SURROGATE-BASED TEST FOR GRANGER CAUSALITY

Temujin Gautama & Marc M. Van Hulle
Laboratorium voor Neuro- en Psychofysiologie
K.U.Leuven, Campus Gasthuisberg
Herestraat 49, B-3000 Leuven, BELGIUM
Tel.: + 32 16 34 59 61 Fax: + 32 16 34 59 60
Email: {temu,marc}@neuro.kuleuven.ac.be

Abstract.

A novel approach for testing the presence of Granger causality between two time series is proposed. The residue of the destination signal after self-prediction is computed, after which a cross-prediction of the source signal over this residue is examined. In the absence of causality, there should be no cross-predictive power, due to which the performance of the cross-prediction system can be used as an indication of causality. The proposed approach uses the surrogate data method, and implements the self- and cross-prediction systems as feedforward neural networks. It is tested on synthetic examples, and a sensitivity analysis demonstrates the robustness of the approach.

INTRODUCTION

The operational definition of Granger causality [3] is based on a comparison between the predictive ability of a restricted and an unrestricted prediction model: x "*Granger causes*" y if x aids in the prediction of y. There exist several tests that compare between predictive models of y based on 1) past values of y only (restricted model, 'self-prediction'), and 2) the combination of past values of y *and* past values of x ('lagged' unrestricted model, 'cross-prediction', see [1, 2]). Alternative versions of these tests examine the '*instantaneous*' Granger causality, which include the present value of x in the unrestricted model.

We propose a slightly different approach for testing the presence of lagged and instantaneous Granger causality: x Granger causes y if past values of x (and in the case of instantaneous causality, also the present values) have predictive power over the residual signal of y after self-prediction. Both the self- and cross-predictions are implemented as feedforward neural networks,

and a bootstrapping approach (surrogate data method) is adopted for the subsequent testing of the predictive ability of the cross-prediction.

METHODS

The proposed approach for testing the presence of $x \rightarrow y$ lagged Granger causality is outlined in Fig. 1. It consists of two stages, namely the *self-prediction* (upper part) and the *cross-prediction* system (lower part). In the case of a test for instantaneous Granger causality, x_k should be included as input to the cross-prediction system. After explaining the two stages, the hypothesis testing and the generation of the surrogate data are described in this section.

Self-Prediction

The self-prediction system is implemented as a feedforward neural network with a single hidden layer of three neurons. It is configured such that it maps 'delay vectors' (DVs) $[y_{k-1}, \ldots, y_{k-m}]$ onto the current sample or 'target', y_k, where m is the input dimensionality. The network is trained in batch mode, using the SuperSAB adaptive learning rate scheme [6], starting at $\eta = 0.001$. Convergence is monitored by means of the variance of the residual signal after self-prediction, $\langle r^2 \rangle = \langle (y - \hat{y})^2 \rangle$, where $\hat{y}$ is the network output and y the desired output. As will be explained in the Discussion Section, the implementation of a cross-validation strategy to prevent overfitting is not necessary in the current framework. After convergence, the network is further trained for 100 iterations using a fixed learning rate of $\eta = 0.001$. Unless otherwise stated, the input dimensionality is set to $m = 4$.

Cross-Prediction

The residual signal after self-prediction, $r = y - \hat{y}$, is now used as the desired output of a neural network with the input consisting of DVs $[x_{k-1}, \ldots, x_{k-m}]$ in the lagged case, and of DVs $[x_k, \ldots, x_{k-m}]$ in the instantaneous case. The network is trained in the same fashion as the self-prediction system, and the variance σ_{o}^2 of the residual signal, $o = r - \hat{r}$, is computed. In the absence of (lagged or instantaneous) $x \rightarrow y$ causality, this residual variance, σ_{o}^2, should be equal to that of the residual signal after self-prediction, σ_{r}^2, and it should be smaller otherwise. In the following, a method is proposed for statistically testing this (in)equality.

Hypothesis Testing

To compare the variances of o and r in a reliable manner, the comparison should be performed in a statistical sense. However, since the analytical form of the distribution of σ_{o}^2 under the assumption of the null hypothesis, H_0,

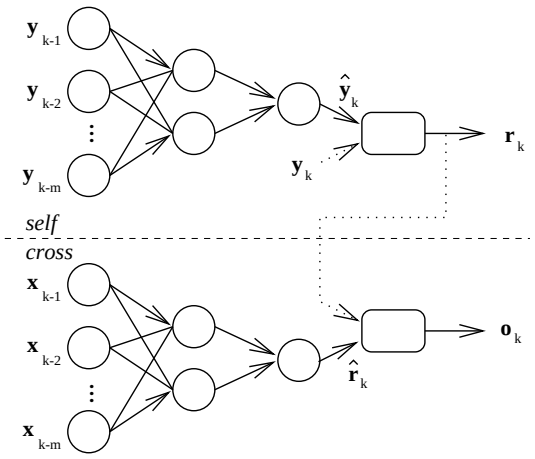

Figure 1: Proposed approach for testing the presence of lagged $x \rightarrow y$ Granger causality. Dashed lines represent desired outputs to the networks. When testing the instantaneous $x \rightarrow y$ Granger causality, x_k should be included as input to the cross-prediction system.

namely the absence of $x \rightarrow y$ Granger causality, is not known, we adopt a bootstrapping procedure for generating an empirical distribution function (EDF) of σ^2_{o}, given H_0. To this cause, the surrogate data methodology [4, 5] is commonly used in the field of signal nonlinearity testing. In the absence of $x \rightarrow y$ Granger causality, a randomisation of x, such as a random permutation of the time samples which retains the signal distribution but makes the signal random otherwise, should not have an effect on the cross-prediction performance. Therefore, if we generate N_{s} such random permutations, or so-called 'surrogates', and retrain the cross-prediction network for each of these, a set of residual variances, $\{\sigma^2_{\mathrm{s},i}\}$, is obtained[1], forming an EDF under the assumption of H_0. The null hypothesis can be subsequently tested at the significance level α using a non-parametric rank-based test, as suggested in Ref. [5], namely by sorting $\{\sigma^2_{\mathrm{o}}, \sigma^2_{\mathrm{s},1}, \ldots, \sigma^2_{\mathrm{s},N_{\mathrm{s}}}\}$ in increasing order and observing the rank of σ^2_{o}. If this rank $r_{\mathrm{o}} \leq \alpha\ (N_{\mathrm{s}} + 1)$, the null hypothesis of the absence of $x \rightarrow y$ Granger causality is rejected.

Iterative Amplitude Adjusted Correlation (iAAC) Surrogates

The rejection of a null hypothesis in a surrogate data framework needs to be interpreted with due caution, as discussed, *e.g.*, in Ref. [4], since there is no information available regarding what aspect of the null hypothesis is violated. In our particular case of instantaneous $x \rightarrow y$ causality, it is well possible that the time series x shows a high degree of correlation to y and r *without* instantaneously Granger causing y. The random permutations of x lose this correlation, possibly resulting in a rejection of the null hypothesis,

[1]The subscripts σ^2_{o} and $\{\sigma^2_{\mathrm{s},i}\}$ refer to 'original' and i–th 'surrogate'.

and, thus, falsely indicating $x \rightarrow y$ Granger causality. In order to compensate for such occurrences, a different type of surrogate is proposed, which retains the signal distribution *and* the correlation to the residual signal r, but is random otherwise.

The proposed method for generating surrogate data, the 'iterative Amplitude Adjusted Correlation' (iAAC) method, is related to the iterative Amplitude Adjusted Fourier Transform (iAAFT) method [4]. In both methods, the signal distribution of the original time series, x, is retained ("Amplitude Adjusted"), but the proposed method retains the (statistical) correlation ("C") to a reference time series, x_{ref}, rather than the amplitude spectrum ("FT" for Fourier Transform) in the iAAFT case. The objective of the iAAC method is to generate a surrogate, x_{s}, which displays the same correlation to x_{ref} as the original data, x. For convenience, a time series is represented as a vector of time values with unit norm and zero mean ($\mathbf{X}$, rather than x), due to which the correlation to x_{ref} is retained by imposing that the dot product $C = (\mathbf{X}_{\mathrm{ref}}^T \mathbf{X})$ equals $(\mathbf{X}_{\mathrm{ref}}^T \mathbf{X}_{\mathrm{s}})$. Note that C can take on any value in $[-1, 1]$, but is set specifically to $(\mathbf{X}_{\mathrm{ref}}^T \mathbf{X})$ in the present application. Furthermore, $(\mathbf{X}_{\mathrm{ref}}^T \mathbf{X}_{\mathrm{s}})$ is only equal to the statistical correlation between $\mathbf{X}_{\mathrm{ref}}$ and $\mathbf{X}_{\mathrm{s}}$ if both have unit norm. The following energy function is used:

$$E = \frac{1}{2}(C - \mathbf{X}_{\mathrm{ref}}^T \mathbf{X}_{\mathrm{s}})^2. \tag{1}$$

Differentiating to $\mathbf{X}_{\mathrm{s}}$ yields the gradient descent update rule at iteration i:

$$\mathbf{X}_{\mathrm{s}}^{i+1} = \mathbf{X}_{\mathrm{s}}^{i} + \eta \, \mathbf{X}_{\mathrm{ref}} \left(C - \mathbf{X}_{\mathrm{ref}}^T \mathbf{X}_{\mathrm{s}}^{i} \right). \tag{2}$$

In our case, where the number of parameters equals the number of time samples and there is only a single 'desired output', namely C, this update rule converges after one iteration for a learning rate of $\eta = 1$. One additional constraint for the dot product to be proportional to the statistical correlation is that the norm $\|\mathbf{X}_{\mathrm{s}}\| = 1$, which could either be included in the energy function (Eq. 1) or as a normalising step following each parameter update. However, the vector $\mathbf{X}_{\mathrm{s}}$ will have a unit norm as a result of the subsequent "Amplitude Adjusting" step, which renders the signal distributions (and, thus, also the norms) of the original, $\mathbf{X}$, and the surrogate, $\mathbf{X}_{\mathrm{s}}$, identical. We use the rank-ordering procedure described in Ref. [4], which sorts the signal values of $\mathbf{X}$ and of $\mathbf{X}_{\mathrm{s}}$ in increasing order, and sets the values of $\mathbf{X}_{\mathrm{s}}$ equal to those of $\mathbf{X}$ with corresponding ranks.

The proposed iAAC method consists of iterating Eq. 2 and the rank-ordering procedure consecutively until E converges, which typically occurs after fewer than 20 iterations. Fig. 2 shows an example in which an iAAC-surrogate is generated with a signal distribution identical to that of a segment of the chaotic laser data from the Santa Fe Competition (x, Fig. 2A; [7]), and a correlation of $C = 0.5$ to the x coordinate of a 1000 time sample realisation of the Lorenz system (x_{ref}, Fig. 2B). The energy function (Eq. 1) during the iterative procedure is plotted in Fig. 2C, showing clear convergence. In our implementation, the iterative procedure is stopped when the decrease in

energy is smaller than 10^{-10}, which in the example case would be after 7 iterations. The obtained surrogate, X_{s}, displays the desired dot product to X_{ref} and has a signal distribution approximately identical to that of the original time series, x. The initial configuration of $\mathbf{X}_{\mathrm{s}}$ is a random permutation of $\frac{\mathbf{X}}{\|\mathbf{X}\|}$, and after convergence, $\mathbf{X}_{\mathrm{s}}$ is scaled by the norm of the original $\mathbf{X}$. Figure 2D and 2E show the obtained iAAC surrogates, with a desired dot product of $C = 0.5$, retaining the signal distribution of the Laser series and of the Lorenz series, respectively (thus, exchanging x and x_{ref}).

SIMULATIONS

The proposed method for testing Granger causality is applied to a number of synthetic benchmark signal pairs (x, y). Both types of surrogates, namely those generated as random permutations and using the proposed iAAC, are evaluated, in the case of lagged and instantaneous Granger causality. The following signals are used, a number of which have been taken from Ref. [2].

BC (x, y) forms a bivariate normal distribution with a correlation of 0.8 between dimensions

BCT identical to BC, but y is shifted forward in time by one time unit

AR1 time series generated by the model:

$$\begin{aligned} x_k &= 0.6\, x_{k-1} + 0.5\, y_{k-1} + \nu_1 \\ y_k &= 0.6\, y_{k-1} + \nu_2, \end{aligned}$$

where ν_1 and ν_2 are $\mathcal{N}(0, 1)$ Gaussian white noise sources

AR2 time series generated by the model:

$$\begin{aligned} x_k &= 0.5\, x_{k-1} + 0.4\, y_{k-1} + \nu_1 \\ y_k &= 0.5 y_{k-1} + 0.4\, x_{k-1} + \nu_2, \end{aligned}$$

where ν_1 and ν_2 are $\mathcal{N}(0, 1)$ Gaussian white noise sources

AR3 x and y consist of coloured noise, generated by the stable AR(4) model: $s_k = 1.79\, s_{k-1} - 1.85\, s_{k-2} + 1.27\, s_{k-3} - 0.41\, s_{k-4} + \nu$

AR4 coloured noise sources identical to AR3, but $(x_k := x_k + 0.3\, y_{k-1})$

NAR1 coloured noise sources of AR3, but $(x_k := x_k + 0.05\, (y_{k-2} + y_{k-1})^3)$

NAR2 starting from (x, y) a bivariate normal distribution with zero mean and unit variance, $x_k := x_k + 0.2\, (x_{k-1} - y_k)^3$

Unidirectional Granger causality $y \to x$ is present in BCT, AR1, AR4, NAR1 and NAR2, and bidirectional $x \leftrightarrow y$ in AR2. For each of the eight processes, 100 realisations are generated consisting of $N = 1000$ time samples.

The results are summarised in Table 1 as the number of rejections of the null hypothesis of non-causality (thus, the number of time series for which $x \to y$ Granger causality is detected) at a significance level of $\alpha = 0.05$,

Table 1: Number of detections of $x \to y$, respectively $y \to x$ Granger causality in the benchmark sets, each containing 100×2 time series, at the significance level of $\alpha = 0.05$. Results are given for the lagged and instantaneous Granger causality, and for the two types of surrogates. The results illustrating the erroneous rejections using the random perm surrogates (see Simulations Section) are underlined.

	lagged				instantaneous			
	iAAC		random perm		iAAC		random perm	
	$x \to y$	$y \to x$	$x \to y$	$y \to x$	$x \to y$	$y \to x$	$x \to y$	$y \to x$
BC	7	7	4	4	3	8	<u>100</u>	<u>100</u>
BCT	3	100	3	100	3	100	1	100
AR1	1	100	1	100	4	100	2	100
AR2	100	100	100	100	100	100	100	100
AR3	4	6	4	5	9	9	<u>23</u>	<u>15</u>
AR4	1	100	1	100	3	95	4	100
NAR1	4	100	4	100	2	81	6	100
NAR2	7	99	3	97	0	99	<u>100</u>	<u>100</u>

for lagged and instantaneous Granger causality, and for both types of surrogates. For each pair of benchmark signals, the time series x and y are interchanged, and the test is repeated, thus, testing $y \to x$. Lagged Granger causality, using both the iAAC and the random permutations surrogates, is correctly judged absent in time series BC, and AR3, unidirectionally present in BCT, AR1, AR4, NAR1 and NAR2, and bidirectionally in AR2. The number of erroneous rejections is slightly above the expected 5% for the BC and NAR2 sets. The BC and NAR2 results for the instantaneous Granger causality (underlined in Table 1) illustrate the false detections using the random permutations surrogates: the instantaneous correlation between x and y is not retained in the random permutations surrogates, and thus leads to a rejection of the null hypothesis, which is not the case when using the iAAC surrogates. Furthermore, the random surrogates yield a high number of erroneous detections for the AR3 set. In contrast to the lagged cases, there are no false detections of $x \to y$ Granger causality in the NAR2 set using the iAAC surrogates. Overall, however, the number of erroneous rejections tend to be slightly higher for the instantaneous iAAFT Granger causality tests, and the number of correct rejections is lower.

Sensitivity Analysis

In this section, the sensitivity of the proposed method to parameter settings is examined. In particular, the number of data points, N, the number of hidden neurons, n_{h}, and the number of input taps, m, are varied separately, keeping the other parameters at the default values described earlier ($N = 1000$, $n_{\mathrm{h}} = 3$, and $m = 4$). For a given parameter setting, 100 realisations of the NAR1 time series are analysed for lagged $x \to y$ and $y \to x$ Granger causality using the iAAC surrogates, and for both cases, the following metric

is evaluated:

$$d = \frac{\sigma_{\mathrm{o}}^2 - \mathrm{mean}(\sigma_{\mathrm{s},i}^2)}{\mathrm{std}(\sigma_{\mathrm{s},i}^2)}, \tag{3}$$

where mean$(\cdot)$ and std$(\cdot)$ denote the mean and standard deviation over the N_{s} possible surrogates $x_{\mathrm{s},i}$. The measure d can be interpreted as the number of standard deviations σ_{o}^2 deviates from the mean. It is expected that this metric is close to zero for the $x \rightarrow y$ case, and negative for the $y \rightarrow x$ case. Note that $\sigma_{\mathrm{s},i}^2$ does not follow a normal distribution, and that, hence, d should only be used as a rough estimate of significance, and not as an absolute measure as such. For convenience, however, the relative measure d will be simply referred to as the 'significance'.

Figures 3A and 3B show the effects of respectively the number of available data points, N, and the number of hidden neurons, n_{h}. It is clear that increasing the number of data points, N, and the number of hidden neurons, n_{h}, improves the significance with which the null hypothesis is correctly rejected for the $y \rightarrow x$ case. Varying these parameters does not induce an increased risk of erroneous rejections, as can be seen by the results for the $x \rightarrow y$ case (dashed curves), namely that d stays close to zero for all parameter values of N and n_{h}. The effect of the number of input taps, m, is shown in Fig. 3C. It can be observed clearly that the results are only reliable for $m \geq 4$, since the measure d for the $x \rightarrow y$ case is too negative otherwise, which is an indication of a high false rejection rate. The minimal value of m which is required for reliable detection of causality is related to the optimal filter size in digital signal processing: if m is too small, the signal variation within the input tap delay line is mostly governed by noise, and m should be increased. On the other hand, an increasing number of input taps *lowers* the significance with which the null hypothesis is rejected for the $y \rightarrow x$ case (d increases as a function of m). This can be attributed to the increasing sparseness of the m-dimensional input data, due to which more data would be required to maintain the level of significance observed for a lower number of input taps (curse of dimensionality). In the current example, the signals are generated from an AR(4)-process, and thus, the optimal order is $m = 4$.

CONCLUSIONS

We have proposed a novel method for testing the presence of causality between two time series, based upon an interpretation of the $x \rightarrow y$ Granger causality concept, which is slightly different from the original one [3]. Rather than checking whether the inclusion of a time series x aids in the (cross-) prediction of y, compared to the self-prediction of y as such, it is tested whether x is able to (cross-) predict the residual signal of y after self-prediction. Both predictions (self- and cross-) are implemented as traditional feedforward neural networks, trained using backpropagation with the SuperSAB adaptive learning rate scheme [6]. Two variants of causality were considered, namely the 'lagged' causality, where only the past values of x are used for the cross-

prediction, and the 'instantaneous' causality, where the past and present values of x are used.

In theory, in the absence of $x \rightarrow y$ Granger causality, x should not have predictive power with respect to the residual of y after self-prediction, $r = y - \hat{y}$, and the variance of the residual signal after cross-prediction, σ_{o}^2 should be equal to that of r. In practice, however, σ_{o}^2 will become gradually smaller when the number of parameters in the neural network is increased due to overfitting. One possible remedy would be cross-validation, thus retaining a portion of the data for validation purposes. We have opted for a different approach in which the complete data set can be used for training, namely that of bootstrapping an empirical distribution function (EDF) of the predictive power, measured by σ_{o}^2, under the assumption of the absence of Granger causality (null hypothesis). If the actual predictive power is statistically different from that expected from this EDF, the null hypothesis is rejected and $x \rightarrow y$ Granger causality is present. The EDF is constructed by generating 'surrogates' of x, either as a random permutation of the time samples or using the proposed iterative Amplitude Adjusted Correlation (iAAC) method, and for each surrogate, x_{s}, a neural network is trained to predict r from x_{s}, and the residual variance is taken as an inverse measure of the predictive power. Furthermore, since both types of surrogates retain the signal distribution of the original time series, effects of possible non-Gaussianity of x are incorporated in the null hypothesis under which the surrogates are generated, and will not influence the test results.

A sensitivity analysis demonstrated the positive effect of the number of available data points, N. The number of hidden neurons, n_{h}, and the number of tap inputs, m, were shown to have different effects on the significance with which a null hypothesis was correctly rejected. Increasing n_{h} increased the rejection significance up to a certain point, after which it remained stable. On the other hand, increasing m initially increased the significance, but decreased it for values of m exceeding a certain value related to the optimal filter size. This could be attributed to the fact that m influenced not only the number of parameters in the prediction networks, but also the dimensionality of the input, which, in turn, could have an effect on training performance, namely when increasing m no longer provides additional information, it only renders the data more sparsely distributed (curse of dimensionality). Therefore, the appropriate choice of m is more crucial than that of n_{h}. It is furthermore expected that a different number of input taps for the self- and the cross-prediction systems could improve the results. The stability with respect to the variation of the number of neurons in the hidden layer is also an indication that the surrogate data methodology compensates for possible overfitting. Indeed, these effects occur both for the original *and* for the surrogates, due to which no cross-validation strategy needs to be employed, and the size of the neural network is less critical than in traditional applications.

A comparative study was performed on a number of data sets, using different types of surrogates, namely random permutations of the time series and surrogates generated using the proposed iterative Amplitude Adjusted

Correlation (iAAC) method. The first retains only the signal distribution of the original time series, x, whereas the latter type also retains the (instantaneous) correlation between the predictor x and the residual signal after self-prediction r. Furthermore, the effect of including the present time sample of the predictor, x, yielding a test for 'instantaneous', rather than 'lagged' Granger causality. The results indicated that a test for 'lagged' causality is to be preferred if applicable, namely if the actual causal relationship is a lagged one. However, if this is not the case (which, in typical applications is not known *a priori*), the test for instantaneous causality should improve the test results. Furthermore, it was found that the iAAC method for generating the surrogates yielded a clear improvement over the random permutations when testing for instantaneous causality.

Acknowledgements

T.G. was supported by a scholarship from the Flemish Regional Ministry of Education (GOA 2000/11) and a research grant from the Fund for Scientific Research (G.0248.03). M.M.V.H. was supported by research grants received from the Fund for Scientific Research (G.0185.96N and G.0248.03), the National Lottery (Belgium) (9.0185.96), the Flemish Regional Ministry of Education (Belgium) (GOA 95/99-06; 2000/11), the Flemish Ministry for Science and Technology (VIS/98/012), and the European Commission, 5th framework programme (QLG3-CT-2000-30161 and IST-2001-32114).

REFERENCES

[1] J. Chao, V. Corradi and N. Swanson, "An Out of Sample Test for Granger Causality," **Macroeconomic Dynamics**, vol. 5, no. 4, pp. 598–620, 2001.

[2] C. Diks and J. DeGoede, "A General Nonparametric Bootstrap Test for Granger Causality," in H. Broer, B. Krauskopf and G. Vegter (eds.), **Global Analysis of Dynamical Systems**, IoP Publishing, London, UK, pp. 391–403, 2001.

[3] C. Granger, "Testing for Causality: A Personal Viewpoint," **Journal of Economic Dynamics and Control**, vol. 2, pp. 329–352, 1980.

[4] T. Schreiber and A. Schmitz, "Surrogate Time Series," **Physica D**, vol. 142, no. 3-4, pp. 346–382, 2000.

[5] J. Theiler and D. Prichard, "Constrained-Realization Monte-Carlo Method for Hypothesis Testing," **Physica D**, vol. 94, pp. 221–235, 1996.

[6] T. Tollenaere, "SuperSAB: Fast Adaptive Back Propagation with Good Scaling Properties," **Neural Networks**, vol. 3, pp. 561–573, 1990.

[7] A. Weigend and N. Gershenfeld (eds.), **Time Series Prediction: Forecasting the Future and Understanding the Past**, Reading, MA: Addison-Wesley, 1994.

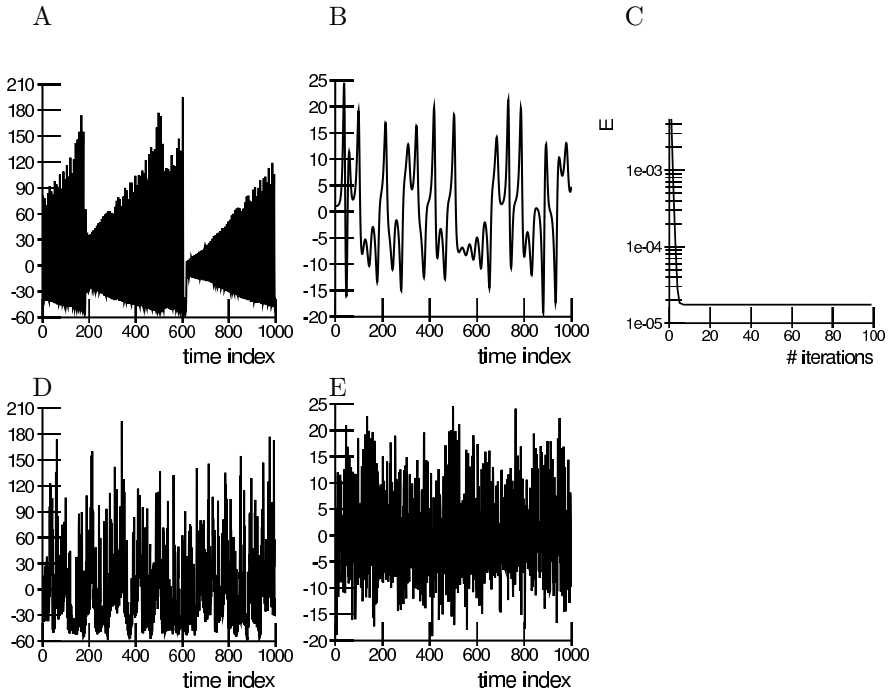

Figure 2: A,B) Laser and Lorenz time series; C) Energy during the iterative iAAC procedure. D) iAAC surrogate with signal distribution identical to that of the Laser time series and correlated to the Lorenz time series; E) iAAC surrogate with signal distribution identical to that of the Lorenz time series and correlated to the Laser time series.

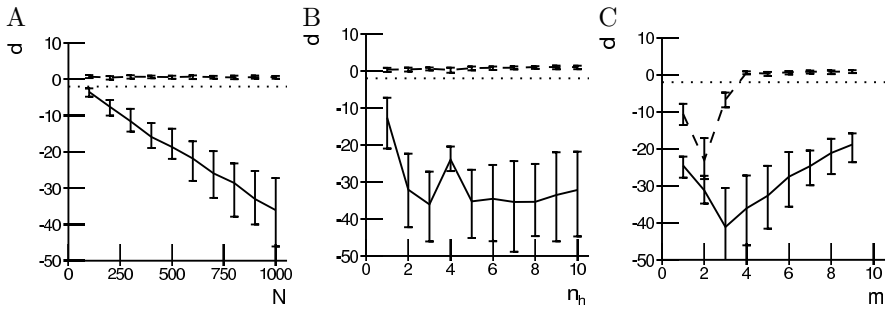

Figure 3: Sensitivity analysis of the proposed method with respect to A) the number of available data points, N, B) the number of hidden neurons, n_{h}, and C) the number of tap inputs, m. The median of the metric d is plotted for the $x \rightarrow y$ (dashed) and $y \rightarrow x$ (solid) Granger causality test applied to the NAR1 data set. For reference, the dotted line represents the $(d = -2)$–level. The error bars denote the upper and lower quartiles computed over the batch of 100 realisations of the NAR1 data set.