

# Comment on Wibral et al. (2013): Measuring Information-Transfer Delays (Correction of the previous comment)

Jakob Runge

Potsdam Institute for Climate Impact Research (PIK), 14473 Potsdam, Germany

Department of Physics, Humboldt University, 12489 Berlin, Germany

January 16, 2014

In their article Wibral et al. (2013), the authors propose a measure of interaction delays rooted in an information-theoretic framework. Their measure, named  $TE_{SPO}$  where  $SPO$  stands for *self-prediction optimality*, is a time-delayed extension to transfer entropy that becomes maximal only at the actual interaction delay as proven in their paper.

Wibral et al. only considered the bivariate coupling case. Prior to their publication, in Runge et al. (2012b,a) the theory of detecting and quantifying causal interactions and their strength for the general multivariate case is discussed. In these publications we use a measure, called information transfer to  $Y$  (ITY), that can be considered as a multivariate generalization of  $TE_{SPO}$ . The general idea is to measure the coupling delay  $X \rightarrow Y$  by excluding past information from  $Y$ . The difference is that Wibral et al. try to infer the past information by reconstructing (vector-valued) states, while in our approach (published earlier than theirs) we use an iterative algorithm coming from the statistics community (Pearl, 2000).

Wibral et al. contrast their measure with a similar information-theoretic measure, the bivariate momentary information transfer (BivMIT), that was introduced before the two main publications (Runge et al., 2012b,a) in Pompe and Runge (2011). Wibral et al. write that “*A major conceptual difference between the Pompe and Runge study and ours is that no formal proof of the maximality of their functional [Biv-]MIT at the correct interaction delay is given, and as we argue below cannot be given.*” Wibral et al. do not provide a proof that the maximality of BivMIT cannot be given, but they construct a model example for which they find the BivMIT unable to serve the purpose of inferring the interaction delay. In a previous version of this comment we had attempted to show that the maximality proof equally holds for the bivariate BivMIT introduced in Pompe and Runge (2011). But this proof contained an error as Wibral et al. brought to our attention. Here we show, that indeed, the BivMIT could be

maximal for a different delay. Interestingly, BivMIT can be regarded to some extent as a derivative of  $TE_{SPO}$  and the maximum, therefore, depends on the decay of  $TE_{SPO}$ .

While the maximality of  $TE_{SPO}$  can be proven for the bivariate case with *one* coupling delay (possibly also in the other direction  $Y \rightarrow X$ ),  $TE_{SPO}$  is not necessarily maximal anymore if there are *two* coupling delays from  $X$  to  $Y$ . These more complicated cases are only captured by the iterative approach introduced in Runge et al. (2012b,a). In these works a two-step approach is proposed. In the first step the causal graph (conditional independence graph) is iteratively reconstructed (Runge et al., 2012b) using the well-established framework of graphical models (Pearl, 2000), where the property to capture the correct causal interaction delays is a consequence of separation properties in the graph. In Runge et al. (2012b) also the underlying assumptions for such an inference are given. In the second step the multivariate MIT (not the bivariate BivMIT introduced in Pompe and Runge (2011)) is used as a measure of the coupling strength solely of the causal links, i.e., the inferred interaction delays, in this graph (Runge et al., 2012a).

Wibral et al. also have overseen that their estimator of conditional mutual information given by their Eq. (19) has already been developed in Frenzel and Pompe (2007). The latter work also discussed the inference of interaction delays.

## Proof that BivMIT is not necessarily maximal at the correct interaction delay and $TE_{SPO}$ not maximal anymore for two coupling delays

BivMIT is defined as

$$I_{X \rightarrow Y}^{\text{BivMIT}}(\delta) \equiv I(X_{t-\delta}; Y_t | Y_t^-, X_{t-\delta}^-), \quad (1)$$

where  $Y_t^- = (Y_{t-1}, \dots)$ ,  $X_{t-\delta}^- = (X_{t-\delta-1}, \dots)$  are the embedding vectors of the past state. I.e., additionally to  $TE_{SPO}$ , the past state of  $X$  is included in the condition.

Wibral et al. prove that their  $TE_{SPO}$  is maximal only for the interaction delay. To this end they use the Markov properties of the process and information theoretic properties of the conditional mutual information like the data processing inequality and the chain rule (Cover and Thomas, 2006) which we also utilize.

First, consider the unidirectional case with only one delay  $X \rightarrow Y$  at lag  $\tau$ . Then  $I_{X \rightarrow Y}^{\text{BivMIT}}(\delta) = 0$  for  $\delta > \tau$  because the condition on  $X_{t-\delta}^- = (X_{t-\delta-1}, \dots)$  excludes all paths coming from  $X_{t-\tau}$ . On the other hand  $TE_{SPO}$  is not zero for smaller lags. In this way BivMIT could be used to measure the coupling delay by checking at which lag it becomes non-zero, but it only holds for unidirectional couplings.

Turning to the more general case, to proof that BivMIT is not necessarily maximal at the correct coupling delay, we simplify the notation by leaving out

the time index  $t$  in the subscripts, i.e.,  $X^- = (X_t, X_{t-1}, \dots)$ ,  $X_\delta = X_{t-\delta}$  and so on. We assume that there is only one coupling delay at  $\tau$  (possibly also in the other direction) and the past state is only the preceding time point, i.e.,  $X_{t-\delta}^- = (X_{t-\delta-1})$ . Now consider the multivariate conditional mutual information with the chain rule applied twice extracting first  $X_\delta$  and then  $X_{\delta+1}$ :

$$\begin{aligned} I(X^-; Y|Y^-) &= \\ &= I(X_\delta; Y|Y^-) + I(X^- \setminus X_\delta; Y|Y^-, X_\delta) \quad (2) \\ &= I(X_\delta; Y|Y^-) + I(X_{\delta+1}; Y|Y^-, X_\delta) + I(X^- \setminus (X_\delta, X_{\delta+1}); Y|Y^-, X_\delta, X_{\delta+1}). \quad (3) \end{aligned}$$

On the other hand, we can also extract the vector  $(X_\delta, X_{\delta+1})$  at once and then extract  $X_{\delta+1}$  and then  $X_\delta$ :

$$\begin{aligned} I(X^-; Y|Y^-) &= \\ &= I((X_\delta, X_{\delta+1}); Y|Y^-) + I(X^- \setminus (X_\delta, X_{\delta+1}); Y|Y^-, X_\delta, X_{\delta+1}) \quad (4) \\ &= I(X_{\delta+1}; Y|Y^-) + I(X_\delta; Y|Y^-, X_{\delta+1}) + I(X^- \setminus (X_\delta, X_{\delta+1}); Y|Y^-, X_\delta, X_{\delta+1}). \quad (5) \end{aligned}$$

Now the last terms in Eqns. (3) and (5) cancel and the BivMIT at lag  $\delta$  and the true coupling delay  $\tau$  are

$$\begin{aligned} I(X_\delta; Y|Y^-, X_{\delta+1}) &= I(X_\delta; Y|Y^-) - I(X_{\delta+1}; Y|Y^-) + I(X_{\delta+1}; Y|Y^-, X_\delta) \quad (6) \end{aligned}$$

$$\begin{aligned} I(X_\tau; Y|Y^-, X_{\tau+1}) &= I(X_\tau; Y|Y^-) - I(X_{\tau+1}; Y|Y^-) + \underbrace{I(X_{\tau+1}; Y|Y^-, X_\tau)}_{= 0 \text{ due to the condition on } X_\tau} \quad (7) \end{aligned}$$

This implies, that BivMIT at  $\tau$  is the difference of  $TE_{SPO}$  at lags  $\tau + 1$  and  $\tau$  and analogously for  $\delta$  where another term is added. This can be regarded as a derivative and depending on the decay rate of  $TE_{SPO}$  with  $\delta$  the maximum shifts. Here, we don't examine further for what kind of processes the maximum shifts, Wibral et al. have given one example. It has a *deterministic* coupling from  $X$  to  $Y$  and randomness only in  $X$ . If even this remaining randomness is set to very small values, only then does the "inability" of MIT appear. Generally, information theory is primarily suited for stochastic processes. As discussed also in the original paper (Pompe and Runge, 2011), deterministic processes generate the source entropy that MIT is based on only in the chaotic case due to quantization effects.

Nevertheless, it cannot be proven for the class of processes sufficing condition (S) in Eichler (2012) that BivMIT reliably infers the correct coupling delay and, as analyzed in detail in the subsequent articles Runge et al. (2012b) and Runge et al. (2012a), we propose the iterative approach for the general, also multivariate, case.

To show that  $TE_{SPO}$  fails if there are two couplings  $\tau_1, \tau_2$  in one direction, consider

$$\begin{aligned} I((X_\delta, X_{\tau_1}, X_{\tau_2}); Y|Y^-) &= \\ &= I(X_\delta; Y|Y^-) + \underbrace{I((X_{\tau_1}, X_{\tau_2}); Y|Y^-, X_\delta)}_{\geq 0} \end{aligned} \quad (8)$$

$$= I((X_{\tau_1}, X_{\tau_2}); Y|Y^-) + \underbrace{I(X_\delta; Y|Y^-, X_{\tau_1}, X_{\tau_2})}_{= 0 \text{ due to the condition on } X_{\tau_1} \text{ and } X_{\tau_2}} \quad (9)$$

$$= I(X_{\tau_1}; Y|Y^-) + I(X_{\tau_2}; Y|Y^-, X_{\tau_1}) \quad (10)$$

$$(= I(X_{\tau_2}; Y|Y^-) + I(X_{\tau_1}; Y|Y^-, X_{\tau_2})). \quad (11)$$

Therefore,  $TE_{SPO}$  given by  $I(X_\delta; Y|Y^-)$  can be proven to be smaller than the multivariate  $I((X_{\tau_1}, X_{\tau_2}); Y|Y^-)$ , but not necessarily smaller than each individual  $I(X_{\tau_i}; Y|Y^-)$ . Again, this more general case requires an iterative approach.

## References

- TM Cover and JA Thomas. *Elements of information theory*. John Wiley & Sons, Hoboken, 2006.
- Michael Eichler. Graphical modelling of multivariate time series. *Probability Theory and Related Fields*, 1:233, February 2012. ISSN 0178-8051.
- S. Frenzel and B. Pompe. Partial Mutual Information for Coupling Analysis of Multivariate Time Series. *Phys. Rev. Lett.*, 99(20):204101, 2007. ISSN 0031-9007.
- Judea Pearl. *Causality: models, reasoning, and inference*. Cambridge University Press, Cambridge, 2000. ISBN 0521773628 9780521773621.
- Bernd Pompe and Jakob Runge. Momentary information transfer as a coupling measure of time series. *Phys. Rev. E*, 83(5):1–12, May 2011. ISSN 1539-3755.
- Jakob Runge, Jobst Heitzig, Norbert Marwan, and Jürgen Kurths. Quantifying Causal Coupling Strength: A Lag-specific Measure For Multivariate Time Series Related To Transfer Entropy. *Phys. Rev. E*, 86(6):1–15, October 2012a.
- Jakob Runge, Jobst Heitzig, Vladimir Petoukhov, and Jürgen Kurths. Escaping the Curse of Dimensionality in Estimating Multivariate Transfer Entropy. *Physical Review Letters*, 108(25):1–4, June 2012b. ISSN 0031-9007.
- P. Spirtes, C.N. Glymour, and R. Scheines. *Causation, prediction, and search*, volume 81. The MIT Press, Boston, 2000.
- Michael Wibral, Nicolae Pampu, Viola Priesemann, Felix Siebenhühner, Hannes Seiwert, Michael Lindner, Joseph T. Lizier, and Raul Vicente. Measuring information-transfer delays. *PloS one*, 8(2), 2013.