Time, frequency, and time-varying Granger-causality measures in neuroscience

Cekic, Sezen; Grandjean, Didier Maurice; Renaud, Olivier

How to cite

CEKIC, Sezen, GRANDJEAN, Didier Maurice, RENAUD, Olivier. Time, frequency, and time-varying Granger-causality measures in neuroscience. In: Statistics in Medicine, 2018, vol. 37, n° 11, p. 1910–1931. doi: 10.1002/sim.7621

This publication URL: https://archive-ouverte.unige.ch/unige:106649

Publication DOI: 10.1002/sim.7621

Received: 20 February 2017 | Revised: 22 December 2017 | Accepted: 8 January 2018

DOI: 10.1002/sim.7621

RESEARCH ARTICLE

WILEY Statistics in Medicine

Time, frequency, and time-varying Granger-causality measures in neuroscience

Sezen Cekic[1] | Didier Grandjean[2] | Olivier Renaud[1]

[1]Methodology and Data Analysis, Department of Psychology, University of Geneva, Geneva, Switzerland

[2]Neuroscience of Emotion and Affective Dynamics Lab, Department of Psychology, University of Geneva, Geneva, Switzerland

Correspondence
Sezen Cekic, University of Geneva, Boulevard du Pont d'Arve 40, CH-1211 Geneva 4, Switzerland.
Email: sezen.cekic@unige.ch; cekicsezen@hotmail.com

Funding information
Swiss National Science Foundation, Grant/Award Number: 100014_156493; Lifebrain, Grant/Award Number: 732592-H2020-SC1-2016-2017/H2020-SC1-2016-RTD

This article proposes a systematic methodological review and an objective criticism of existing methods enabling the derivation of time, frequency, and time-varying Granger-causality statistics in neuroscience. The capacity to describe the causal links between signals recorded at different brain locations during a neuroscience experiment is indeed of primary interest for neuroscientists, who often have very precise prior hypotheses about the relationships between recorded brain signals. The increasing interest and the huge number of publications related to this topic calls for this systematic review, which describes the very complex methodological aspects underlying the derivation of these statistics. In this article, we first present a general framework that allows us to review and compare Granger-causality statistics in the time domain, and the link with transfer entropy. Then, the spectral and the time-varying extensions are exposed and discussed together with their estimation and distributional properties. Although not the focus of this article, partial and conditional Granger causality, dynamical causal modelling, directed transfer function, directed coherence, partial directed coherence, and their variant are also mentioned.

KEYWORDS

Granger causality, nonstationarity, nonparametric estimation, review, spectral domain, time domain, transfer entropy, vector autoregressive

1 | INTRODUCTION

The investigation of the dynamical causal relationships between neuronal populations is a very important step towards the overall goal of understanding the links between functional cerebral aspects and their underlying brain mechanisms. This investigation requires statistical methods able to capture not only functional connectivities (eg, symmetrical relationships) but also, and probably more importantly, effective connectivities (eg, directional or causal relationships) between brain activities recorded during a specific task or stimuli exposure.

The Granger-causality concept relies on passing from causality to predictability. It was provided in the 1960s by the economist Clive Granger. According to Granger,[1] if a signal X is causal for another signal Y in the Granger sense, then the history of X should contain information that helps to predict Y above and beyond the information contained in the history of Y alone. It is the axiomatic imposition of a temporal ordering that allows us to interpret such dependence as causal: "The arrow of time imposes the structure necessary."[2, p. 139] The presence of this relation between X and Y will be referred to "Granger causality" throughout the text.

Granger[1] adapted the definition of causality proposed by Wiener[3] into a practical form, and since then, Granger causality has been widely used in economics and econometrics. It is however only during the past few years that it has become popular in neuroscience.

Since its causal nature relies on prediction, Granger causality does not necessarily mean "true causality." Indeed, 2 distinct problems may occur. If the the causal connection from the first to the second variable is completely mediated by a third variable, one might reject the null hypothesis of non-Granger causality between signals although manipulation of one of them would not change the other, which contradicts what "true causality" would have implied. This is called the sequential driving problem.[4] The second problem, called different delay driving, occurs when the first variable drives both the second and the third variables, but the driving of the second variable has a smaller delay than the driving of the third one. Then, samples of the history of the second variable contain information that helps predict future samples of the third variable. A bivariate analysis might spuriously reject the null hypothesis of non-Granger causality from the second to the third variable, because it cannot distinguish that this is an indirect causality scheme.[4]

More generally, Granger causality may also produce misleading results when the true causal relationship involves more variables than those that have been selected and so the accuracy of its causal interpretation relies on a suitable preliminary variable selection procedure.[5]

If we concentrate on just 2 signals, the problem is twofold: The first part is the choice of a suitable causality statistic that can easily be interpreted and that answers the question of interest. This said, the statistic needs to rely on a model that intrinsically includes this prediction or Granger-causality principle, and so the second part of the problem is to define and properly estimate this fundamental statistical model. A wrong statistical model indeed may lead to a wrong causality inference.

The scope of this article is to review and describe existing Granger-causality statistics in the time and frequency domains and then to focus on their time-varying extensions. We will describe existing estimation methods for time-varying Granger-causality statistics, in order to give the reader a global overview and some insight on the pertinence of using a given method depending on the research question and the nature of the data.

In Sections 4 and 5, we will present time and frequency-domain Granger-causality statistics in the stationary case. In Section 6, we will discuss their time-varying extensions in terms of time-varying causal model estimation. In Section 7, we will outline existing toolboxes allowing us to derive time-varying frequency-specific Granger-causality statistics and then discuss the limitations and the potential application of these statistics in neuroscience in Section 8.

There exist already a couple of very complete works reviewing the application of Granger causality to neuroscience data in time and frequency domain. Pereda et al[6] and Bressler and Seth,[7] for instance, reviewed the issue of the application of Granger causality to neural data in time and frequency domain, and Porta and Faes[8] added the issue of information-theoretic approaches for Granger-causality analysis and their specific applications in neuroscience. The principles of Granger-causality application in neuroscience and neuroimaging are surveyed in Seth et al,[9] and Kleinberg and Hripcsak[10] reviewed the core concepts in understanding and identifying causality, as well as graphical models and Granger-causality approaches for inference in health sciences. Amblard and Michel[11] reviewed the conceptual and theoretical links between Granger causality and directed information theory, and Müller et al[12] explained and reviewed some of the most important coupling measures (correlation, Granger causality–based tools, entropy-based techniques, nonlinear prediction measures, and symbolic dynamics) and classified them according to their origin and capabilities in the light of physiological analyses. Finally, Sameshima and Baccala[13] collect surveys of time, frequency, and time-variant approaches for Granger causal inference in neuroscience.

The scope of the present article, as opposed to these excellent works, is to give a systematic methodological review and objective criticism of existing methods that lead to time-varying Granger-causality statistics. The increasing interest reflected by the number of publications related to this topic in neuroscience justifies this literature review undertaken from a statistical viewpoint.

2 | OTHER EXISTING APPROACHES

Computing effective connectivity is a challenging task, and various methods have been proposed in order to solve this issue. First, the methods are based on intervention causality.[14] The underlying principle is that a cause-and-effect relationship should persist if the cause is manipulated without directly affecting any other variables, whereas any non-causal associations should disappear. Another proposed technique is the so-called dynamic causal modelling (DCM).[15,16] Dynamic causal modelling is a model-based Bayesian generalization of the "covariance structural equation modelling"

approach.[17] It assigns effective connection strengths to anatomical model that best match observed covariance structure based on nonlinear input-state-output systems and bilinear approximation of dynamic interactions. The DCM results strongly rely on prior connectivity specifications and especially on the assumption of stationarity. The lack of reference to the DCM methodology here is therefore explained by its unsuitability in the context of nonstationarity. Another approach is based on asymmetric coupling measures (see, for instance, Sugihara et al,[18] for an application of the "convergent cross mapping" approach in a context where Granger causality is not applicable). Finally, graph theory, especially the theory of directed graphs, is of special interest for computing effective connectivity, as it provides a framework for the analysis of the dependence structure of a time series.[19]

This article does not discuss either symmetric functional connectivity statistics such as correlation and coherence. The reader is referred to Delorme et al[20] and Pereda et al[6] for an overall review of these statistics in the time and frequency domains. This symmetric connectivity aspect is also very important and carries a lot of information, but its presentation is beyond the scope of this article, which proposes a review of all existing methods, allowing us to derive a time-varying Granger-causality statistic.

3 | STATIONARITY

Many Granger-causality models rely on the assumption that the system analysed is covariance stationary. Covariance stationarity (also known as weak- or wide-sense stationarity) requires that the first moment and the covariance of the system do not vary with respect to time.

A random process $\mathbf{Z}_t$ is covariance stationary if it satisfies the following restrictions on its mean function,

$$\mathbb{E}[\mathbf{Z}(t)] = m_{\mathbf{Z}}, \qquad \forall t \in \mathbb{R}, \tag{1}$$

and on its autocovariance function,

$$\mathbb{E}[(\mathbf{Z}(t_1) - m_{\mathbf{Z}})(\mathbf{Z}(t_2) - m_{\mathbf{Z}})] = C_{\mathbf{Z}}(t_1, t_2) = C_{\mathbf{Z}}(\tau), \text{ where } \tau = t_1 - t_2, \quad \forall t_1, t_2 \in \mathbb{R}. \tag{2}$$

The first property implies that the mean function $m_{\mathbf{Z}}$ is constant with respect to t. The second property implies that the covariance function depends only on the difference between t_1 and t_2. The variance is consequently constant as well.

4 | TIME-DOMAIN CAUSALITY

4.1 | General model

As mentioned in the introduction, Granger causality is based on prediction, and its fundamental axiom is that "the past and present may cause the future but the future cannot cause the past."[1] The origin of Granger-no-causality was stated by Wiener in 1956 and then adapted and defined into practical form by Granger. As we will see, Granger restates Wiener's principle in the context of autoregressive models.[1] In particular, the main idea lies in the fact that if a signal X is causal for another signal Y in the Granger sense, then past values of X should contain information that helps to predict Y better than merely using the information contained in past values of Y.[1]

This concept of predicting better with an additional variable can be linked to significance tests in multiple linear regression, where an independent variable is declared significant if the full model explains (predicts) the dependent variable better than the model that does not contain this variable. In many fields, these tests are called marginal and are linked to the so-called type III sum of squares in analysis of variance.

The general criterion of causality is if the prediction error of a first series given its own past is significantly bigger than its prediction error given its own past plus the past of a second series, then this second series causes the first, in the Granger sense.[1,2,21]

As Chamberlain,[22] Florens,[23] and Chicharro[24] point out, the most general criterion of Granger noncausality can be defined based on the equivalence of 2 conditional densities:

$$f_t(Y_t|Y_{t-1}^{t-p}) = f_t(Y_t|Y_{t-1}^{t-p}, X_{t-1}^{t-p}), \tag{3}$$

where X_t and Y_t are the 2 recorded time series; Y_{t-1}^{t-p} and X_{t-1}^{t-p} denote the history from time $t-1$ to $t-p$ of Y and X, respectively (ie, $[Y_{t-1}, \dots, Y_{t-p}]$, and $[X_{t-1}, \dots, X_{t-p}]$); and p is a suitable model order. This general criterion is expressed

in terms of the distributions only, so it does not rely on any model assumptions.[25] Note that in this general definition, $f_t(.)$ can be different for each time, and therefore, the general criterion in Equation 3 includes nonstationary models.

Any existing method for assessing Granger causality can be viewed as a restricted estimation procedure allowing us to estimate the densities in Equation 3 and to derive a causality statistic in order to test their difference.

4.2 | Linear, Gaussian, and stationary case

We will first discuss the simplest case of Granger causality, which is defined in the time domain. For linear Gaussian autoregressive models, the assumptions are Gaussianity, homoscedasticity, and linearity. It is important to note that this requires that the data are stationary. The quantities in Equation 3 become an autoregressive model of order p (AR(p)) for the left-hand side,

$$f_t(Y_t|Y_{t-1}^{t-p}) = \phi\left(Y_t; \mu = \sum_{j=1}^{p} \vartheta_{1(j)} Y_{t-j}, \sigma^2 = \Sigma_1\right), \tag{4}$$

and a vector autoregressive model of order p (VAR(p)) for the right-hand side,

$$f_t(Y_t|Y_{t-1}^{t-p}, X_{t-1}^{t-p}) = \phi\left(Y_t; \mu = \sum_{j=1}^{p} \vartheta_{11(j)} Y_{t-j} + \sum_{j=1}^{p} \vartheta_{12(j)} X_{t-j}, \sigma^2 = \Sigma_2\right), \tag{5}$$

where ϕ stands for the Gaussian probability density function.

In the next sections, we will present the 2 widely used approaches for testing hypotheses (3) in the linear Gaussian context. The first one is based on an F statistic expressed as the ratio of the residual variances of models[26] on Equations 4 and 5. The second one is based on a Wald statistic and tests the significance of the causal VAR coefficients.[27,28]

4.3 | Granger-causality criterion for linear Gaussian processes based on variances

The original formulation of Granger causality is expressed in terms of comparing the innovation variances of the whole (Equation 5) and the restricted (Equation 4) linear Gaussian autoregressive models.[1,21,26] Granger[1] proposed the following quantity to quantify this variance comparison:

$$F_{X\to Y} = \ln\left(\frac{\Sigma_1}{\Sigma_2}\right). \tag{6}$$

In Hesse et al[29] and Goebel et al,[30] this quantity is estimated by replacing the 2 variances by estimates. A test based on resampling this statistic is used for assessing the significance.

Geweke[26,31] made several other important statements for (6). He showed first that the total interdependence between 2 variables can be decomposed in terms of their 2 reciprocal causalities plus an instantaneous feedback term. Secondly, he showed that under fairly general conditions, $F_{X\to Y}$ can be decomposed additively by frequency (see Section 5). Lastly, he pointed out that it is possible to extend Granger causality to include other series. On the basis of the conditional densities, the null hypothesis would write

$$f_t(Y_t|Y_{t-1}^{t-p}, \mathbf{W}_{t-1}^{t-p}) = f_t(Y_t|Y_{t-1}^{t-p}, X_{t-1}^{t-p}, \mathbf{W}_{t-1}^{t-p}), \tag{7}$$

where $\mathbf{W}_{t-1}^{t-p}$ represents a set of variables that are controlled for when assessing the causality from X to Y. In the literature, this extension bears the name conditional Granger causality.[21]

As explained in Bressler and Seth[7] and Geweke,[26] comparing the innovation variances of the whole and restricted linear Gaussian autoregressive models amounts to evaluating the hypothesis

$$\text{Ho:} \quad \Sigma_1 = \Sigma_2, \tag{8}$$

which can be assessed through the statistic

$$F = \frac{\frac{RSS_r - RSS_{ur}}{p}}{\frac{RSS_{ur}}{T-2p-1}}. \tag{9}$$

RSS_r and RSS_{ur} are the residual sum of squares of the linear models in Equations 4 and 5, needed to estimate Σ_1 and Σ_2, respectively; p is the model order; and T is the total number of observations used to estimate the unrestricted model.

This statistic follows approximately an F distribution with degrees of freedom p and $T - 2p - 1$. A significant F may reasonably be interpreted as an indication that the unrestricted model provides a better prediction than does the restricted one, and so that X causes Y in the Granger sense.

4.4 | Granger-causality criterion for linear Gaussian processes based on coefficients

Another way to test for causality between 2 series under the same conditions as in section 4.3 is to estimate model (5) only and to directly test the significance of the VAR coefficients of interest.[27,28] Let us first define the complementary equation of Equation 5,

$$f_t(X_t|X_{t-1}^{t-p}, Y_{t-1}^{t-p}) = \phi\left(X_t; \mu = \sum_{j=1}^{p} \vartheta_{22(j)} X_{t-j} + \sum_{j=1}^{p} \vartheta_{21(j)} Y_{t-j}, \sigma^2 = \Sigma_3\right), \tag{10}$$

and the variance-covariance matrix of the whole system,

$$\Sigma = \begin{pmatrix} \Sigma_2 & \Gamma_{23} \\ \Gamma_{23} & \Sigma_3 \end{pmatrix}, \tag{11}$$

where the off-diagonal elements may or may not be equal to zero. Testing whether X causes Y in the Granger sense amounts to testing the hypotheses,

$$\vartheta_{12(1)} = \vartheta_{12(2)} = \vartheta_{12(3)} = \cdots = \vartheta_{12(p)} = 0, \tag{12}$$

and testing whether Y causes X in the Granger sense amounts to testing,

$$\vartheta_{21(1)} = \vartheta_{21(2)} = \vartheta_{21(3)} = \cdots = \vartheta_{21(p)} = 0. \tag{13}$$

In the context of linear Gaussian autoregressive models, the 2 null hypotheses (8) and (12) are equivalent.

We can observe that the approach using hypothesis (8) requires the computation of 2 models (an AR model and a VAR model), whereas a single VAR model is sufficient for the approach using hypothesis (12).

Under joint normality and finite variance-covariance assumptions, the Wald statistic is defined as

$$W = (\hat{\boldsymbol{\vartheta}}_{\mathbf{12}})' \left(\text{var}(\hat{\boldsymbol{\vartheta}}_{\mathbf{12}})\right)^{-1} (\hat{\boldsymbol{\vartheta}}_{\mathbf{12}}), \tag{14}$$

where $\boldsymbol{\vartheta}_{\mathbf{12}}$ contains all the parameters $\vartheta_{12(j)}$, for $j = 1, \ldots, p$. As T increases, this statistic asymptotically follows a χ^2 distribution with p degrees of freedom.[28] A significant Wald statistic suggests that at least one of the causal coefficients is different from zero and, in that sense, that X is causal for Y in the Granger sense. See Sato et al[32] for an example of application of this statistic in neuroscience.

The time-domain Granger-causality statistics in Equations 9 and 14 are derived from AR and VAR modelling of the data. Their relevance therefore relies on the quality of the fitted models. The first issue is the selection of the model order p. Traditional criteria used in time series are the Akaike information criterion and the Bayesian information criterion.[33,34] For the first statistic, in Equation 9, it is advisable to select the same p for the 2 models. The second issue is probably often overlooked but of utmost importance. In practice, and particularly for neuroscience data, the plausibility of the assumptions behind these models must be checked before interpreting the resulting tests. This includes analysis of the residuals from the fitted model.

4.5 | Transfer entropy

Transfer entropy (TE) is a functional statistic developed in information theory.[35] It can be used to test the null hypothesis (3) in terms of the distributions themselves and thus does not rely on the linear Gaussian assumption. It is constructed from the Kullback-Leibler distances between the conditional distributions $f(Y_t|Y_{t-1}^{t-p}, X_{t-1}^{t-p})$ and $f(Y_t|Y_{t-1}^{t-p})$[24,35]:

$$T_{X\to Y} = \int \cdots \int f(y_t, y_{t-1}^{t-p}, x_{t-1}^{t-p}) \ln \frac{f(y_t|y_{t-1}^{t-p}, x_{t-1}^{t-p})}{f(y_t|y_{t-1}^{t-p})} \, \mathrm{d}y_t \mathrm{d}y_{t-1}^{t-p} \mathrm{d}x_{t-1}^{t-p}, \tag{15}$$

where the integrals over y_{t-1}^{t-p} and x_{t-1}^{t-p} are both of dimension p, and so the overall integral in Equation 15 is of dimension $\{2p + 1\}$. An even more general definition would allow the distributions $f(.)$ to depend on time, letting the TE statistic be time dependent.

It has been shown that for stationary linear Gaussian autoregressive models (4) and (5), the indices (15) and (6) are equivalent.[24,36]

In its general form, TE is a functional statistic, free from any parametric assumption on the 2 densities $f(Y_t|Y_{t-1}^{t-p})$ and $f(Y_t|Y_{t-1}^{t-p}, X_{t-1}^{t-p})$. Applications of TE in neuroscience can be found for instance in Chàvez et al,[37] Garofalo et al,[38] Vicente et al,[39] Wibral et al,[40] Lizier et al,[41] and Besserve et al.[42,43] The difficulties arise when trying to estimate and compute the conditional densities in Equation 15.

The estimation can be preformed parametrically or nonparametrically. Parametric estimators impose a parametric density to the data (see, for example, the linear estimator in Montalto et al[44]). Moreover, several nonparametric estimators exist to estimate $f(Y_t|Y_{t-1}^{t-p})$ and $f(Y_t|Y_{t-1}^{t-p}, X_{t-1}^{t-p})$, and the performance of each of them strongly depends on the characteristics of the data. For a general review of nonparametric estimation methods in information theory, see Vicente et al,[39] Hlaváčková-Schindler et al,[45] and Wibral et al.[46] For simple discrete processes, the probabilities can be determined by computing the frequencies of occurrence of different states. For continuous processes, which are those of interest for neuroscience, it is more delicate to find a reliable nonparametric density estimation. Following Hlaváčková-Schindler et al,[45] there exist 3 main classes of nonparametric estimator. First, the partition-based estimators estimate the probability densities by counting how many samples fall into each division of a certain partition of the data (see, for example, the binning estimator[44]). Secondly, the plug-in estimators are based on consistent estimates for the probability densities that are plugged into the corresponding functional (kernel-based estimator is among the most popular plug-in estimator[35,39,47,48]). Finally, the metric-based estimators rely on the fact that the larger the distance between 1 point to its nearest neighbour, the lower the local density around that point (see, for example, the nearest-neighbour estimator[44]).

The major limitation of nonparametric estimation is due to the dimension of the densities in (15) that can rapidly be too large for any nonparametric approach (known as the curse of dimensionality problem). Furthermore, in the present case, the estimation of $f(Y_t|Y_{t-1}^{t-p})$ and $f(Y_t|Y_{t-1}^{t-p}, X_{t-1}^{t-p})$ implies an integration in dimension $2p+1$ in Equation 15. Even for a moderate dimension p, a huge number of observations would therefore be required. Typically, Schreiber[35] proposes to choose the minimal p, meaning $p = 1$, for the above reasons.[35, p. 462]

A supplementary parameter, called the embedding delay (τ), which represents the lag in time between each observation of the past values of variables X and Y, has to be estimated. Equation 15 then becomes

$$T_{X\to Y} = \int \cdots \int f(y_t, y_{t-1\tau}^{t-p\tau}, x_{t-1\tau}^{t-p\tau}) \ln \frac{f(y_t|y_{t-1\tau}^{t-p\tau}, x_{t-1\tau}^{t-p\tau})}{f(y_t|y_{t-1\tau}^{t-p\tau})} \mathrm{d}y_t \mathrm{d}y_{t-1\tau}^{t-p\tau} \mathrm{d}x_{t-1\tau}^{t-p\tau}. \tag{16}$$

The selection of the model order p (called the embedding dimension in this context) and the embedding delay τ is a delicate issue. As discussed in Lindner et al,[49] if p is chosen too small, the causal structure may not be captured, and thus, the TE statistic will be incorrect. On the other hand, using an embedding dimension that is higher than necessary will lead to an increase of variability in the estimation, in addition to a considerable increase in computation time. As explained in Montalto et al,[44] the large majority of the estimation methods used a uniform conditioned embedding schemes for the selection of the embedding dimension and the embedding delay, where the components of the histories $y_{t-1\tau}^{t-p\tau}$ and $x_{t-1\tau}^{t-p\tau}$ to be included in the embedding vectors are selected a priori and separately for each series.

The non-uniform embedding schemes are a convenient alternative to the uniform embedding issue.[44] This approach is based on a stepwise selection of $y_{t-1\tau}^{t-p\tau}$ and $x_{t-1\tau}^{t-p\tau}$, considered up to a maximum lag p_{lag}, by considering the quantities that are most significant in terms of predictive information for the target variable (see Kugiumtzis[50] for a non-uniform embedding scheme approach and Faes et al[51] for a non-uniform embedding approach coupled with a nearest-neighbour estimation technique).

Other accurate techniques are devised to deal with the curse of dimensionality. Runge et al[52] proposed to overcome it by embedding the TE statistic into the framework of graphical models. Kugiumtzis[50] and Faes et al[53] moreover propose some ad hoc techniques for non-uniform embedding, which circumvent the issue of the embedding parameters estimation.

A toolbox named TRENTOOL provides the computation of TE and the estimation of $f(y_t|y_{t-1\tau}^{t-p\tau})$ and $f(y_t|y_{t-1\tau}^{t-p\tau}, x_{t-1\tau}^{t-p\tau})$ through kernel-based estimation.[49]

The model order p is optimized simultaneously with the embedding delay τ through 2 implemented criteria. The first is the "Cao criterion," which selects τ on an "ad hoc" basis and p through a false neighbour criterion.[49,54] The second is the "Ragwitz criterion," which selects τ and p simultaneously by minimizing the prediction error of a local predictor.[35] Typically, Wibral et al[40] select the value of p as the maximum determined by the Cao criterion from $p = 1$ to 4, and choose the value of τ following a popular ad hoc option, as the first zero of the autocorrelation function of the signal.

TRENTOOL allows us moreover to compute the distribution of the TE statistic under the null hypothesis through a permutation method. The data are shuffled in order to break the links between the signals, and then the TE statistic is recomputed on each surrogate dataset (eg, Wibral et al[40] use 1.9×10^5 permutations for assessing the significance of the TE statistic). Analyses with TRENTOOL are limited so far to bivariate systems.

MuTE is a more recent and complete toolbox that estimates the TE together with the embedding parameters.[44] Uniform and non-uniform schemes are implemented for the embedding parameter selection and 3 different types of estimators, namely, the linear, the binning, and the nearest-neighbour estimator, are implemented for the estimation of $f(y_t|y_{t-1\tau}^{t-p\tau})$ and $f(y_t|y_{t-1\tau}^{t-p\tau}, x_{t-1\tau}^{t-p\tau})$. This therefore allows us to perform 6 different estimations of the TE.

MuTE toolbox moreover allows us to compute the significance tests associated with the 3 different estimators implemented in the toolbox. The statistical significance of the TE estimated through the linear estimator is assessed by a parametric F test.[44] For the TE estimated through the binning estimator, statistical significance is assessed based on a surrogate data procedure. Transfer entropy statistic is compared to the null distribution computed by calculating TE on each surrogate dataset, computed by shifting the original series by a randomly selected lag. As for the binning estimator, the statistical significance of the TE estimated through the nearest-neighbour estimator exploited the method of surrogate data implemented by the time-shift procedure.[44]

The formulation of causality based on the conditional independence in Equation 3 was later used and theoretically refined in Chamberlain[22] and Florens.[23] Although less general, the statistics given in Equations 6 and 14 are much easier to implement and are testable. This probably explains why they have received considerably more attention in applied work.

5 | FREQUENCY-DOMAIN CAUSALITY

5.1 | Geweke-Granger-causality statistic

As mentioned in section 4.3, an important advance in developing the Granger-causality methodology was to provide a spectral decomposition of the time-domain statistics.[26,55]

For completeness, we give below the mathematical details of this derivation. The Fourier transform of Equations 5 and 10 for a given frequency ω (expressed as a system of equations) is

$$\begin{pmatrix} \vartheta_{11}(\omega) & \vartheta_{12}(\omega) \\ \vartheta_{21}(\omega) & \vartheta_{22}(\omega) \end{pmatrix} \begin{pmatrix} Y(\omega) \\ X(\omega) \end{pmatrix} = \begin{pmatrix} \varepsilon_1(\omega) \\ \varepsilon_2(\omega) \end{pmatrix}, \tag{17}$$

where $Y(\omega)$ and $X(\omega)$ are the Fourier transforms of Y_1^T and X_1^T at frequency ω, and $\varepsilon_1(\omega)$ and $\varepsilon_2(\omega)$ are the Fourier transforms of the errors of the models (5) and (10) at frequency ω. The components of the matrix are

$$\vartheta_{lm}(\omega) = \delta_{lm} - \sum_{j=1}^{p} \vartheta_{lm}(j) e^{(-i2\pi\omega j)}, \quad \text{where} \quad \begin{cases} \delta_{lm} = 0, & l = m, \\ \delta_{lm} = 1, & l \neq m, \end{cases} \quad l, m = 1, 2.$$

Rewriting Equation 17 as

$$\begin{pmatrix} Y(\omega) \\ X(\omega) \end{pmatrix} = \begin{pmatrix} H_{11}(\omega) & H_{12}(\omega) \\ H_{21}(\omega) & H_{22}(\omega) \end{pmatrix} \begin{pmatrix} \varepsilon_1(\omega) \\ \varepsilon_2(\omega) \end{pmatrix}, \tag{18}$$

we have

$$\begin{pmatrix} H_{11}(\omega) & H_{12}(\omega) \\ H_{21}(\omega) & H_{22}(\omega) \end{pmatrix} = \begin{pmatrix} \vartheta_{11}(\omega) & \vartheta_{12}(\omega) \\ \vartheta_{21}(\omega) & \vartheta_{22}(\omega) \end{pmatrix}^{-1}, \tag{19}$$

where H is the transfer matrix. The spectral matrix $S(\omega)$ can now be derived as

$$S(\omega) = H(\omega)\Sigma H^*(\omega), \tag{20}$$

where the asterisk denotes matrix transposition and complex conjugation. Σ is the matrix[21] defined in Equation 11. The spectral matrix $S(\omega)$ contains cross-spectra terms, $S_{12}(\omega)$ and $S_{21}(\omega)$, and auto-spectra terms, $S_{11}(\omega)$ and $S_{22}(\omega)$. If X and Y are independent, the cross-spectra terms are equal to zero.

Let us now write the auto-spectrum of Y as

$$S(\omega)_{11} = H(\omega)_{11}\Sigma_2 H^*(\omega)_{11} + 2\Gamma_{23} Re(H(\omega)_{11}) H^*(\omega)_{12}) + H(\omega)_{12}\Sigma_3 H^*(\omega)_{12}. \tag{21}$$

In the following derivation, we will suppose that Γ_{23}, the off-diagonal element of the Σ matrix in Equation 11, is equal to zero. The nonfulfilment of this condition is an indication of instantaneous causality (or correlation) between the 2 time series, and methods to deal with these instantaneous effects incorporating them in the VAR model have been proposed.[56,57] In the case of instantaneous effects, a more complex derivation is required (see Ding et al,[21] Hyvärinen et al,[56] and Faes et al[57] for further details).

However, if this independence condition is fulfilled, the auto-spectrum reduces to 2 terms,

$$S(\omega)_{11} = H(\omega)_{11}\Sigma_2 H^*(\omega)_{11} + H(\omega)_{12}\Sigma_3 H^*(\omega)_{12}. \tag{22}$$

The first term, $H(\omega)_{11}\Sigma_2 H^*(\omega)_{11}$, only involves the variance of the signal of interest and thus can be viewed as the intrinsic part of the auto-spectrum. The second term, $H(\omega)_{12}\Sigma_3 H^*(\omega)_{12}$, only involves the variance of the second signal and thus can be viewed as the causal part of the auto-spectrum.

In Geweke spectral formulation, the derivation of the spectral measure $f_{X\to Y}$ requires the fulfilment of the following properties. The measures have to be non-negative, and the sum over all frequencies of the spectral Granger-causality components has to equal the time-domain Granger-causality quantity (6):

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} f_{X\to Y}(\omega)d\omega = F_{X\to Y}. \tag{23}$$

The 2 conditions together imply the desirable property

$$F_{X\to Y} = 0 \iff f_{X\to Y}(\omega) = 0, \quad \forall\omega. \tag{24}$$

The third condition is that the spectral statistics have an empirical interpretation. The spectral Granger-causality statistic proposed by Geweke fulfils all 3 requirements. For a given frequency ω and scalar variables X and Y, it is defined as

$$f_{X\to Y}(\omega) = \frac{S_{11}(\omega)}{H_{11}(\omega)\Sigma_2 H_{11}^*(\omega)}, \tag{25}$$

where Σ_2 is the variance defined in Equation 5, $S_{11}(\omega)$ is the auto-spectrum of Y, and $H_{11}(\omega)$ is the $(1,1)$ element of the transfer matrix in Equation 19. The form of Equation 25 provides an important interpretation: The causal influence depends on the relative size of the total power $S_{11}(\omega)$ and the intrinsic power $H_{11}(\omega)\Sigma_2 H_{11}^*(\omega)$. Since the total power is the sum of the intrinsic and the causal powers (see Equation 22), the spectral Geweke-Granger-causality (GGC) statistic is zero when the causal power is zero (ie, when the intrinsic power equals the total power). The statistic increases as the causal power increases.[21] Given the requirements imposed by Geweke, the measure $f_{X\to Y}(\omega)$ has a clear interpretation: It represents the portion of the power spectrum associated with the innovation process of model (5). However, this interpretation relies on the VAR model because the innovation process is only well-defined in this context (see Brovelli et al,[58] Chen et al,[59,60] and Bressler et al[61] for examples of application in neuroscience).

The estimation of the parameters and the model order selection procedure is the same as in section 4.4, because the frequency-domain VAR model in Equation 17 is directly derived from the time-domain VAR model. The model order selection has to be performed within the time-domain model estimation procedure.[58,62]

Lin et al[62] showed that under the null hypothesis $f_{X\to Y}(\omega) = 0$ and based on (25), one can derive a statistic that follows an F distribution with degrees of freedom $(p, T-2p)$ when the number of observations tends to infinity (it was first derived in Brovelli et al[58] and Gourévitch et al[63]).

5.2 | Directed transfer function and partial directed coherence

The directed transfer function (DTF) and the partial directed coherence (PDC) are alternative measures also derived from VAR estimated quantities that are closely related to the GGC statistic. Differences between these 2 measures as well as the relation between DTF and directed coherence and the equivalence of these methods for bivariate time series are discussed in Baccalà et al,[64] Faes et al,[65] and Faes and Nollo[66]

The DTF is a frequency-domain measure of causal influence based on the elements of the transfer matrix $H(\omega)$ in Equation 19. It has both normalized and non-normalized forms.[67,68] The PDC is derived from the matrix of the Fourier transformation of the estimated VAR coefficients[69] in Equation 17. See Schelter et al[70] for a renormalized version of PDC and Schelter et al[71] for an example of application in neuroscience.

The DTF is expressed as

$$\mathrm{DTF}_{X\to Y}(\omega) = \sqrt{\frac{|H_{12}(\omega)|^2}{|H_{11}(\omega)|^2 + |H_{12}(\omega)|^2}}. \tag{26}$$

As explained in Bořil and Sovka,[4] outcomes of DTF analyses must be interpreted carefully. Indeed, although DTF decomposes the causal relations in the frequency domain, it does not distinguish direct from indirect connections. The direct directed transfer function (dDTF) statistic (see Bořil and Sovka[4] and references therein) claims to improve the DTF in this sense, but as it is shown in Bořil and Sovka,[4] the dDTF may not be able to distinguish direct relations from indirect ones in some cases. The PDC solves this problem by evaluating the direct connections only.[4] The PDC statistic is defined as

$$\mathrm{PDC}_{X\to Y}(\omega) = \frac{\vartheta_{12}(\omega)}{\boldsymbol{\vartheta}_2^*(\omega)\boldsymbol{\vartheta}_2(\omega)}, \tag{27}$$

where $\vartheta_{12}(\omega)$ represents the Fourier-transformed VAR coefficient (ie, the causal influence from X to Y at frequency ω) and $\boldsymbol{\vartheta}_2(\omega)$ represents all outflows from X.

The PDC is normalized, but in a different way from the DTF. Indeed, the PDC represents the outflow from X to Y, normalized by the total amount of outflows from X. The normalized DTF however represents the inflow from X to Y, normalized by the total amount of inflows to Y. However, the normalization used in the PDC formulation prevents to compare the strength of the coupling among variables. A lower value of the PDC may instead correspond to a stronger relation and conversely. The generalized partial directed coherence (GPDC) statistic (see Bořil and Sovka[4] and references therein) modifies the PDC statistic using an additional normalization that make values more comparable. The square modulus of GPDC is often interpreted as the PDC but, as it is derived in Bořil and Sovka,[4] this analogy is correct only in the case of 1-unidirectional causal relation between 2 variables. Indeed, even for a bivariate model, the GPDC statistic does not take into account the feedback effect in a bidirectional causal relation case (see Bořil and Sovka[4] and references therein for a complete insight of DTF dDTF, PDC, and GPDC statistics).

Comparisons between the GGC statistic, the DTF, and the PDC are discussed in Eichler,[72] Baccalà and Sameshima,[69] Gourévitch et al,[63] Pereda et al,[6] Winterhalder et al,[73,74] and more recently in the context of information theory in Chicharro.[24] The causal interpretation of the PDC and the GGC, at least in the bivariate case, relies on Granger's definition of causality.[64-66] For the directed coherence and the DTF, the causal interpretation is different, as it relies on Sim's definition of causality.[75] See Chamberlain[22] and Kuersteiner[25] for a global overview and comparison of these 2 definitions of causality. Finally, Winterhalder et al[73] conducted a simulation-based comparison of the DTF and the PDC (and other statistics) in a neuroscience context.

The statistical properties of these spectral measures are very complex. For instance, the influence of signal preprocessing (eg, smoothing and filtering) is a crucial issue that has been studied, for example, in Florin et al[76] and Barnett and Seth.[77] The latter study showed that filtering can be a useful preprocessing step allowing us to remove artefacts and improve stationarity but is inappropriate for isolating causal influences within a specific frequency band.

5.2.1 | Assessment of significance

Theoretical distributions for DTF and PDC have been derived and are listed below. They are all based on the asymptotic normality of the estimated VAR coefficients. Therefore, they can be used and interpreted only if the assumptions behind this model hold.

Schelter et al[71] showed that the PDC statistic asymptotically follows a χ^2 distribution with 1 degree of freedom. Furthermore, Schelter et al[70] showed that a renormalized form of PDC can be related to a χ^2 distribution with 2 degrees of freedom. Finally, Winterhalder et al[73] provide simulations that suggest that this χ^2 distribution even works well if the true model order is strongly overestimated. Note that the asymptotic distribution of the 3 main forms of the PDC statistic (PDC, GPDC, and iPDC [information PDC]) have been provided by the proposers of these metrics themselves.[78]

Eichler[72] showed that the DTF quantity can be compared to a χ^2 distribution with 1 degree of freedom. This property is also based on the asymptotic normality of estimated VAR coefficients, and its accuracy is evaluated through simulations.

In Faes et al[79] and Hesse et al[29] propose to compute the distribution under the null hypothesis of, respectively, the PDC and the time-varying Granger causality (estimated by the generalized recursive least squares[80]) through suitable surrogate methods.

For the PDC as well as for the DTF asymptotic distributions, Schelter[81] and Eichler[72] state that a major drawback is that there are a lot of tests—one for each frequency. It is well known that when many tests are produced, caution has to be taken in interpreting those that are significant. For example, even under the null hypothesis of no information flow, there is a high probability that for a few frequencies, the test will be significant.

6 | TIME-VARYING GRANGER CAUSALITY

Neuroscience data are nonstationary in most cases. The specificity (task or stimulus related) of the increase or decrease and/or local field potential implies this nonstationarity, which is of primary interest. A Granger-causality statistic that is time specific or time and frequency specific is desirable, as it would capture the evolution of Granger causality through time.

Since the original statistics are based on AR and VAR models, and therefore on assumptions assuming that the autocorelation does not vary along the time, these models have to be extended to cases assuming changing autocorelation structure in order to suitably extract a Granger-causality statistic.

Practically, getting a statistic to assess the causality between 2 series for each time requires the estimation of the densities $f_t(Y_t|Y_{t-1}^{t-p})$ and $f_t(Y_t|Y_{t-1}^{t-p},X_{t-1}^{t-p})$ separately for each time t. There are 2 additional difficulties to keep in mind. The first is the necessity of an objective criterion for time-varying model order selection, and the second is the difficulty of incorporating all the recorded data (meaning all the trials) in the estimation procedure.

6.1 | Nonparametric statistics

6.1.1 | Wavelet-based statistic

In the context of neuroscience, Dhamala et al[82] proposed to bypass the nonstationarity problem by nonparametrically estimating the quantities that allow us to derive the spectral GGC statistic (25). They derived an evolutionary spectral density through the continuous wavelet transform of the data and then derived a quantity related to the transfer function (by spectral matrix factorization). On the basis of this quantity, they obtain a GGC statistic that can be interpreted as a time-varying version of the GGC statistic defined in (25).

This approach bypassed the delicate step of estimating $f_t(Y_t|Y_{t-1}^{t-p})$ and $f_t(Y_t|Y_{t-1}^{t-p},X_{t-1}^{t-p})$ separately for each time. However, this method presents several drawbacks in terms of interpretation of the resulting quantity. The GGC statistic is indeed derived from a VAR model and its interpretation directly follows from the causal nature of the VAR coefficients. The nonparametric wavelet spectral density however does not have this Granger-causality interpretation. Therefore, attention must be paid when interpreting this proposed evolutionary causal GGC statistic derived from spectral quantities that are not based on a VAR model.

6.1.2 | Local TE

Lizier et al[41,83] and Prokopenko et al[84] proposed a time-varying version of the TE (15), in order to detect dynamical causal structure in a functional magnetic resonance imaging (fMRI) study context. The "global" TE defined in Equation 15 can be expressed as a sum of "local transfer entropies" at each time:

$$T_{X\to Y}=\frac{1}{T}\sum_{t=1}^{T} f(y_t,y_{t-1}^{t-p},x_{t-1}^{t-p})\ln\frac{f(y_t|y_{t-1}^{t-p},x_{t-1}^{t-p})}{f(y_t|y_{t-1}^{t-p})}, \tag{28}$$

where each summed quantity can be interpreted as a single "local transfer entropy":

$$t_{x\to y}(t)=\ln\frac{f(y_t|y_{t-1}^{t-p},x_{t-1}^{t-p})}{f(y_t|y_{t-1}^{t-p})}. \tag{29}$$

The step from Equations 15 to 28 is obtained by replacing the joint density $f(Y_t,Y_{t-1}^{t-p},X_{t-1}^{t-p})$ with its empirical version.

As explained in Wollstadt et al,[85] local TE (28) localizes information transfer in time evaluating the probability density functions $f(.)$ of a "stationary" process at each time. Indeed, time-independent PDFs $f(.)$ in (15) are estimated, and then these densities $f(.)$ are evaluated locally in order to give information on the TE dynamic.

Note that Prokopenko et al[84] or Lizier et al[41,83] do not provide an objective criterion for model order selection.

Wollstadt et al[85] propose a time-varying version of the TE with the local PDFs $f_t(.)$ estimated differentially for each time. The estimation procedure requires an ensemble of realizations, as it is the case with the multiple trials in neuroscience.

With this local TE formulation, the directed information transfer can be analysed in terms of average information transfer (28) or, locally, by computing (29) individually for each time.

Attention must be paid to the fact that, even if the overall quantity in Equation 15 can be suitably expressed as a sum as in Equation 29, its causal nature does not necessarily remain in each part. As such, we should not directly interpret these individuals parts as causal local measures of causality, even if the sum of them gives an overall quantity that has an intrinsic causal meaning.

6.2 | Time-varying VAR model

As seen before in Equations 9, 14, 25, 26 and 27, parametric Granger-causality statistics in the time and frequency domains are derived from AR and VAR modelling of the data (Equations 4 and 5, respectively). One way to extend these statistics to the nonstationary case amounts to allowing the AR and VAR parameters to evolve in time. In addition to the difficulties related to model order selection and the fact that we have to deal with several trials, time-varying AR and VAR models are difficult to estimate since the number of parameters is most of the time considerable compared to the available number of observations. To overcome the dimensionality of this problem, Chen[86] propose to make 1 of the 3 following assumptions, local stationarity of the process,[87] slowly varying nonstationary characteristics,[88] and slowly varying parameters for nonstationary models.[89] In practice, it is difficult to distinguish between these assumptions, but they all allow nonstationarity. Chen[86] asserts that if one of the above assumptions is fulfilled, the estimate of a signal at some specific time can be approximated and inferred using the neighbourhood of this time point. Probably all time-varying methods proposed in the literature are based on one of these characteristics.

We will discuss now the 2 widely used approaches that deal with this type of nonstationarity: the windowing approach, based on the locally stationary assumption, and the adaptive estimation approach, based on slowly varying parameters.

6.2.1 | Windowing approach

A classical approach to adapt VAR models to the nonstationary case is windowing. This methodology consists in estimating VAR models in short temporal sliding windows where the underlying process is assumed to be (locally) stationary. See Ding et al[90] for a methodological tutorial on windowing estimate in neuroscience and Long et al[91] and Hoerzer et al[92] for some applications in neuroscience.

The segment or window length is a trade-off between the accuracy of the parameter estimates and the resolution in time. The shorter the segment length, the higher the time resolution but also the larger the variance of the estimated coefficients. The choice of the model order is a related very important issue. With a short segment, the model order is limited, especially since we do not have enough residuals to check the quality of the fit in each window. Some criteria have been proposed in order to simultaneously optimize the window length and model order.[62,91,93] This windowing methodology was extensively analysed and commented in Cekic.[94] This method can easily incorporate several recorded trials in the analysis by combining all of them for the parameter estimate.[90]

In Cekic,[94] we found that this windowing methodology has several limitations. First, increasing the time resolution implies short time windows and thus too few residuals to assess the quality of the fit. Second, the size of the temporal windows is somehow subjective (even if it depends on a criterion), as is the overlap between the time windows. The order of the model in turn depends on the size of the windows, and so the quality of the estimate strongly relies on several subjective parameters.

6.2.2 | Adaptive estimation method

A second existing methodology for estimating time-varying AR and VAR models is adaptive algorithms. They consist in estimating a different model at each time and not inside overlapped time windows. The principle is always the same: The observations at time t are expressed as a linear combination of the past values with coefficients evolving slowly over time plus an error term. The difference between the methods lies in the form of transition and update from coefficients at time

t to those at time $t+1$. This transition is always based on the prediction error at time t.[95] The scheme is

$$\begin{cases} \boldsymbol{\varphi}_{t+1} = f(\boldsymbol{\varphi}_t, \boldsymbol{w}_t) \\ Z_t = C_t\boldsymbol{\varphi}_t + \boldsymbol{\nu}_t \end{cases} \quad \text{with} \quad \begin{cases} \boldsymbol{\varphi}_t = \text{vec}[\boldsymbol{\vartheta}_{1(t)}, \boldsymbol{\vartheta}_{2(t)}, .., \boldsymbol{\vartheta}_{p(t)}]', \\ Z_t = (Y_t\ X_t)', \\ C_t\boldsymbol{\varphi}_t = \sum_{j=1}^{p} \vartheta_{j(t)}(Y_{t-j}\ X_{t-j})', \end{cases} \tag{30}$$

where $\boldsymbol{\vartheta}_{j(t)}$ is the time-varying VAR coefficients at lag j for time t, $\boldsymbol{\nu_t}$ is the error of the time-varying VAR equation at time t, and $\boldsymbol{w_t}$ is the error of the Markovian update of the time-varying VAR coefficients from time t to time $t+1$.

There are several recursive algorithms to estimate this kind of model. They are based on the least-mean-squares approach,[96] the recursive least-squares approach (see Mainardi et al,[97] Patomaki et al,[98,99] and Akay[100] for basic developments; Möller et al[80] for an extension to multivariate and multitrial data; and Astolfi et al,[101,102] Hesse et al,[29] Tarvainen et al,[103] and Wilke et al[104] for examples of application in neuroscience), and the recursive AR approach.[105] They are all described in detail in Schlögl.[95]

All these adaptive estimation methods depend on a free quantity that acts as a tuning parameter and defines the relative influence of $\boldsymbol{\varphi}_t$ and $\boldsymbol{w_t}$ on the recursive estimate of $\boldsymbol{\varphi}_{t+1}$. Generally, this free tuning parameter determines the speed of adaptation, as well as the smoothness of the time-varying VAR parameter estimates. The sensitivity of the least-mean-squares, RLS, and recursive AR algorithms to this tuning parameter was investigated in Schlögl,[95] and estimation quality strongly depends on it. The ad hoc nature of these procedures does not allow for proper statistical inference.

Finally, as for the previous models, the model order has to be selected. It is often optimized in terms of mean square error, in parallel with tuning parameter selection.[106,107]

6.2.3 | Kalman filter and the state space model

Kalman[108] presented the original idea of the Kalman filter. Meinhold and Singpurwalla[109] provided a Bayesian formulation.

A Kalman filtering algorithm can be used to estimate time-varying VAR models if it can be expressed in a state space form with the VAR parameters evolving in a Markovian way. This leads to the system of equations

$$\begin{cases} \boldsymbol{\varphi}_{t+1} = A\boldsymbol{\varphi}_t + \boldsymbol{w}_t & \boldsymbol{w}_t \sim N(0, Q) \\ Z_t = C_t\boldsymbol{\varphi}_t + \boldsymbol{\nu}_t & \boldsymbol{\nu}_t \sim N(0, R) \end{cases} \quad \text{with} \quad \begin{cases} \boldsymbol{\varphi}_t = \text{vec}[\boldsymbol{\vartheta}_{1(t)}, \boldsymbol{\vartheta}_{2(t)}, \ \dots\ , \boldsymbol{\vartheta}_{p(t)}]', \\ Z_t = (Y_t\ X_t)', \\ C_t\boldsymbol{\varphi}_t = \sum_{j=1}^{p} \vartheta_{j(t)}(Y_{t-j}\ X_{t-j})', \end{cases} \tag{31}$$

where the vector $\boldsymbol{\varphi}_t$ contains the time-varying VAR coefficients that are adaptively estimated through the Kalman filter equations. The matrix Q represents the variance-covariance matrix of the state equation that defines the Markovian process of the time-varying VAR coefficients. The matrix R is the variance-covariance matrix of the observed equation containing the time-varying VAR model equation.

With known parameters A, Q, and R, the Kalman smoother algorithm gives the best linear unbiased estimator for the state vector, which here contains the time-varying VAR coefficients of interest.[108]

In the engineering and neuroscience literature, the matrix A is systematically chosen as the identity matrix, and Q and R are often estimated through some ad hoc estimation procedures. These procedures and their relative references are listed in Tables 1 and 2, which are based on Schlögl.[95]

There are many applications of these estimation procedures in the neuroscience literature.[29,39,80,101,102,116,117] For an extension to several trials, the reader is referred to Milde et al[110,118] and to Havlicek et al[119] for an extension to forward and backward filter estimation procedure.

Any given method must provide a way to estimate the parameter matrices A, Q, and R simultaneously with the state vector $\boldsymbol{\varphi}_{t+1}$, while selecting the model order in a suitable way. The procedure must also manage models based on several trials.

In the statistics literature, it has been known for a long time that the matrices A, Q, and R can be obtained through a maximum likelihood Expectation-Maximization (EM) based approach (see Shumway and Stoffer[120] and Cassidy and Penny[121] for a Bayesian extension of this methodology).

Cekic et al[122] proposed a multiscale fully Bayesian implementation of the state space model in (31), providing a global estimation of all the model parameters A, Q, and R, a time-varying frequency-specific estimation of the φ_t's together

TABLE 1 Variants for estimating the covariance matrix R_t based on Schlögl[95]

Type	Estimate of R_t	References
Univariate	$R_t = (1 - UC)R_{t-1} + UCe_t{}^2$	Schack et al[96]
One trial	$e_t = y_t - C_t x_t$	
Multivariate	$R_0 = I_d$	Milde et al[110]
Multiple trial	$\overline{R_t} = \overline{R_{t-1}}(1 - UC) + UCe'e/(K-1)$	
Univariate	$R_t = 1$	Isaksson et al[111]
One trial		
Univariate	$R_t = 1 - UC$	Patomaki et al[98]
One trial		Patomaki et al[99]
		Astolfi et al[112]
		Akay[100]
Univariate	$q_t = Y'_{t-1}A_{t-1}Y_{t-1}$	Jazwinski[113]
One trial	$R_t{}^+ = \begin{cases} (1 - UC)R_{t-1}{}^+ + UC(e_t - q_t) & \text{if } e_t{}^2 > q_t \\ R_{t-1}{}^+ & \text{if } e_t{}^2 \le q_t \end{cases}$	
	$R_t = R_t{}^+$	
Univariate	Same as Jazwinski[113] except that	Penny and Roberts[114]
One trial	$R_t = R_{t-1}{}^+$	
Univariate	$R_t = 0$	Kalman[108]
One trial		Kalman and Bucy[115]

Note. UC acts as tuning parameters that must be choosen between 0 and 1.

TABLE 2 Variants for estimating the covariance matrix Q_t based on Schlögl[95]

Type	Estimate of Q_t	References
Univariate	$Q_t = UCx_t$	Akay[100]
One trial		Haykin et al[112]
Univariate	$x_t = (I - k_t)y'_{t-1}A_{t-1}$	Isaksson et al[111]
One trial	$Q_t = UC^2 I$	
Univariate	$K_t = y'_{t-1}x_{t-1}y'_{t-1} + R_t$	Jazwinski[113]
One trial	$L_t = (1 - UC)L_{t-1} + \frac{UC*(e_t{}^2 - K_t)}{y_{t-1}{}'y_{t-1}}$	Penny and Roberts[114]
	$Q_t = \begin{cases} L_t I & \text{if } L_t > 0 \\ 0 & \text{if } L_t \le 0 \end{cases}$	

with an objective criterion for the model order selection and finally a time-varying frequency-specific Granger-causality statistic.

6.2.4 | Wavelet dynamic vector autoregressive model

To derive a dynamic Granger-causality statistic in an fMRI experiment context, Sato et al[32] proposed another time-varying VAR model estimation procedure based on a wavelet expansion. They allow a time-varying structure for the VAR coefficients as well as for the variance-covariance matrix, in a linear Gaussian context. Their model is expressed as

$$f_t(Y_t|Y_{t-1}^{t-p}, X_{t-1}^{t-p}) = \phi\left(Y_t; \mu = \sum_{j=1}^{p} \vartheta_{11(j)}(t)Y_{t-j} + \sum_{j=1}^{p} \vartheta_{12(j)}(t)X_{t-j}, \sigma(t)^2 = \Sigma(t)\right), \tag{32}$$

where $\vartheta_{11(j)}(t)$ and $\vartheta_{12(j)}(t)$ are the time-varying VAR coefficients at time t and $\Sigma(t)$ is the time-varying variance-covariance matrix at time t. These are both unknown quantities that have to be estimated.

They make use of the wavelet expansion of functions in order to estimate the time-varying VAR coefficients and the time-varying variance-covariance matrix. As any function can be expressed as a linear combination of wavelet functions, Sato et al[32] consider the dynamic VAR coefficient vector $\boldsymbol{\vartheta}(t)$ and the dynamic covariance matrix Σ_t as functions of time, and so expressed them as a linear combination of wavelet functions.

They proposed a 2-step iterative generalized least square estimation procedure. The first step consists in estimating the coefficients of the expanded wavelet functions using a generalized least squares procedure. In the second step, the squared residuals obtained in the previous step are used to estimate the wavelet expansion functions for the covariance matrix Σ_t (see Sato et al[32] for further details).

The authors gave asymptotic properties for the parameter estimates, and statistical assessment of Granger-causal connectivities is achieved through a time-varying Wald-type statistic as described in Equation 14. An application in the context of gene expression regulatory network modelling can be found in Fujita et al.[123]

This wavelet-based dynamic VAR model estimation methodology has the advantage of avoiding both stationarity and linearity assumptions. However there is, surprisingly, no mention of a model order selection criterion, and the question how to take into account all the recorded trials in the estimation procedure is not addressed.

7 | EXISTING TOOLBOXES

Several toolboxes to analyse neuroscience data have been made available in recent years. We will only list those providing estimate of time-varying VAR models and Granger-causality statistics. Tables 3 and 4 present a list of these toolboxes, with references and details of their content. The description of the content is not exhaustive, and all of them contain utilities beyond (time-varying) VAR model estimate and Granger-causality analysis.

8 | DISCUSSION

8.1 | Limitations

An important topic not highlighted here is the estimation procedure and interpretation of Granger-causality statistics in a multivariate context. As discussed in section 5.2, by their relative normalization, the DTF and PDC statistics take into account the influence of other information flows when testing for a causal relationship between 2 signals. Another measure is conditional Granger causality, which was briefly mentioned in Equation 7. Indeed, when 3 or more simultaneous brain areas are recorded, the causal relation between any 2 of the series may either be direct, or be mediated by a third, or a combination of both. These cases can be addressed by conditional Granger causality, which has the ability to determine whether the interaction between 2 time series is direct or mediated by another one. Conditional Granger causality in time and frequency domains is described in Ding et al,[21] based on previous work of Geweke.[55] However, straightforward

TABLE 3 List of available toolboxes for estimating time-varying VAR models and Granger-causality statistics

Toolbox	Software	TV-VAR implemented estimation method	Implemented statistics of causality
BSMART	Matlab	Windowing approach based on Ding et al[90]	Geweke-spectral Granger
Brain System for Multivariate Autoregressive Time series Cui et al[124]		Implemented for single and multiple trials	statistic (25)
BioSig	Matlab	Kalman filter estimation type (mvaar.m Matlab function)	No causality statistic implemented
Schlögl and Brunner[125]		Implemented for single trial only Variants for estimating the covariance matrices R_t and Q_t are implemented based on Schlögl[95]	
GCCA (Granger Causal Connectivity Analysis)	Matlab	Windowing approach based on Ding et al[90]	Geweke-spectral Granger-causality statistic (25)
Seth[126]		Implemented for single and multiple trials	Partial Granger causality[7,134] Granger autonomy[127,128] Causal density[129,130]

Abbreviation: TV, time varying.

TABLE 4 List of available toolboxes for estimating time-varying VAR models and Granger-causality statistics

Toolbox	Software	TV-VAR implemented estimation method	Implemented statistic of causality
eConnectome	Matlab	Kalman filter estimation type (same mvaar.m Matlab function as BioSig toolbox)	Directed transfer function (26)
He et al[131]		Implemented for single trial only	Adaptive version of directed transfer function[104]
		Variants for estimating the covariance matrices R_t and Q_t are implemented based on[95]	
SIFT (Source Information Flow Toolbox[20])	Matlab	Windowing approach based on Ding et al[90]	Partial directed coherence (27)
		Implemented for single and multiple trials	Generalized partial directed coherence[132]
			Renormalized partial directed coherence[70]
		Kalman filter estimation type (same mvaar.m Matlab function as BioSig toolbox)	Directed transfer function (26)
		Implemented for single trial only	Full frequency directed transfer function[133]
			Geweke-Granger-causality (25)
GEDI (Gene expression data interpreter)	R	Wavelet dynamic vector autoregressive estimation method (section6.2.4)	Granger-causality criterion 2 (12) and Wald statistic (14)[123]
Fujita et al.[123]			
MSGranger	Matlab	Bayesian Multiscale state-space model	time-varying frequency-specific Granger-causality
Cekic et al.[122]			

transformation of conditional Granger causality into the frequency domain is problematic, as it may contain negative values with no meaning in terms of causality (as pointed out in Ding et al[21] and partially solved for 3 variables in Chen et al[60] with the proposed "partition matrix technique").

Finally, an important extension is partial Granger causality. As described in Bressler and Seth,[7,126] all brain connectivity analyses involve variable selection, in which the relevant set of recording brain regions is selected for the analysis. In practice, this step may exclude some relevant variables. The lack of exogenous and latent inputs in the model can lead to the detection of apparent causal interactions that are actually spurious. The response of Guo et al[134] to this challenge is what is called partial Granger causality. This is based on the same intuition as partial coherence, namely, that the influence of exogenous and/or latent variables on a recorded system will be highlighted by the correlations among residuals of the VAR modelling of the selected measured variables. Guo et al[134] also provide an extension in the frequency domain. An alternative to the development of partial Granger-causality methods, already mentioned in Section 5, is the methods allowing us to deal with instantaneous effects between time series.[56,57] Faes et al,[57] moreover, provide an extension to these methods in the frequency domain.

8.2 | EEG and fMRI application

The application of Granger-causality methods to fMRI data is very promising, given the high spatial resolution of the fMRI blood-oxygen-level dependent (BOLD) signal.[7,126]

However, as explained in Seth et al,[9] Granger-causality analysis of fMRI data has been highly controversial due to the indirect relationship between the fMRI BOLD signal and the underlying neural processes. Functional magnetic resonance imaging responses are indeed a convolution with an hemodynamic response function (HRF), which implies long delays compared to real neural activity and which moreover may have significant interregional and interindividual variability. Several findings indicate that the BOLD signal might also be biased for specific kinds of neuronal activities (higher BOLD response for gamma range compared to lower frequencies for example[135]). However, because the HRF acts as a filter,

Granger-causality analysis should be invariant to its variability. But, for this invariance to apply in practice, the sampling rate of the signal should be of the same order as the neuronal delays, which is not currently feasible with fMRI data.[136] Seth et al[136] showed that Granger-causality analysis of downsampled and convolved data can lead to increasing type I and II errors. Wen et al,[137] however, showed that a monotonic relationship is preserved between Granger-causality results observed at the neural level and in simulated BOLD signals, and that under several convolution and sampling rates parameters in a bivariate situations. Seth et al,[9] however, argued that although Granger-causality analysis is invariant to HRF variability given sufficiently fast sampling rate and low measurement noise, current applications of Granger causality to fMRI should be treated cautiously .

The very high time resolution offered by magnetoencephalography (MEG), scalp electroencephalography (EEG), or intracranial electroencephalography (iEEG) methods allows the application of Granger causality to be very powerful.[7] An application of spectral Granger-causality statistics for discovering causal relationships at different frequencies in MEG and scalp EEG data can be found for example in Astolfi et al,[138] Bressler et al,[61] and Brovelli et al[58] A key problem with the application of Granger-causality methods to MEG, scalp EEG, or iEEG data is the introduction of causal artefacts during the preprocessing. Bandpass filtering, for example, can cause severe confounding in Granger-causality analysis by introducing temporal correlations in the data and including future information in present (smoothed) data.[76,126]

Scalp EEG also poses the problem of volume conduction. Kaminski and Blinowska[139] argued that the DTF measure (see section 5.2) is not influenced by the volume conduction because it is a measure of phase difference between 2 channels, and that for the same reasons, preprocessing procedures are not needed. On the contrary, Brunner et al[140] argued that both the DTF and the PDC are adversely affected by volume conduction from multiple sources to the scalp electrodes and, therefore, that in general, application of connectivity measures to scalp EEG signals without correcting for the volume conduction effect does not allows a clear interpretation in terms of underlying source dynamics. Going in the same direction, Van de Steen et al[141] argued that time-domain Granger causality and DTF applied on scalp EEG time series do not allow interpretation in terms of interacting brain sources due to the volume conduction and to the fact that spurious connectivity can occur between sensors. Indeed, they showed that mixing effects due to volume conduction can lead to spurious causal connections. They concluded that time-domain GC and DTF should therefore be computed at the source level (eg, on iEEG data) or derived within an analysis framework that takes into account the volume conduction effect.[141] Finally, in their recent article, Barnett and Seth[142] conducted a simulation study to investigate how volume conduction affects the causal connectivity estimation on scalp EEG with temporal GC, DTF, PDC and phase-slope index.[143] Their interesting results showed that only the PSI statistic is able to correctly detect significant information flows between signals. However, the problem with PSI is that it cannot detect bidirectional (feedback) causal relations, because of the nature of the PSI definition itself.

8.3 | Neuroscience data specificities

As described in Vicente et al,[39] neuroscience data have specific characteristics that complicates their analysis in terms of effective connectivity.

First, as highlighted in Barnett and Seth,[142] there are numerous problems associated with Granger-causal inference from subsampled data. Subsampling may indeed induce type I and II errors and therefore lead to wrong Granger causality inference (assessment of spurious causality, where Granger causality is absent at the finer time scale but assessed nonzero for the subsampled process and undetectable causality, where Granger causality is present at the finer time scale and assessed nonsignificant for the subsampled process). The reader is refereed to Barnett and Seth[142] and references therein for further reading on Granger causality and associated subsampling problems.

Furthermore, the causal interaction between 2 signals may not be instantaneous but delayed over a certain time interval (υ), so the history of the variables Y and X in Equation 5 has to be taken from time $t-\upsilon-1$ to $t-\upsilon-p$, instead of from time $t-1$ to $t-p$, depending on the research hypothesis.

The last very important issue reported here is the time-lag τ between the data points in the history of Y and X, which permits more parsimonious models. Choosing a certain time-lag parameter means that the causal history of variables Y and X should be selected by taking the time-points from $t-\upsilon-1$ to $t-\upsilon-\tau p$, all of them being spaced by a lag τ. This is a very useful tool for dealing with high- or low-frequency modulations of the data, as high frequency phenomena needs a small time lag and conversely for low-frequency phenomena. This time-lag parameter τ has a clear and interpretable influence on Granger-causality statistics in the time domain, which directly relies on the estimated VAR parameters. It is however very difficult to see what its impact is on the frequency-domain causality statistics, where the time-domain parameter estimates are Fourier transformed and only then interpreted as a causality measure at each

frequency. Barnett and Seth[142] carefully analysed the interactions and links between the time-lag parameter τ (called the embedding delay) and the sampling frequency, and they were able to identify critical relationships between causal delay, sampling interval, and detectability of Granger causality (see Barnett and Seth[142] and references therein for further reading).

8.4 | Asymptotic distributions

As we have seen in Sections 4 and 5, time-domain Granger-causality statistics in Equations 9 and 14 asymptotically follow F and χ^2 distributions. Frequency-domain causality statistics in Equations 26 and 27 are both asymptotically related to a χ^2 distribution. "Asymptotic" here means when the number of observations T goes to infinity.

These distributions have the advantage of requiring very little computational time compared to bootstrap or permutation surrogate statistics. However, one has to be aware that all these properties are derived from the asymptotic properties of the VAR estimated coefficients. They are thus accurate only if the assumptions behind VAR modelling are fulfilled. They also may be very approximate when the number of sample points is not large enough.

Since in neuroscience causal hypotheses are often numerous (in terms of number of channels or/and number of specific hypothesis to test), these distributions can nonetheless provide a very useful tool allowing us to rapidly check for statistical significance of several causality hypotheses. They thus offer a quick overview of the overall causal relationships.

Another important aspect is that the tests based either on the asymptotic distributions or on resampling are only pointwise significance tests. Therefore, when jointly testing a collection of values for a complete time or frequency or time-frequency connectivity map, it is important to suitably correct the significance threshold for multiple comparisons.

9 | CONCLUSION

Neuroscience hypotheses are often relatively complex, such as asking about time-varying causal relationships specific to certain frequency bands and even sometimes between different frequency bands (so-called cross-frequency coupling).

Granger causality is a promising statistical tool for dealing with some of these complicated research questions about effective connectivity. However, the postulated models behind have to be suitably estimated in order to derive accurate statistics.

In this article, we have reviewed and described existing Granger-causality statistics and focused on model estimation methods that possess a time-varying extension. Time-varying Granger causality is of primary interest in neuroscience since recorded data are intrinsically nonstationary. However, its implementation is not trivial as it depends on the complex estimate of time-varying densities. We reviewed existing methods providing time-varying Granger-causality statistics and discussed their qualities, limits, and drawbacks.

ACKNOWLEDGEMENTS

The authors would like to thank the 2 referees for countless suggestions that improved the manuscript and gratefully acknowledge the Swiss National Science Foundation, project number 100014_156493, and Lifebrain, grant number 732592 - H2020-SC1-2016-2017/H2020-SC1-2016-RTD.

ORCID

Sezen Cekic http://orcid.org/0000-0001-6767-5465

REFERENCES

1. Granger CWJ. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*. 1969;37(3):424-438. https://doi.org/10.2307/1912791
2. Granger CWJ. Testing for causality: a personal viewpoint. *J Econ Dyn Control*. 1980;2:329-352. https://doi.org/10.1016/0165-1889(80)90069-X
3. Wiener N. The theory of prediction. In: Beckenbach EF, Ed. *Modern Mathematics for Engineers*. New York: McGraw Hill; 1956:165-190.
4. Boňil T, Sovka P. Autoregressive causal relation: digital filtering approach to causality measures in frequency domain. *Digit Signal Process*. 2013;23(5):1756-1766. https://doi.org/10.1016/j.dsp.2013.04.006
5. Pearl J. Causal inference in statistics: an overview. *Stat Surv*. 2009;3:96-146. https://doi.org/10.1214/09-SS057
6. Pereda E, Quiroga RQ, Bhattacharya J. Nonlinear multivariate analysis of neurophysiological signals. *Prog Neurobiol*. 2005;77(12):1-37. https://doi.org/10.1016/j.pneurobio.2005.10.003

7. Bressler SL, Seth AK. Wiener–Granger Causality: a well established methodology. *NeuroImage*. 2011;58(2):323-329. https://doi.org/10.1016/j.neuroimage.2010.02.059
8. Porta A, Faes L. Wiener-Granger causality in network physiology with applications to cardiovascular control and neuroscience. *Proc IEEE*. 2016;104(2):282-309. https://doi.org/10.1109/JPROC.2015.2476824
9. Seth AK, Barrett AB, Barnett L. Granger causality analysis in neuroscience and neuroimaging. *J Neurosci*. 2015;35(8):3293-3297. https://doi.org/10.1523/JNEUROSCI.4399-14.2015
10. Kleinberg S, Hripcsak G. A review of causal inference for biomedical informatics. *J Biomed Inform*. 2011;44(6):1102-1112. https://doi.org/10.1016/j.jbi.2011.07.001
11. Amblard P-O, Michel OJJ. The relation between Granger causality and directed information theory: a review. *Entropy*. 2012;15(1):113-143. https://doi.org/10.3390/e15010113
12. Müller A, Kraemer JF, Penzel T, Bonnemeier H, Kurths J, Wessel N. Causality in physiological signals. *Physiol Meas*. 2016;37(5):R46-R72. https://doi.org/10.1088/0967-3334/37/5/r46
13. Sameshima K, Baccala LA. *Methods in Brain Connectivity Inference Through Multivariate Time Series Analysis*. Boca Raton, FL: CRC press; 2014.
14. Pearl J. Causal diagrams for empirical research. *Biometrika*. 1995;82(4):669-688. https://doi.org/10.1093/biomet/82.4.669
15. Friston KJ, Harrison L, Penny W. Dynamic causal modelling. *NeuroImage*. 2003;19(4):1273-1302. https://doi.org/10.1016/s1053-8119(03)00202-7
16. Friston K, Moran R, Seth AK. Analysing connectivity with Granger causality and dynamic causal modelling. *Curr Opin Neurobiol*. 2013;23(2):172-178. https://doi.org/10.1016/j.conb.2012.11.010
17. McLntosh AR, Gonzalez-Lima F. Structural equation modeling and its application to network analysis in functional brain imaging. *Hum Brain Mapp*. 1994;2(1-2):2-22. https://doi.org/10.1002/hbm.460020104
18. Sugihara G, May R, Ye H, et al. Detecting causality in complex ecosystems. *Science*. 2012;338(6106):496-500. https://doi.org/10.1126/science.1227079
19. Eichler M. Granger causality and path diagrams for multivariate time series. *J Econom*. 2007;137(2):334-353. https://doi.org/10.1016/j.jeconom.2005.06.032
20. Delorme A, Mullen T, Kothe C, et al. EEGLAB, SIFT, NFT, BCILAB, and ERICA: new tools for advanced EEG processing. *Comput Intell Neurosci*. 2011;2011:1-12. https://doi.org/10.1155/2011/130714
21. Ding M, Chen Y, Bressler SL. Granger causality: basic theory and application to neuroscience. eprint arXiv:q-bio/0608035; 2006.
22. Chamberlain G. The general equivalence of Granger and Sims causality. *Econometrica*. 1982;50(3):569-581. https://doi.org/10.2307/1912601
23. Florens J-P. Some technical issues in defining causality. *J Econom*. 2003;112(1):127-128. https://doi.org/10.1016/s0304-4076(02)00155-0
24. Chicharro D. On the spectral formulation of Granger causality. *Biol Cybern*. 2011;105(5-6):331-347. https://doi.org/10.1007/s00422-011-0469-z
25. Kuersteiner GM. Granger-Sims causality. In: Durlauf SN, Blume LE, Eds. *The New Palgrave Dictionary of Economics*. Basingstoke: Palgrave Macmillan; 2010:119-134. https://doi.org/10.1057/9780230280830_14
26. Geweke J. Measurement of linear dependence and feedback between multiple time series. *J Am Stat Assoc*. 1982;77(378):304-313. https://doi.org/10.2307/2287238
27. Hamilton JD. *Time Series Analysis*. Princeton, NJ: Princeton University Press; 1994.
28. Lütkepohl H. *New Introduction to Multiple Time Series Analysis*. Cambridge, UK: Cambridge University Press; 2005. https://doi.org/10.1007/978-3-540-27752-1
29. Hesse W, Möller E, Arnold M, Schack BÃ. The use of time-variant EEG Granger causality for inspecting directed interdependencies of neural assemblies. *J Neurosci Methods*. 2003;124(1):27-44. https://doi.org/10.1016/s0165-0270(02)00366-7
30. Goebel R, Roebroeck A, Kim D-S, Formisano E. Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping. *Magn Reson Imaging*. 2003;21(10):1251-1261. https://doi.org/10.1016/j.mri.2003.08.026
31. Geweke J. Inference and causality in economic time series models. *Handb Econom*. 1984;2:1101-1144. https://doi.org/10.1016/s1573-4412(84)02011-0
32. Sato JR, Junior EA, Takahashi DY, de Maria Felix M, Brammer MJ, Morettin PA. A method to produce evolving functional connectivity maps during the course of an fM R I experiment using wavelet-based time-varying Granger causality. *NeuroImage*. 2006;31(1):187-196. https://doi.org/10.1523/JNEUROSCI.4399-14.2015
33. Akaike H. A new look at the statistical model identification. *IEEE Trans Autom Control*. 1974;19(6):215-222. https://doi.org/10.1007/978-1-4612-1694-0_16
34. Schwarz G. Estimating the dimension of a model. *Ann Stat*. 1978;6(2):461-464. https://doi.org/10.1214/aos/1176344136
35. Schreiber T. Measuring information transfer. *Phys Rev Lett*. 2000;85(2):461-464. https://doi.org/10.1103/physrevlett.85.461
36. Barnett L, Barrett AB, Seth AK. Granger causality and transfer entropy are equivalent for Gaussian variables. *Phys Rev Lett*. 2009;103(23):238701. https://doi.org/10.1103/physrevlett.103.238701
37. Chàvez M, Martinerie J, Le Van Quyen M. Statistical assessment of nonlinear causality: application to epileptic EEG signals. *J Neurosci Methods*. 2003;124(2):113-128. https://doi.org/10.1016/s0165-0270(02)00367-9
38. Garofalo M, Nieus T, Massobrio P, Martinoia S. Evaluation of the performance of information theory-based methods and cross-correlation to estimate the functional connectivity in cortical networks. *PLoS ONE*. 2009;4(8):e6482. https://doi.org/10.1371/journal.pone.0006482

39. Vicente R, Wibral M, Lindner M, Pipa G. Transfer entropy—a model-free measure of effective connectivity for the neurosciences. *J Comput Neurosci*. 2011;30(1):45-67. https://doi.org/10.1007/s10827-010-0262-3
40. Wibral M, Rahm B, Rieder M, Lindner M, Vicente R, Kaiser J. Transfer entropy in magnetoencephalographic data: quantifying information flow in cortical and cerebellar networks. *Prog Biophys Mol Biol*. 2011;105(12):80-97. https://doi.org/10.1016/j.pbiomolbio.2010.11.006
41. Lizier JT, Heinzle J, Horstmann A, Haynes J-D, Prokopenko M. Multivariate information-theoretic measures reveal directed information structure and task relevant changes in fMRI connectivity. *J Comput Neurosci*. 2011;30(1):85-107. https://doi.org/10.1007/s10827-010-0271-2
42. Besserve M, Martinerie J. Extraction of functional information from ongoing brain electrical activity. *IRBM*. 2011;32(1):27-34. https://doi.org/10.1016/j.irbm.2011.01.001
43. Besserve M, Schölkopf B, Logothetis N, Panzeri S. Causal relationships between frequency bands of extracellular signals in visual cortex revealed by an information theoretic analysis. *J Comput Neurosci*. 2010;29(3):547-566. https://doi.org/10.1007/s10827-010-0236-5
44. Montalto A, Faes L, Marinazzo D. MuTE: A MATLAB toolbox to compare established and novel estimators of the multivariate transfer entropy. *PLOS ONE*. October 2014;9(10):1-13. https://doi.org/10.1371/journal.pone.0109462
45. Hlaváčková-Schindler K, Paluš M, Vejmelka M, Bhattacharya J. Causality detection based on information-theoretic approaches in time series analysis. *Phys Rep*. 2007;441(1):1-46. https://doi.org/10.1016/j.physrep.2006.12.004
46. Wibral M, Vicente R, Lizier JT. *Directed Information Measures in Neuroscience*. Berlin: Springer; 2014.
47. Victor JD. Binless strategies for estimation of information from neural data. *Phys Rev E*. 2002;66(5):051903. https://doi.org/10.1103/physreve.66.051903
48. Kaiser A, Schreiber T. Information transfer in continuous processes. *Physica D: Nonlinear Phenom*. 2002;166(1-2):43-62. https://doi.org/10.1016/s0167-2789(02)00432-3
49. Lindner M, Vicente R, Priesemann V, Wibral M. TRENTOOL: a Matlab open source toolbox to analyse information flow in time series data with transfer entropy. *BMC Neurosci*. 2011;12(1):119. https://doi.org/10.1186/1471-2202-12-119
50. Kugiumtzis D. Direct-coupling information measure from nonuniform embedding. *Phys Rev E*. 2013;87(6):062918. https://doi.org/10.1103/physreve.87.062918
51. Faes L, Kugiumtzis D, Nollo G, Jurysta F, Marinazzo D. Estimating the decomposition of predictive information in multivariate systems. *Phys Rev E*. 2015;91(3):032904. https://doi.org/10.1103/physreve.91.032904
52. Runge J, Heitzig J, Petoukhov V, Kurths J. Escaping the curse of dimensionality in estimating multivariate transfer entropy. *Phys Rev Lett*. June 2012;108:258701. https://doi.org/10.1103/PhysRevLett.108.258701
53. Faes L, Nollo G, Porta A. Information-based detection of nonlinear Granger causality in multivariate processes via a nonuniform embedding technique. *Phys Rev E*. 2011;83(5):051112. https://doi.org/10.1103/physreve.83.051112
54. Cao L. Practical method for determining the minimum embedding dimension of a scalar time series. *Physica D: Nonlinear Phenom*. 1997;110(1–2):43-50. https://doi.org/10.1016/s0167-2789(97)00118-8
55. Geweke JF. Measures of conditional linear dependence and feedback between time series. *J Am Stat Assoc*. 1984;79(388):907-915. https://doi.org/10.2307/2288723
56. Hyvärinen A, Zhang K, Shimizu S, Hoyer PO. Estimation of a structural vector autoregression model using non-Gaussianity. *J Mach Learn Res*. May 2010;11:1709-1731.
57. Faes L, Erla S, Porta A, Nollo G. A framework for assessing frequency domain causality in physiological time series with instantaneous effects. *Philos Trans Royal Soc London A Math Phys Eng Sci*. 2013;371(1997). https://doi.org/10.1098/rsta.2011.0618.
58. Brovelli A, Ding M, Ledberg A, Chen Y, Nakamura R, Bressler SL. Beta oscillations in a large-scale sensorimotor cortical network: Directional influences revealed by Granger causality. *Proc Nat Acad Sci U.S.A.* 2004;101(26):9849-9854. https://doi.org/10.1073/pnas.0308538101
59. Chen Y, Bressler SL, Ding M. Dynamics on networks: assessing functional connectivity with Granger causality. *Computational and Mathematical Organization Theory*. 2009;15(4):329-350.
60. Chen Y, Bressler SL, Ding M. Frequency decomposition of conditional Granger causality and application to multivariate neural field potential data. *J Neurosci Methods*. 2006;150(2):228-237. https://doi.org/10.1016/j.jneumeth.2005.06.011
61. Bressler SL, Richter CG, Chen Y, Ding M. Cortical functional network organization from autoregressive modeling of local field potential oscillations. *Stat Med*. 2007;26(21):3875-3885. https://doi.org/10.1002/sim.2935
62. Lin F-H, Hara K, Solo V, et al. Dynamic Granger-Geweke causality modeling with application to interictal spike propagation. *Hum Brain Mapp*. 2009;30(6):1877-1886. https://doi.org/10.1002/hbm.20772
63. Gourévitch B, Bouquin Jeannès R, Faucon G. Linear and nonlinear causality between signals: methods, examples and neurophysiological applications. *Biol Cybern*. 2006;95(4):349-369. https://doi.org/10.1007/s00422-006-0098-0
64. Baccalà LA, Sameshima K, Ballester G, Do Valle AC, Timo-Iaria C. Studying the interaction between brain structures via directed coherence and Granger causality. *Appl Signal Process*. 1998;5:40-48. https://doi.org/10.1007/s005290050005
65. Faes L, Erla S, Nollo G. Measuring connectivity in linear multivariate processes: definitions, interpretation, and practical analysis. *Comput Math Methods Med*. 2012;2012:Article ID 140513. https://doi.org/10.1155/2012/140513
66. Faes L, Nollo G. Measuring frequency domain Granger causality for multiple blocks of interacting time series. *Biol Cybern*. 2013;107(2):217-232. https://doi.org/10.1007/s00422-013-0547-5
67. Kamiński M, Ding M, Truccolo WA, Bressler SL. Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. *Biol Cybern*. 2001;85(2):145-157. https://doi.org/10.1007/s004220000235
68. Kamiński M. Multichannel data analysis in biomedical research, *Handbook of Brain Connectivity*. Berlin: Springer; 2007:327-355. https://doi.org/10.1007/978-3-540-71512-2_11

69. Baccalà LA, Sameshima K. Partial directed coherence: a new concept in neural structure determination. *Biol Cybern*. 2001;84(6):463-474. https://doi.org/10.1007/pl00007990
70. Schelter B, Timmer J, Eichler M. Assessing the strength of directed influences among neural signals using renormalized partial directed coherence. *J Neurosci Res*. 2009;179(1):121-130. https://doi.org/10.1016/j.jneumeth.2009.01.006
71. Schelter B, Winterhalder M, Eichler M, et al. Testing for directed influences among neural signals using partial directed coherence. *J Neurosci Res*. 2006;152(1):210-219. https://doi.org/10.1523/JNEUROSCI.4399-14.2015
72. Eichler M. On the evaluation of information flow in multivariate systems by the directed transfer function. *Biol Cybern*. 2006;94(6):469-482. https://doi.org/10.1007/s00422-006-0062-z
73. Winterhalder M, Schelter, Hesse W, et al. Comparison of linear signal processing techniques to infer directed interactions in multivariate neural systems. *Signal Process*. 2005;85(11):2137-2160. https://doi.org/10.1016/j.sigpro.2005.07.011
74. Winterhalder M, Schelter B, Hesse W, et al. Detection of directed information flow in biosignals. *Biomed Tech*. 2006;51(5–6):281-287. https://doi.org/10.1515/bmt.2006.058
75. Sims CA. Money, income, and causality. *Am Econ Rev*. 1972;62(4):540-552. https://doi.org/10.2307/1806097
76. Florin E, Gross J, Pfeifer J, Fink GR, Timmermann L. The effect of filtering on Granger causality based multivariate causality measures. *NeuroImage*. 2010;50(2):577-588. https://doi.org/10.1016/j.neuroimage.2009.12.050
77. Barnett L, Seth AK. Behaviour of Granger causality under filtering: theoretical invariance and practical application. *J Neurosci Methods*. 2011;201(2):404-419. https://doi.org/10.1016/j.jneumeth.2011.08.010
78. Baccalá LA, de Brito CSN, Takahashi DY, Sameshima K. Unified asymptotic theory for all partial directed coherence forms. *Philos Trans Royal Soc London A: Math Phys Eng Sci*. 2013;371(1997). https://doi.org/10.1098/rsta.2012.0158.
79. Faes L, Porta A, Nollo G. Testing frequency-domain causality in multivariate time series. *IEEE Trans Biomed Eng*. August 2010;57(8):1897-1906. https://doi.org/10.1109/TBME.2010.2042715
80. Möller E, Schack BÃ, Arnold M, Witte H. Instantaneous multivariate EEG coherence analysis by means of adaptive high-dimensional autoregressive models. *J Neurosci Res*. 2001;105(2):143-158. https://doi.org/10.1016/s0165-0270(00)00350-2
81. Schelter B. *Quantification of Directed Signal transfer within Neural Networks by Partial Directed Coherence: A Novel Approach to Infer Causal Time-depending Influences in Noisy, Multivariate Time Series*, DFG-Schwerpunktprogramm 1114, Mathematical methods for time series analysis and digital image processing: Zentrum für Technomathematik; 2005.
82. Dhamala M, Rangarajan G, Ding M. Analyzing information flow in brain networks with nonparametric Granger causality. *NeuroImage*. 2008;41(2):354-362. https://doi.org/10.1016/j.neuroimage.2008.02.020
83. Lizier JT, Prokopenko M, Zomaya AY. Local information transfer as a spatiotemporal filter for complex systems. *Phys Rev E*. 2008;77(2):026110. https://doi.org/10.1103/PhysRevE.77.026110
84. Prokopenko M, Lizier J, Price D. On thermodynamic interpretation of transfer entropy. *Entropy*. 2013;15(2):524-543. https://doi.org/10.3390/e15020524
85. Wollstadt P, Martïnez-Zarzuela M, Vicente R, Dïaz-Pernas FJ, Wibral M. Efficient transfer entropy analysis of non-stationary neural time series. *PLOS ONE*. July 2014;9(7):1-21. https://doi.org/10.1371/journal.pone.0102833
86. Chen L. Vector time-varying autoregressive (TVAR) models and their application to downburst wind speeds. *Ph.D. Thesis*; Texas Tech University; 2005.
87. Dahlhaus R. Fitting time series models to nonstationary processes. *Ann Stat*. 1997;25(1):1-37. https://doi.org/10.1214/aos/1034276620
88. Priestley MB. Evolutionary spectra and non-stationary processes. *J R Stat Soc Series B Stat Methodol*. 1965;27(2):204-237.
89. Ledolter J. Recursive estimation and adaptive forecasting in ARIMA models with time varying coefficients. *Applied Time Series Analysis, II (Tulsa, Okla)*. New York-London: Academic Press; 1981:449-471. https://doi.org/10.1016/b978-0-12-256420-8.50020-4
90. Ding M, Bressler SL, Yang W, Liang H. Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modeling: data preprocessing, model validation, and variability assessment. *Biol Cybern*. 2000;83(1):35-45. https://doi.org/10.1007/s004229900137
91. Long CJ, Brown EN, Triantafyllou C, Aharon I, Wald LL, Solo V. Nonstationary noise estimation in functional MRI. *NeuroImage*. 2005;28(4):890-903. https://doi.org/10.1016/j.neuroimage.2005.06.043
92. Hoerzer GM, Liebe S, Schloegl A, Logothetis NK, Rainer G. Directed coupling in local field potentials of macaque V4 during visual short-term memory revealed by multivariate autoregressive models. *Front Comput Neurosci*. 2010;4:12. https://doi.org/10.3389/fncom.2010.00014
93. Solo V, Purdon P, Weisskoff R, Brown E. A signal estimation approach to functional MRI. *IEEE Trans Med Imaging*. 2001;20(1):26-35. https://doi.org/10.1109/42.906422
94. Cekic S. Lien entre activité neuronale des sites cérébraux de l'amygdale et du cortex orbito-frontal en réponse à une prosodie émotionnelle: investigation par la Granger-causalité. *Master Thesis*; University of Geneva; 2010.
95. Schlögl A. The electroencephalogram and the adapdative autoregressive model: theory and applications. *Ph.D. Thesis*. Graz; 2000.
96. Schack B, Witte H, Griessbach G. Parametrische methoden der dynamischen spektralanalyse und ihre Anwendung in der biosignalanalyse. *Biomed Tech Eng*. 1993;38(s1):79-80. https://doi.org/10.1515/bmte.1993.38.s1.79
97. Mainardi LT, Bianchi AM, Baselli G, Cerutti S. Pole-tracking algorithms for the extraction of time-variant heart rate variability spectral parameters. *IEEE Trans Biomed Eng*. 1995;42(3):250-259. https://doi.org/10.1109/10.364511
98. Patomaki L, Kaipio JP, Karjalainen PA, Juntunen M. Tracking of nonstationary EEG with the polynomial root perturbation. In: Proceedings of the 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 1996. Bridging Disciplines for Biomedicine, Vol. 3. 1996; Amsterdam, Netherlands. 939-940. https://doi.org/10.1109/IEMBS.1996.652650

99. Patomaki L, Kaipio JP, Karjalainen PA. Tracking of nonstationary EEG with the roots of ARMA models. In: IEEE 17th Annual Conference Engineering in Medicine and Biology Society 1995. Vol. 2. 1995; Montreal, QC, Canada. 877-878 vol.2. https://doi.org/10.1109/IEMBS.1995.579249.
100. Akay M. *Biomedical Signal Processing*. San Diego, CA: Academic Press; 1994.
101. Astolfi L, Cincotti F, Mattia D, et al. Tracking the time-varying cortical connectivity patterns by adaptive multivariate estimators. *IEEE Trans Biomed Eng*. 2008;55(3):902-913. https://doi.org/10.1109/tbme.2007.905419
102. Astolfi L, Cincotti F, Mattia D, et al. Time-varying cortical connectivity estimation from noninvasive, high-resolution EEG recordings. *J Psychophysiol*. 2010;24(2):83-90. https://doi.org/10.1027/0269-8803/a000017
103. Tarvainen MP, Hiltunen JK, Ranta aho PO, Karjalainen PA. Estimation of nonstationary EEG with Kalman smoother approach: an application to event-related synchronization (ERS). *IEEE Trans Biomed Eng*. 2004;51(3):516-524. https://doi.org/10.1109/tbme.2003.821029.
104. Wilke C, Ding L, He B. Estimation of time-varying connectivity patterns through the use of an adaptive directed transfer function. *IEEE Trans Biomed Eng*. 2008;55(11):2557-2564. https://doi.org/10.1109/TBME.2008.919885
105. Bianchi AM, Mainardi LT, Meloni C, Chierchiu S, Cerutti S. Continuous monitoring of the sympatho-vagal balance through spectral analysis. *IEEE Eng Med Biol Mag*. 1997;16(5):64-73. https://doi.org/10.1109/51.620497
106. Costa AH, Hengstler S. Adaptive time-frequency analysis based on autoregressive modeling. *Sig Proc*. 2011;91(4):740-749. https://doi.org/10.1016/j.sigpro.2010.07.020
107. Schlögl A, Roberts SJ, Pfurtscheller G. A criterion for adaptive autoregressive models. In: Proceedings of the 22nd IEEE International Conference on Engineering in Medicine and Biology; 2000; Chicago, IL. 1581-1582.
108. Kalman RE. A new approach to linear filtering and prediction problems. *J Basic Eng*. 1960;82(1):35-45. https://doi.org/10.1115/1.3662552
109. Meinhold RJ, Singpurwalla ND. Understanding the Kalman filter. *Am Stat*. 1983;37(2):123-127. https://doi.org/10.1080/00031305.1983.10482723
110. Milde T, Leistritz L, Astolfi L, et al. A new Kalman filter approach for the estimation of high-dimensional time-variant multivariate AR models and its application in analysis of laser-evoked brain potentials. *NeuroImage*. 2010;50(3):960-969. https://doi.org/10.1016/j.neuroimage.2009.12.110
111. Isaksson A, Wennberg A, Zetterberg L-H. Computer analysis of EEG signals with parametric models. *Proc IEEE*. 1981;69(4):451-461. https://doi.org/10.1109/PROC.1981.11988
112. Haykin S, Sayed AH, Zeidler JR, Yee P, Wei PC. Adaptive tracking of linear time-variant systems by extended RLS algorithms. *IEEE Trans Signal Process*. 1997;45(5):1118-1128. https://doi.org/10.1109/78.575687
113. Jazwinski AH. Adaptive filtering. *Automatica*. 1969;5(4):475-485. https://doi.org/10.1016/0005-1098(69)90109-5
114. Penny WD, Roberts SJ. Dynamic linear models, recursive least squares and steepest descent learning, Department of Electrical and Electronic Engineering, Imperial College of Science and Technology and Medicine; 1998.
115. Kalman RE, Bucy RS. New results in linear filtering and prediction theory. *J Basic Eng*. 1961;83(1):95-108. https://doi.org/10.1115/1.3658902
116. Roebroeck A, Formisano E, Goebel R. Mapping directed influence over the brain using Granger causality and fMRI. *NeuroImage*. 2005;25(1):230-242. https://doi.org/10.1016/j.neuroimage.2004.11.017
117. Arnold M, Milner XHR, Witte H, Bauer R, Braun C. Adaptive AR modeling of nonstationary time series by means of Kalman filtering. *IEEE Trans Biomed Eng*. 1998;45(5):553-562. https://doi.org/10.1109/10.668741
118. Milde T, Putsche P, Schwab K, Wacker M, Eiselt M, Witte H. Dynamics of directed interactions between brain regions during interburst-burst EEG patterns in quiet sleep of full-term neonates. *Neurosci Lett*. 2011;488(2):148-153. https://doi.org/10.1016/j.neulet.2010.11.018
119. Havlicek M, Jan J, Brazdil M, Calhoun VD. Dynamic Granger causality based on Kalman filter for evaluation of functional network connectivity in fMRI data. *NeuroImage*. 2010;53(1):65-77. https://doi.org/10.1016/j.neuroimage.2010.05.063
120. Shumway RH, Stoffer DS. An approach to time series smoothing and forecasting using the EM algorithm. *J Time Ser Anal*. 1982;3(4):253-264. https://doi.org/10.1111/j.1467-9892.1982.tb00349.x
121. Cassidy W, Penny MJ. Bayesian nonstationary autoregressive models for biomedical signal analysis. *IEEE Trans Biomed Eng*. 2002;49(10):1142-1152. https://doi.org/10.1109/tbme.2002.803511
122. Cekic S, Grandjean D, Renaud O. Multiscale Bayesian state space model for Granger causality analysis of brain signal. ArXiv e-prints; 2017. https://arxiv.org/abs/1704.03177
123. Fujita A, Sato JR, Garay Malpartida HM, Morettin PA, Sogayar MC, Ferreira CE. Time-varying modeling of gene expression regulatory networks using the wavelet dynamic vector autoregressive method. *Bioinformatics*. 2007;23:1623-1630. https://doi.org/10.1093/bioinformatics/btm151
124. Cui J, Xu L, Bressler SL, Ding M, Liang H. BSMART: A Matlab/C toolbox for analysis of multichannel neural time series. *Neural Netw*. 2008;21(8):1094-1104. https://doi.org/10.1016/j.neunet.2008.05.007
125. Schlögl A, Brunner C. BioSig: a free and open source software library for BCI research. *Computer*. 2008;41(10):44-50. https://doi.org/10.1109/mc.2008.407
126. Seth AK. A MATLAB toolbox for Granger causal connectivity analysis. *J Neurosci Methods*. 2010;186(2):262-273. https://doi.org/10.1016/j.jneumeth.2009.11.020
127. Bertschinger N, Olbrich E, Ay N, Jost J. Autonomy: an information theoretic perspective. *Biosystems*. 2008;91(2):331-345. https://doi.org/10.1016/j.biosystems.2007.05.018

128. Seth AK. Measuring autonomy and emergence via Granger causality. *Artif Life*. 2010;16(2):179-196. https://doi.org/10.1162/artl.2010.16.2.16204
129. Seth AK. Causal connectivity of evolved neural networks during behavior. *Netw Comput Neural Syst*. 2005;16(1):35-54. https://doi.org/10.1080/09548980500238756
130. Seth AK. Causal networks in simulated neural systems. *Cogn Neurodyn*. 2008;2(1):49-64. https://doi.org/10.1007/s11571-007-9031-z
131. He B, Dai Y, Astolfi L, Babiloni F, Yuan H, Yang L. eConnectome: a MATLAB toolbox for mapping and imaging of brain functional connectivity. *J Neurosci Methods*. 2011;195(2):261-269. https://doi.org/10.1016/j.jneumeth.2010.11.015
132. Baccalà LA, de Medicina F. Generalized partial directed coherence. In: 2007 15th International Conference on Digital Signal Processing; 2007; Cardiff, UK. 163-166. https://doi.org/10.1109/icdsp.2007.4288544.
133. Korzeniewska A, Mańczak M, Kamiński M, Blinowska KJ, Kasicki S. Determination of information flow direction among brain structures by a modified directed transfer function (dDTF) method. *J Neurosci Methods*. 2003;125(1):195-207. https://doi.org/10.1016/s0165-0270(03)00052-9
134. Guo S, Seth AK, Kendrick KM, Zhou C, Feng J. Partial Granger causality—eliminating exogenous inputs and latent variables. *J Neurosci Methods*. 2008;172(1):79-93. https://doi.org/10.1016/j.jneumeth.2008.04.011
135. Niessing J, Ebisch B, Schmidt KE, Niessing M, Singer W, Galuske RAW. Hemodynamic signals correlate tightly with synchronized gamma oscillations. *Science*. 2005;309(5736):948-951. https://doi.org/10.1126/science.1110948
136. Seth AK, Chorley P, Barnett LC. Granger causality analysis of fMRI BOLD signals is invariant to hemodynamic convolution but not downsampling. *NeuroImage*. 2013;65:540-555. https://doi.org/10.1016/j.neuroimage.2012.09.049
137. Wen X, Rangarajan G, Ding M. Is Granger causality a viable technique for analyzing fMRI data? *PLOS ONE*. July 2013;8(7):1-11. https://doi.org/10.1371/journal.pone.0067428
138. Astolfi L, Cincotti F, Mattia D, et al. Comparison of different cortical connectivity estimators for high-resolution EEG recordings. *Hum Brain Mapp*. 2007;28(2):143-157. https://doi.org/10.1002/hbm.20263
139. Kaminski M, Blinowska KJ. Directed transfer function is not influenced by volume conduction—inexpedient pre-processing should be avoided. *Front Comput Neurosci*. 2014;8. https://doi.org/10.3389/fncom.2014.00061
140. Brunner C, Billinger M, Seeber M, Mullen TR, Makeig S. Volume conduction influences scalp-based connectivity estimates. *Front Comput Neurosci*. 2016;10. https://doi.org/10.3389/fncom.2016.00121
141. Van de Steen F, Faes L, Karahan E, Songsiri J, Valdes-Sosa PA, Marinazzo D. Critical comments on EEG sensor space dynamical connectivity analysis. *Brain Topogr*. 2016:1-12. https://doi.org/10.1007/s10548-016-0538-7
142. Barnett L, Seth AK. Detectability of Granger causality for subsampled continuous-time neurophysiological processes. *J Neurosci Methods*. 2017;275:93-121. https://doi.org/10.1016/j.jneumeth.2016.10.016
143. Nolte G, Ziehe A, Nikulin VV, et al. Robustly estimating the flow direction of information in complex physical systems. *Phys Rev Lett*. 2008;100:234101. https://doi.org/10.1103/PhysRevLett.100.234101

How to cite this article: Cekic S, Grandjean D, Renaud O. Time, frequency, and time-varying Granger-causality measures in neuroscience. *Statistics in Medicine*. 2018;37:1910–1931. https://doi.org/10.1002/sim.7621