34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Revisiting Wiener's principle of causality — interaction-delay reconstruction using transfer entropy and multivariate analysis on delay-weighted graphs

Michael Wibral[1], Patricia Wollstadt[2], Ulrich Meyer[3], Nicolae Pampu[4], Viola Priesemann[5] and Raul Vicente[6]

***Abstract*— To understand the function of networks we have to identify the structure of their interactions, but also interaction *timing*, as compromised timing of interactions may disrupt network function. We demonstrate how both questions can be addressed using a modified estimator of transfer entropy. Transfer entropy is an implementation of Wiener's principle of observational causality based on information theory, and detects arbitrary linear and non-linear interactions. Using a modified estimator that uses delayed states of the driving system and independently optimized delayed states of the receiving system, we show that transfer entropy values peak if the delay of the state of the driving system equals the true interaction delay. In addition, we show how reconstructed delays from a bivariate transfer entropy analysis of a network can be used to label spurious interactions arising from cascade effects and apply this approach to local field potential (LFP) and magnetoencephalography (MEG) data.**

I. INTRODUCTION

Many complex phenomena, such as traffic systems, gene regulatory networks, and neural circuits can be best understood in terms of network analysis, describing the ways the nodes of the network interact. Accordingly, neuroscience has focused on discovering the *interaction structure* in such networks in terms of quantifying deviations from independence between the activities measured at each node, using a variety of linear and nonlinear techniques, ranging from simple analyses of cross-correlations to fitting of autoregressive linear [2] or dynamic models [8] or the use of model-free measures from information theory [5], [16]. However, for a better understanding of network function – and dysfunction – parameters other than the interaction structure are also important. For example, to understand deficits present in multiple sclerosis – a common neurological disorder where axons loose their insulating myelin sheath, which slows down conduction velocities and eventually leads to axonal disruption –, we have to look at the *interaction delays* between neurons. This is because interaction structure remains topologically intact as long as the axons still exist and carry action potentials. However, the drastically slowed conduction velocity after loss of myelinisation [4] may already severely disrupt network function by interfering with the precisely tuned timing of neural activity [9]. Here, we first demonstrate how an information theoretic measure of directed interactions referred to as transfer entropy [5] can be modified to reconstruct interaction delays. We then present a graph-based method that uses these reconstructed delays to identify putative spurious interactions between two network nodes that are in fact mediated via one or more intermediate nodes in the network ('cascade effects'), but do appear as an interaction in bivariate analyses. This is important, as the problem of spurious interactions frequently arises in analyses using information theoretic tools, because the limited amount of available data limits analyses to the bivariate case. The proposed graph-based method is evaluated on simulated data, LFP and human MEG data.

II. METHODS

A. Transfer entropy

Most measures of directed interactions, including transfer entropy, are based on Nobert Wiener's principle of *observational* causality which states that a time series X is called causal to a second time series Y, if knowledge about the past of X and Y together allows one to predict the future of Y better than knowledge about the past of Y alone [1]. In many networks the interactions of interest are nonlinear, e.g. the all or none mechanisms of action potential generation and the shunting mechanisms of inhibitory coupling in neural networks. The infinite number of possible nonlinear interactions requires that the question of directed interactions is addressed in a way that is free of a model of the interaction.This can be achieved by reformulating Wiener's principle in terms of a conditional mutual information for Markov processes X, Y [5], [16]:

$$TE(X \rightarrow Y) = I(Y^+; \mathbf{X}^- | \mathbf{Y}^-) \qquad (1)$$

Where Y^+ is a future value of Y, whereas $\mathbf{X}^-$ and $\mathbf{Y}^-$ denote suitably chosen past *states* of the processes X and Y,

*This work was supported by LOEWE-Grant Neuronale Koordination Forschungsschwerpunkt Frankfurt (NeFF)

[1]M. Wibral is with the MEG Unit, Brain Imaging Center, Goethe University, 60528 Frankfurt, Germany, wibral at bic.uni-frankfurt.de

[2] P. Wollstadt is with the MEG Unit, Brain Imaging Center, Goethe University, 60528 Frankfurt, Germany, Patricia.Wollstadt at gmx.de

[2] U. Meyer is with the Algorithm Engineering Group, Institute of Computer Science, Goethe University, 60325 Frankfurt, Germany, umeyer at cs.uni-frankfurt.de

[4] N. Pampu is with the Center for Cognitive and Neural Studies (Coneural), 400487 Cluj-Napoca, Romania, nicolae.pampu at ael.utcluj.ro

[5] V. Priesemann is with the Max Planck Institute for Brain Research, 60528 Frankfurt, Germany, viola.priesemann at brain.mpg.de

[6] R. Vicente is with the Frankfurt Institute for Advanced Studies (FIAS), 60438 Frankfurt, Germany, vicente at fias.uni-frankfurt.de

respectively. The corresponding quantity has been described several times (e.g. [6], [5]) and is most often referred to as transfer entropy [5].

B. Interaction-delay reconstruction

To specify explicitly the 'suitably chosen' past in Eq. 1 we suggest here to write:

$$TE(X \to Y, u) = I(Y(t); \mathbf{X}(t-u)|\mathbf{Y}(t-u_0)) \ , \quad (2)$$

where u is a parameter introduced to represent a finite interaction delay that we call the *source delay* here, and u_0 is a parameter that is chosen to ensure that the conditioning of the mutual information is optimal in Wiener's sense, i.e. we first try to predict the future of Y optimally from past embedding states $\mathbf{Y}(t-u_0)$ by optimizing u_0 and remove this influence by conditioning on this optimal past state. In practical terms this can be done using a criterion suggested by Ragwitz and Kantz [7]. To perform the necessary reconstruction of states of the signals we use Taken's delay embedding [18] and write the states of the systems as delay vectors of the form:

$$\mathbf{x}_t^d = (x(t), x(t-\tau), x(t-2\tau), \ldots, x(t-(d-1)\tau)) \ , \quad (3)$$

where d and τ denote the embedding dimension and the embedding delay,respectively. These parameters can also be optimized using Ragwitz' criterion [7]. Using the states obtained by delay embedding we can rewrite transfer entropy in the form of four Shannon (differential) entropies as:

$$TE(X \to Y, u) = S\left(\mathbf{y}_{t-u_0}^{d_y}, \mathbf{x}_{t-u}^{d_x}\right) - S\left(y_t, \mathbf{y}_{t-u_0}^{d_y}, \mathbf{x}_{t-u}^{d_x}\right) + S\left(y_t, \mathbf{y}_{t-u_0}^{d_y}\right) - S\left(\mathbf{y}_{t-u_0}^{d_y}\right) . \quad (4)$$

These differential Shannon entropies can be estimated using nearest-neighbor techniques in the embedding space [3]. As the dimensionality of the spaces involved in Eq. 4 can differ largely across terms, bias problems may arise and we used the Kraskov-Grassberger-Stögbauer estimator which handles this problem by fixing the number of neighbors in the highest dimensional space and by projecting the resulting distances to the lower dimensional spaces as the range to look for neighbors [10]. After adapting this technique to the TE formula, the estimator can be written as:

$$TE(X \to Y, u) = \psi(k) + \langle \psi\left(n_{y_t^{d_y}} + 1\right) - \psi\left(n_{y_t y_{t-u_0}^{d_y}} + 1\right) - \psi\left(n_{y_{t-u_0}^{d_y} x_{t-u}^{d_x}}\right)\rangle_t \quad (5)$$

where the distances to the k-th nearest neighbor in the highest dimensional space (spanned by $y_t, y_{t-u_0}^{d_y}, x_{t-u}^{d_x}$) define the radius of the spheres for the counting of points n_Z in all the marginal spaces Z involved. ψ denotes the digamma function, while the angle brackets ($\langle\cdot\rangle_t$) indicate an averaging over different time points.

Finally, for a reconstruction of delays we scan different values of the parameter u, identifying the value that maximizes information transfer from X to Y as the interaction delay, based on the Markov properties of X, Y.

Due to ppossible residual bias, TE values have to be compared against suitable surrogate data using non-parametric statistical testing to infer the presence or absence of directed interactions [16]. To this end we construct surrogate data by shifting the time series of one of the two signals of a pair by one experimental epoch or simulated block of data, preserving as many data features as possible and use permutation testing to assess significance. All applied estimators and statistical procedures are available as part of the TRENTOOL toolbox [14].

C. Graph based approach to the detection of cascade effects

If we are interested in more than two interacting processes, we have to extend Wiener's principle such that the past of *all* processes except the one source process we are investigating for its potential influence on a given target, are taken into account. Modifying Eq. 1 accordingly we obtain:

$$TE(X \to Y|V) = I(Y^+; \mathbf{X}^-|\mathbf{Y}^-, \mathbf{V}^-), \quad (6)$$

where $\mathbf{V}^-$ is the past state of all processes other than X, Y. This quantity is called the 'complete transfer entropy' [12], and in principle correctly describes the information transfer from X to Y in the presence of other influences. If the bivariate formulation from Eq. 1 is used instead, we may detect spurious information transfer from X to Y, although the actual interactions pass from X to a third node V_i (or more) and then from V_i to Y ('cascade effect'). While the complete transfer entropy solves this problem, it may be hard to estimate from finite data, because of the high dimensionality of the resulting embedding spaces – although novel embedding techniques [13] may to a certain degree ameliorate this problem. Here, we propose an approximate solution to the problem by exploiting the fact that in cascade effects the actual delays d (and intra-node) processing times accumulating in a multi-node pathway of interaction (e.g. $X \to \ldots \to V_i \to \ldots \to Y$) have to sum up to the apparent delay reconstructed for the pathway from X to Y for a spurious interaction. Reversing this argument, we can safely assume that a link does not arise from cascade effects if this condition is not met within the precision of our delay reconstruction.

Algorithmically our task is thus to identify for any significant transfer entropy value between two nodes i, j in the network, that is associated with an interaction delay d_{ij}, whether there is an alternative pathway between the two nodes with an equal sum of interaction delays. To this end we define the graph $G = (V, E)$ with vertices V denoting the nodes of the network and edges E representing significant transfer entropy values between two nodes. Edges are weighted by the respective delays between two nodes x and y as $w_{(x,y)} = d_{xy}$. A multi-node pathway $v_0 \rightsquigarrow v_l$ is described as a sequence of vertices $\langle v_0, \ldots, v_i, \ldots, v_l\rangle$, where $l-1$ is the length (number of edges) of the pathway.

The total weight of the pathway is the sum of the individual weights of all edges comprising the path, $\sum_i w_{(v_i,v_{i+1})}$.

To identify all alternative paths for a given edge (x, y), the edge is removed from the network and its respective nodes x and y are entered into the algorithm. Alternative paths are then detected in two steps: (1) a memoized dynamic programming approach [17] is used to determine, whether any path $x \rightsquigarrow y$ of a given total weight exists; (2) a modified depth first search (DFS) is used to reconstruct paths from the solution obtained in step (1) to reject paths that contain loops and to allow for further analysis.

Dynamic programming allows for the solution of a complex problem by decomposing it into easily solvable subproblems. By starting with trivial base cases, subproblems are solved iteratively by taking recourse to solutions to previous (more simple) subproblems to reduce computational demand. Solutions to previous subproblems are tabulated and are used for the solution of successive subproblems. Hence, in a first step the presented problem of finding a path $x \rightsquigarrow y$ of weight $w_{(x,y)}$ is decomposed into subproblems of the form $\exists Path_x[w_l, v_i]$ (asking whether any path $x \rightsquigarrow v_i$ with a certain weight w_l exists). This is answered for all combinations of nodes $v_j \in V$ and path weights $w_l = 1, \ldots, m$ (m denotes an upper limit for path weights that is calculated as $w_{(x,y)} + k$, where k denotes a user-defined correction, that accounts for imprecisions in measurements and estimation). The algorithm terminates when the solution to the initial problem ($\exists Path_x[m, y]$, a path $x \rightsquigarrow y$ of length m) can be computed from all previously solved subproblems.

In the second step, paths are reconstructed from the stored solutions to the previously defined subproblems, using a modified DFS [17], that starts with node y and recursively expands the first predecessor of a node until node x is reached. The reconstructed path is kept and the current recursion is aborted, which leads to backtracking of the algorithm until the most recent, not yet expanded node is found. To avoid loops, visited nodes are marked in a boolean array, such that visiting a node a second time during a recursion leads to a preemptive backtracking of the search.

Backtracking is conducted for all paths $x \rightsquigarrow y$, that have a total edge weight which lies within an interval $[n, m]$ (calculated as $w_{(x,y)} \pm k$). As the first part of the algorithm iterates over all path lengths smaller than m, path lengths from the interval $[n, m]$ are automatically solved as 'subproblems' of the initial problem and can thus be reconstructed by the second part of the algorithm. If a path $x \rightsquigarrow y$ with a total weight within the interval exists, the edge (x, y) is flagged as potentially spurious.

III. Validation

A. Bidirectionally coupled Lorenz systems

We considered two bidirectionally, quadratically coupled Lorenz systems:

$$\begin{aligned}
\dot{X}_i(t) &= \sigma(Y_i(t) - X_i(t)) \\
\dot{Y}_i(t) &= X_i(t)(\rho_i - Z_i(t)) - Y_i(t) + \gamma_{ij} Y_j^2(t - \delta_{ij}), \\
\dot{Z}_i(t) &= X_i(t)Y_i(t) - \beta Z_i(t)
\end{aligned}$$

where $i, j = 1, 2$ $j \neq i$; σ, ρ and β, are the *Prandtl number*, the *Rayleigh number*, and a geometrical scale; γ_{ij} represent the coupling strengths , and δ_{ij} the delays of the bidirectional system. Simulation parameters were: $\sigma = 10$, $\rho_1 = 25$, $\rho_2 = 28$ and $\beta = 8/3$, $\gamma_{12} = 0.1$ and $\gamma_{21} = 0.05$. The interaction delays were set to $\delta_{12} = 3$ and $\delta_{21} = 5$. Solutions were computed using the $dde23$ solver in MATLAB and results were resampled such that the original delays δ_{12}, δ_{21} amounted to 45 and 75 samples of resampled time. Analyses were performed in TRENTOOL. The Ragwitz criterion was used to determine the embedding dimension and lag τ. We used a significance level of 0.05, corrected for multiple comparisons via false dicovery rate to assess significance of the coupling. To identify interaction delays we scanned the source delay parameter u from 25 to 95 time steps in steps of 1 sample. Transfer entropy values peaked a $u = 45$

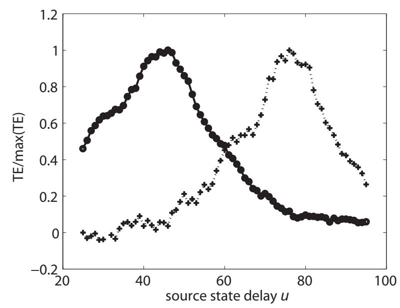

Fig. 1. Transfer entropy values versus source delay u for two bidirectionally coupled, chaotic Lorenz systems.

and $u = 75$ samples for the interaction from system 1 to 2 and 2 to 1, respectively; these were also the true interaction delays used for simulation.

B. Local field potential data

To demonstrate that spurious interactions due to cascade effects are correctly identified using our graph-based algorithm we reanalyzed LFP recordings from the retina and tectum of the turtle brain (Pseudemys scripta elegans) stimulated with random light flashes. Procedures have been described in [14]. In short, stimulus light intensity, LFPs in the tectum and the electroretinogram were recorded. Physical interactions in this system exist from the light source (LS) to the retina (R), and from the retina to the tectum (T). In a bivariate transfer entropy analysis, a spurious interaction is detected from LS to T [14]. For this spurious interaction, the *reconstructed* interaction delays from the LS to the R and from the R to T should sum up to the *reconstructed* interaction delay between LS and T. We reconstructed the interaction delays by scanning the source delay u using TRENTOOL. The reconstructed delays were: $d_{(\text{LS,R})} = 28\ ms$, $d_{(\text{R,T})} = 13\ ms$, $d_{(\text{LS,T})} = 43\ ms$, meaning that the reconstructed delays added up with an error of 2 ms or 4%. Using the proposed graphical algorithm with a suitable precision parameter therefore indeed identified the spurious interaction between light source and tectum.

C. MEG Data

Empirical datasets were obtained from MEG measurements on 30 healthy participants, who had to complete a face detection task [11]. Data were preprocessed and time courses of active sources in the brain differentiating successful from unsuccessful face recognition were extracted using a beamformer approach implemented in Fieldtrip [15]. Transfer entropy values and interactions delays were reconstructed using TRENTOOL [14]. Figure 2 presents the transfer entropy graphs before (left panel) and after pruning (right panel) with the algorithm presented above. The correction term k was set to 3 ms, such that paths were considered alternative paths to the apparent edge (x, y), whenever the weight of the path lay within the interval $w_{(x,y)} \pm 3$ms. On average this led to the removal of 52% of the edges due to 'cascade effects'.

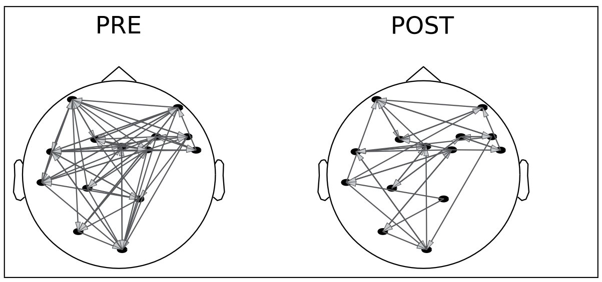

Fig. 2. Transfer entropy networks before and after pruning of potentially spurious edges.

IV. Discussion

The proposed modified transfer entropy estimator identified the true interaction delays in simulated data and resulted in physiologically plausible values in neural data. The proposed graph-based algorithm was able to identify cascade effects in the test case of local field potential data, and it revealed that cascade effects play a major role in results from bivariate analysis of source activity from human MEG recordings. Nevertheless, three important facts should not be overlooked: (1) The algorithm can only label potentially spurious interactions. If indeed two pathways - one direct, and one via intermediate nodes, exist that have the same timing, then we label a true interaction as potentially spurious. Hence, arguments should build on positive results, i.e. only those interactions where we can safely assume that they are not due to spurious interactions. (2) The related problem of 'common drive' where one source node drives two or more target nodes with different delays, was not addressed in this study. It is however accessible using a similar methodology, (3) Results depend on the precision of the match of summed delay times we ask for. If no initial guess about a suitable precision parameter exists, results may be misleading. Again this suggests to interpret only those interactions that are positively identified as not being due to cascade effects.

Acknowledgment

MW,RV,VP received support from the grant LOEWE Neuronale Koordination Forschungsschwerpunkt Frankfurt NeFF. UM was partially supported by the DFG grant ME 3250/1-3, and by MADALGO Center for Massive Data Algorithmics, a center of the Danish National Research Foundation.The authors thank Joseph T. Lizier from the MPI for Mathematics in the Sciences, Leipzig, Germany, for stimulating discussions.

References

[1] N. Wiener. The theory of prediction. In E. F. Beckmann, editor, *In Modern Mathematics for the Engineer*. McGraw-Hill, New York, 1956.

[2] C. W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37:424–438, 1969.

[3] L.F. Kozachenko and N.N. Leonenko. Sample estimate of entropy of a random vector. *Probl. Inform. Transm.*, 23:95–100, 1987.

[4] P. A. Felts, T. A. Baker, and K. J. Smith. Conduction in segmentally demyelinated mammalian central axons. *J Neurosci*, 17(19):7267–7277, Oct 1997.

[5] T. Schreiber. Measuring information transfer. *Phys Rev Lett*, 85(2):461–464, Jul 2000.

[6] M. Paluš. Synchronization as adjustment of information rates: detection from bivariate time series. *Phys. Rev. E*, 63:046211, 2001.

[7] M. Ragwitz and H. Kantz. Markov models from data by simple nonlinear time series predictors in delay embedding spaces. *Phys Rev E Stat Nonlin Soft Matter Phys*, 65(5 Pt 2):056201, May 2002.

[8] K. J. Friston, L. Harrison, and W. Penny. Dynamic causal modelling. *Neuroimage*, 19(4):1273–1302, Aug 2003.

[9] E. J. Lang and Jack Rosenbluth. Role of myelination in the development of a uniform olivocerebellar conduction time. *J Neurophysiol*, 89(4):2259–2270, Apr 2003.

[10] A. Kraskov, H. Stoegbauer, and P. Grassberger. Estimating mutual information. *Phys Rev E Stat Nonlin Soft Matter Phys*, 69(6 Pt 2):066138, Jun 2004.

[11] C. Grützner, P. J. Uhlhaas, E. Genc, A. Kohler, W. Singer, and Michael Wibral. Neuroelectromagnetic correlates of perceptual closure processes. *J Neurosci*, 30(24):8342–8352, Jun 2010.

[12] J. Lizier and M. Prokopenko. Differentiating information transfer and causal effect. *Eur. Phys. J. B*, 73:605–615, 2010.

[13] L. Faes, G. Nollo, and A. Porta. Information-based detection of nonlinear granger causality in multivariate processes via a nonuniform embedding technique. *Phys Rev E Stat Nonlin Soft Matter Phys*, 83(5 Pt 1):051112, May 2011.

[14] M. Lindner, R. Vicente, V. Priesemann, and M. Wibral. Trentool: A matlab open source toolbox to analyse information flow in time series data with transfer entropy. *BMC Neurosci*, 12(1):119, Nov 2011.

[15] R. Oostenveld, P. Fries, E. Maris, and J.-M. Schoffelen. Fieldtrip: Open source software for advanced analysis of meg, eeg, and invasive electrophysiological data. *Comput Intell Neurosci*, 2011:156869, 2011.

[16] R. Vicente, M. Wibral, M. Lindner, and G. Pipa. Transfer entropy – a model-free measure of effective connectivity for the neurosciences. *J Comput Neurosci*, 30(1):45–67, 2011.

[17] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to Algorithms*. The MIT Press, 2 edition, 2001.

[18] F. Takens. *Dynamical Systems and Turbulence, Warwick 1980*, volume 898 of *Lecture Notes in Mathematics*, chapter Detecting Strange Attractors in Turbulence, pages 366–381. Springer, 1981.