A parsimonious Granger causality formulation for capturing arbitrarily long multivariate associations

Andrea Duggento¹* , **Gaetano Valenza²** , **Luca Passamonti³,⁴*** , **Salvatore Nigro⁵** , **Maria Giovanna Bianco⁶** , **Maria Guerrisi¹** , **Riccardo Barbieri⁷** , **Nicola Toschi¹,⁸**

1. Department of Biomedicine and Prevention, University of Rome Tor Vergata, Rome, Italy
2. Department of Information Engineering and Research Centre "E. Piaggio", University of Pisa, Pisa, Italy
3. Institute of Bioimaging and Molecular Physiology, National Research Council, Milano, Italy
4. Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK
5. Department of Experimental and Clinical Medicine, Magna Graecia University, Catanzaro, Italy
6. Department of Health Sciences, Magna Graecia University, Catanzaro, Italy
7. Department of Electronics, Informatics and Bioengineering, Politecnico di Milano, Milano, Italy
8. Department of Radiology, Martinos Center for Biomedical Imaging and Harvard Medical School, Boston, US
* Correspondence: duggento@med.uniroma2.it, lp337@medschl.cam.ac.uk

Abstract: High-frequency neuroelectric signals like electroencephalography (EEG) or magnetoencephalography (MEG) provide a unique opportunity to infer causal relationships between local activity of brain areas. While causal inference is commonly performed through Classical Granger causality (GC) based on multivariate autoregressive models, this method may encounter important limitations (e.g. data paucity) in the case of high dimensional data from densely connected systems like the brain. Additionally, physiological signals often present long-range dependencies which commonly require high autoregressive model orders / number of parameters. We present a generalization of autoregressive models for GC estimation based on Wiener-Volterra decompositions with Laguerre polynomials as basis functions. In this basis, the introduction of only one additional global parameter allows to capture arbitrary long dependencies without increasing model order, hence retaining model simplicity, linearity and ease of parameters estimation. We validate our method in synthetic data generated from families of complex, densely connected networks and demonstrate superior performance as compared to classical GC. Additionally, we apply our framework to studying the directed human brain connectome through MEG data from 89 subjects drawn from the Human Connectome Project (HCP) database, showing that it is able to reproduce current knowledge as well as to uncover previously unknown directed influences between cortical and limbic brain regions.

Keywords: Granger Causality; Directed brain connectivity; MEG connectivity, Laguerre polynomials

---

1. Introduction

Granger causality is a widely applied model-based tool to quantify directed information transfer between signals originating from (possibly) complex networks in a number of disciplines, ranging from econometrics to neuroscience. In particular, the study of the so-called brain 'connectome' (i.e. the entirety of all multivariate functional interactions between brain areas that can be possibly estimated starting from neuromonitoring time-series data such as Magnetoencephalography (MEG), electroencephalography (EEG) or functional magnetic resonance imaging (fMRI)) has provided fertile ground for the application of most diverse causality-based methods [1–6]. Often, connectomics studies in general, and causality estimation in neuroscience in particular, are planned in a network-discovery fashion, i.e. no strong *a priori* hypotheses exist about which causal links or sub-networks should be expected and/or studied. However, typical neuromonitoring data are often characterized by data paucity as compared to the number of channels necessary for whole-brain coverage. While fMRI represents an extreme case of this problem (with hundreds of thousands of voxels, i.e. signal sources, for a few thousand data points), neuroelectric signals like EEG or MEG (which are commonly recorded

with much higher frequencies as compared to fMRI) are not immune from the same issue. Indeed, signal nonstationarity and the high density of MEG and EEG artifacts often mandates estimation within short signal epochs. At the same time the physiological relevance of long-range correlations in brain signals is becoming more and more evident [7–11].

In the context of large multivariate systems, the classical multivariate autoregressive Granger causality-based formulation (MVAR-GC) may therefore suffer from important limitations due to the approximate quadratic dependence of the number of parameters on the system's dimensionality, especially in relation to the number of available datapoints. In these cases, approaches like e.g. "partial conditioning" [12–14], may be more relevant/useful than full conditioning. Additionally, methods dealing with data redundancy, possibly determined by an influence that is external to the system, have been devised [15]. Another important aspect is the possible presence of long range correlations (see above) and/or putative non-instantaneous, possibly delayed influences between regional brain signals. Here, approaches able to estimate or take into account characteristic interaction at different time-scales have been proposed, based on e.g. state-space [16,17] or on wavelet formulations [18]. When such long-range correlations are present, the optimal autoregressive order in MVAR estimation (and hence the number of model parameters) will have to grow further in order to be able to capture enough information from the system's past to faithfully represent signal dynamics.

A number of analytical and algorithmic methods have been targeted specifically to the reduction of model parameters. Such approaches are generally based on strategies to optimally select a subset of possible alternative models, and employ more or less sophisticated search strategies such as stepwise methods [19,20], genetic algorithms [21], particle swarm optimization [22,23], or least absolute shrinkage and selection operators (LASSO) [24–26]. Still, the complexity of these methods decreases generality, calls for application-specific optimization and generally comes with high computational demands.

Other approaches have tackled the problem from a data reduction perspective, i.e. rather than shrinking model parameter space, signal space is compressed / embedded into a component subspace which is assumed to represent most of the non-redundant information contained in the original signals [2,27]. While this may lead to efficient model estimation, when applying such approaches the estimated causal relationships will be related to this reduced network of (latent) components rather than to the original signals, possibly hampering interpretability.

In view of the widespread application of Granger causality to biosignals and neuroscience in general, in the rest of this paper we will refer to "causality" in the classical Granger-sense. Still, it is important to note that the definition of causality provided by Granger is still controversial, and that there is a lively scientific debate about (possibly) alternative definitions and ideas about causal inference based on more recent and/or sophisticated frameworks. As an example, the ideas introduced though Pearl's interventionist do-calculus [28] (a leading concept of contemporary philosophy in which causality is represented through the concepts of action-intervention-manipulation) recently facilitated the development of modern approaches to estimating causality based on e.g. symbolic dynamics and algorithmic probability [29]. These applications are particularly relevant when e.g. developing a machine-oriented tool [30] based on reasonable functional assumptions stemming from structural equation models (like e.g. in econometrics or molecular biology - please see Discussion for additional details).

In this study, we aimed to overcome the above issues while remaining within the realm of linear models, for which parameter estimation remains a convex problem. To this end, we introduce the use of a Wiener-Volterra decomposition with Laguerre polynomials as basis functions. This choice allows to build parsimonious MVAR models [31–35] while capturing arbitrarily long past dependencies without increasing the number of parameters. Specifically, the orthonormal basis of the discrete-time Laguerre functions expand the Wiener-Volterra kernels, accounting for the long-term information in correlated time series with a possible heterogeneous delay structure in their interdependence [31,33,35]. We then reformulate classic causality estimators in terms of this decomposition, validate our method in
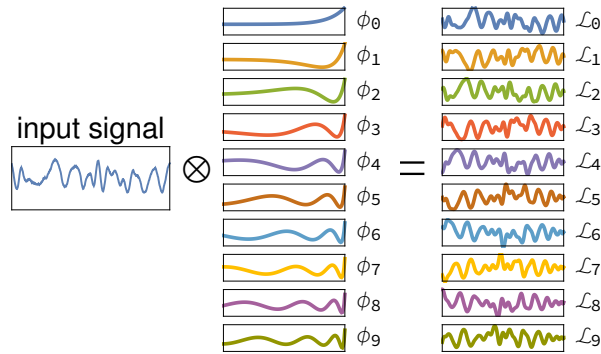
Figure 1. Example of the use of Laguerre basis functions in compactly representing signals: the input signal x_i is convolved in time with functions ϕ_m to obtain the function $\mathcal{L}_i^m$, which is then used in autoregressive modeling

synthetic data and report an example application MEG data drawn from the large Human Connectome Project (HCP) database.

2. Theory

When formulating the problem of detecting causality from variable Y_t to variable X_t ($Y \rightarrow X$) one tests the null hypothesis that the knowledge about the past of Y_t provides no added prediction power when forecasting the future of X_t. We reformulate the classical "restricted" model (RM) and "unrestricted" model (UM) which are at the basis of causality estimation using a Laguerre expansion. We call this approach Laguerre-based globally conditioned (LGC) causality (an earlier version has been presented in [36]). The RM for X_t, includes the past of X_t itself and Z_t. The latter term includes all other variables except for Y_t (i.e. the model is "conditioned"). Further, the UM includes all variables X_t, Y_t, Z_t [4]. Both models are fitted using MVAR systems which, in this paper, are defined over the components of a Volterra-Wiener expansion with Laguerre polynomials:

$$x_{t+1} = \sum_{k=1}^{p} a_k \mathcal{L}_t^{(k)}(x) + c_k \mathcal{L}_t^{(k)}(z) + \varepsilon_t \qquad \text{(RM)} \qquad (1)$$

$$x_{t+1} = \sum_{k=1}^{p} \tilde{a}_k \mathcal{L}_t^{(k)}(x) + \tilde{b}_k \mathcal{L}_t^{(k)}(y) + \tilde{c}_k \mathcal{L}_t^{(k)}(z) + \tilde{\varepsilon}_t \qquad \text{(UM)} \qquad (2)$$

(see Figure 1) where p is the autoregressive model order, a, $\tilde{a}$, $\tilde{b}$, c, $\tilde{c}$ are autoregressive coefficients, ε, $\tilde{\varepsilon}$ are uncorrelated white noise processes, and the discrete-time Volterra-Wiener decomposition with Laguerre polynomials $\mathcal{L}^{(m)}(\cdot)$ over the discrete time signal x_t is defined as:

$$\mathcal{L}_t^{(m)}(x) = \sum_{n=0}^{N} \phi_m(n)(x_{t-n}) \qquad (3)$$

where the mth-order, discrete time Laguerre polynomial $\phi_m(n)$ is defined as [34,37]:

$$\phi_m(n) = \alpha^{\frac{n-m}{2}} (1-\alpha)^{\frac{1}{2}} \sum_{j=0}^{m} (-1)^j \binom{n}{j} \binom{m}{j} \alpha^{m-j} (1-\alpha)^j . \qquad (4)$$

Here, the parameter α ($0 \leq \alpha < 1$) determines the rate of exponential asymptotic decay of $\phi_m(n)$ [38] (see Figure 1 for a visual example). Notably, the null-hypothesis that Y_t does not cause X_t (conditioned to Z_t), can then be expressed as the an equality between the expectation of x_t

$$E(x_t|Z_{t-k}) = E(x_t|Y_{t-k}, Z_{t-k}) \qquad (5)$$

and therefore rejected if the f_{ratio} of the residual sum of squares (RSS):

$$f_{\text{ratio}} = \frac{RSS_{\text{r}} - RSS_{\text{ur}}}{RSS_{\text{ur}}} \frac{N_{\text{obs}} - 2m}{m} \tag{6}$$

is extreme with respect of its parent distribution, i.e. the Fisher–Snedecor distribution with $N_{\text{obs}} - 2m$ and m degrees of freedom [3]. In this context, it is common practice to employ the logarithm of the ratio of average squared residuals as a measure of GC strength as follows [39]:

$$s_{j \to i} = \log\left(\frac{RSS_{\text{r}}}{RSS_{\text{un}}}\right) \, . \tag{7}$$

Also, the *p*-value related to the above hypotheses test can be expressed analytically if needed [40]. Still, the use of Eq.(7) has several advantages: i) in the limit of large strengths it is proportional to -$\log(p\text{-value})$, which, in turns is normality distributed with a mean proportional to the number of independent observations [41]; and ii) it can be easily be generalized to the case of multivariate GC [4,39].

3. Methods

In this paper we explore the performance of the above described LGC, and compare it to the classical MVAR-GC employing massive multivariate synthetic data simulation from random networks. In-house developed code employed for LGC is available at: https://github.com/andreaduggento/Laguerre-GrangerCausality.

3.1. Network generation

In order to compare the performances of LGC and MVAR-GC in detecting true causal connections within complex directed networks, we performed synthetic data simulation by generating data from families of 17-node, ground-truth networks described by a binary, zero-diagonal, asymmetrical adjacency matrix **A**, whose elements A_{ij} represent the direct influence of node j on node i. Pairs of nodes with bidirectional connections as well as loops are explicitly allowed. The total number of "edges" n_{e} depends on the desired network density d_{n} which, for a network with L nodes, is defined as $n_{\text{e}}(L(L-1))$. We generated graph families at 19 different densities, where the values of d_n are chosen to be approximately equidistant on a logarithmic scale between 0.01 and 0.5. For each value of d_{n} we generate 32 different random networks to account for possible fluctuations of our estimator with respect to network topology. Examples of the generated networks and their corresponding adjacency matrices for exemplary density values are shown in Fig. 2.

3.2. Synthetic simulations

Each node in the network families described above is assigned a system which is coupled to other nodes through the adjacency matrix described above (see below for details), and evolves as a stochastic variable characterized by long range correlations. The coupling terms are characterized through integral relationships which include random delays (see below). Each stochastic variable x_i is a sum of three components: i) a noise term which exhibits long-range correlation behaviour ξ_i; ii) a cubic integral term; iii) a coupling term which feeds from other nodes:

$$x_i(t) = d\xi_i(t) - h\left(\int_{t-h}^{t} \frac{x_i^3(s)}{3} ds + c_{ij}\left(\eta_j(t - \tau_j) - x_i(t)\right)\right) \tag{8}$$

where τ_j is a delay term which is randomly sampled at each realization from a uniform distribution in the range 0-12 s, $d = 0.1$, and the time step is set to $h = 0.25$ s. Each coupling coefficient c_{ij} is defined by a global, overall coupling strength $w = 0.2$ as well as by the the ground-truth adjacency matrix **A**.
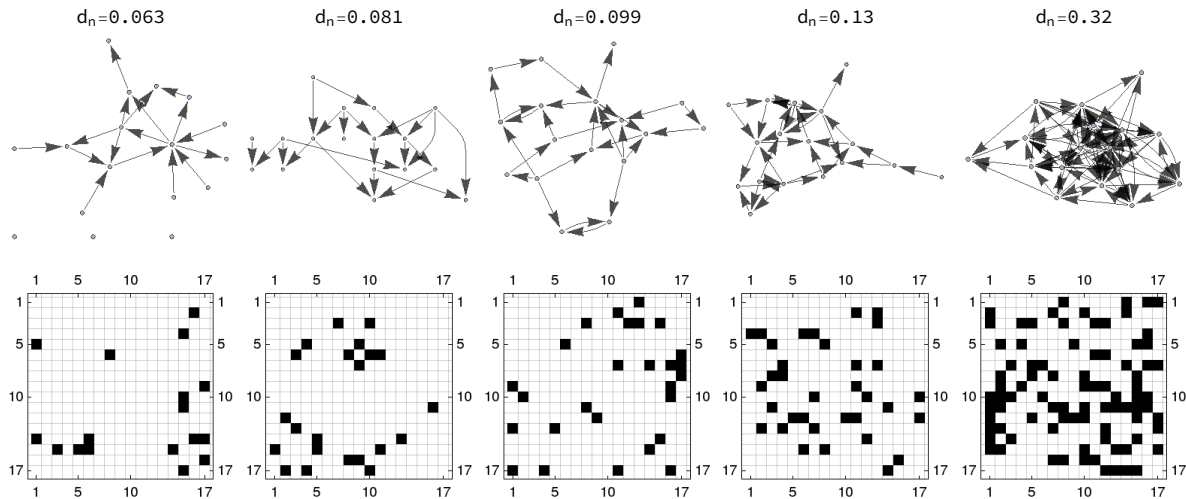
Figure 2. Example networks (for a subset of density values) used to generate synthetic data. Top row: topological directed representation; bottom row: corresponding adjacency matrices (white fields are zeros, black fields are ones).

Specifically, for the i-th node, if for all j $A_{ij} = 0$, then $c_{ij} = 0$; otherwise c_{ij} is equal to w normalized by the number of incoming connections $c_{ij} = w / \sum_j A_{ij}$.

The noise term is the main determinant of the the statistical properties of x_i and it is designed to include long-range correlation properties [42]

$$\xi_i(t) = \sum_{l=1}^{w} -a_l \xi_i(t-l) + d\epsilon_i(t) \tag{9}$$

$$a_l = \left(l - 1 - \frac{\beta}{2}\right)\frac{a_{l-1}}{l} \tag{10}$$

where w is the noise-model size (set to $w = 3$ in this paper), $a_0 = 1$, $\epsilon(t) \sim \mathcal{N}(0,1)$ is a normally distributed random variable, and β characterizes the noise spectrum. The latter follows a power law $1/f^\beta$ and does not have a characteristic timescale [43]. Also, the cubic integral term x^3 guarantees that the variable x_i remains symmetrically bounded around 0. Finally, the coupling term $c_{ij}\left(\eta_j(t - \tau_j) - x_i(t)\right)$ is defined by the coupling coefficient c_{ij} and an integral term delayed by τ_j for of the jth variable:

$$\frac{d}{dt}\eta_j(t) = -\eta_j/\tau_0 + bx_j. \tag{11}$$

In our simulations we set the time decay constant $\tau_0 = 8$ and $b = 0.1$. Each network was evolved for a total of 10000 time-points, and the generated multivariate time series were employed for caulsality analysis, using both the LGC and MVAR-GC approaches. For the MVAR-based GC method, the autoregressive model order was chosen according to the Akaike information criterion [44] resulting in an optimal order $p = 10$. The autoregressive order was employed to model all variables of the system. An example of signal generated from Eq.s (8-11) is presented in Fig.(3). For each set of multivariate signals (i.e. for each network realization) and corresponding estimated adjacency matrix, the ability of LGC and MVAR-GC to detect causal links (i.e. belonging to the ground truth network) while rejecting false causal links was quantified as the area under the receiver operating characteristic (ROC) curve (AUC). The AUC was computed by varying the threshold in causality strength $s_{j \to i}$ in Eq.(7) which determined, for every pair of nodes i and j, whether their causal connection should be accepted as "true" or rejected. Additionally, estimation was performed while varying parameter α within the interval $\alpha \in [0, 0.7]$ in order to investigate the relationship between α and AUC.
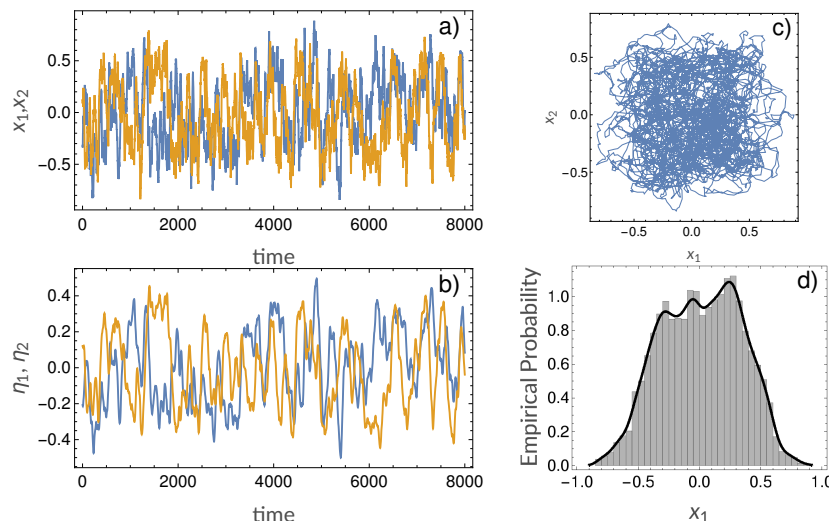
Figure 3. Example realizations of our model system: a) example signals x_i; b) example signals η_i; c) phase plot of x_1 vs x_2; d) typical distribution of x_i.

3.3. Directed brain connectivity estimation in MEG data

As an example application of our method to biological signlas, we use resting-state magnetoencephalography (rMEG) data from 89 subjects made available by the Human Connectome Project (HCP) [45] as part of the S1200 release (https://www.humanconnectome.org/study/hcp-young-adult/document/1200-subjects-data-release).

Every subject included in this dataset underwent 3 sessions of 6-minutes rMEG scans, along with physiological data were also acquired, including electrooculography (EOG), electrocardiography (ECG), and electromyography (EMG). These data were employed by the HCP consortium in denoising and preprocessing the data. In brief, the MEG preprocessing pipeline begins with several steps of data sanity check and quality assurance tasks (see https://www.humanconnectome.org/storage/app/media/documentation/s1200/HCP_S1200_Release_Reference_Manual.pdf for details).

Data were purged of artifacts resulting in excessive signal amplitude, and "bad" channels were selected through the estimation of the correlation between each channel and its neighbors [46]. Raw MEG data was then separated into brain-related and noise components through an Independent Component Analysis (ICA) pipeline which included an automatic classification step using a manually trained expert classifer. The ICA pipeline included i) band-pass filters; ii) ECG and EOG preprocessing; iii) iterative ICA decompositions (20 iterations of FastICA with deflation approach run from different initializations); iv) power spectrum density and power time course estimation; v) parameter estimations to allow classification of environmental and instrumental artifacts. Successively, source estimation is performed using a single shell volume conduction model defined in a MEG-system based head coordinates obtained by segmentation of the anatomical magnetic resonance imaging (MRI) scan for each subject. Finally, time-varying estimates of the band-limited (8 frequency bands) power envelope of each source are generated (final signal frequency: 40 Hz). The frequency ranges for each band are summarized in Table 1. Successively, the power envelopes are aggregated according to the 17 cortical parcels described in [47]. This resulted in 136 (17 parcels x 8 bands) 6-minute power envelopes per subjects. The physiological interpretation of the functional networks/parcels employed are summarized in Table 2. Given the notable skewness of the data, each power envelope was log-transformed prior to multivariate LGC estimation. The latter was performed subject- and band- wise, after which the 3rd percentile in strength across all subject was extracted for visualization purposes.

Frequency band name	Frequency band ranges
delta	[1.3 4.5] Hz
theta	[3 9.5] Hz
alpha	[6.3 16.5] Hz
beta low	[12.5 29] Hz
beta high	[22.5 39] Hz
gamma low	[30 55] Hz
gamma mid	[45 82] Hz
gamma high	[70 125] Hz

Table 1. Frequencies band resulting from MEG preprocessing pipeline executed by the HCP consortium

	Network name	Physiological interpretation
1	VIS-1	Visual
2	VIS-2	
3	MOT-1	Motor
4	MOT-2	
5	DAN-2	Dorsal Attention
6	DAN-1	
7	VAN-1	Ventral Attention
8	FP-1	Frontoparietal
9	LIM-1	Limbic
10	LIM-2	
11	FP-2	Frontoparietal
12	FP-3	
13	FP-4	
14	MOT-3	Motor
	DMN-3	Default Mode Network
16	DMN-1	
17	DMN-2	

Table 2. Legend of the 17 functional network from Yeo resting state network map along with their physiological interpretation.

4. Results

Figure 4 shows an example of the comparison between the ROC curves obtained when using LGC vs MVAR-GC for density $d_n = 0.0625$ (approximately 0.9 vs 0.75 in AUC), as well as the gain in performance (i.e. AUC) obtained when using LGC as opposed to classical MVAR-GC as a function of network density and α. For this system, a qualitatively optimal (in the sense of performance gain over MVAR-GC) region of α centered around α=0.6 emerges almost across the whole density range.

Additionally, Figure 5 shows AUC values (along with interquartile ranges) as a function of density for both LGC and MVAR-GC. Here, the value for α was chosen according to Fig. 4. Performance gain is also shown on the right as a function of density. As expected, the discrimination performances degrade monotonically when network density (and therefore the complexity of the problem) increases. Both classical MVAR-GC and LGC reach AUC=1 (100% accuracy) for very low network densities, and both methods approach AUC=0.5 (chance-level discrimination accuracy) for very densely connected networks. However, for every density tested, LGC performs better than MVAR-GC, and its discrimination performance appears to degrade less steeply as network density increases. The maximum performance gain is approximately 0.17 (difference in AUC), which is attained at a network density of approximately 0.1 (which, incidentally, corresponds to a typical threshold values employed in connectomic studies). Finally, Fig. 6 shows the results of employing LGC to quantify the directed, MEG-based connectome in the high quality HCP sample. For each frequency band, only the top 3% connections (median strength across subjects) are shown. The 3% threshold was arbitrary chosen in
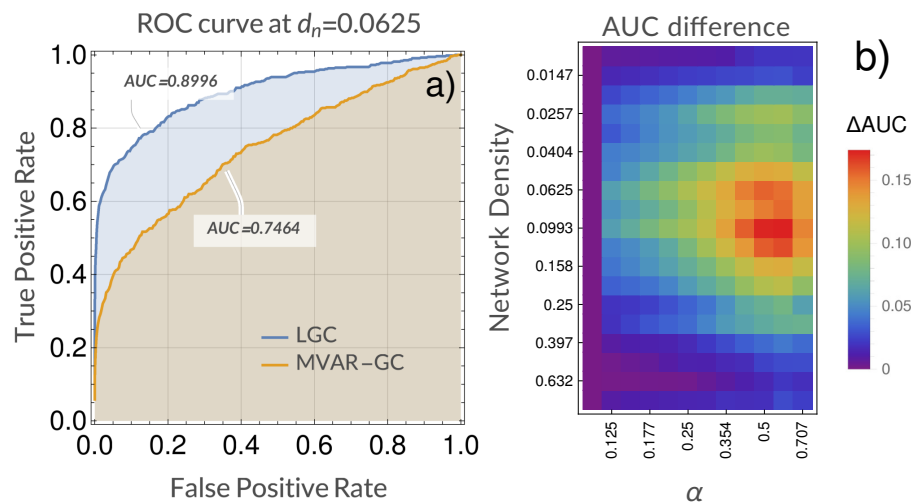
Figure 4. Comparison of detection performance between MVAR-GC and LGC for our model system. a) Example AUCs resulting from using both classical MVAR-GC and LGC (α=0.595) for a single network density. ROC curves shown on the left were built over the prediction of $(17^2 - 17) \times 32 = 8704$ links relative to 32 random networks with density $d_\mathrm{n} = 0.0625$. b) Difference between AUCs (defined as ΔAUC=AUC(LGC) - AUC(MVAR-GC) as a function of network density and α. As in the ROC curves on the left, every ΔAUC value (corresponding to every pair of density and α values) in the figure on the right is built over all links belonging to 32 random networks.

order to compromise between including physiologically-relevant information while minimizing the number of possibly false positive causal links.

5. Discussion and Conclusion

In this paper, we have proposed a method to compactly represent and model signals in a linear autoregressive framework through a novel orthogonal basis based on Laguerre polynomials. This novel formulation allows to attain model parsimony while retaining the ability to represent long-range correlation, a fundamental requirement when modeling high-frequency brain signals. This approach was employed to revise classical, MVAR model based Granger causality and validated in large scale simulation using synthetic data generated from families of complex networks at varying densities. A clear advantage of the LGC methods was shown across all densities. This is in line with the idea that Laguerre polynomials are smooth basis functions, able to capture damped multiple-frequency oscillations with fewer parameters when compared to classical MVAR models. In order to mimic the complexity of biological signals, our synthetic simulations were based on a nonlinear system with integral couplings as well as random delays, however without the presumption that it could serve as a thorough synthetic generative model for MEG data. It is thus possible to hypothesize that one of the several nonlinear causality estimation methods presented in the literature [48–52] could deliver even better performance. However, such methods require articulate parameter optimization, and their identification may not be univocal. It should be noted that the advantages of LGC with respect of MVAR-GC may only apply to certain situations, e.g. when significant delays or slow couplings at multiple timescales are present. However, as mentioned above, this is very often the case with physiological signals derived from the brain as well as from other physiological subsystems. Also, interestingly, MVAR system defined through Laguerre base functions can be simply viewed as a generalization of the classical MVAR system; indeed, the latter is simply recovered by setting $\alpha = 0$. In this sense, this work could be viewed as proposing and extension to classical, linear multivariate causality by introducing the additional choice of the optimal α (with $0 \leq \alpha < 1$), which only introduces one additional degree of freedom with a large gain in signal representation ability. In
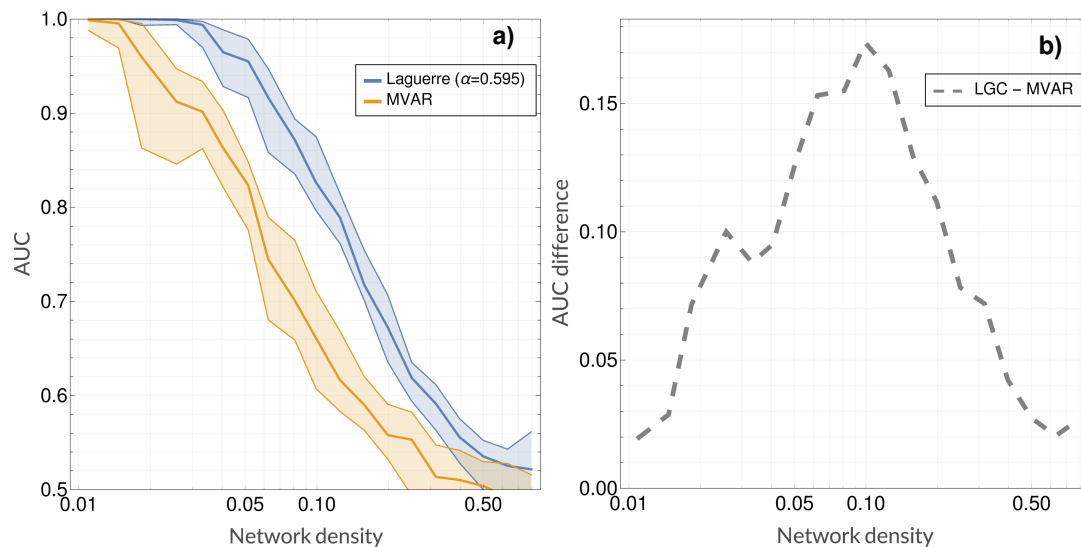
Figure 5. median AUC values (along with interquartile ranges calculated across 32 random networks for each density) as a function of density for both LGC and MVAR-GC. Here, the value for α was chosen according to 4. Performance gain is also shown on the right as a function of density.
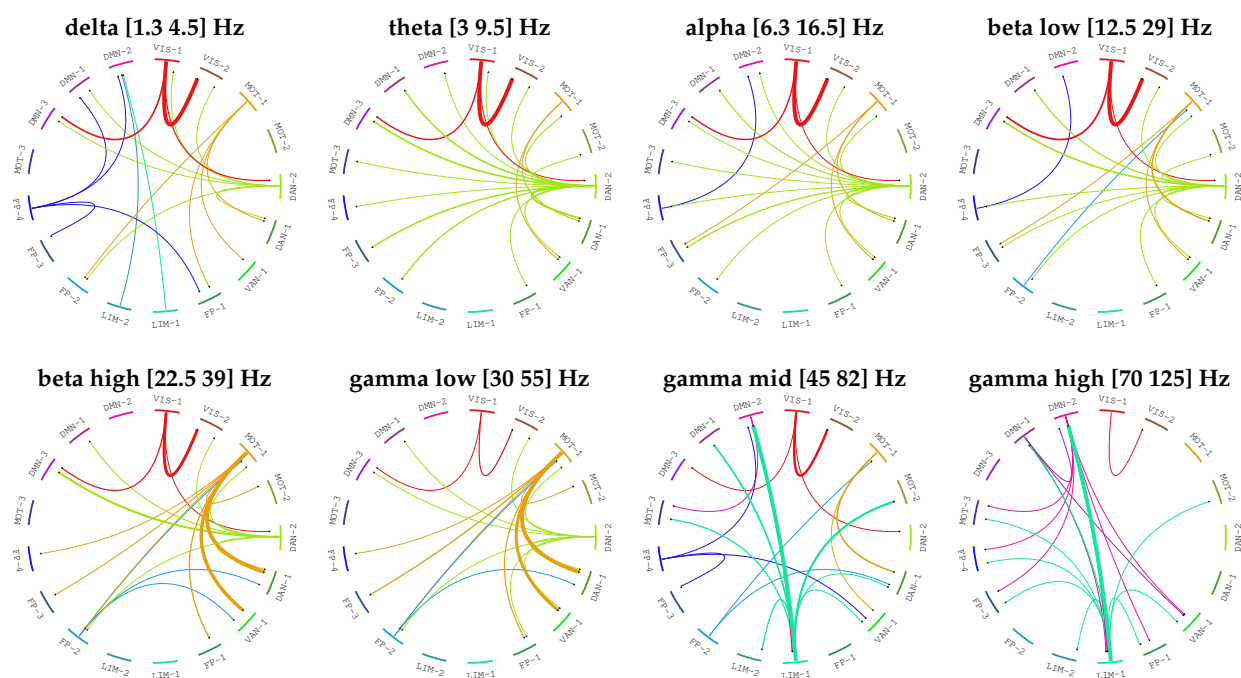
Figure 6. Results of employing LGC to quantify the directed, MEG-based connectome in the high quality HCP sample. For each band, only the top 3% connections (median strength across subjects) are shown. Every edge (i.e. connection) is colored according to the node (i.e. network) which is driving the other node. The width of each edge if proportional to the strength of the connection.

this sense, introducing an $\alpha>0$ could be viewed as an attractive alternative to incrementing the model autoregressive order p.

It is important to note that Granger-style approach might not be the most suitable tool to accurately investigate the complex causal interplay between the phenomena underlying observed dynamics, and that recent literature points towards a revised definition fo the term 'causal' which does not necessarily include Granger approaches. For example, recently the ϵ-transducer has been exploited to provide a

structural representation of the minimal optimal predictor of one process from another [53,54]. Also, based on the recent theoretical formulation of Partial Information Decomposition (PID) by Williams and Beer [55] the scientific community is currently debating what the most appropriate definitions of redundant, unique and synergistic components of mutual information should be (for a review, see [56,57] and references therein). While redundancy measures a form of equivalence [58], alternative definitions have emerged both for shared information [59–61] and for unique information [62]. It is interesting to note that several recent tool which have been presented in order to tackle the ideas of redundancy and synergy in information within time series rely on autoregressive representations, and may therefore benefit from the representational basis we have presented in this paper in order to minimize parameter count at implementation time

Following synthetic data validation, we applied LGC to estimating the directed connectome of the human brain through a specific 17-dimensional parcellation of a high-quality MEG dataset provided by the HCP consortium. This was intended as a data-driven, whole brain exploration of the directed, MEG-based connectome of the human brain. In view of the peculiar characteristics of the MEG signal (presence of long-range correlations), this application is well matched to the circumstances for which our method was designed. Also, while this approach, like any time-series analysis tool applied to biological systems, can be susceptible to higher order biological as well as data-driven confounders, it serves as a discovery tool which can help formulate more specific hypotheses (e.g. to be tested in targeted, tasks-based paradigms). Additionally, it is important to note that a) each data channel employed in this paper contained approximately 14000 data points, ensuring the feasibility of full conditioning, and b) given the dangers of incurring in false positives when performing whole-brain, data driven explorations, only the top 3% connections resulting from our full multivariate analysis are presented and interpreted. Although the neurobiological basis and neuroanatomical localization of the different MEG bands remains a matter of debate, we found that our results are highly consistent with previous findings [63] showing that the theta band localizes to dorsal prefrontal networks, while the alpha and beta band respectively relate to neuronal activity in posterior occipital regions and motor cortices. The origin of gamma bands is controversial but has been linked to prefrontal networks and cognitive control [63].

More importantly, our data clearly show that the key brain regions previously implicated in specific neuronal rhythms are the same areas that causally drive the connectivity in other circuits ('nodes') in each distinct band. For example, the primary occipital network (VIS-1) dominates the casually-driven connections in the alpha band, although an effect of the dorsal attentional network is also evident for the alpha and theta rhythms. Likewise, the motor circuit (MOT-1) directs the MEG-derived functional connectivity in ventral and dorsal attentional networks in the beta frequency domain. Increased beta activity over the motor cortex have been associated to voluntary movement suppression and can be dependent on the experimental procedures adopted in the study (i.e. resting-state MEG).

Finally, it is of particular interest to note the preponderant role of a limbic network (LIM-1) in mediating the activity in other circuits (especially the default mode network) at the gamma bands, particularly the highest frequencies. Previous studies have localized the origin of gamma bands to prefrontal regions and have suggested that gamma rhythms would reflect a synchronized firing of several circuits that would be critical to associate internal inputs (e.g. memories) with external (e.g. visual) ones to form coherent perception. Our findings support this idea by showing that the directed activity from limbic circuits to areas within the default mode network might play an important role in such high-level cognitive associative functions.

In conclusion, we have extended classical MVAR-GC to incorporate long-range past information and coupling while retaining model simplicity and linearity. In an application to real-world brain data, we have shown how our method is able to reproduce current knowledge as well as to uncover novel directed influences between brain regions. While the latter observations warrant additional validation through specific task-based investigation, in general we have shown that the LGC method

is able to detect in vivo functional interactions and causal dynamics across multiple neural networks while delivering superior performance as compared to classical, linear MVAR-based Granger causality methods.

Conflicts of Interest: The authors declare no personal, professional or financial relationships that could potentially be construed as a conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AUC	Area Under ROC Curve
ECG	Electrocardiography
EEG	Electroencephalography
EMG	Electromyography
EOG	Electrooculography
fMRI	functional Magnetic Resonance Imaging
HCP	Human Connectome Project
ICA	Independent Component Analysis
LASSO	Least Absolute Shrinkage and Selection Operator
LGC	Laguerre Granger Causality
MEG	Magnetoencephalography
MVAR-GC	MultiVARiate Granger Causality
RM	Restricted Model
rMEG	resting-state Magnetoencephalography
ROC	Receiver Operating Characteristic
RSS	Residual Sum of Squares
UM	Enrestricted Model

Acknowledgements Luca Passamonti is funded by the Medical Research Council (MRC) (MR/P01271X/1) at the University of Cambridge

Author Contributions: Conceptualization, Andrea Duggento and Nicola Toschi; Data curation, Salvatore Nigro and Maria Giovanna Bianco; Formal analysis, Andrea Duggento; Funding acquisition, Nicola Toschi and Luca Passamonti; Investigation, Andrea Duggento; Methodology, Andrea Duggento and Gaetano Valenza; Project administration, Nicola Toschi; Resources, Nicola Toschi; Software, Salvatore Nigro and Maria Giovanna Bianco; Supervision, Nicola Toschi; Validation, Andrea Duggento and Luca Passamonti; Visualization, Andrea Duggento; Writing – original draft, Andrea Duggento, Luca Passamonti and Nicola Toschi; Writing – review and editing, Andrea Duggento, Gaetano Valenza, Luca Passamonti, Salvatore Nigro, Maria Giovanna Bianco, Maria Guerrisi, Riccardo Barbieri and Nicola Toschi.

1. Astolfi, L.; Cincotti, F.; Mattia, D.; Babiloni, C.; Carducci, F.; Basilisco, A.; Rossini, P.; Salinari, S.; Ding, L.; Ni, Y.; others. Assessing cortical functional connectivity by linear inverse estimation and directed transfer function: simulations and application to real data. *Clinical neurophysiology* **2005**, *116*, 920–932.

2. Deshpande, G.; LaConte, S.; James, G.A.; Peltier, S.; Hu, X. Multivariate Granger causality analysis of fMRI data. *Human brain mapping* **2009**, *30*, 1361–1373.

3. Bressler, S.L.; Seth, A.K. Wiener–Granger causality: a well established methodology. *Neuroimage* **2011**, *58*, 323–329.

4. Seth, A.K.; Barrett, A.B.; Barnett, L. Granger causality analysis in neuroscience and neuroimaging. *J Neurosci* **2015**, *35*, 3293–7. doi:10.1523/JNEUROSCI.4399-14.2015.

5. Duggento, A.; Passamonti, L.; Valenza, G.; Barbieri, R.; Guerrisi, M.; Toschi, N. Multivariate Granger causality unveils directed parietal to prefrontal cortex connectivity during task-free MRI. *Scientific reports* **2018**, *8*, 5571.

6. Anzolin, A.; Presti, P.; Van De Steen, F.; Astolfi, L.; Haufe, S.; Marinazzo, D. Quantifying the effect of demixing approaches on directed connectivity estimated between reconstructed EEG sources. *Brain topography* **2019**, pp. 1–20.

7. Botcharova, M.; Berthouze, L.; Brookes, M.J.; Barnes, G.R.; Farmer, S.F. Resting state MEG oscillations show long-range temporal correlations of phase synchrony that break down during finger movement. *Frontiers in physiology* **2015**, *6*, 183.

8. Palva, J.M.; Zhigalov, A.; Hirvonen, J.; Korhonen, O.; Linkenkaer-Hansen, K.; Palva, S. Neuronal long-range temporal correlations and avalanche dynamics are correlated with behavioral scaling laws. *Proceedings of the National Academy of Sciences* **2013**, *110*, 3585–3590.

9. Bornas, X.; Fiol-Veny, A.; Balle, M.; Morillas-Romero, A.; Tortella-Feliu, M. Long range temporal correlations in EEG oscillations of subclinically depressed individuals: their association with brooding and suppression. *Cognitive neurodynamics* **2015**, *9*, 53–62.

10. Nikulin, V.V.; Brismar, T. Long-range temporal correlations in electroencephalographic oscillations: relation to topography, frequency band, age and gender. *Neuroscience* **2005**, *130*, 549–558.

11. Nolte, G.; Aburidi, M.; Engel, A.K. Robust calculation of slopes in detrended fluctuation analysis and its application to envelopes of human alpha rhythms. *Scientific reports* **2019**, *9*, 6339.

12. Guo, S.; Seth, A.K.; Kendrick, K.M.; Zhou, C.; Feng, J. Partial Granger causality—eliminating exogenous inputs and latent variables. *Journal of neuroscience methods* **2008**, *172*, 79–93.

13. Marinazzo, D.; Pellicoro, M.; Stramaglia, S. Causal information approach to partial conditioning in multivariate data sets. *Computational and mathematical methods in medicine* **2012**, *2012*.

14. Stramaglia, S.; Cortes, J.M.; Marinazzo, D. Synergy and redundancy in the Granger causal analysis of dynamical networks. *New Journal of Physics* **2014**, *16*, 105003.

15. Lizier, J.T.; Rubinov, M. Multivariate construction of effective computational networks from observational data. Preprint, 2012.

16. Faes, L.; Nollo, G.; Stramaglia, S.; Marinazzo, D. Multiscale granger causality. *Physical Review E* **2017**, *96*, 042150.

17. Faes, L.; Pereira, M.A.; Silva, M.E.; Pernice, R.; Busacca, A.; Javorka, M.; Rocha, A.P. Multiscale information storage of linear long-range correlated stochastic processes. *Physical Review E* **2019**, *99*, 032115.

18. Stramaglia, S.; Bassez, I.; Faes, L.; Marinazzo, D. Multiscale Granger causality analysis by à trous wavelet transform. 2017 7th IEEE International Workshop on Advances in Sensors and Interfaces (IWASI). IEEE, 2017, pp. 25–28.

19. Scheines, R.; Spirtes, P.; Glymour, C.; Meek, C.; Richardson, T. The TETRAD project: Constraint based aids to causal model specification. *Multivariate Behavioral Research* **1998**, *33*, 65–117.

20. Pearl, J. *Causality*; Cambridge university press, 2009.

21. Balcombe, K.G. Model selection using information criteria and genetic algorithms. *Computational Economics* **2005**, *25*, 207–228.

22. Yu, S.; Wang, K.; Wei, Y.M. A hybrid self-adaptive Particle Swarm Optimization–Genetic Algorithm–Radial Basis Function model for annual electricity demand prediction. *Energy Conversion and Management* **2015**, *91*, 176–185.

23. Huang, C.M.; Huang, C.J.; Wang, M.L. A particle swarm optimization to identifying the ARMAX model for short-term load forecasting. *IEEE Transactions on Power Systems* **2005**, *20*, 1126–1133.

24. Shojaie, A.; Michailidis, G. Discovering graphical Granger causality using the truncating lasso penalty. *Bioinformatics* **2010**, *26*, i517–i523.

25. Tang, W.; Bressler, S.L.; Sylvester, C.M.; Shulman, G.L.; Corbetta, M. Measuring Granger causality between cortical regions from voxelwise fMRI BOLD signals with LASSO. *PLoS computational biology* **2012**, *8*, e1002513.

26. Furqan, M.S.; Siyal, M.Y. Random forest Granger causality for detection of effective brain connectivity using high-dimensional data. *Journal of integrative neuroscience* **2016**, *15*, 55–66. doi:10.1142/s0219635216500035.

27. DSouza, A.M.; Abidin, A.Z.; Leistritz, L.; Wismüller, A. Exploring connectivity with large-scale Granger causality on resting-state functional MRI. *Journal of neuroscience methods* **2017**, *287*, 68–79.

28. Pearl, J. *Causality: Models, Reasoning and Inference*; Cambridge University Press, 2000.

29. Solomonoff, R.J. A formal theory of inductive inference. *Information and control* **1964**, *7*, 1–22.

30. Zenil, H.; Kiani, N.A.; Zea, A.A.; Tegnér, J. Causal deconvolution by algorithmic generative models. *Nature Machine Intelligence* **2019**, *1*, 58.

31. Valenza, G.; Citi, L.; Barbieri, R. Instantaneous nonlinear assessment of complex cardiovascular dynamics by laguerre-volterra point process models. Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE. IEEE, 2013, pp. 6131–6134.

32. Akay, M. *Nonlinear Biomedical Signal Processing Vol. II: Dynamic Analysis and Modeling*; Wiley-IEEE Press, 2000.

33. Valenza, G.; Citi, L.; Scilingo, E.P.; Barbieri, R. Point-process nonlinear models with laguerre and volterra expansions: Instantaneous assessment of heartbeat dynamics. *Signal Processing, IEEE Transactions On* **2013**, *61*, 2914–2926.

34. Marmarelis, V. Identification of nonlinear biological system using Laguerre expansions of kernels. *Ann. Biomed. Eng.* **1993**, *21*, 573–589.

35. Valenza, G.; Citi, L.; Scilingo, E.P.; Barbieri, R. Using Laguerre expansion within point-process models of heartbeat dynamics: a comparative study. Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE. IEEE, 2012, pp. 29–32.

36. Duggento, A.; Valenza, G.; Passamonti, L.; Guerrisi, M.; Barbieri, R.; Toschi, N. Reconstructing multivariate causal structure between functional brain networks through a Laguerre-Volterra based Granger causality approach. 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2016, pp. 5477–5480.

37. Wahlberg, B. System identification using Laguerre models. *IEEE Transactions on Automatic Control* **1991**, *36*, 551–562.

38. Watanabe, A.; Stark, L. Kernel method for nonlinear analysis: Identification of a biological control system. *Mathematical Biosciences* **1975**, *27*, 99–108.

39. Geweke, J.F. Measures of conditional linear dependence and feedback between time series. *Journal of the American Statistical Association* **1984**, *79*, 907–915.

40. Geweke, J. Measurement of Linear Dependence and Feedback between Multiple Time Series. *Journal of the American Statistical Association* **1982**, *77*, 304–313. doi:10.1080/01621459.1982.10477803.

41. Lambert, D.; Hall, W. Asymptotic lognormality of p-values. *The Annals of Statistics* **1982**, pp. 44–64.

42. Kasdin, N.J. Discrete simulation of colored noise and stochastic processes and 1/f/sup/spl alpha//power law noise generation. *Proceedings of the IEEE* **1995**, *83*, 802–827.

43. Nakamura, T.; Small, M.; Tanizawa, T. Long-range correlation properties of stationary linear models with mixed periodicities. *Physical Review E* **2019**, *99*, 022128.

44. Shibata, R. Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika* **1976**, *63*, 117–126.

45. Van Essen, D.C.; Smith, S.M.; Barch, D.M.; Behrens, T.E.; Yacoub, E.; Ugurbil, K.; Consortium, W.M.H.; others. The WU-Minn human connectome project: an overview. *Neuroimage* **2013**, *80*, 62–79.

46. Winter, W.R.; Nunez, P.L.; Ding, J.; Srinivasan, R. Comparison of the effect of volume conduction on EEG coherence with the effect of field spread on MEG coherence. *Statistics in medicine* **2007**, *26*, 3946–3957.

47. Thomas Yeo, B.; Krienen, F.M.; Sepulcre, J.; Sabuncu, M.R.; Lashkari, D.; Hollinshead, M.; Roffman, J.L.; Smoller, J.W.; Zöllei, L.; Polimeni, J.R.; others. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of neurophysiology* **2011**, *106*, 1125–1165.

48. Ancona, N.; Marinazzo, D.; Stramaglia, S. Radial basis function approach to nonlinear Granger causality of time series. *Physical Review E* **2004**, *70*, 056221.

49. Marinazzo, D.; Pellicoro, M.; Stramaglia, S. Kernel method for nonlinear Granger causality. *Physical review letters* **2008**, *100*, 144103.

50. Faes, L.; Nollo, G.; Porta, A. Information-based detection of nonlinear Granger causality in multivariate processes via a nonuniform embedding technique. *Physical Review E* **2011**, *83*, 051112.

51. Benhmad, F. Modeling nonlinear Granger causality between the oil price and US dollar: A wavelet based approach. *Economic Modelling* **2012**, *29*, 1505–1514.

52. Tank, A.; Covert, I.; Foti, N.; Shojaie, A.; Fox, E. Neural granger causality for nonlinear time series. *arXiv preprint arXiv:1802.05842* **2018**.

53. Barnett, N.; Crutchfield, J.P. Computational Mechanics of Input–Output Processes: Structured Transformations and the ϵ-Transducer. *Journal of Statistical Physics* **2015**, *161*, 404–451.

54. James, R.G.; Barnett, N.; Crutchfield, J.P. Information flows? A critique of transfer entropies. *Physical review letters* **2016**, *116*, 238701.

55.   Williams, P.L.; Beer, R.D. Nonnegative decomposition of multivariate information. *arXiv preprint arXiv:1004.2515* **2010**.

56.   Faes, L.; Marinazzo, D.; Stramaglia, S. Multiscale information decomposition: exact computation for multivariate Gaussian processes. *Entropy* **2017**, *19*, 408.

57.   Lizier, J.; Bertschinger, N.; Jost, J.; Wibral, M. Information decomposition of target effects from multi-source interactions: Perspectives on previous, current and future work, 2018.

58.   Barrett, A.B. Exploration of synergistic and redundant information sharing in static and dynamical Gaussian systems. *Physical Review E* **2015**, *91*, 052802.

59.   Harder, M.; Salge, C.; Polani, D. Bivariate measure of redundant information. *Physical Review E* **2013**, *87*, 012130.

60.   Bertschinger, N.; Rauh, J.; Olbrich, E.; Jost, J. Shared information–New insights and problems in decomposing information in complex systems. Proceedings of the European conference on complex systems 2012. Springer, 2013, pp. 251–269.

61.   Griffith, V.; Koch, C. Quantifying synergistic mutual information. In *Guided Self-Organization: Inception*; Springer, 2014; pp. 159–190.

62.   Bertschinger, N.; Rauh, J.; Olbrich, E.; Jost, J.; Ay, N. Quantifying unique information. *Entropy* **2014**, *16*, 2161–2183.

63.   Florin, E.; Baillet, S. The brain's resting-state activity is shaped by synchronized cross-frequency coupling of neural oscillations. *Neuroimage* **2015**, *111*, 26–35.