examples (3) and (4), one of the output nodes fits the "lower" part of the estimated function and the other output node fits the "upper" part of the estimated function (see Fig. 1).

In addition to the standard hyperparameters of the prediction model (in our case, the multi-output model), we have one more hyperparameter – the number of outputs. In the Logistic map example, the correct number of output nodes is obvious, but for real data (such as the German electricity consumption data [13] used in Section B) we choose the number of output nodes based on a validation set.

We consider two different validation metrics for choosing the number of output nodes. The natural first choice is the AMSE, as defined in (6). It turns out that this metric often decreases with the number of outputs, so we (heuristically) choose the number of outputs using an elbow diagram [14].

The second metric we propose is the *adaptive mean absolute error (AMAE)*, given by

$$\sum_{p=1}^{P} \frac{1}{|A_p|} \sum_{a \in A_p} a, \qquad (7)$$

where for $p = 1, \dots, P$, we define the sets A_p as

$$A_p = \Big\{ \big|\hat{\mathbf{Y}}[i,p] - \mathbf{Y}[i]\big| : p = \underset{p' \in \{1,\dots,P\}}{\arg\min}\, \big|\hat{\mathbf{Y}}[i,p'] - \mathbf{Y}[i]\big| \Big\}.$$

This metric is similar to the AMSE, but based on absolute error instead of squared error. We find that this metric does not prefer a large number of outputs (like the AMSE does), so we choose the optimal number by minimising the validation AMAE. An intuitive argument for why AMSE prefers a large number of outputs and AMAE does not is that for P outputs, the absolute error is roughly of order $1/P$, while the squared error is of order $1/P^2$. If you sum over p in (7) and (6), we find that AMAE is of order 1 and AMSE is of the order $1/P$.

Table 4. Mean and standard deviation for repeated (over 100 repetitions) experiment with multiple-output nodes models for validation data from German daily electricity consumption.

	AMAE	AMSE
output nodes	mean $\pm\sqrt{\text{var}}$	mean $\pm\sqrt{\text{var}}$
1	0.2733± 0.0094	0.1833 ± 0.0117
2	0.2071± 0.0106	0.0409 ± 0.0075
3	0.2501± 0.0082	0.0378 ± 0.0019
4	0.2692± 0.0140	0.0330 ± 0.0046

B. Examples

We will illustrate our method with two examples of irreversible data. The first one is the chaotic Logistic map:

$$x(t+1) = 3.7x(t)\,(1-x(t)) \tag{8}$$

and the second is the real data example of country-wide totals of daily German electricity consumption for the years 2006-2017 [13].

We used the same approach as described in Section A, except that we used the multi-output model and instead of the MSE objective function we used the AMSE objective function.

Logistic map

For the Logistic map, we generated a time series X of length $n = 12000$ according to (8). We modelled the function f:

$$x(t-3) = f(x(t), x(t-1), x(t-2)),$$

with a neural network (5) with two hidden layers and ten nodes per layer. The number of inputs was three, the number of outputs was two, and the activation function was the hyperbolic tangent. The objective function was AMSE. For optimisation, we used stochastic gradient descent with a learning rate of 0.1 and a momentum of 0.9. We performed 5000 optimisation steps. We trained the model with the training data and then evaluated it with the test data. The results can be found in Section C.

German daily electricity consumption data

We used German electricity consumption measurements not in a causal context, but only as an example of a single real-time series where backward prediction requires consideration of multiple outputs. For these data, we considered time series with a length of 4383. We normalised the data to zero mean and unit variance, reverted the time, and split the data into a training set (2500), a validation set (500), and a test set (1383). We varied the number of output nodes from 1 to 4 and selected the best number based on the validation AMAE. Otherwise, we used the same experimental setup as described above and repeated the entire experiment 100 times for each number of output nodes. The results can be found in Section C.

C. Results of multi-output prediction

Logistic map

In Fig. 1, we present the predictions for each output node of both the two-output model and single-output model and compare them to the ground truth. The multi-output model fits both parts of the curve, forming an upper and a lower envelope of the data. Additionally, this example shows that our method can handle "imbalanced" data, where the different parts of the series have different sizes. As expected, the single-output model focused on the upper branch (see the left bottom part of Fig. 1), but could not fit the non-bijective part of the function (the lower branch and values around 0.7). The multi-output model fits both the upper part (red - second output) and the lower part (yellow - first output) of the function. For the causality detection process, we use the smaller prediction errors of the two output nodes.

German daily electricity consumption data

The results for the electricity consumption data can be found in Fig. 2. We present the predictions for each output node of both the two-output model and the single-output model and compare them to the ground truth. The results for the electricity consumption data in Table 4 show the validation AMSE and AMAE for the different number of outputs. We report the mean and standard deviation over 100 repetitions. The lowest AMAE value (and the greatest difference in AMSE) value is obtained for the two-output models, indicating that this type of model is preferable for the particular dataset. For the visualisation of the test data, we selected the two-output model with the lowest AMAE value; the results are shown in Fig. 2. The single-output model fails to predict some lower values (see the left upper part of Fig. 2). The multi-output model fits both the upper part (red - second output) and the lower part (yellow - first output) of the function.

4. Results of backward causal analysis using multi-output prediction

In contrast to the single-output prediction errors in Table 3 (see backward results of causality detection), the multi-output prediction extension and AMSE resulted in very low ($< 10^{-4}$) prediction errors even for systems with many-to-one functions (Logistic map).

As a result, the conclusions of forward and backward causal analysis for the GC-like methods became the same $X \perp\!\!\!\perp Y$, because for many-to-one functions, the backward self-prediction error of multi-output models would be as small as the forward self-prediction error and incorporating additional information from other variables would not enhance the predictions.

The results from CM-like methods would also be the same $X \to Y$ for forward and backward analysis since multi-output models make finding transformations for many-to-one functions equally good in the backward and forward directions.

Thus, with multi-output models, we move away from the inconsistent causal conclusions obtained with classical pre-

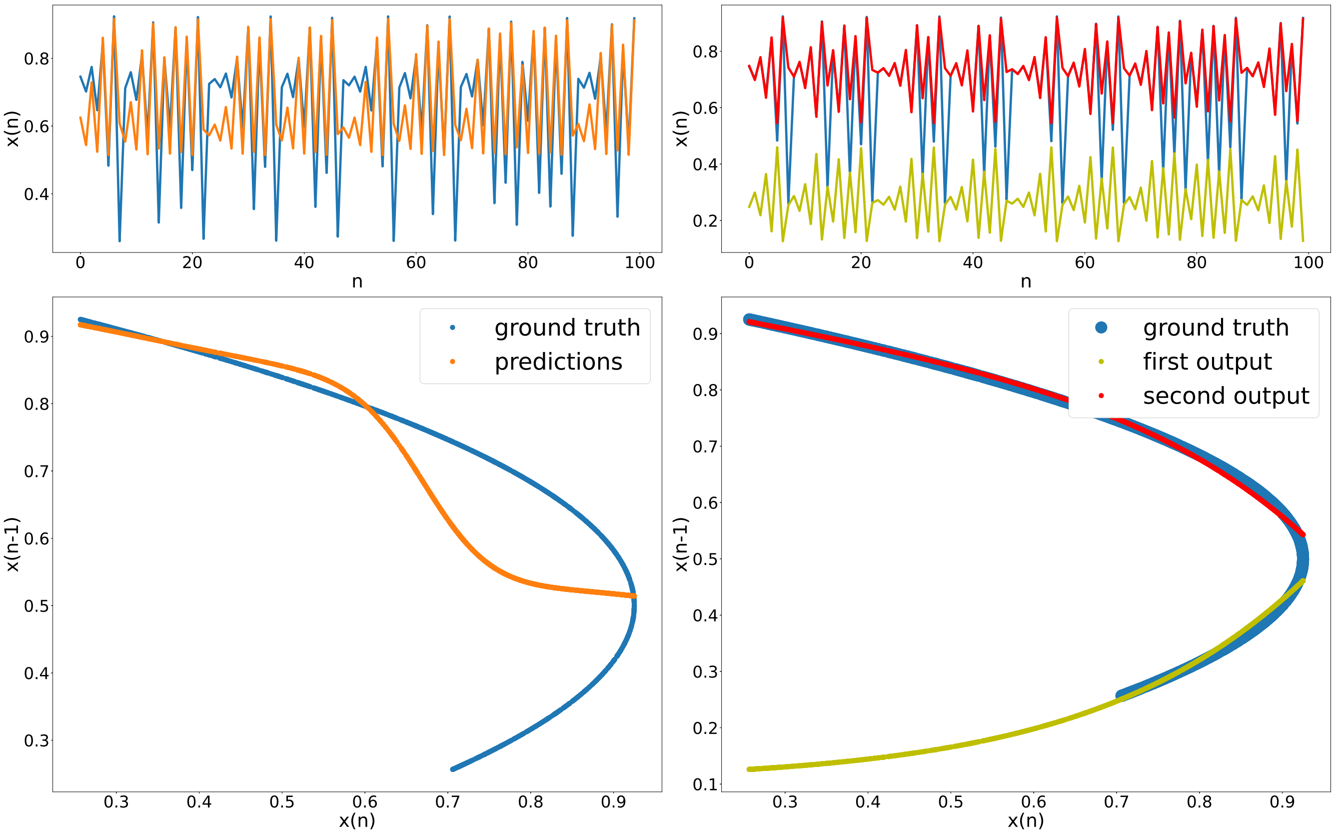

Fig. 1. **Logistic map** (Section B). Visualisation of the time series and the backward predictions corresponding to the single-output (left top) and two-output nodes (right top), and the one-to-many relationship between $x(t)$ and $x(t-1)$, and how it is fitted by the single-output (left bottom) and two-output nodes (right bottom).

Table 5. Forward and backward causality analysis of unidirectional coupling $X \to Y$ of deterministic dynamical systems. Outcomes of GC-like and CM-like methods when using the multi-output prediction method.

	GC-like	CM-like
forward	$X \perp\!\!\!\perp Y$	$X \to Y$
backward	$X \perp\!\!\!\perp Y$	$X \to Y$

diction methods (Tables 1 and 2) and get to the causal detections given in Table 5.

5. Conclusion

The study of causality in time-reversed data is related to questions surrounding the so-called arrow of time and efforts to use Granger's first principle of causality for more reliable causality detection. The nature of these questions leads us to focus on the study of unidirectionally connected processes and their temporal reversals. However, before accepting the rationale for such efforts, we need to clarify whether the direction of detected causality should actually reverse after the data has been reversed.

It turns out that this cannot be unambiguously expected even with the standard Granger analysis of AR processes. Although the change in causality has been shown to occur in the backward analysis of AR models where there is stochasticity and thus irreversibility of processes, the original causal link is usually observed [6] as well.

As we have shown here, the situation is different in the case of the time series originating in deterministic dynamical systems $X \to Y$. GC-like methods are unsuitable for deterministic systems and produce erratic results unrelated to true causality (Tables 1 and 2). On the other hand, with CM-like methods applied to unidirectionally linked deterministic systems, it should be theoretically possible to correctly detect $X \to Y$ in the forward direction. The question remains as to what we can expect in the backward analysis. There, we may face the problem of being unable to predict irreversible processes, which is related to predicting one-to-many relationships. In Section A, we propose an approach to address this issue based on the multi-output models and AMSE. Our results show that the improvement in prediction (as measured by AMAE) can be quite significant. For deterministic data, even for complex nonlinear data, retrospective prediction used to detect backward causality is only about overcoming technical problems.

Thanks to the presented multi-output prediction technique, even irreversible time series of deterministic systems, including dissipative chaotic systems, become practically tractable. As a consequence, for unidirectionally interconnected $X \to Y$ deterministic processes, the direction of the detected causal-

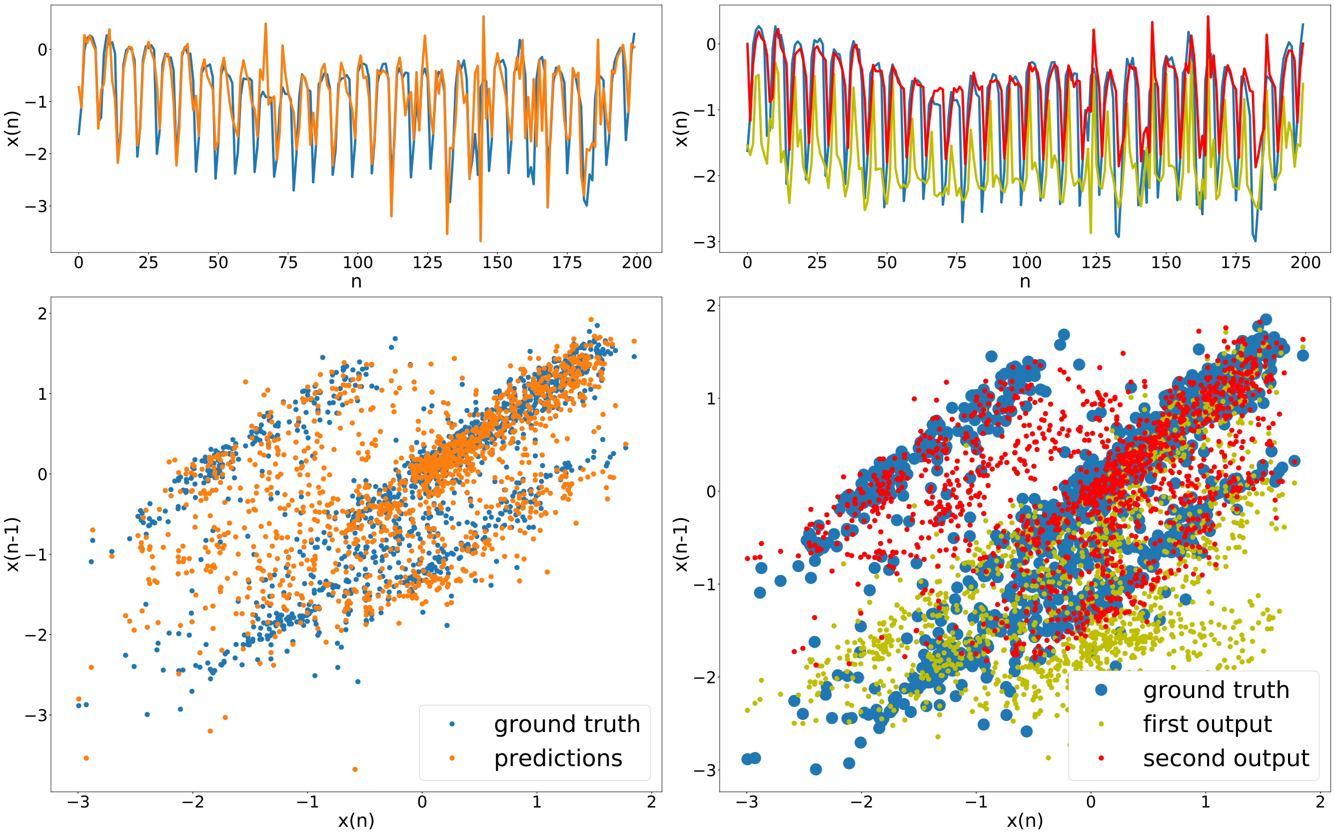

Fig. 2. **German daily electricity consumption data** (Section B). Visualisation of the time series and the backward predictions corresponding to the single-output (left top) and two-output nodes (right top), and the one-to-many relationship between $x(t)$ and $x(t-1)$, and how it is fitted by the single-output (left bottom) and two-output nodes (right bottom)

ity does not change after a data reversal. If we use CM-like methods with improved prediction techniques, we can consistently detect $X \rightarrow Y$ in both forward and backward causal analysis.

The main conclusion can be summarised as follows. If $X \rightarrow Y$ is valid for the time series, we should not expect GC-like and CM-like methods to detect $X \leftarrow Y$ after the time reversal of the data. In fact, for AR modelable data and Granger causality test, we should expect $X \leftrightarrow Y$ [6] and for reversible deterministic time series and CM-like methods, we should expect $X \rightarrow Y$, just as for forward analysis.

In this study, we have drawn attention to symmetry, which is that deterministic time series can be well predicted back and forth. Consequently, for deterministic systems, causal methods based on prediction errors do not help to detect the arrow of time. If we want to look for asymmetry, we can, for example, try to examine the computational complexity of backward prediction compared to forward prediction. But that is a topic for future research.

In future work, it should also be possible to extend the theoretical results to continuous dynamical systems using Takens's theorem [15], which can possibly stand as expressibility and self-expressibility.

Acknowledgments

The work was supported by the Scientific Grant Agency of the Ministry of Education of the Slovak Republic and the Slovak Academy of Sciences, projects APVV-21-0216, VEGA 2-0096-21 and VEGA 2-0023-22.

References

[1] Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society* 424–438. https://doi.org/10.2307/1912791.

[2] Paluš, M., Krakovská, A., Jakubík, J., Chvosteková, M. (2018). Causality, dynamical systems and the arrow of time. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 28(7), 075307. https://doi.org/10.1063/1.5019944.

[3] Haufe, S., Nikulin, V. V., Müller, K.-R., Nolte, G. (2013). A critical assessment of connectivity measures for eeg data: a simulation study. *Neuroimage* 64, 120–133. https://doi.org/10.1016/j.neuroimage.2012.09.036.

[4] Winkler, I., Panknin, D., Bartz, D., Müller, K.-R., Haufe, S. (2016). Validity of time reversal for testing granger causality. *IEEE Transactions on Signal Processing* 64(11), 2746–2760. https://doi.org/10.1109/TSP.2016.2531628.

[5] Riek, R. (2020). Entropy derived from causality. *Entropy* 22(6), 647. https://doi.org/10.3390/e22060647.

[6] Chvosteková, M., Jakubík, J., Krakovská, A. (2021). Granger causality on forward and reversed time series. *Entropy* 23(4), 409. https://doi.org/10.3390/e23040409.

[7] Kořenek, J., Hlinka, J. (2021). Causality in reversed time series: Reversed or conserved? *Entropy* 23(8), 1067. https://doi.org/10.3390/e23081067.

[8] Marinazzo, D., Pellicoro, M., Stramaglia, S. (2008). Kernel method for nonlinear granger causality. *Physical review letters* 100(14), 144103. https://doi.org/10.1103/PhysRevLett.100.144103.

[9] Krakovská, A., Hanzely, F. (2016). Testing for causality in reconstructed state spaces by an optimized mixed prediction method. *Physical Review E* 94(5), 052203. https://doi.org/10.1103/PhysRevE.94.052203.

[10] Chicharro, D., Andrzejak, R. G. (2009). Reliable detection of directional couplings using rank statistics. *Physical Review E* 80(2), 026217. https://doi.org/10.1103/PhysRevE.80.026217.

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

[12] Krakovská, A., Jakubík, J. (2020). Implementation of two causal methods based on predictions in reconstructed state spaces. *Physical Review E* 102(2), 022203. https://doi.org/10.1103/PhysRevE.102.022203.

[13] Walker, J. (2019). Tutorial: Time series analysis with pandas. https://www.dataquest.io/blog/tutorial-time-series-analysis-with-pandas/.

[14] Thorndike, R. L. (1953). Who belongs in the family. In *Psychometrika*. https://doi.org/10.1007/BF02289263.

[15] Takens, F. (1981). Detecting strange attractors in turbulence. In *Dynamical systems and turbulence, Warwick 1980*. Springer, 366–381.

Received January 16, 2023
Accepted July 26, 2023