



# Article Minimum Sample Size for Reliable Causal Inference Using Transfer Entropy

# Antônio M. T. Ramos \* and Elbert E. N. Macau

National Institute for Space Research, São José dos Campos 12227-010, Brazil; elbert.macau@inpe.br \* Correspondence: antonio.ramos@inpe.br; Tel.: +55-12-3208-6562

Academic Editors: António M. Lopes and J. A. Tenreiro Machado Received: 20 February 2017; Accepted: 29 March 2017; Published: 31 March 2017

**Abstract:** Transfer Entropy has been applied to experimental datasets to unveil causality between variables. In particular, its application to non-stationary systems has posed a great challenge due to restrictions on the sample size. Here, we have investigated the minimum sample size that produces a reliable causal inference. The methodology has been applied to two prototypical models: the linear model autoregressive-moving average and the non-linear logistic map. The relationship between the Transfer Entropy value and the sample size has been systematically examined. Additionally, we have shown the dependence of the reliable sample size and the strength of coupling between the variables. Our methodology offers a realistic lower bound for the sample size to produce a reliable outcome.

**Keywords:** transfer entropy; multiple comparison analysis; bias analysis; coupled logistic maps; coupled autoregressive-moving-average (ARMA) model

# 1. Introduction

Transfer Entropy (TE) [1] is an information-theoretical functional able to detect a causal association between two variables [2–5]. TE identifies the potential driver and driven variable by statistically quantifying the flow of information from one to the other distinguishing the directionality due to its asymmetric property. Backshifting one variable at some particular lag, it returns a nonzero value indicating information flow between variables, or zero otherwise. However, in real data application, it may assume a nonzero value simply due to bias associated with the finiteness of the sample, and not because of an actual coupling [6,7].

Recently, sophisticated algorithms using transfer entropy have improved causal detection [8,9]. However, even these advanced methods require sufficient sample points to avoid false positives due to data-shortage bias. Previous studies addressed the entropy bias by estimating the probability distribution of the underlying processes [10–12]. However, from an inference-based approach, there is still a lack of rigorous computation regarding the appropriate minimum size of the sample for a reliable TE outcome. To overcome such an issue, a statistical hypothesis test is required to distinguish independence from a causal relation. The outcome of testing depends on the sample size. This raises concern regarding the minimum sample size required for TE to provide a reliable inference of causal relation.

To answer this, we employ TE as a statistical test and search for the lower bound of the sample size that produces a reliable true positive. For the purpose of consistency, the investigation is carried out in a stationary regime, so the sample range does not result in a different interaction behavior. To this end, paradigmatic models are used in the analysis: two coupled linear *autoregressive-moving-average* (ARMA) models and two coupled non-linear *logistic maps*. The former is used in model-driven approaches to detect Granger causality through parametric estimation [13]. The latter is a typical example of a nonlinear system that challenges causal inference due to its chaotic properties.

The results show that the minimum sample size for reliable outcome depends on the strength of the coupling. The larger the coupling, the smaller the sample size required to provide a reliable true positive. The relationship between coupling strength and sample size varies according to the model. We pinpoint the lower bound of the sample size at the limit of high coupling for each model.

The contents of this paper are organized as follows. In Sections 2 and 3, the models and the methodology are described, in Section 4 the main findings are presented, and in Section 5, the results are discussed.

## 2. Models

The models are constructed to describe two coupled variables *X* and *Y*; *X* influences *Y* after a lag  $\tau$ , with coupling strength regulated by the parameter  $\eta$ . In the following subsection, the models are defined.

## 2.1. Coupled ARMA[p,q]

The ARMA model is a linear regression model used as a toolkit in the Granger causality approach. Two coupled ARMA[p, q] models X and Y are defined; X influences Y after  $\tau = 15$  time steps, that is,

$$x_{i} = \rho \sum_{k=1}^{p} x_{i-k} + \epsilon \sum_{k=1}^{q} \xi_{i-k}^{x},$$
  

$$y_{i} = (1 - \eta) \left( \rho \sum_{k=1}^{p} y_{i-k} + \epsilon \sum_{k=1}^{q} \xi_{i-k}^{y} \right) + \eta x_{i-15},$$
(1)

where i = 0, ..., n so that n is much bigger than any sample size N;  $\xi$  is a uniformly distributed random number;  $\rho$  and  $\epsilon$  are parameters that regulate the state transition and stochasticity respectively; finally,  $\eta$  is the coupling parameter. For  $\eta = 0$ , the two variables are uncoupled, while if  $\eta = 1$ , the Y variable is completely defined by X. The parameter  $\rho = 0.8$  is fixed throughout the whole analysis. In the methodology section, we investigate the role of  $\epsilon$  and  $\eta$  in detecting  $N^*$ .

#### 2.2. Couple Logistic Maps

Inspired by the studies of both Gyllenber [14] and Hastings [15] to describe migration dynamics within two populations, we define the coupled logistic maps with delay as follows

$$x_{i} = rx_{i-1}(1 - x_{i-1}),$$
  

$$y_{i} = (1 - \eta)ry_{i-1}(1 - y_{i-1}) + \eta x_{i-15}.$$
(2)

The dynamics of *X* is the standard logistic map, and the dynamics of *Y* has two terms. The first is a local logistic growing rate, while the other represents a migration effect, where individuals from the *X* migrate to *Y* after  $\tau = 15$  time steps. We assume that both populations have the same growth rate r = 3.9. The parameter  $\eta$  is the coupling between *X* and *Y* and represents the migration influx controlling the growth of one population to the detriment of the other.

## 3. Methodology

Consider two given sets of discrete variable  $X = \{x_i\}$  and  $Y = \{y_i\}$ . Any causal relationship between them should be restrained in the temporal order of the *X* and *Y* elements. Let us suppose that *X* influences *Y* after  $\tau$  units of time. One can detect this relationship using the TE functional. TE evaluates the reduction in the uncertainty of the *past* of a possible driver  $X^{\tau}$  due to the knowledge of the *present* of a possible driven *Y* when the *past* of  $Y^{\tau}$  is given. Figure 1 illustrates the idea behind the sampling for TE calculation. The TE from *X* to *Y* is evaluated as follows

$$I(X^{\tau}; Y|Y^{\tau}) = H(X^{\tau}, Y^{\tau}) + H(Y, Y^{\tau}) - H(Y^{\tau}) - H(X^{\tau}, Y, Y^{\tau}).$$
(3)

Whenever X and Y are infinite independent random variables, the expression (3) is zero. However, in the case of two finite datasets, TE will be greater than zero, even if X and Y are supposed to be independent. This bias is due to a finite size effect, and it has to be addressed via a statistical inference procedure.



**Figure 1.** Diagrammatic representation of the sampled values used to calculate the Transfer Entropy (TE) between *X* and *Y*. The solid black box represents the *period of interest* and the other two dashed boxes represent the lagged intervals for analysis.

Let us assume a null hypothesis stating that *X* and *Y* are independent for a particular lag  $\tau$ . At this lag, an ensemble of surrogate data is generated from the original data. A confidence interval is defined as a percentile of the surrogate's TE distribution. If the TE calculated from the original dataset (original TE) is significantly higher than the confidence interval, then the null hypothesis is rejected and causality is detected. However, for a small dataset, the bias is considerable when compared to the original TE [16]. The small-sample issue leads to a type II error, e.g., *X* and *Y* are not independent, but the hypothesis is falsely accepted.

The minimum number of sample points  $N^*$  that prevent such bias has been estimated. Assuming the processes as being independent and with each having equiprobable *B* partitions, one can find a preliminary lower bound for the proper sample size such that  $N \gg N^* = B^3$ .

Our computational approach improves this naive estimation by investigating the dependence of the TE value with the sample size N. It searches for the smallest  $N = N^*$  so the original TE is significantly higher than the confidence interval. The trustworthiness of TE is tested for the models presented in Section 2, contemplating its particularities.

The methodology is described as follows. The reliability of the true positives is tested against several sample sizes. For each sample size tested, 100 runs with different initial conditions are performed. For each run, an ensemble of 2000 surrogates is obtained from the original data, and the respective TE is calculated at a fixed  $\tau = 15$ . A significance level of  $\alpha = 0.0005$  is chosen, so the 99.95 percentile of the surrogate ensemble is defined as the upper confidence bound, which for the sake of simplicity we have called the threshold  $I_{thr}$ . Whenever the original TE is higher than  $I_{thr}$ , the null hypothesis is rejected. Otherwise it is accepted. This produces a comparison-wise error rate of  $\pi = 0.049$  when considering the multiple comparisons between the 100 runs of different initial conditions. In other words, a significance level of 4.9% is obtained when all the 100 TEs calculated from the original dataset are significantly higher than all the respective  $I_{thr}$ . In the analysis, we consider the most coarse-grained measure, i.e., a bipartition of the variable's dynamics domain using two ordinal patterns. Appendix A presents the details of the probability estimation while Appendix B explains the upper bound of Transfer Entropy.

## 4. Results

In this section, the minimum sample size  $N^*$  required to avoid type II error is pinpointed with a significance level of 4.9%. Figures 2–5 show the boxplot regarding the 100 TE values versus the sample size N. The ends of the whiskers represent the minimum and maximum TE values. The TE calculated from the original time series is seen in black, and the threshold  $I_{thr}$  (family-wise error rate of 0.0005) is seen in red. The minimum reliable sample size  $N^*$  is defined as the smallest N therefore the minimum TE value calculated from the original time series is higher than the maximum TE calculated from the threshold. All statements have been made according to a confidence level of 4.9% (comparison-wise error rate).

### 4.1. Coupled ARMA[p,q]

Figure 2 shows the results for the two coupled ARMA[1,1] models. The reliable minimum sample size  $N^*$  depends on the coupling strength of the model. Figure 2a shows results for coupling strength  $\eta = 0.9$ ; the TE method yields a reliable true positive if one uses  $N > N^* = 40(5)$ . Figure 2b shows results for coupling strength  $\eta = 0.5$ ; the TE method yields a reliable true positive if one uses  $N > N^* = 40(5)$ . Figure 2b shows results for coupling strength  $\eta = 0.5$ ; the TE method yields a reliable true positive if one uses  $N > N^* = 165(5)$ . Finally, Figure 2c shows results for coupling strength  $\eta = 0.1$ ; no minimum sample size  $N^*$  is identified (up to  $N^* = 400$ ). This means that there is a higher chance of obtaining false negatives. Figure 2d–f show the TE of  $Y \rightarrow X$  versus the sample size. In this case, the methodology presents a negligible chance of type I error, which is the incorrect rejection of a true null hypothesis.



**Figure 2.** The black boxplots refer to the time series of the coupled autoregressive-moving-average (ARMA)[1,1] model, and the red boxplots refer to its surrogates. Left: Transfer Entropy of  $X \rightarrow Y$  versus the sample size for the following coupling parameter (**a**)  $\eta = 0.9$ ; (**b**)  $\eta = 0.5$ ; (**c**)  $\eta = 0.1$ . Right: Transfer Entropy of  $Y \rightarrow X$  versus the sample size for the following coupling parameter (**d**)  $\eta = 0.9$ ; (**e**)  $\eta = 0.5$ ; (**f**)  $\eta = 0.1$ .

In the particular case of the ARMA[p, q] model, Figure 3 shows that the reliable minimum sample size  $N^*$  does not depend significantly on the memory of the autoregressive term. The results refer to the autoregressive window size p = 1, 4 and 8.

We also investigated the relationship between the reliable minimum sample size  $N^*$  and the parameter  $\epsilon$  that regulates the stochasticity. Figure 4a,b show the results for  $\epsilon = 0.5$  and  $\epsilon = 1.0$ . One can notice that  $N^*$  depends on the ARMA parameter  $\epsilon$ . For  $\epsilon = 0.5$ , the minimum reliable sample size is  $N^* = 110(5)$ , and for  $\epsilon = 1.0$ , the minimum reliable sample size is  $N^* = 165(5)$ . So the higher the parameter  $\epsilon$  is, the larger the  $N^*$  value must be to obtain a reliable true positive.



**Figure 3.** Transfer Entropy of  $X \rightarrow Y$  versus the sample size for different autoregressive window sizes (a) p = 1, (b) p = 4 and (c) p = 8.



**Figure 4.** Transfer Entropy of  $X \rightarrow Y$  versus the sample size for different stochastic parameters (a)  $\epsilon = 0.5$  and (b)  $\epsilon = 1.0$ .

#### 4.2. Couple Logistic Maps

Figure 5 shows that the  $N^*$  value here changes more abruptly according to the coupling strength in the nonlinear coupled logistic map if compared with the ARMA model. Figure 5a shows the results for coupling strength of  $\eta = 0.9$ . In this case, one requires  $N > N^* = 35(5)$  sample points for the proper detection of the influence of X on Y. This lower bound is very close to the one identified in the ARMA model with the same coupling strength. Figure 5b shows the results for the coupling strength  $\eta = 0.5$ ; in this case, the number of sample points for a reliable true positive is  $N > N^* = 45(5)$ . For the coupling strength  $\eta = 0.1$ , no minimum sample size  $N^*$  is identified (up to N = 400). Figure 5c shows a high probability of obtaining false negatives, so no lower bound is identified. Figure 5d–f show that the TE method presents a negligible chance of type I error.

Figure 6 shows the relationship between the reliable minimum sample size  $N^*$  and the coupling parameter  $\eta$ . The larger the coupling parameter  $\eta$  is, the smaller the  $N^*$  that is required, but the way  $N^*$  decays with  $\eta$  is different for the two models. The relationship between  $N^*$  and  $\eta$  for the ARMA[1,1] model presents an exponential decay, whereas for the coupled logistic map,  $N^*$  presents an abrupt decay when  $\eta < 0.4$  and a saturation behavior around  $N^* \approx 35$  for  $\eta > 0.4$ .



**Figure 5.** The black boxplots refer to the time series of the coupled logistic maps, and the red boxplots refer to its surrogates. Left: Transfer Entropy of  $X \rightarrow Y$  versus the sample size for the following coupling parameter (**a**)  $\eta = 0.9$ ; (**b**)  $\eta = 0.5$ ; (**c**)  $\eta = 0.1$ . Right: Transfer Entropy of  $Y \rightarrow X$  versus the sample size for the following coupling parameter (**d**)  $\eta = 0.9$ ; (**e**)  $\eta = 0.5$ ; (**f**)  $\eta = 0.1$ .



**Figure 6.** Reliable minimum sample size  $N^*$  versus the coupling parameter  $\eta$ , using a binary partition for probability estimation.

The result regarding the coupled logistic map depicted in Figure 6 can be explained as follows: The saturation happens because around  $\eta = 0.4$ , the driven synchronizes its dynamics with the driver. Figure 7a shows the behavior of the system for  $\eta = 0.3$ , i.e., below the synchronization transition. Furthermore, Figure 7b shows the behavior when the system is very close to the synchronized state between the systems with  $\eta = 0.6$ . We stress that this result is important and unexpected because it shows that, even in an almost synchronized state, the underlying methodology can detect interactions correctly.



**Figure 7.** Behavior of the coupled logistic maps according to the coupling strength  $\eta$ . (**a**) Shows the behavior when  $\eta = 0.3$ ; and (**b**) shows the synchronized behavior when  $\eta = 0.6$ .

Moreover, at the high coupling limit, the minimum sample size of the ARMA[1,1] model is  $N^* = 30(5)$ , somewhat small if compared with the logistic maps, namely  $N^* = 35(5)$ . Despite the fact that this coupling analysis considers only ideal and simplified models, this number can be thought of as the lower bound for the entropy transfer using a significance level of 5% and two ordinal patterns.

#### 4.3. Three Ordinal Pattern

A larger amount of sample points is needed to obtain a reliable outcome using three ordinal patterns. Figure 8 shows the relationship between TE and *N* using three ordinal patterns and  $\eta = 1.0$ . One can see that the ARMA[1,1] model requires at least  $N^* = 65(5)$  sample points, while the logistic model needs at least  $N^* = 80(5)$  sample points.



**Figure 8.** Transfer Entropy of  $X \rightarrow Y$  versus sample size *N* using three ordinal patterns and coupling parameters  $\eta = 1.0$ . (a) Coupled ARMA[1,1] model; (b) Couple logistic maps.

#### 5. Conclusions

In this paper, we have presented a quantitative analysis to find the minimum sample size  $N^*$  that produces a reliable true positive outcome using Transfer Entropy. We have tested two paradigmatic models: the linear ARMA model, commonly used in Granger causality approaches; and the well known nonlinear logistic map. The models are constructed to describe two coupled variables *X* and *Y*; *X* influences *Y* after a lag  $\tau$  with coupling strength regulated by the parameter  $\eta$ .

The analysis shows that the size of  $N^*$  depends on the coupling strength  $\eta$ ,  $\eta$  being the larger and  $N^*$  the smaller. This result is expected since the information flow increases with the coupling [5,17], therefore it is reasonable to conclude that the larger the coupling is, the smaller the sample size required to infer causality.

However, the relationship between  $N^*$  and  $\eta$  depends on the model. For the coupled ARMA model,  $N^*$  decreases exponentially as  $\eta$  increases whereas, for the coupled logistic map,  $N^*$  decays abruptly for  $\eta < 0.4$  followed by a saturation as  $\eta$  increases.

Furthermore, at the high coupling limit, the value of  $N^*$  approaches 35(5) sample points. This result establishes the lower bound of the reliable minimum sample size for inference-based causality testing using Transfer Entropy. For this particular case, we use a probability estimation through bipartition of the variable domain. We have shown that the higher the partitioning is, the higher the  $N^*$  value is. This methodological procedure can be reproduced for different models, control experimental data and also probability estimation.

Acknowledgments: This research was supported by the São Paulo Research Foundation FAPESP (Grant No. 2015/50122-0, 2014/14229-2 and 2015/07373-2).

Author Contributions: Antônio M. T. Ramos and Elbert E. N. Macau conceived and designed the numerical experiments; Antônio M. T. Ramos performed the experiments; Antônio M. T. Ramos and Elbert E. N. Macau analyzed the data; Antônio M. T. Ramos wrote the paper. Both authors have read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

#### Appendix A. Probability Estimation

The probability is estimated using the *ordinal partitions* symbolization [18,19]. The procedure transforms an experimental sample into a sequence of *n*-tuples called *order patterns*  $\pi$  of size *n*. Each  $\pi$  contains *n* indexes of the original sample. The indexes are arranged in such a way that their respective values follow an ascending order.

The order patterns are generated by sampling *n* values with delay *d*, so that the *i*-sample is defined as  $(x_i, x_{i+d}, x_{i+2d}, ..., x_{i+(n-1)d})$ . In the context of phase-space reconstruction of dynamical systems, *n* is also known as the embedding dimension and *d* as the embedding delay [20,21]. In this study, we test n = 2, 3, 4 and d = 1.

The two kinds of models considered here are discrete recurrence relation, so there is no slow response time in the dynamics, therefore no scatter sampling across the time series is necessary. Besides, an embedding delay equal to one is enough for the ordinal patterns to capture the intrinsic features of the model's dynamics.

## Appendix B. Upper Bound of Transfer Entropy

The upper bound of the TE can be obtained through the following consideration: Equations (1) and (2) are devised to make *X* and *Y* uncouple when  $\eta = 0$ , and conversely, if  $\eta = 1$  then *Y* is entirely described by *X*. In the latter case,  $H(X^{\tau}, Y, Y^{\tau}) = H(Y, Y^{\tau})$  which leads us to  $I(X^{\tau}; Y|Y^{\tau}) = H(Y, Y^{\tau}) - H(Y^{\tau}) = H(Y|Y^{\tau})$ . Assuming *Y* and *Y*<sup> $\tau$ </sup> to be independent, then  $H(Y|Y^{\tau}) = H(Y)$ . This finally leads us to the upper bound of the TE as  $I(X^{\tau}; Y|Y^{\tau}) = \log_2 B$  for equiprobable *B* partitions in the variable's domain.

## References

- 1. Schreiber, T. Measuring information transfer. Phys. Rev. Lett. 2000, 85, 461–464.
- Balasis, G.; Donner, R.V.; Potirakis, S.M.; Runge, J.; Papadimitriou, C.; Daglis, I.A.; Eftaxias, K.; Kurths, J. Statistical mechanics and information-theoretic perspectives on complexity in the earth system. *Entropy* 2013, 15, 4844–4888.
- 3. Li, Z.; Li, X. Estimating temporal causal interaction between spike trains with permutation and transfer entropy. *PLoS ONE* **2013**, *8*, e70894.
- 4. Runge, J.; Riedl, M.; Müller, A.; Stepan, H.; Kurths, J.; Wessel, N. Quantifying the causal strength of multivariate cardiovascular couplings with momentary information transfer. *Physiol. Meas.* **2015**, *36*, 813–825.
- 5. Cafaro, C.; Ali, S.A.; Giffin, A. Thermodynamic aspects of information transfer in complex dynamical systems. *Phys. Rev. E* **2016**, *93*, 022114.
- 6. Miller, G.A. Note on the Bias of Information Estimates. In *Information Theory in Psychology: Problems and Methods II–B*; Free Press: Glencoe, IL, USA, 1955; pp. 95–100.
- 7. Green, B.F.; Rogers, M.S. The Moments of Sample Information When the Alternatives Are Equally Likely. In *Information Theory in Psychology: Problems and Methods II–B*; Free Press: Glencoe, IL, USA, 1955; pp. 101–108.
- 8. Sun, J.; Cafaro, C.; Bollt, E.M. Identifying the coupling structure in complex systems through the optimal causation entropy principle. *Entropy* **2014**, *16*, 3416–3433.
- 9. Runge, J.; Heitzig, J.; Petoukhov, V.; Kurths, J. Escaping the curse of dimensionality in estimating multivariate transfer entropy. *Phys. Rev. Lett.* **2012**, *108*, 258701.
- 10. Wolpert, D.H.; Wolf, D.R. Estimating functions of probability distributions from a finite set of samples. *Phys. Rev. E* **1995**, *52*, 6841–6854.
- 11. Holste, D.; Große, I.; Herzel, H. Bayes' estimators of generalized entropies. J. Phys. A Math. Gen. 1998, 31, 2551–2566.
- 12. Bonachela, J.A.; Hinrichsen, H.; Munoz, M.A. Entropy estimates of small data sets. *J. Phys. A Math. Theor.* **2008**, *41*, 202001.
- 13. Barnett, L.; Seth, A.K. Granger causality for state-space models. Phys. Rev. E 2015, 91, 040101.
- 14. Gyllenberg, M.; Söderbacka, G.; Ericsson, S. Does migration stabilize local population dynamics? Analysis of a discrete metapopulation model. *Math. Biosci.* **1993**, *118*, 25–49.
- 15. Hastings, A. Complex interactions between dispersal and dynamics: Lessons from coupled logistic equations. *Ecology* **1993**, *74*, 1362–1372.
- 16. Roulston, M.S. Estimating the errors on measured entropy and mutual information. *Physica D* **1999**, 125, 285–294.
- 17. Sun, J.; Bollt, E.M. Causation entropy identifies indirect influences, dominance of neighbors and anticipatory couplings. *Physica D* 2014, 267, 49–57.
- Bandt, C.; Pompe, B. Permutation entropy: A natural complexity measure for time series. *Phys. Rev. Lett.* 2002, *88*, 174102.

- 19. Bandt, C.; Shiha, F. Order patterns in time series. J. Time Ser. Anal. 2007, 28, 646–665.
- 20. Takens, F. Detecting Strange Attractors in Turbulence; Springer: Berlin/Heidelberg, Germany, 1981; pp. 366–381.
- 21. Packard, N.H.; Crutchfield, J.P.; Farmer, J.D.; Shaw, R.S. Geometry from a time series. *Phys. Rev. Lett.* **1980**, 45, 712–716.



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).