Letter

Disentangling high-order effects in the transfer entropy

Sebastiano Stramaglia,[1] Luca Faes,[2] Jesus M. Cortes,[3,4,5,6] and Daniele Marinazzo[7]

[1]*Dipartimento Interateneo di Fisica, Università degli Studi di Bari Aldo Moro, and INFN Sezione di Bari, 70126 Bari, Italy*
[2]*Dipartimento di Ingegneria, Università di Palermo, 90128 Palermo, Italy*
[3]*Biocruces-Bizkaia Health Research Institute, Barakaldo, Spain*
[4]*Biomedical Research Doctorate Program, University of the Basque Country, Leioa, Spain*
[5]*Department of Cell Biology and Histology, University of the Basque Country, Leioa, Spain*
[6]*IKERBASQUE Basque Foundation for Science, Bilbao, Spain*
[7]*Department of Data Analysis, Ghent University, Ghent, Belgium*

(Received 13 February 2024; accepted 17 May 2024; published 9 July 2024)

Transfer entropy (TE), the primary method for determining directed information flow within a network system, can exhibit bias—either in deficiency or excess—during both pairwise and conditioned calculations, owing to high-order dependencies among the dynamic processes under consideration and the remaining processes in the system used for conditioning. Here, we propose a novel approach. Instead of conditioning TE on all network processes except the driver and the target, as in its fully conditioned version, or not conditioning at all, as in the pairwise approach, our method searches for both the multiplets of variables that maximize information flow and those that minimize it. This provides a decomposition of TE into unique, redundant, and synergistic atoms. Our approach enables the quantification of the relative importance of high-order effects compared to pure two-body effects in information transfer between two processes, while also highlighting the processes that contribute to building these high-order effects alongside the driver. We demonstrate the application of our approach in climatology by analyzing data from El Niño and the Southern Oscillation.

DOI: 10.1103/PhysRevResearch.6.L032007

Introduction. A central task in analyzing complex systems is to understand the joint dynamics of its components. Granger causality (GC) [1] and transfer entropy (TE) [2,3] are widely used tools to detect and quantify statistical relationships between random processes mapping the evolution of coupled dynamic systems over time in terms of reduction of variance and surprise, respectively. For Gaussian systems, GC and TE are equivalent [4]; a symbolic version of TE has been developed in Ref. [5].

The target properties of these statistical dependencies have been called "information flow" or "causality," names which are used in this paper to indicate a measured effect [6]; this clarification pairs an important distinction that needs indeed to be made between mechanisms and behaviors in this context [7].

In recent years, alongside the growing interest in high-order interactions [7,8], there has been increasing attention devoted to the emergent properties of complex systems, which manifest through high-order behaviors sought in observed data, moving beyond traditional dyadic descriptions. A key framework in this literature is the partial information decomposition (PID) [9] and its subsequent developments [10], which utilize information-theoretic tools to reveal high-order dependencies among groups of three or more random variables and describe their synergistic or redundant nature. Within this framework, redundancy refers to information retrievable from multiple sources, while synergy refers to statistical relationships existing within the whole system that cannot be observed in its individual parts. Importantly, while the PID was originally proposed for sets of random variables, it was then generalized to random processes [11].

It is worth mentioning that in Ref. [12], a critique of TE has been raised: if one believes that a dyadic network accurately models a complex system, then one implicitly assumes that polyadic relationships are either unimportant or nonexistent. However, assessing the relative importance of these high-order interactions compared to dyadic relations remains an open problem. Moreover, in Ref. [13], the influence of synergy and redundancy on the inference of information flow between two subsystems of a complex network, in terms of GC, has been studied. This research demonstrates that both pairwise and fully conditioned GC analyses encounter challenges in the presence of synergy or redundancy in time series data, indeed pairwise GC fails to reveal synergistic effects whilst fully conditioned GC may fail to reveal redundant effects.

The question we address here is: how can we compare the pure dyadic influence with the many-body effects due to the remaining processes in a network, given a driving random process and a target process? Within the framework of partial conditioning in multivariate data sets [14], we propose a straightforward decomposition of the maximal information flow from the driver to the target into three positive components: unique, redundant, and synergistic TEs. The magnitudes of these components quantify the relative importance of high-order effects and pure dyadic effects in the influence from the driver to the target. Moreover, the proposed analysis highlights the processes that, along with the driver, contribute to synergistic (or redundant) effects on the target.

To introduce our approach, we start observing the following fundamental fact behind any approach to PID: for any three (scalar or vector) random variables a, b, and c, with H being the entropy and I being the mutual information, we will always have $H(a) \geqslant H(a|c)$, but we will not necessarily have $I(a;b) \geqslant I(a;b|c)$. In fact, conditioning on c can either reduce the information shared between a and b or increase it: the two cases are related to redundancy $[I(a;b) > I(a;b|c)]$ or to synergy $[I(a;b) < I(a;b|c)]$ among the three variables. Now, suppose that c is the set of all the variables at hand that describe the *environment* around the pair (a, b); the quantity $I(a;b) - I(a;b|c)$ has been proposed to assess the redundant (if positive) or synergistic (if negative) nature of the interaction among a, b, and the full environment [15,16]. Nevertheless, considering only a subset of all the variables in c, it is intuitive that searching for $c_{\min}$ minimizing $I(a;b|c)$ should better capture the amount of redundancy R that the environment shares with the pair, i.e., $R = I(a;b) - I(a;b|c_{\min})$. On the other hand, searching for $c_{\max}$ maximizing $I(a;b|c)$ leads to the amount of synergy S that the *environment* provides in terms of the increase of mutual information $S = I(a;b|c_{\max}) - I(a;b)$. The definitions given above for R and S constitute the core of our approach. In the next section we apply this heuristics to the transfer entropy.

Method. Given a zero-mean stationary two-dimensional Markov process consisting of the scalar processes X and Y, we aim at evaluating the transfer entropy $X \to Y$. Denoting the present state of the target as Y_t, the vector of the target's past variables as $Y_{<t} = [Y_{t-1} \dots Y_{t-p}]$, and the vector of the driver's past variables as $X_{<t} = [X_{t-1} \dots X_{t-p}]$, with p being the order of the process, the pairwise TE, T_p, is defined as

$$T_p = T_{X\to Y} = I(Y_t; X_{<t}|Y_{<t}), \tag{1}$$

where $I(\cdot;\cdot|\cdot)$ indicates the conditional mutual information. Suppose now that we also simultaneously measure n other processes $\mathbf{Z} = \{Z_1, Z_2, \dots, Z_n\}$. The fully conditioned TE, T_f, is defined as

$$T_f = T_{X\to Y|\mathbf{Z}} = I(Y_t; X_{<t}|Y_{<t}, Z_{1,<t}, \dots, Z_{n,<t}). \tag{2}$$

Now, instead of fully conditioning on all processes, we condition only on a subset in $\mathbf{Z}$; i.e., we compute $T_\alpha = T_{X\to Y|\mathbf{Z}_\alpha}$, where $\mathbf{Z}_\alpha$, $\alpha \subset \{1, \dots, n\}$, is an element of the powerset of $\{Z_1, Z_2, \dots, Z_n\}$. We denote $\mathbf{Z}_m$ as the subset of processes in $\mathbf{Z}$ which minimizes T_α, and the corresponding value of the TE is denoted as T_m:

$$T_m = T_{X\to Y|\mathbf{Z}_m} = I(Y_t; X_{<t}|Y_{<t}, \mathbf{Z}_{m,<t}). \tag{3}$$

Similarly, $\mathbf{Z}_M$ represents the subset of processes in $\mathbf{Z}$ that maximizes T_α, with the corresponding value of the transfer entropy denoted as T_M:

$$T_M = T_{X\to Y|\mathbf{Z}_M} = I(Y_t; X_{<t}|Y_{<t}, \mathbf{Z}_{M,<t}). \tag{4}$$

This decomposition is illustrated in Fig. 1(a).

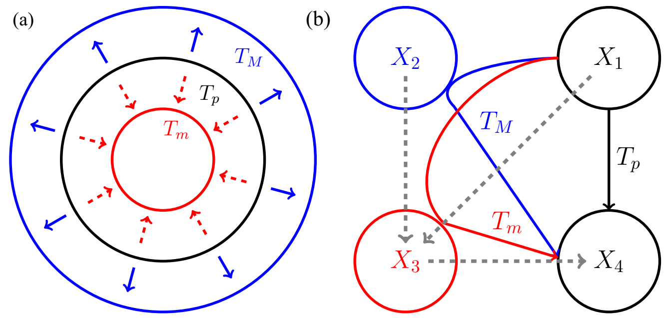

FIG. 1. (a) Representation of the pairwise transfer entropy T_p and of the values obtained maximizing and minimizing T_α, called T_M and T_m, respectively. Redundant influences are depicted by red arrows, and synergistic influences are depicted by blue arrows. (b) Pairwise, synergistic, and redundant conditioning paths are represented on a simple model with four nodes, with simulated couplings indicated by the dashed gray arrows.

Given that the reduction of TE is associated with redundancy (R), we express this relationship as $R = T_p - T_m$, where the unique information (U) flowing from X to Y is identified with T_m. On the other hand, the synergistic information flow (S) signifies the increase in transfer entropy when additional variables are included in the set of conditioning variables. This relationship can be expressed as $S = T_M - T_p$. It follows that

$$T_M = S + R + U. \tag{5}$$

In other words, the maximal information flow from X to Y can be decomposed into the sum of a unique contribution (U), representing a pure two-body effect, and synergistic and redundant contributions that describe higher-order components of the interdependence between X and Y. In the Appendix we describe the assumption for the PID which corresponds to these definitions of R, S, and U in the case of three variables.

Conducting an exhaustive search for subsets $\mathbf{Z}_m$ and $\mathbf{Z}_M$ becomes unfeasible for large n. Therefore, we employ a greedy search strategy, wherein firstly we perform a search over all the processes for the first process to be tentatively used as a conditioner. Subsequently, one process is added at a time, to the previously selected ones, to construct the set of conditioning processes that either maximize or minimize the TE. The criterion for terminating the greedy search for conditioning processes, minimizing (maximizing) the TE, is to stop when the corresponding decrease (increase) of the TE can be explained as being due to chance. Therefore, one can estimate the probability that the increase in T_α is lower (higher) than that corresponding to the inclusion of a process sharing the individual statistical properties of the selected one but being otherwise uncoupled from X and Y (realizations of the selected Z process are obtained as iterative amplitude-adjusted fourier transform [17] surrogates). When such a probability is lower than a given threshold, after correction for multiple comparisons, the selected process is thus added to the multiplet of conditioning processes.

As a toy model (which in the distinction between mechanism and behavior would constitute a model for the former, but which should be reasonably echoed by the latter), consider a system of four random processes, with X_1 and X_2 influencing X_3, whilst X_3 influences X_4, as depicted by the gray dashed arrows in Fig. 1(b):

$$\begin{aligned} X_{1,t} &= \epsilon_{1,t}, \\ X_{2,t} &= \epsilon_{2,t}, \\ X_{3,t} &= 0.5(X_{1,t-1} + X_{2,t-1}) + 0.1\ \epsilon_{3,t}, \\ X_{4,t} &= 0.9\ X_{3,t-1} + 0.1\ \epsilon_{4,t}, \end{aligned} \tag{6}$$

where ϵ's are unit variance zero mean i.i.d. Gaussian processes. Taking X_1 as the driver and X_4 as the target, we easily obtain $T_p = 0.32$ nats, $T_m = 0$ nats (with $\mathbf{Z}_m$ coinciding with X_3), and $T_M = 1.25$ nats (with $\mathbf{Z}_M$ coinciding with X_2); it follows that $U = 0$ nats (there is no pure two-body influence), while $R = 0.32$ nats and $S = 0.93$ nats are the redundant and synergistic TEs, respectively. Noting that X_1 and X_2 are colliders for X_3, and that the chain $X_1 \to X_3 \to X_4$ is a redundant circuit, it is easy to realize that these results are what one should expect for this example. We remark that the output, on this toy model, for the fully conditioned TE is $T_f = 0$ nats, whilst for the pairwise TE it is $T_p = 0.32$ nats: both approaches fail to highlight many-body effects.

Application to climate science. As an example of an application to a real dataset, we consider a case study in climate science, i.e., the influence of NINO34 (the East Central Tropical Pacific sea surface temperature anomaly, also called El Niño) on the SOI (Southern Oscillation Index, the standardized difference in surface air pressure between Tahiti and Darwin). These two indexes are crucial for the description of El Niño and the Southern Oscillation (ENSO), a periodic fluctuation in sea surface temperature and the air pressure of the overlying atmosphere across the equatorial Pacific Ocean. ENSO is considered the most prominent interannual climate variability on Earth [18]. Since the exact initiating causes of an ENSO warm or cool event are not fully understood, it is important to analyze the statistical relation between the two components of ENSO—atmospheric pressure (SOI) and sea surface temperature (NINO34). The question we consider is: what fraction of the total information flow NINO34 → SOI is due to pure two-body effects?

The other climatic indexes that we consider here are AIR (All Indian Rainfall), AMO (Atlantic Multidecadal Oscillation), GMT (Global Mean Temperature anomaly), HURR (total number of hurricanes or named tropical storms in a given month in the Atlantic region), NOA (North Atlantic Oscillation of pressure anomalies over the Atlantic), NP (North Pacific pattern of sea level pressure), NTA (North Tropical Atlantic), PDO (Pacific Decadal Oscillation), QBO (Quasi-Biennial Oscillation), Sahel (Sahel Standardized Rainfall), and TSA (Tropical Southern Atlantic Index). All time series have been detrended and deseasonalized; globally, these 13 monthly sampled time series coincide with those analyzed using an approach based on a linear approximation of the pairwise transfer entropy in Ref. [19]. It is worth stressing that these variables are not guaranteed to measure separate processes and that latent factors are likely to be present. This further motivates an analysis that takes shared information into account.

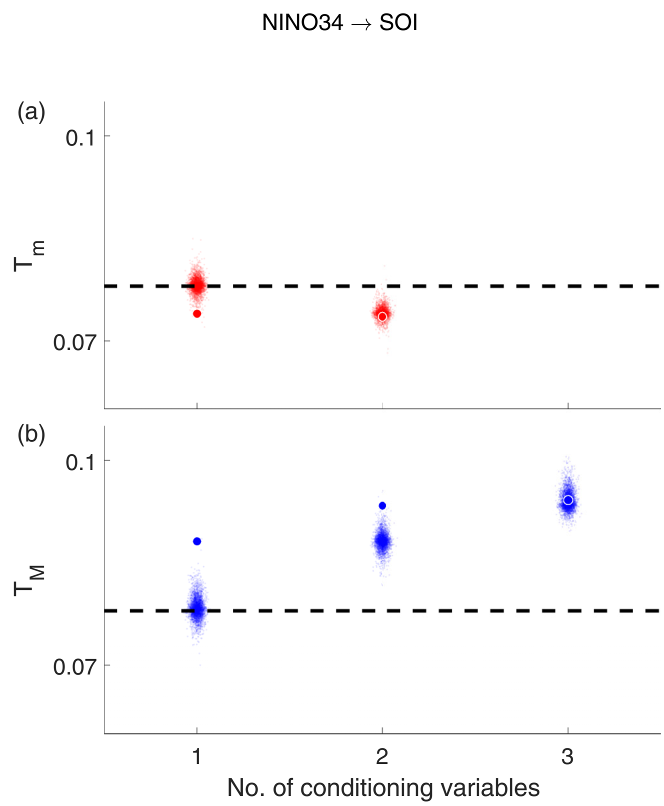

FIG. 2. (a) The transfer entropy NINO34 → SOI is depicted for various numbers of conditioning variables along the greedy search to minimize the transfer entropy (T_m), together with the values obtained on 4000 surrogates of the added time series. Conditioning to a second variable results in values compatible with the null hypothesis that the decrease is due to chance; hence, the set $\mathbf{Z_m}$ reduces to the first variable (Pacific Decadal Oscillation). The value of the pairwise transfer entropy (T_p) is indicated by the dashed black line. (b) As in panel (a), but referring to the greedy search that maximizes the transfer entropy, resulting in T_M. Conditioning to a third variable results in values compatible with the null hypothesis; hence, the set $\mathbf{Z_M}$ reduces to the first and second variables (North Tropical Atlantic and Tropical Southern Atlantic Index).

We consider the period 1950–2016 (i.e., 792 data points) for which all the values of the 13 time series are available, and we adopt the assumption of Gaussianity so as to identify transfer entropy with GC (we remark that also in Ref. [19] the Gaussian assumption was adopted). We find that the pairwise TE, NINO34 → SOI, is $T_p = 0.078$ nats, using $p = 2$ as the order the model fixed by minimum description length, a widely used criterion for model selection [20]. Then we find $T_m = 0.074$ nats [with $\mathbf{Z_m}$ equal to PDO, see Fig. 2(a)] and $T_M = 0.094$ nats [with $\mathbf{Z_M}$ equal to the pair NTA and TSA, see Fig. 2(b)]: therefore, in this case we have $S = 0.016$ nats, $R = 0.004$ nats, and $U = 0.074$ nats. Since $\frac{(R+S)}{T_M} = 0.2$, we conclude that about 20% of the total information flow from NINO34 to SOI can be ascribed to many-body effects, and 80% can be ascribed to pure two-body effects. Moreover, we note that in Ref. [19], using a pairwise approach, eight drivers of SOI were identified among the 13 time series, including

NINO34, NTA, and TSA; moreover PDO was found to be target for both NINO34 and SOI. The proposed approach allows one to identify which drivers play a role in producing high-order effects on SOI in cooperation with NINO34.

Conclusions. In summary, we have introduced a decomposition of transfer entropy (TE) that separates pure dyadic dependence from many-body effects resulting from interactions with other variables. Applying this methodology to the interplay between NINO34 and SOI, two components of ENSO, revealed the presence of non-negligible redundant and synergistic high-order effects. We posit that this approach serves as a bridge between dyadic and polyadic methodologies, offering a complementary perspective to those focusing on the assessment of high-order effects [9,15,21]: indeed, these approaches aim at decomposing the total information about the target from many sources, whilst our framework's focus is on assessing how the environment changes the transfer entropy between two variables. In other words, our goal here is to provide a tool to anatomize the transfer entropy into a pure two-body component and two many-body components, and not to rigorously define redundancy and synergy for the whole group of drivers at hand (we note that the number of atoms in PID explodes exponentially with the number of sources). It follows that the quantities R and S, here introduced, should be considered only in relation with the driver-target pair whose transfer entropy is under evaluation. Moreover, the computational demand of the proposed methodology is not intensive, especially when employing a greedy search strategy to identify the optimal multiplet of conditioning variables. Further work will be devoted to assessing how R, S, and U are affected by influencing factors typically encountered in empirical studies (e.g., coupling strengths, bidirectional couplings, number of data points in time series, noise, mixed processes), especially in the case of a large number of conditioning variables.

The code to simulate and analyze data is available from Ref. [22]. Climate data as described in Ref. [19] can be downloaded at the NOAA website [23], with the exception of AIR, which is available via the Indian Institute of Tropical Meteorology [24].

Acknowledgments. L.F. and S.S. were supported by the project "HONEST - High-Order Dynamical Networks in Computational Neuroscience and Physiology: An Information-Theoretic Framework," Italian Ministry of University and Research (funded by MUR, PRIN 2022, code 2022YMHNPY, CUP: B53D23003020006). S.S. was supported by the project "Higher-order complex systems modeling for personalized medicine," Italian Ministry of University and Research (funded by MUR, PRIN 2022-PNRR, code P2022JAYMH, CUP: H53D23009130001). We thank Yves Rosseel and Niels Van Santen (Ghent University) for useful comments.

Appendix: Relation with partial information decomposition. In order to clarify the relation between the proposed decomposition and the PID [9], we analyze the case of only three stochastic variables: a, b, and c. For simplicity we consider here the decomposition of the mutual information, as described in the Introduction; a similar reasoning holds for the transfer entropy. The PID of the mutual information from the pair of sources $\{a, c\}$ to the target b is as follows:

$$\begin{aligned} I(b; a, c) &= R + S + U_a + U_c, \\ I(b; a) &= R + U_a, \\ I(b; c) &= R + U_c. \end{aligned} \tag{A1}$$

It follows that

$$\begin{aligned} I(b; a|c) &= S + U_a, \\ I(b; a) &= R + U_a. \end{aligned} \tag{A2}$$

Now, in the proposed frame the variable c is tentatively used as a conditioning variable If $I(b; a|c) > I(b; a)$, then $c_{\max} = \{c\}$ and $c_{\min} = \emptyset$; therefore,

$$\begin{aligned} S &= I(b; a|c) - I(b; a) = -II, \\ R &= 0, \end{aligned} \tag{A3}$$

where II is the interaction information, i.e., the classical three-body measure of information [25]. On the other hand, if $I(b; a|c) < I(b; a)$, then $c_{\min} = \{c\}$ and $c_{\max} = \emptyset$; therefore,

$$\begin{aligned} S &= 0, \\ R &= I(b; a) - I(b; a|c) = II. \end{aligned} \tag{A4}$$

Note that one always has $II = R - S$, i.e., also in the proposed decomposition the interaction information is the balance between redundancy and synergy; however, we note that one between R and S always vanishes in this frame. We conclude that, in the case of three variables, the proposed decomposition is equivalent to the PID with the requirement that either S or R should vanish. This is different, e.g., from the minimum mutual information prescription [26], which requires that $R = \min\{I(b; a), I(b; c)\}$.

We remark, however, that unlike PID, our framework is not designed to provide a decomposition of the total information about the target from many sources: its focus is rather on the way the environment influences the transfer entropy between two variables, with practical applicability also in the case of a large number of variables.

[1] C. W. J. Granger, Investigating causal relations by econometric models and cross-spectral methods, Econometrica **37**, 424 (1969).

[2] T. Schreiber, Measuring information transfer, Phys. Rev. Lett. **85**, 461 (2000).

[3] T. Bossomaier, L. Barnett, M. Harré, and J. T. Lizier, *An Introduction to Tranfer Entropy* (Springer, Berlin, 2016).

[4] L. Barnett, A. B. Barrett, and A. K. Seth, Granger causality and transfer entropy are equivalent for Gaussian variables, Phys. Rev. Lett. **103**, 238701 (2009).

[5] M. Staniek and K. Lehnertz, Symbolic transfer entropy, Phys. Rev. Lett. **100**, 158101 (2008).

[6] A. B. Barrett and L. Barnett, Granger causality is designed to measure effect, not mechanism, Front. Neuroinf. **7**, 6 (2013).

[7] F. Rosas, P. Mediano, A. Luppi, T. Varley, J. Lizier, S. Stramaglia, H. Jensen, and D. Marinazzo, Disentangling high-order mechanisms and high-order behaviours in complex systems, Nat. Phys. **18**, 476 (2022).

[8] F. Battiston, E. Amico, A. Barrat, G. Bianconi, G. Ferraz, Arruda, B. Franceschiello, I. Iacopini, S. Kéfi, V. Latora, Y. Moreno, M. Murray, T. Peixoto, F. Vaccarino, and G. Petri, The physics of higher-order interactions in complex systems, Nat. Phys. **17**, 1093 (2021).

[9] P. L. Williams and R. D. Beer, Nonnegative decomposition of multivariate information, arXiv:1004.2515.

[10] J. T. Lizier, N. Bertschinger, J. Jost, and M. Wibral, Information decomposition of target effects from multi-source interactions: Perspectives on previous, current and future work, Entropy **20**, 307 (2018).

[11] L. Faes, D. Marinazzo, and S. Stramaglia, Multiscale information decomposition: Exact computation for multivariate Gaussian processes, Entropy **19**, 408 (2017).

[12] R. G. James, N. Barnett, and J. P. Crutchfield, Information flows? A critique of transfer entropies, Phys. Rev. Lett. **116**, 238701 (2016).

[13] S. Stramaglia, J. M. Cortes, and D. Marinazzo, Synergy and redundancy in the Granger causal analysis of dynamical networks, New J. Phys. **16**, 105003 (2014).

[14] D. Marinazzo, M. Pellicoro, and S. Stramaglia, Causal information approach to partial conditioning in multivariate data sets, Comput. Math. Methods Med. **2012**, 303601 (2012).

[15] F. E. Rosas, P. A. M. Mediano, M. Gastpar, and H. J. Jensen, Quantifying high-order interdependencies via multivariate extensions of the mutual information, Phys. Rev. E **100**, 032305 (2019).

[16] G. Mijatovic, L. Sparacino, Y. Antonacci, M. Javorka, D. Marinazzo, S. Stramaglia, and L. Faes, Assessing high-order links in cardiovascular and respiratory networks via static and dynamic information measures, IEEE Open J. Eng. Med. Biol. **1**, 1 (2024).

[17] T. Schreiber and A. Schmitz, Improved surrogate data for nonlinearity tests, Phys. Rev. Lett. **77**, 635 (1996).

[18] M. J. McPhaden, S. E. Zebiak, and M. H. Glantz, ENSO as an integrating concept in earth science, Science **314**, 1740 (2006).

[19] R. Silini, G. Tirabassi, M. Barreiro, L. Ferranti, and C. Masoller, Assessing causal dependencies in climatic indices, Clim. Dyn. **61**, 79 (2023).

[20] M. H. Hansen and B. Yu, Model selection and the principle of minimum description length, J. Am. Stat. Assoc. **96**, 746 (2001).

[21] S. Stramaglia, T. Scagliarini, B. C. Daniels, and D. Marinazzo, Quantifying dynamical high-order interdependencies from the O-information: An application to neural spiking dynamics, Front. Physiol. **11**, 595736 (2021).

[22] GitHub, https://github.com/danielemarinazzo/TE_decomposition.

[23] NOAA website, https://psl.noaa.gov/data/climateindices/list/).

[24] Indiana Institute of Tropical Meteorology, https://www.tropmet.res.in/.

[25] W. J. McGill, Multivariate information transmission, Psychometrika **19**, 97 (1954).

[26] A. B. Barrett, Exploration of synergistic and redundant information sharing in static and dynamical Gaussian systems, Phys. Rev. E **91**, 052802 (2015).