

Effective connectivity: Influence, causality and biophysical modeling

Citation for published version (APA):

Valdes-Sosa, P. A., Roebroeck, A. F., Daunizeau, J., & Friston, K. (2011). Effective connectivity: Influence, causality and biophysical modeling. *Neuroimage*, 58(2), 339-361.
<https://doi.org/10.1016/j.neuroimage.2011.03.058>

Document status and date:

Published: 01/01/2011

DOI:

[10.1016/j.neuroimage.2011.03.058](https://doi.org/10.1016/j.neuroimage.2011.03.058)

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Taverne

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.



Comments and Controversies

Effective connectivity: Influence, causality and biophysical modeling

Pedro A. Valdes-Sosa^{a,*}, Alard Roebroeck^b, Jean Daunizeau^{c,d}, Karl Friston^c^a Cuban Neuroscience Center, Ave 25 #15202 esquina 158, Cubanacan, Playa, Cuba^b Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, The Netherlands^c The Wellcome Trust Centre for Neuroimaging, Institute of Neurology, UCL, 12 Queen Square, London, WC1N 3BG, UK^d Laboratory for Social and Neural Systems Research, Institute of Empirical Research in Economics, University of Zurich, Zurich, Switzerland

ARTICLE INFO

Article history:

Received 22 September 2010

Revised 15 March 2011

Accepted 23 March 2011

Available online 6 April 2011

Keywords:

Granger Causality

Effective connectivity

Dynamic Causal Modeling

EEG

fMRI

ABSTRACT

This is the final paper in a Comments and Controversies series dedicated to “The identification of interacting networks in the brain using fMRI: Model selection, causality and deconvolution”. We argue that discovering effective connectivity depends critically on state-space models with biophysically informed observation and state equations. These models have to be endowed with priors on unknown parameters and afford checks for model Identifiability. We consider the similarities and differences among Dynamic Causal Modeling, Granger Causal Modeling and other approaches. We establish links between past and current statistical causal modeling, in terms of Bayesian dependency graphs and Wiener–Akaike–Granger–Schweder influence measures. We show that some of the challenges faced in this field have promising solutions and speculate on future developments.

© 2011 Elsevier Inc. Open access under [CC BY license](http://creativecommons.org/licenses/by/3.0/).

Introduction

Following an empirical evaluation of effective connectivity measurements (David et al., 2008) and a primer on its implications (Friston, 2009a), the Comments and Controversy (C&C) exchange, initiated by Roebroeck et al. (2011b-this issue) and continued by David (2011-this issue), Friston (2011b-this issue), and Roebroeck et al. (2011a-this issue), has provided a lively and constructive discussion on the relative merits of two current techniques, Granger Causal Modeling (GCM)¹ and Dynamic Causal Modeling (DCM), for detecting effective connectivity using EEG/MEG and fMRI time series. The core papers of the C&C have been complemented by authoritative contributions (Bressler and Seth, 2011-this issue; Daunizeau et al., 2011a-this issue; Marinazzo et al., 2011-this issue) that clarify the state of the art for each approach.

This final paper in the series attempts to summarize the main points discussed and elaborate a conceptual framework for the analysis of effective connectivity (Figs. 1 and 2). Inferring effective connectivity comprises the successive steps of model specification, model identification and model (causal) inference (see Fig. 1). These steps are common to DCM, GCM and indeed any evidence-based inference. We will look at the choices made at each stage to clarify current areas of agreement and disagreement, of successes and shortcomings.

* Corresponding author at: Ave 25 #15202 esquina 158, Cubanacan, Playa, Apartado 6648 Habana 6 CP 10600, Cuba.

E-mail address: peter@cneuro.edu.cu (P.A. Valdes-Sosa).

¹ Note that GCM is also used as an acronym for Granger Causal Mapping (Roebroeck et al., 2005). To avoid confusion we shall use GCM mapping or the abbreviation GCMMap.

This entails a selective review of key issues and lines of work. Although an important area, we will not consider models that are just used to measure statistical associations (i.e. functional connectivity). In other words, we limit our focus to discovering effective connectivity (Friston, 2009a); that is causal relations between neural systems. Importantly, we hope to establish a clear terminology to eschew purely semantic discussions, and perhaps dispel some confusion in this regard. While preparing this material, we were struck with how easy it is to recapitulate heated arguments in other fields (such as econometrics), which were resolved several decades ago. We are also mindful of the importance of referring to prior work, to avoid repeating past mistakes² and to identify where more work is needed to address specific problems in the neurosciences.

We shall emphasize several times in this paper that causality is an epistemological concept that can be particularly difficult to capture with equations. This is because one's intuitive understanding of causality becomes inherently constrained whenever one tries to model it. In brief, one can think of causality in at least two distinct ways:

- Temporal precedence, i.e.: causes precede their consequences;
- Physical influence (control), i.e.: changing causes changes their consequences.

² “Progress, far from consisting in change, depends on retentiveness. When change is absolute there remains nothing to improve and no direction is set for possible improvement: and when experience is not retained, as among savages, infancy is perpetual. Those who cannot remember the past are condemned to repeat it.” George Santayana, *The Life of Reason, Volume 1, 1905*.

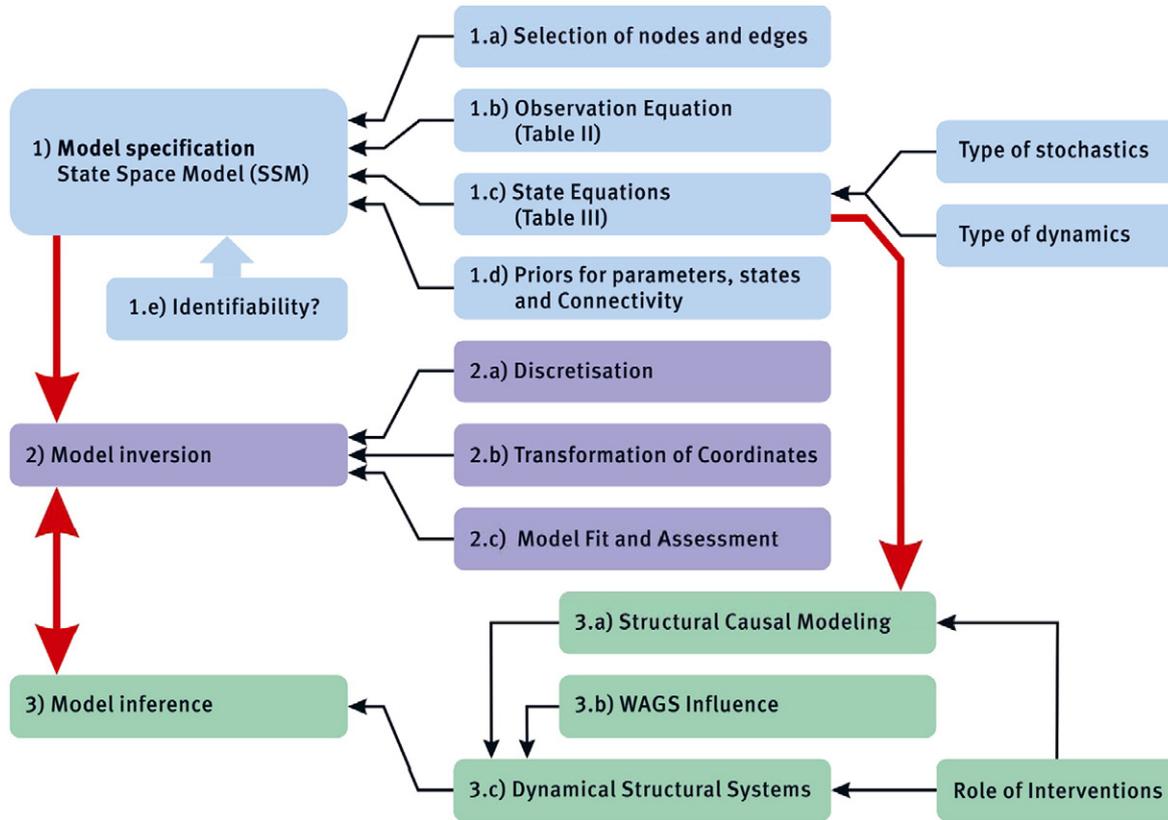


Fig. 1. Overview of causal modeling in Neuroimaging. Overall view of conceptual framework for defining and detecting effective connectivity in Neuroimaging studies.

This distinction is important, since it is the basis for any statistical detection of causal influence. In the context of brain connectivity, identifying causal relationships between two regions in the brain thus depends upon whether one tests for improvement in predictive capacity between temporally distinct neural events or one assesses the distal effect of (experimentally controlled) interventions.

Temporal precedence is the basis for Granger-like (what we call WAGS influence, see [WAGS influence](#) section) inferences about causality. In its simplest form, the idea is the following: A WAGS-causes B if one reduces the uncertainty about the future of B given the

past of A. Statistical tests of WAGS-causality thus rely upon information theoretic measures of predictability (of B given A).

In contradistinction, physical influence speaks to the notion of intervention and control, which has been formalized using a probabilistic framework called causal calculus (Pearl, 2000) ([Structural causal modeling: graphical models and Bayes-Nets](#) section). Observing (or estimating) activity at a network node potentially provides information about its effects at remote nodes. However, physically acting upon (e.g., fixing) this activity effectively removes any other physical influence this node receives. This means that inferences based on the effects of an intervention are somewhat different in nature from those based on purely observational effects. Generally speaking, inference on structural causality rests on modeling the effects of (controlled) experimental manipulations of the system, c.f. the popular quote ‘no causes in, no causes out’ (Cartwright, 2007). As we shall see later, these two approaches can be combined ([Dynamic structural causal modeling](#) section).

The structure of the paper is as follows. We first review the types of models used for studying effective connectivity. We then touch briefly on the methods used to invert and make inferences about these models. We then provide a brief summary of modern statistical causal modeling, list some current approaches in the literature and discuss their relevance to brain imaging. Finally, we list outstanding issues that could be addressed and state our conclusions.

Model specification

State-space models of effective connectivity

From the C&C discussion, there seems to be a consensus that discovering effective connectivity in Neuroimaging is essentially a comparison of generative models based on state-space models (SSM) of controllable (i.e., “causal” in a control theory sense) biophysical processes that have hidden neural states and possibly exogenous

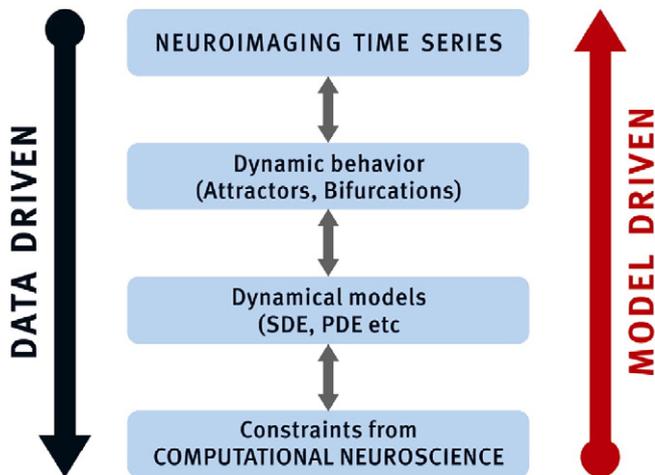


Fig. 2. Data and model driven approaches to causal modeling. Data driven approaches look for nonparametric models that not only fit the data but also describe important dynamical properties. They complement hypothesis driven approaches that are not only constrained by having to explain dynamical behavior but also provide links to computational models of brain function.

input. While having a long history in engineering (Candy, 2006; Kailath, 1980), SSM was only introduced recently for inference on hidden neural states (Valdes-Sosa et al., 1999; Valdes-Sosa et al., 1996; Valdés-Sosa et al., 2009a). For a comprehensive review of SSM and its application in Neuroscience see the forthcoming book (Ozaki, 2011).

Neural states describe the activity of a set of “nodes” that comprise a graph, the purpose of causal discovery being the identification of active links (edges or connections) in the graph. The nodes can be associated with neural populations at different levels; most commonly at the macroscopic (whole brain areas) or mesoscopic (sub-areas to cortical columns) level. These state-space models have unknown parameters (e.g., effective connectivity) and hyperparameters (e.g., the amplitude of random fluctuations). The specific model, states, parameters, hyperparameters and observables chosen determines the type of analysis and the nature of the final inference about causality. These choices are summarized in Fig. 1 (Step 1).

Given a set of observations or brain measurements, the first problem is: which data features are relevant for detecting causal influences? The most efficient way to address this question is to specify a *generative model*, i.e. a set of equations that quantify how observed data are affected by the presence of causal links. Put simply, this model translates the assumption of (i) *temporal precedence* or (ii) *physical influence* into how data should appear, given that (i) or (ii) is true. By presenting the data to generative models, model comparison can then be used to decide whether some causal link is likely to be present (by comparing models with and without that link). We now turn to the specification of generative models, in the form of a SSM.

Nodes and random variables

The first things we consider are the basic units or nodes, among which one wants to find causal links. These are usually modeled as macroscopic, coarse grained, ensembles of neurons, whose activity is summarized by a time varying state vector $x_r(t)$ or $x(r, t)$: $r \in R$. For example $x(t)$ could be instantaneous (ensemble average) post-synaptic membrane depolarization or pre-synaptic firing rate of neurons. The set R of nodes is usually taken as a small number of neural masses corresponding to pre-selected regions of interest (ROI) as is typical in both DCM and GCM. However, there has been recent interest in making R a continuous manifold (i.e. the cortex) that is approximated by a very high dimensional representation at the voxel level. We denote the complete set of random variables associated with each node as $X = \{X_i, X_i\}$ whose joint distribution is described using a generative model. X_i is the set of nodes without node i and $p(x) \triangleq p(X=x)$.

The observation equation

Any model always includes an explicit or implicit observation equation that generally varies with the imaging modality. This equation specifies how hidden (neuronal) states $x_r(t)$ produce observable data $y_q(t_k)$: $q \in Q$. This is the sensor data sampled at discrete time points $t_k = k\Delta$:

$$y_q(t_k) = g(x_r, t) + e(t_k) : r \in R_r, t \in [t_k, t_{k-1}] \quad (1)$$

for $k=1 \dots K$. It is important to note that observations at a given sensor q only reflect neural states from a subset of brain sites, modified by the function g over a time interval determined by the sampling period Δt and corrupted by instrumental noise $e(t_k)$. When the sampling period is not considered explicitly, the observations are denoted by $y_q(k)$. In most cases, this mapping does not need to be dynamic since there is no physical feedback from observed data to brain processes. In this special case, the observation equation reduces to an instantaneous transformation: $Y(t) = \tilde{g}(X(t))$, where \tilde{g} is

derived from g and any retarded (past) hidden states have been absorbed in $X(t)$ (e.g., to model hemodynamic convolutions).

A selected collection of observation equations used in Neuroimaging is provided in Table 1. The observation equation is sometimes simplified by assuming that observed data is a direct measurement of neural states (with negligible error). While this might be an acceptable assumption for invasive electrophysiological recordings, it is inappropriate in many other situations: for example, much of the activity in the brain is reflected in the EEG/MEG via the lead field with a resultant spatial smearing. For the BOLD signal, the C&C articles have discussed exhaustively the need to account for temporal smearing produced by the hemodynamic response function (HRF) when analyzing BOLD responses. This is important for fMRI because the sampling period is quite large with respect to the time course of neural events (we shall elaborate on this below).

Instrumental or sensor noise can seriously affect the results of causal analyses. One simple approach to causal modeling is to take the observation equation out of the picture by inverting the observation equation (i.e., mapping from data to hidden states). The estimated states can then be used for determining effective connectivity. This approach has been taken both for the EEG (Supp et al., 2007) and fMRI (David et al., 2008). However, this is suboptimal because it assumes that the causal modeling of hidden states is conditionally independent of the mapping from data. This is generally not the case (e.g., non-identifiability between observation and evolution processes described below). The optimal statistical procedure is to invert the complete generative model, including the observation and state equations modeling the evolution of hidden states. This properly accommodates conditional dependencies between parameters of the observer and state equations. A nice example of this is DCM for EEG and MEG, in which a SSM of coupled neuronal sources and a conventional electromagnetic forward model are inverted together. This means the parameters describing the spatial deployment of sources (e.g., dipole orientation and location) are optimized in relation to parameters controlling the effective connectivity among hidden sources. This sort of combined estimation has been described for simple noise models (Table 1-#2 by Nalatore et al. (2007)). For fMRI, DCM models the hemodynamic response with hidden physiological states like blood flow and volume and then uses a nonlinear observer function to generate BOLD responses (Table 2-#4). Early applications of GCM did not model the HRF but in recent years a number of papers have included explicit observation models in GCM (Ge et al., 2009; Havlicek et al., 2010), which have even incorporated the full nonlinear HRF model used in DCM (Havlicek et al., 2009; Havlicek et al., 2011).

The state equation

The evolution of the neuronal states is specified by the dynamical equations:

$$x_r(t) = f\left(x_{r' \in R_r}(\tau), u(\tau), \xi_{r' \in R_r}(\tau)\right) : \tau \in (t, t-t_0]. \quad (2)$$

This equation³ expresses, $x_r(t)$, the state vector of node r at time t as a function of:

- the states of nodes $x_{r'}(\tau)$: $r' \in R_r \subseteq R$
- exogenous inputs, $u(\tau)$ and a
- stochastic process $\xi_{r'}(\tau)$.

Note that the dependence of the current states at node r may be contingent on values of other variables from an arbitrary past from $t-t_0$ to just before t . The time dependence of Eq. (2) is important because it allows to model feedback processes within the network.

³ We use the following conventions for intervals, [a,b) indicates that the left endpoint is included but not the right one and that b precedes a.

Table 1
 Observation equations. Examples of observation equations used for causal modeling of effective connectivity in the recent literature. Abbreviations: discrete (D), continuous (C), white noise (WN). Note for Models #5 and #6 the observation equation is considered as all the equations except for the (neural) state equations. Strictly speaking, the observer function is just the first equality (because the subsequent equations of motion are part of the state equation); however, we have presented the equations like this so that one can compare instantaneous observation equations that are a function of hidden states, convolution operators or a set of differential equations that take hidden neuronal states as their inputs.

Model	Observation equation	Measurement	Space	Time	Equation type	Kind of stochastic process
1 None (Bressler and Seth, 2010)	$y(r, k) = x(r, k)$	EEG/fMRI	D	D	Identity	none
2 Added noise (Nalatore et al., 2007)	$y(r, k) = x(r, k) + e(r, t)$	fMRI	D	D	Linear regression	WN
3 Spatial smearing (Riera et al., 2006)	$y(q, t) = \int_{r \in R} k(r, r') x(r', t) dr' + e(r, t)$	EEG/MEG	D	C	Volterra integral equation with noise	none
4 Convolution with linear HRF (Glover, 1999)	$y(r, k) = \int_{-\infty}^{t-k} h(\tau) x(r, t-\tau) d\tau + e(r, k)$	fMRI	D	C	Temporal convolution	WN
5 Nonlinear HRF function (Friston et al., 2000)	$y_t = V_0(a_1(1-q_t) - a_2(1-v_t))$ $\dot{v}_t = \frac{1}{\tau_0} (f_t - v_t^{1/\alpha})$ $\dot{q}_t = \frac{1}{\tau_0} \left(\frac{f_t(1 - (1-E_0)^{1/f_t})}{E_0} - \frac{q_t}{v_t^{1-1/\alpha}} \right)$ $\dot{s}_t = \varepsilon u_t - \frac{1}{\tau_s} s_t - \frac{1}{\tau_f} (f_t - 1)$ $f_t = s_t$	fMRI	C	C	Nonlinear differential equation	none
6 Nonlinear HRF function (Valdes-Sosa et al., 2009a)	$\begin{cases} \dot{g}_e(t) = s_e(t) \\ \dot{s}_e(t) = \frac{a_e}{\tau_e} (u_e(t - \delta_e) - 1) - \frac{2}{\tau_e} s_e(t) - \frac{1}{\tau_e^2} (g_e(t) - 1) \\ \dot{g}_i(t) = s_i(t) \\ \dot{s}_i(t) = \frac{a_i}{\tau_i} (u_i(t - \delta_i) - 1) - \frac{2}{\tau_i} s_i(t) - \frac{1}{\tau_i^2} (g_i(t) - 1) \end{cases}$ $x = \frac{1}{1 + e^{-c(g_e(t) - d)}}$ $\begin{cases} \dot{f}(t) = s_f(t) \\ \dot{s}_f(t) = \varepsilon (u_e(t - \delta_f) - 1) - \frac{s_f(t)}{\tau_s} - \frac{f(t) - 1}{\tau_f} \end{cases}$ $m_i(t) = g_i(t), m_e(t) = \frac{2-x}{2-x_0} g_e(t), m(t) = \frac{\gamma m_e(t) + m_i(t)}{\gamma + 1}$ $\begin{cases} \dot{v}(t) = \frac{1}{\tau_0} (f(t) - f_{out}(v, t)) \\ \dot{q}(t) = \frac{1}{\tau_0} \left(m(t) - f_{out}(v, t) \frac{q(t)}{v(t)} \right), f_{out}(v, t) = v^{\frac{1}{\alpha}} \end{cases}$ $y(t) = V_0(a_1(1-q) - a_2(1-v))$	EEG/fMRI	C	C	Nonlinear random differential algebraic equation	none

Many specific forms have been proposed for Eq. (2); some examples are listed in Table 2, which is just a selection to illustrate different points discussed below. Some types of equations, to our knowledge, have not been yet used for the analysis of effective connectivity. Several general observations emerge from these examples:

Discrete versus continuous time modeling: The state equations for GCM have been for the most part discrete time recurrence models (Bressler and Seth, 2010). Those for DCM are based on continuous time models (differential equations) (Friston, 2009a). The latter have advantages in dealing with the problem of temporal aggregation and sub-sampling as we shall see below. In fact, DCM is distinguished from general SSM by the fact it is based on differential equations of one sort or another.

Discrete versus continuous spatial modeling: GCM has been applied to continuous space (neural fields) though limited to discrete time (Galka et al., 2004; Valdes-Sosa, 2004). DCM has mainly been developed for discrete-space (ROIs) and, as mentioned above, continuous time. State space models that are continuous in space and time have recently been looked at in the context of neural field equations (Daunizeau et al., 2009c; Galka et al., 2008).

Type of equation: GCM has been predominantly based on linear stochastic recurrence (autoregressive) models (Bressler and Seth, 2010). DCM on the other hand has popularized the use of deterministic ordinary differential equations (ODE). These range from simple bilinear forms for fMRI that accommodate interactions between the input and the state variables (Friston, 2009a) to complicated nonlinear equations describing the ensemble dynamics of neural mass models. In their most comprehensive form, these models can be formulated as Hierarchical Dynamical Models (HDM) (Friston, 2008a,b). HDM uses nonlinear random differential equations and static nonlinearities, which can be deployed hierarchically to reproduce most known parametric models. However, as noted in the C&C, GCM is not limited to linear models. GCM mapping (Roebroek et al., 2005) uses an (implicit) bilinear model, because the Autoregressive coefficients depend on the stimulus; this bilinearity is explicit in GCM on manifolds (Valdés-Sosa et al., 2005) GCM has also been extended to cover nonlinear state-equations (Freiwald et al., 1999; Marinazzo et al., 2011). The type of models used as state equations are very varied (and are sometimes equivalent). One can find (for discrete spatial nodes) recurrence equations, ordinary differential equations, and (for neural fields) differential-integral and partial differential equations.

Table 2

State equations. Examples of the state equations used in the recent literature for causal modeling of effective connectivity. Abbreviations: C (continuous), D (discrete), WN (white noise).

Model	State equation	Space	Time	Equation type	Stochastic process
Linear GCM (Bressler and Seth, 2010)	$x(r, k) = \sum_{r'=1}^{N_r} \sum_{l=1}^T a_l(r, r') x(r', k-l) + \xi(r, k)$	D	D	Linear multivariate linear autoregressive (VAR)	WN
2 Nonlinear GCM (Freiwald et al., 1999)	$x(r, k) = \sum_{r'=1}^{N_r} \sum_{l=1}^T a[l, r, r'; x(r', k-l)] x(r', k-l) + \xi(r, k)$	D	D	Nonlinear nonparametric VAR (NNp_MVAR)	WN
3 Linear bivariate GCM mapping (Roebroeck et al., 2005)	$\begin{bmatrix} x(r, k) \\ x(ROI, k) \end{bmatrix} = \sum_{r'=1}^{N_I} \begin{bmatrix} a_l(r, r) & a_l(r, ROI) \\ a_l(ROI, r) & a_l(ROI, ROI) \end{bmatrix} \begin{bmatrix} x(r, k-l) \\ x(ROI, k-l) \end{bmatrix} + \begin{bmatrix} \xi(r, k) \\ \xi(ROI, k) \end{bmatrix}$ $\forall r \in R \quad x(ROI, k) = \int_{r \in R} x(r, k) dr$	D	D	VAR since $a_l(r, r')$ Implicitly bi-linear changes with state. (GCMMap)	WN
4 Linear GCM on spatial manifold (Valdés et al., 2006)	$x(r, k) = \sum_{l=1}^{N_I} \int_{r \in R} a_l(r, r') x(r', k-l) dr' + \xi(r, k)$	C	D	Implicitly bi-linear VAR as in 3	WN
5 Nonlinear DCM (Stephan et al., 2008)	$\dot{x}(r, t) = \sum_{r'=1}^{N_x} a(r, r') x(r', t) + \sum_{i=1}^{N_u} u(i, t) \sum_{r'=1}^{N_x} b(r, r') x(r', t) + \sum_{r'=1}^{N_x} \sum_{r''=1}^{N_x} d(r, r', r'') x(r', t) x(r'', t) + \sum_{i=1}^{N_u} c(r, i) u(i, t)$	D	D	Differential equation bilinear in both states and inputs (DE)	None
6 Neural mass model (Valdes et al., 1999)	$\dot{x}(r, t) = f(x(r, t)) + \xi(r, t)$	C	C	Ito stochastic differential (SDE)	WN as formal derivative of Brownian motion
7 Hierarchical dynamic causal model (Friston, 2008a,b)	$\dot{x}(r, t) = f(x(r, t), u(t)) + \xi(r, t)$	D	C	General nonlinear (HDM)	Analytic, non-Markovian
8 Neural field (Jirsa et al., 2002)	$\left(\frac{\partial^2}{\partial t^2} + 2\omega_0 \frac{\partial}{\partial t} + \omega_0^2 - v^2 \nabla^2\right)^{3/2} x(r, t) = \left(\omega_0^3 + \omega_0^2 \frac{\partial}{\partial t}\right) S[x(r, t) + \xi(r, t)]$	C	C	Stochastic fractional partial differential (StPDE)	WN
9 Modified neural field (P. A. Valdes-Sosa et al., 2009a)	$\ddot{x}(r, t) = f(\dot{x}(r, t), x(r, t)) + S(z(r, t)) + \xi(r, t)$ $z(r, t) = \int_R a(r, r') x(r, \tau(r, r')) dr'$ $\tau(r, r') = t - \frac{ r-r' }{v}$	C	C	Random differential-algebraic-equation (RDE)	General

To underscore the variety of forms for effective connectivity, we note entry #8 in Table 2 which boasts a fractional differential operator! Fractional operators arise in the context of neural fields in more than one dimension; they result from the Fourier transform of a synaptic connection density that is a continuous function of physical distance. However, the ensuing fractional differential operators are usually replaced by ordinary (partial) differential operators, when numerically solving the neural wave propagation equation given in Table 2; see Bojak and Liley (2010) and Coombes et al. (2007) for the so-called 'long wavelength approximation'.

Among other things, it can be important to include time delays in the state equation; this is usually avoided when possible to keep the numerics simple (delay differential equations are infinite dimensional) and are generally considered unnecessary for fMRI. However, delays are crucial when modeling electromagnetics data, since they can have a profound effect on systems dynamics (Brandt et al., 2007). For example, delayed excitatory connections can have an inhibitory instantaneous effect. In fact starting with Jansen and Rit (1995) it has been common practice to include time delays. This can be implemented within the framework of ODEs; David et al. (2006) describe an ODE approximation to delayed differential equations in the context of DCM for EEG and MEG. An example of the potential richness of model structures is found in Valdes-Sosa et al. (2009a) in a neural field forward model for EEG/fMRI fusion, which includes anatomical connections and delays as algebraic constraints. This approach (of including algebraic constraints) affords the possibility of building complex models from nonlinear components, using simple interconnection rules—something that has been developed for control theory (Shampine and Gahinet, 2006). Note that algebraic constraints may be added to any of the aforementioned forms of state equation.

Type of stochastics: for GCM-type modeling with discrete-time models, Gaussian White Noise (GWN) is usually assumed for the random fluctuations (state noise) or driving forces (innovations) for the SSM and poses no special difficulties. However in continuous time the problem becomes more intricate. A popular approach is to treat the innovation as nowhere differentiable but continuous Gaussian White Noise (the “derivative” of Brownian motion (i.e., a Wiener process)). When added to ordinary differential equations we obtain “stochastic differential equations” (SDE) as described in Medvegyev (2007) and used for connectivity analysis of neural masses in Riera et al., (2007a,b), Riera et al. (2006). Wiener noise is also central to the theory of Stochastic Partial Differential Equations (SPDE) (Holden et al., 1996), which may play a similar role in neural field theory as SDEs have played for neural masses (Shardlow, 2003).

Despite the historical predominance of the classical SDE formulation in econometrics (and SSM generally), we wish to emphasize the following developments, which may take us (in the biological sciences) in a different direction:

1. The first is the development of a theory for “random differential equations” (RDE) (Jentzen and Kloeden, 2009). Here randomness is not limited to additive Gaussian white noise because the parameters of the state equations are treated as stochastic. RDE are treated as deterministic ODE, in the spirit of Sussmann (1977), an approach usable to great advantage in extensive neural mass modeling (Valdes-Sosa et al., 2009a) that is implicitly a neural field.
2. The second development, also motivated by dissatisfaction with classical SDE was introduced in Friston and Daunizeau (2008). In that paper, it was argued that DCMs should be based on stochastic processes, whose sample paths are infinitely differentiable—in other words, analytic and non-Markovian.

Though overlooked in the heyday of SDE theory, this type of process was described very early by Belyaev (1959).⁴ In fact any band-limited stochastic process is an example of an analytic random process; a stochastic process with a spectrum that decreases sharply with frequency, has long memory, and is non-Markovian (Łuczka, 2005). The connection between analytic stochastic processes and RDE can be found in Calbo et al. (2010). An interesting point here is that for the process to be analytic its successive derivatives must have finite variances, as explained in Friston and Daunizeau (2008). This leads to the generalization of classical SSM into generalized coordinates of motion that model high-order temporal derivatives explicitly. As pointed out in Friston (2008a,b), it is possible to cast an RDE as a SDE by truncating the temporal derivatives at some suitably high order (see also Carbonell et al., 2007). However, this is not necessary because the theory and numerics for RDEs in generalized coordinates are simpler than for the equivalent SDE (and avoid the unwieldy calculus of Markovian formulations, due to Ito and Stratonovich).

3. The third development is the recognition that non-Markovian processes may be essential for neurobiological modeling. This has been studied for some time in physics (Łuczka, 2005) but has only recently been pointed out by Friston (2008a,b) in a neuroscience setting. In fact, Faugeras et al. (2009) provide a constructive mean-field analysis of multi-population neural networks with random synaptic weights and stochastic inputs that exhibits, as a main characteristic, the emergence of non-Markovian stochastics.
4. Finally the fourth development is the emergence of neural field models (Coombes, 2010; Deco et al., 2008), which not only poses much larger scale problems but also the use of integral equations, differential–integral equations, and partial differential equations which have yet to be exploited by DCM or GCM.

Biophysical versus non-parametric motivation: As discussed above, there is an ever increasing use of biophysically motivated neural mass and field state equations and, in principle, these are preferred when possible because they bring biophysical constraints to bear on model inversion and inference. When carrying out exploratory analyses with very large SSM, it may be acceptable to use simple linear or bilinear models as long as basic aspects of modeling are not omitted.

Further generalizations: We want to end this subsection by mentioning that there is a wealth of theory and numerics for other stochastic (point) processes (Aalen and Frigessi, 2007; Commenges and Gégout-Petit, 2009) that have not yet been, to our knowledge, treated formally in Neuroimaging. Spike trains, interictal-spikes, and random short-timed external stimuli may be treated as point processes and can be analyzed in a unified framework with the more familiar continuous time series. This theory even encompasses mixtures of slow wave and spike trains. Causal modeling depends very specifically on the temporal and spatial scales chosen and the implicit level of granularity chosen to characterize functional brain architectures. For example, if we were to study the interaction of two neural masses and model the propagation of activity between them in detail, we would have to make use of the PDE that describes the propagation of nerve impulses. If we eschew this level of detail, we may just model the fact that afferent activity arrives at a neural mass with a conduction delay and use delay differential equations. In short, the specification of the appropriate SSM depends on the spatial and temporal scale that one is analyzing. For example, in concurrent EEG/fMRI analysis of resting state oscillations

(Martínez-Montes et al., 2004) the temporal scale of interesting phenomena (fluctuations of the EEG spectrum) is such that one may convolve the EEG signal and do away with the observation equation! This is exactly the opposite of the deconvolution approach mentioned above. The purpose of Tables 1 and 2 is to highlight the variety of forms that both state and observation equations can take; for example, in Table 2–#6 key differential equations are transformed into differential algebraic equations to great computational advantage (Valdes-Sosa et al., 2009a).

Specification of priors

It is safe to say that the Neuroimaging (and perhaps generally) modeling can be cast as Bayesian inference. This is just a euphemism for saying that inference rests on probability theory. The two key aspects of Bayesian inference we will appeal to in this article are (i) the importance of prior beliefs that form an explicit part of the generative model; and (ii) the central role of Bayesian model evidence in optimizing (comparing and selecting) models to test hypotheses. In terms of priors, it was very clear in an early state space model for EEG connectivity (Valdes-Sosa et al., 1996) that without prior assumptions about the spatial and temporal properties of the EEG, it was not possible to even attempt source reconstruction. Indeed the whole literature on ill-posed inverse problems rests on regularization that can be cast in terms of prior beliefs.

In the SSM formulation, priors may be placed upon parameters in the observation and state equations, and the states themselves (e.g., through priors on the higher-order motion of states or state-noise). Sometimes, it may be necessary to place priors on the priors (hyperpriors) to control model complexity. There has been an increasing use of priors in fMRI research, as clearly formulated in the DCM and HDM framework (Friston, 2008a,b). In connectivity analyses, in addition to the usual use of priors to constrain the range of parameters quantitatively; formal or structural priors are crucial for switching off subsets of connections to form different (alternative) models of observed data. Effectively, this specifies the model in terms of its adjacency matrix, which defines allowable connections or conditional dependencies among nodes. Conditional independence (absence of an edge or anti-edge) is easy to specify by using a prior expectation of zero and with zero variance. This is an explicit part of model specification in DCM and is implicit in Granger tests of autoregressive models, with and without a particular autoregression coefficient.

Crucially, formal priors are not restricted to the parameters of a model; they can also be applied to the form of the prior density over parameters. These can be regarded as formal hyperpriors. An important example here is the prior belief that connections are distributed sparsely (with lots of small or absent connections and a small number of strong connections). This sort of hyperprior can be implemented by assuming the prior over parameters is sparse. A nice example of this can be found in Valdés-Sosa (2004), Valdés-Sosa et al. (2005, 2006), and Sánchez-Bornot et al. (2008).

The essential features of their model are shown in Fig. 3. The authors analyzed slow fluctuations in resting state EEG. In this situation, convolving these electrophysiological fluctuations with a HRF affords (convolved) EEG and BOLD signals on the same time scale, permitting lag-based inference. An example is presented in Fig. 4, which shows the results of GCM Mapping for 579 ROIs from an EEG inverse solution and concurrent BOLD signals. The EEG sources were obtained via a time resolved VARETA inverse solution (Bosch-Bayard et al., 2001) at the peak of the alpha rhythm. The graphs present the result of inverting a (first order) multivariate vector autoregression model, where a sparse l_1 norm penalty was imposed on the parameters (coefficient matrix). The implications of these results will be further discussed in Conclusion and suggestions for further work section below.

⁴ With suggestion by A.N. Kolmogorov.

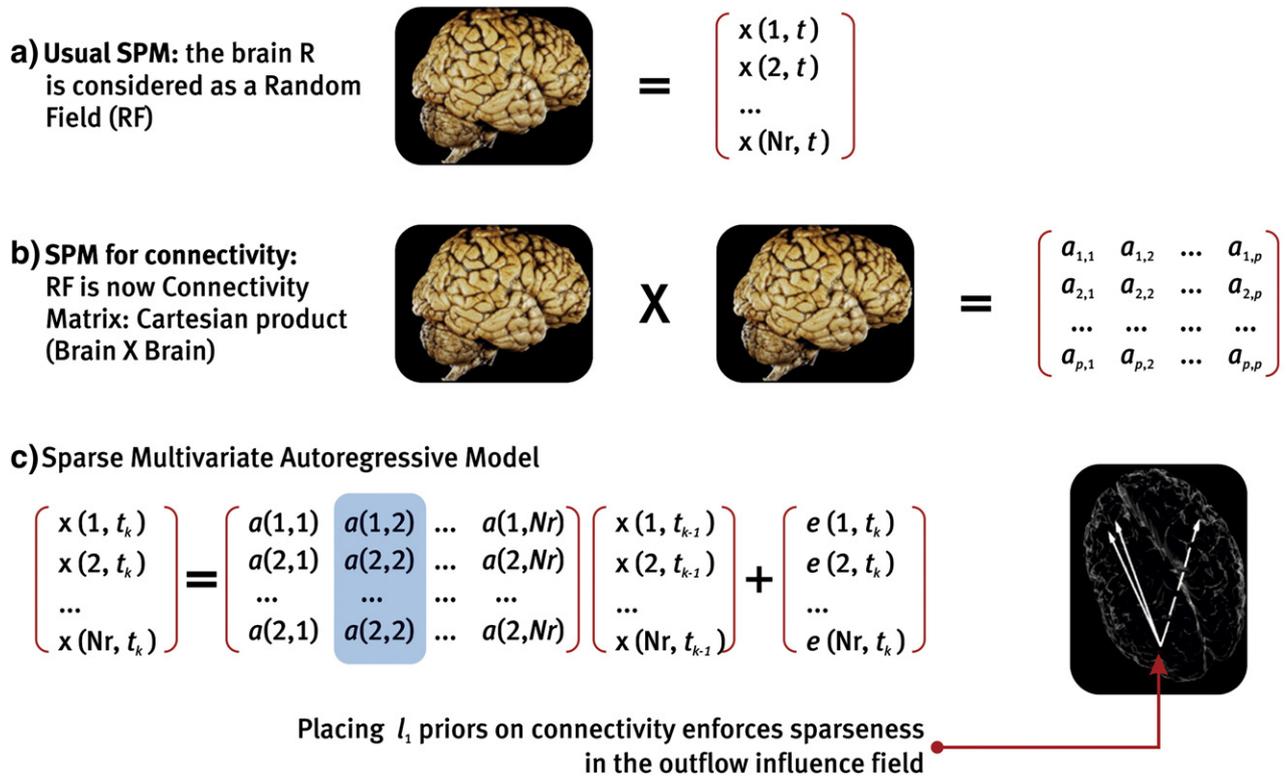


Fig. 3. Bayesian inference on the connectivity matrix as a random field. a) Causal modeling in Neuroimaging has concentrated on inference on neural states $x(r, t) \in \mathbb{R}$ defined on a subset of nodes in the brain. However, spatial priors can be used to extend models into the spatial domain (cf., minimum norm priors over current source densities in EEG/MEG inverse problems). b) In connectivity analysis, attention shifts to the AR (connectivity) matrix (or function) $a(r, r')$, where the ordered pairs (r, r') belong to the Cartesian product $R \times R$. For this type of inference, priors are now placed on the connectivity matrix. c) Sparse multivariate autoregression obtains by penalizing the columns of a full multivariate autoregressive model (Valdés-Sosa et al., 2005) thus forcing the columns of the connectivity matrix to be sparse. The columns of the connectivity matrix are the “outfields” that map each voxel to the rest of the brain. This is an example of using sparse (spatial) hyperpriors to regularize a very difficult inverse problem in causal modeling.

Model comparison and Identifiability

As we have seen, the SSMs considered for EEG and fMRI analysis are becoming increasingly complex, with greater spatial or temporal coverage and improved biological realism. A fundamental question arises: *Are these models identifiable?* That is to say, are all states and parameters uniquely determined by a given set of data? This is a basic issue for all inverse problems, and indeed we are faced with a dynamical inverse problem of the greatest importance. For example, recent discussions about whether lag information can be derived from the fMRI signal (in spite of heavy smoothing by the HRF and the subsequent sub sampling) can be understood in terms of the identifiability of delays in the corresponding SSM. It is striking that, in spite of much classical work on the Identifiability of SSMs (see for example Ljung and Glad, 1994), a systematic treatment of identification has not been performed for Neuroimaging models (but see below). An example of the type of problem encountered is the complaint that a model with many neural masses and different configurations or parameter values can produce traces that “look the same as an observed response”.

Identifiability has been addressed in bioinformatics, where much theory for nonlinear SSM has been developed (Anguelova and Wennberg, 2010; August and Papachristodoulou, 2009). Of particular note is DAISY, a computer algebra system for checking nonlinear SSM Identifiability (Saccomani et al., 2010). Another framework for modeling and fitting systems defined by differential equations in bioinformatics is “Potters Wheel” (Maiwald and Timmer, 2008), which uses a profile likelihood approach (Raue et al., 2009) to explore “practical identifiability” in addition to structural (theoretical) Identifiability. So why has Neuroimaging not developed similar schemes?

In fact, it has. In a Bayesian setting the issue of model (and parameter) identifiability is resolved though Bayesian model com-

parison. If two models generate exactly the same data with the same number of parameters (complexity), then their evidence will be identical. This means there is no evidence for one model over the other and they cannot be distinguished. We will refer a lot to model evidence in what follows: model evidence is simply the probability of the data given the model. It is the marginal likelihood that obtains from marginalizing the likelihood over unknown model parameters. This is useful to remember because it means the likelihood of a model (the probability of data given a model and its parameters) is a special case of model evidence that results when we ignore uncertainty about the parameters. In the same way, classical likelihood ratio tests of two models are special cases of Bayes Factors used in Bayesian model comparison. In this context, identifiability is a particular aspect of model comparison. Identifiability mandates that changing a component of a model changes the model evidence. This is the basic idea behind the profile likelihood approach (Raue et al., 2009), which is based on the profile of the evidence for models with different parameter values. There are other examples that can be regarded as special cases of model comparison; for example, the Kullback–Leibler information criterion proposed for model identification (Chen et al., 2009). The evidence can be decomposed into an accuracy and complexity term (see Penny et al., 2004). Interestingly, the complexity term is the Kullback–Leibler divergence between the posterior and prior densities over parameters. This means that in the absence of informative priors, model evidence reduces to accuracy; and identifiability reduces to a (nontrivial) change in the accuracy or fit when changing a model or parameter.

The Bayes–Net literature (see below) has dealt with the problem of Identifiability for graphical causal models at its inception (Spirtes et al., 2000). It can be shown that a given data set can be compatible not with a single causal model but with an equivalence class of models

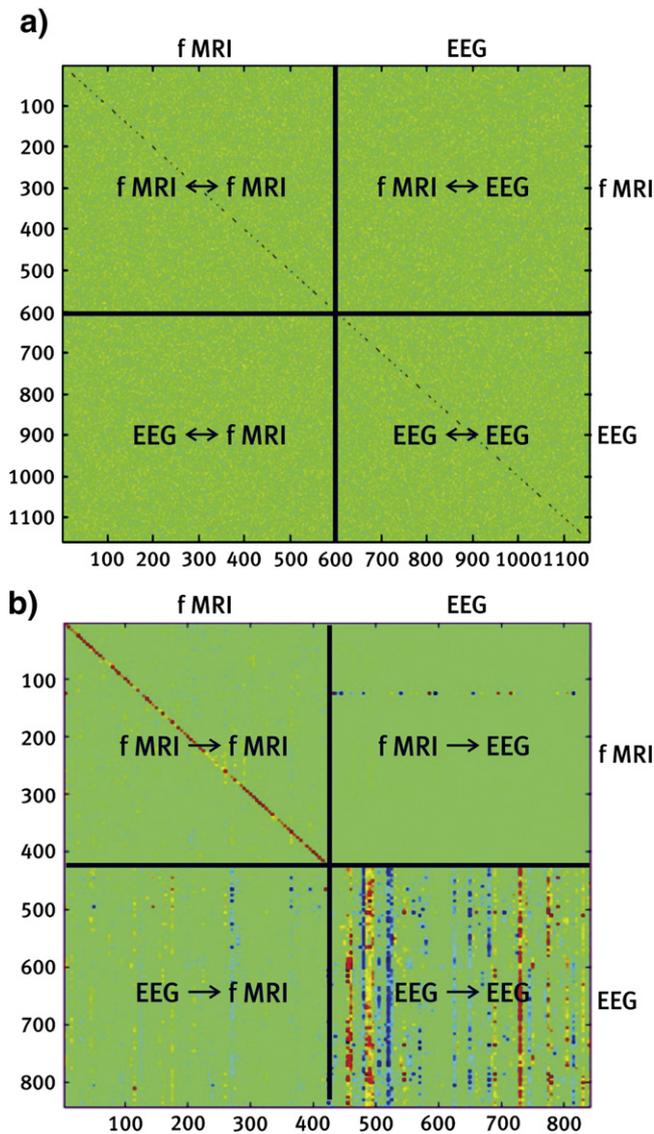


Fig. 4. Sparse multivariate autoregression of concurrent EEG/fMRI recordings. Intra and inter modality connectivity matrix for a concurrent EEG/fMRI recordings. The data analyzed here were the time courses of the average activity in 579 ROI: for BOLD (first half of data vector) and EEG power at the alpha peak. A first-order sparse multivariate autoregressive model was fitted with an l_1 norm (hyper) prior on the coefficient matrix. The t-statistics of the autoregression coefficients were used for display. The color bar is scaled to the largest absolute value of the matrix, where green codes for zero. a) the innovation covariance matrix reflecting the absence of contemporaneous influences: b) t-statistics for the lag 1 AR coefficients.

(that all have the same evidence). The implications of this for Neuroimaging have been considered in Ramsey et al. (2010). From this discussion, it becomes clear that the ability to measure model evidence (or some proxy) is absolutely essential to make sensible

Table 3
 Classification of observation and state equations used in Neuroimaging state-space models. Generic models lack specific biophysical constraints but are widely applicable. Biophysically informed models are hypothesis driven and may afford more efficient inference (if correct). The term parametric refers to models with a small enough parameter set to be identifiable without additional priors but that may yield biased estimators. Nonparametric models are richly parameterized and therefore require prior distributions to be estimable but are generally unbiased.

	Observation model		State model	
	Parametric	Non-parametric	Parametric	Non-parametric
Generic	Linear canonical HRF (Glover, 1999)	Linear spline HRF (Marrelec et al., 2003)	GCM (Bressler and Seth, 2010) Switching VAR (Smith et al., 2010a) bilinear discrete DCM (Penny et al., 2005)	GCM (Roebroeck et al., 2005)
Biophysically informed	DCM nonlinear HRF (Friston et al., 2000)	-	Neural mass models (Valdes et al., 1999) Biophysical DCM (Moran et al., 2008)	Neural fields (Daunizeau et al., 2009c)

inferences about models or architectures generating observed data. This is at the heart of evidence-based inference and DCM.

Summary

State space models for Neuroimaging come in an ever increasing variety of forms (Tables 1 and 2). It is useful to classify the types of models used in terms of their observation and state equations, as in Table 3. Here, we see a distinction between models that are fairly generic (in that they are not based on biophysical assumptions) and those that correspond to biologically informed models. The canonical HRF model is an example of generic HRF. Conventional GCM is based on a generic model for neural states: the VAR model and has been extended to switching VAR and bilinear models, the latter used in some forms of DCM. Being generic is, at the same time, a strength and weakness; biophysical models allow much more precise and informed inference—but only if the model is right or can be optimized in terms of its evidence. We have also seen the key role that model evidence plays in both making causal inferences by comparing models and (implicitly) establishing their identifiability. The evidence for a model depends on both accuracy and complexity and the complexity of the model depends on its priors.

Another distinction between models is their complexity (e.g., number of parameters they call on). It is clear that without prior beliefs, one cannot estimate more parameters than the degrees of freedom in the data available. However, modern statistical learning has gone beyond low dimensional parametric models to embrace non-parametric models with very high dimensional parameter spaces. The effective number of degrees of freedom is controlled by the use of priors. DCM has been concerned mainly with hypothesis driven parametric models, as has conventional GCM. However, nonparametric models, such as smoothness priors in the time domain have been used to estimate the HRF (Marrelec et al., 2003). Another example is the use of spatial priors to estimate the connectivity matrix in GCM (Valdes-Sosa, 2004). Finally, when choosing a State Space model, it is useful to appreciate that there are two agendas when trying to understanding the connectivity of complex systems:

1. A data driven exploratory (discovery) approach that tries to scan the largest model space possible, identifying robust phenomena or candidates that will serve as constraints for more detailed modeling. This type of approach generally uses nonparametric or simply parameterized models for knowledge discovery. Prior knowledge is generally nonspecific (e.g., connections are sparse) but relatively non-restrictive.
2. A model driven confirmatory approach that is based on specific hypothesis driven models that incorporate as much biophysical prior knowledge as possible. Generally, the priors entail specific hypothesis about connectivity that can be resolved using model comparison.

These two approaches are shown in Fig. 2 (modified from Valdés-Sosa et al., 1999). In both cases, modeling is constrained by the data, by biophysical plausibility and ultimately the ability to establish links with computational models (hypotheses) of information processing

in the brain. Table 3 shows that at one extreme the model-driven approach is epitomized by Generic Nonparametric Models. Here, modeling efforts are constrained by data and the attempt to disclose emergent behavior, attractors and bifurcations (Breakspear et al., 2006) that can be checked against biophysically motivated models. An example of this approach is searching the complete brain times brain connectivity space (Fig. 3) with GCM mapping (Valdes-Sosa, 2004; Roebroek et al., 2005). At the other end we have the parametric and biophysically informed approach that DCM has emphasized (Chen et al., 2008). Having said this, as evidenced by this paper and companion papers, there is convergence of the two approaches, with a gradual blurring of the boundaries between DCM and GCM.

Model inversions and inference

In this section, we look at the problem of model identification or inversion; namely, estimating the states and parameters of a particular model. It can be confusing when there is discussion of a new model that claims to be different from previous models, when it is actually the same model but with a different inversion or estimation scheme. We will try to clarify the distinction between models and highlight their points of contact when possible. Our main focus here will be on different formulations of SSM and how these formulations affect model inversion.

Discrete or continuous time?

One (almost) always works with discretely sampled data. When the model is itself discrete, then the only issue is matching the sampling times of the model predictions and the data predicted. However, when starting from a continuous time model, one has to model explicitly the mapping to discrete time.

Mapping continuous time predictions to discrete samples is a well-known topic in engineering and (probably from the early 50s) has been solved by linearization of the ODEs and integration over discrete time steps; a method known as the Exponential Euler method for reasons we shall see below: see Minchev and Wright (2005) for a historical review. For a recent review, with pointers to engineering toolboxes, see Garnier and Wang (2008).

One of the most exciting developments in the 60s, in econometrics was the development of explicit methods for estimating continuous models from sampled data, initiated by Bergstrom (1966).⁵ His idea was essentially the following. Consider 3 time series $X_1(t)$, $X_2(t)$, and $X_3(t)$ where we know the values at time t :

$$\begin{pmatrix} dX_1(t) \\ dX_2(t) \\ dX_3(t) \end{pmatrix} = A \begin{pmatrix} X_1(t) \\ X_2(t) \\ X_3(t) \end{pmatrix} dt + \sum^{1/2} dB(t). \quad (3)$$

Then the explicit integration⁶ over the interval $[t + \Delta t, t]$ is

$$\begin{pmatrix} X_1(t + \Delta t) \\ X_2(t + \Delta t) \\ X_3(t + \Delta t) \end{pmatrix} = \exp(A\Delta t) \begin{pmatrix} X_1(t) \\ X_2(t) \\ X_3(t) \end{pmatrix} + e(t + \Delta t) \quad (4)$$

$$e(t + \Delta t) = \int_0^{\Delta t} \exp(sA) \sum^{1/2} dB(t-s)$$

$$\Sigma_{discrete} = \int_0^{\Delta t} \exp(sA) \Sigma \exp(sA^T) ds$$

$$e(t + \Delta t) \sim \mathcal{N}(0, \Sigma_{discrete}).$$

⁵ Who, in fact, did this not for SDE (ODE driven by Brownian noise) but for linear ODE driven by random measures, as reviewed in Bergstrom (1984).

⁶ Note, once again, that we use the convention $[t + \Delta t, t]$ for the time interval that goes from t in the past to $t + \Delta t$ in the present; while not the conventional usage this will make later notation clearer.

The noise of the discrete process now has the covariance matrix $\Sigma_{discrete}$. It is immediately evident from the equation above that the lag zero covariance matrix $\Sigma_{discrete}$ will show contemporaneous covariance even if the continuous covariance matrix Σ is diagonal. In other words, the discrete noise becomes correlated over the three time-series (e.g., channels). This is because the random fluctuations ‘persist’ through their influence on the motion of the states. Rather than considering this a disadvantage Bergstrom (1984) and Phillips (1974) initiated a line of work studying the estimation of continuous time Autoregressive models (Mccrorie and Chambers, 2006), and continuous time Autoregressive Moving Average Models (Chambers and Thornton, 2009). This approach tries to use both lag information (the AR part) and zero-lag covariance information to identify the underlying linear model.

The extension of the above methods to nonlinear stochastic systems was proposed by Ozaki (1992) and has been extensively developed in recent years, as reviewed in Valdes-Sosa et al. (2009a). Consider a nonlinear system of the form:

$$dX(t) = f(X(t))dt + \sum^{1/2} dB(t)$$

$$X(t) = \begin{pmatrix} X_1(t) \\ X_2(t) \\ X_3(t) \end{pmatrix}. \quad (5)$$

The essential assumption in local linearization (LL) of this nonlinear system is to consider the Jacobian matrix $A = \partial f / \partial X$ as constant over the time period, $[t + \Delta t, t]$. This Jacobian plays the same role as the matrix of autoregression coefficient in the linear systems above. Integration over this interval follows as above, with the solution:

$$X(t + \Delta t) = X(t) + A^{-1}(\exp(A\Delta t) - I)f(X(t)) + e(t + \Delta t)^7 \quad (6)$$

where I is the identity matrix. This solution is locally linear but crucially it changes with the state at the beginning of each integration interval; this is how it accommodates nonlinearity (i.e., a state-dependent autoregression matrix). As above, the discretised noise shows instantaneous correlations. Examples of inverting nonlinear continuous time neural models using this procedure are described in Valdes-Sosa et al. (1999), Riera et al. (2007b), Friston and Daunizeau (2008), Marreiros et al. (2009), Stephan et al. (2008), and Daunizeau et al. (2009b). Local linearization of this sort is used in all DCMs, including those formulated in generalized coordinates of motion.

There are several well-known technical issues regarding continuous model inversion:

1. The econometrics literature has been very much concerned with identifiability in continuous time models—an issue raised by one of us in the C&C series (Friston, 2009b) due to the non-uniqueness of the inverse mapping of the matrix exponential operator (matrix logarithm) for large sampling periods Δt . This is not a problem for DCM, which parameterizes the state-equation directly in terms of the connectivity A . However, autoregressive models (AR) try to estimate $A = \exp(A\Delta t)$ directly, which requires a mapping $A = \frac{1}{\Delta t} \ln(A)$ to get back to the underlying connectivity. Phillips noted in the 70s that A is not necessarily invertible, unless one is sampling at twice the highest frequency of the underlying signal (the Nyquist frequency) (Phillips, 1973); in other words, unless one samples quickly, in relation to the fluctuations in hidden states. In econometrics, there are several papers that study the conditions in which under-sampled systems can avoid an implicit aliasing problem (Hansen and Sargent, 1983; Mccrorie and Chambers,

⁷ Note integration should not be computed this way since it is numerically unstable, especially when the Jacobian is poorly conditioned. A list of robust and fast procedures is reviewed in Valdes-Sosa et al. (2009a).

2006; Mccrorie, 2003). This is not a problem for electrophysiological models because sampling is fast relative to the underlying neuronal dynamics. However, for fMRI this is not the case and AR models provide connectivity estimates, $A = \frac{1}{\Delta t} \ln(A) \in \mathbb{C}^{N \times N}$ that are not necessarily unique (a phenomenon known as “aliasing” as discussed below). We will return to this problem in the next section, when considering the mediation of local (direct) and global (indirect) influences over time. Although this “missing time” problem precludes inference about coupling between neuronal states that fluctuate quickly in relation to hemodynamics, one can use AR models to make inferences about slow neuronal fluctuations based on fMRI (e.g., the amplitude modulation of certain frequencies; see Fig. 4). Optimal sampling for AR models has been studied extensively in the engineering literature—the essential point being that sampling should not be below or even much above the optimal choice that matches the natural frequencies (time-constants) of the hidden states (Astrom, 1969; Larsson et al., 2006).

- When the sampling period Δt is sufficiently small, the AR model is approximately true. What is small? We found very few practical recommendations, with the exception of Sargan (1974), who uses heuristic arguments and Taylor expansions to suggest that a sampling frequency 1.5 times faster than the Nyquist frequency allows the use of a bilinear (or Tustin) approximation in (two stage non-recursive) autoregression procedures. As shown in the references cited above, it might be necessary to sample at several times the Nyquist frequency to use AR models directly. However, an interesting “Catch 22” emerges for AR models: The aliasing problem mandates fast sampling, but fast sampling violates Markovian (e.g., Gaussian noise) assumptions, if the true innovations are real (analytic) fluctuations.
- A different (and a more complicated) issue concerns the identifiability of models of neural activity actually occurring at rates much higher than the sampling rates of fMRI, even when a DCM is parameterized in terms of neuronal coupling. This is an inverse problem that depends on prior assumptions. There are lessons to be learned from the EEG literature here: Linear deconvolution methods for inferring neural activity from EEG proposed by Glover (1999) and Valdes-Sosa et al. (2009a) correspond to a temporal version of the minimum norm and LORETA spatial inverse solutions respectively. Riera et al. (2007a) and Riera et al. (2006), proposed a nonlinear deconvolution method. In fact, every standard SPM analysis of fMRI data is effectively a deconvolution, where the stimulus function (that is convolved with an assumed HRF) provides a generative model whose inversion corresponds to deconvolution. In the present context, the stimulus function provides the prior expectations about neuronal activity and the assumed HRF places priors on the ensuing hemodynamics. In short, model inversion or deconvolution depends on priors. The extent to which identifiability will limit inferences about neuronal coupling rests on whether the data supports evidence for different models of neuronal activity. We already know that there is sufficient information in fMRI time series to resolve DCMs with different neuronal connectivity architectures (through Bayesian model comparison), provided we use simple bilinear models. The issue here is whether we can make these models more realistic (cf., the neural mass models used for EEG) and still adjudicate among them, using model evidences: When models are too complex for their data, their evidence falls and model selection (identification) fails. This is an unresolved issue.

As one can see from these points, the issue of inference from discretised data depends on the fundamental frequencies of fluctuations in hidden states, data sampling rate, the model, and the prior information we bring to the inferential problem. When writing these lines, we were reminded of the dictum, prevalent in the first years of

EEG source modeling, that one could “only estimate a number of dipoles that was less than or equal to a sixth of the number of electrodes”. Bayesian modeling has not increased the amount of information in data but it has given us a principled framework to optimize generative or forward models (i.e., priors) in terms of their complexity, by choosing priors that maximize model evidence. This has enabled advances in distributed source modeling and the elaboration of better constraints (Valdés-Sosa et al., 2009b). One might anticipate the same advances in causal modeling over the next few years.

Time, frequency or generalized coordinates?

A last point to mention is that (prior to model inversion) it may be convenient to transform the time domain data to a different coordinate system, to facilitate computations or achieve a theoretical objective. In particular transformation to the frequency domain has proved quite useful.

- This was proposed first for generic linear models in both continuous and discrete time (Robinson, 1991). More recently a nonparametric frequency domain approach has been proposed for Granger Causality (Dhamala et al., 2008).
- A recent stream of EEG/MEG effective connectivity modeling has been introduced by Nolte et al. (2008), Marzetti et al. (2008), Nolte et al. (2009), and Nolte et al. (2006) with the realization that time (phase) delays are reflected in the imaginary part of the EEG/MEG cross-spectra, whereas the real part contains contemporaneous contributions due to volume conduction.
- Linearised versions of nonlinear DCMs have also been transformed successfully to the frequency domain (Moran et al., 2008; Robinson et al., 2008).

As noted above Friston (2008a,b) has proposed a transformation to generalized coordinates, inspired by their success in physics. This involves representing the motion of the system by means of an infinite sequence of derivatives. The truncation of this sequence provides a summary of the time-series, in much the same way that a Fourier transform provides a series of Fourier coefficients. In classical time series analysis, the truncation is based on frequencies of interest. In generalized coordinates, the truncation is based on the smoothness of the time series. This use of generalized coordinates in causal modeling is predicated on the assumption that real stochastic processes are analytic (Belyaev, 1959).

Model inversion and inference

There are many inversion schemes to estimate the states, parameters and hyperparameters of a model. Some of the most commonly used are variants of the Kalman Filter, Monte-Carlo methods and variational methods (see e.g., Daunizeau et al., 2009b for a variational Bayesian scheme). As reviewed in Valdes-Sosa et al. (2009a) the main challenges are how to scale the numerics of these schemes for more realistic and extensive modeling. The one thing all these schemes have in common is that they (implicitly or explicitly) optimize model parameters with respect to model evidence. In this sense model inversion and inference on models *per se* share a common objective; namely to maximize the evidence for a model.

Selecting or optimizing a model for effective connectivity ultimately rests on model evidence used in model comparison or averaging. The familiar tests for GCM (i.e. Dickey–Fuller test) are based on likelihood comparisons. As noted above, the likelihood (the probability of the data given a model and its parameters) is the same as model evidence (the probability of the data given a model), if we ignore uncertainty about the model parameters. However, the models considered in this paper, that include qualitative prior beliefs call for measures of goodness that balance accuracy (expected log-likelihood)

with model complexity. All of these measures (AIC, BIC, GCV, and variational free energy) are approximations to the model evidence (Friston, 2008a,b). Model evidence furnishes the measure used for the final probabilistic inference about a causal architecture (i.e., causal inference). Clearly, to carry out model comparison one must have an adequate set of candidates. Model Diagnostics are useful heuristics in this context that ensure that the correct models have been chosen for comparison. An interesting example that can be used to perform a detailed check of the adequacy of models is to assess the spatial and temporal whiteness of the residual innovation of the model which is illustrated in (Galka et al., 2004). More generally, the specification and exploration of model sets (spaces) probably represents one of the greatest challenges that lie ahead in this area.

Summary

In summary, we have reviewed the distinction between autoregression (AR) models and models formulated in continuous time (DCM). We have touched upon the important role of local linearisation in mapping from continuous dynamics of hidden states to discrete data samples and the implications for sampling under AR models. In terms of model inversion and selection, we have highlighted the underlying role played by model evidence and have cast most of the cores issues in model identifiability and selection in terms of Bayesian model comparison. This subsumes questions about the complexity of models that can be supported by fMRI data; through to ultimate inferences about causality, in terms of which causal model has the greatest evidence. This section concludes our review of pragmatic issues and advances in the causal modeling of effective connectivity. We now turn to more conceptual issues and try to link the causal modeling for Neuroimaging described in this section to classical constructs that have dominated the theoretical literature over the past few decades.

Statistical causal modeling

In this section, we review some key approaches to statistical causality. At one level, these approaches have had relatively little impact on recent developments in causal modeling in Neuroimaging, largely because they based on classical Markovian (and linear) models or ignore dynamics completely. However, this field contains some deep ideas and we include this section in the hope that it will illuminate some of the outstanding problems we face when modeling brain connectivity. Furthermore, it may be the case that bringing together classical temporal precedence treatments with structural causal modeling will finesse these problems and inspire theoreticians to tackle the special issues that attend the analysis of biological time series.

Philosophical background

Defining, discovering and exploiting causal relations have a long and enduring history (Bunge, 2009). Examples of current philosophical debates about causality can be found in Woodward (Woodward, 2003) and Cartwright (2007). An important concept, stressed by Woodward, is that a cause is something that “makes things happen”. Cartwright, on the other hand (Cartwright, 2007), argues for the need to separate the definition, discovery and use of causes; stresses the pluralism of the concept of cause and argues for the use of “thick causal concepts”. An example of what she calls a “thin causal claim” would be that “activity in the retina causes activity in V1”—represented as a directed arrow from one structure to the other. Instead, it might be more useful to say that the Retina is mapped via a complex logarithmic transform to V1 (Schwartz, 1977). A “thick causal” explanation tries to explain how information is actually transmitted. For a different perspective see Glymour (2009). It may be that both thin and thick causal concepts are useful when characterizing complex systems.

Despite philosophical disagreements about the study of causality, there seems to be a consensus that causal modeling is a legitimate statistical enterprise (Cox and Wermuth, 2004; Frosini, 2006; Pearl, 2003). One can clearly differentiate two current streams of statistical causal modeling; one based on Bayesian dependency graphs or graphical models which has been labeled as “Structural Causal Modeling” by White and Lu (2010). The other, apparently unrelated, approach rests on some variant of Granger Causality for which we prefer the terms WAGS influence⁸ for reasons stated below. WAGS influence modeling appeals to an improved predictability of one time series by another. We will describe these two streams of modeling, which leads us to anticipate their combination in a third line of work, called Dynamic Structural Systems (White and Lu, 2010).

Structural causal modeling: graphical models and Bayes-Nets

Structural Causal Modeling originated with Structural Equation Modeling (SEM) (Wright, 1921) and is characterized by the use of graphical models, in which direct causal links are encoded by directed edges in the graph (Lauritzen, 1996; Pearl, 2000; Spirtes et al., 2000). Ideally these edges can be given a mechanistic interpretation (Machamer et al., 2000). Using these graphs, statistical procedures then discover the best model (graph) given the data (Pearl, 2000; 2003; Spirtes et al., 2000). As explained in the previous section, the “best” model has the highest evidence. There may be many models with the same evidence; in this case, the statistical search produces an equivalence class of models with the same explanatory power. With regard to effective connectivity, the multiplicity of possibly equivalent models has been highlighted by Ramsey et al. (2010).

This line of work has furnished Statistical Causal Modeling with a rigorous foundation and specific graphical procedures such as the “Back-door” and “Front-door” criteria, to decide whether a given causal model explains observational data. Here, causal architectures are encoded by the structure of the graph. In fMRI studies these methods have been applied by Ramsey et al. (2010) to estimate directionality in several steps, first looking for “unshielded colliders” (paths of the form $A \rightarrow B \leftarrow C$) and then finding out what further dependencies are implied by these colliders. We now summarize Structural Causal Modeling, as presented by Pearl (2000).

One of the key concepts in Pearl’s causal calculus is *interventional probabilities*, which he denotes $p(x_i | do(X_i = x_i))$ or more simply $p(x_i | do(x_i))$, which are distinct from conditional probabilities $p(x_i | X_i = x_i)$. Pearl highlights the difference between the *action* $do(X_i = x_i)$ and the *observation* $X_i = x_i$. Note that observing $X_i = x_i$ provides information both about the children and parents of X_i in a directed acyclic graph (DAG⁹). However, whatever relationship existed between X_i and its parents prior to action, this relationship is no longer in effect when we perform the action $do(X_i = x_i)$. X_i is held fixed by the action $do(X_i = x_i)$, and therefore cannot be influenced. Thus, inferences based on evaluating $do(X_i = x_i)$ are different in nature from the usual conditional inference. Interventional probabilities are calculated via a *truncated factorization*; i.e. by conditioning on a “mutilated graph”, with the edges (links) from the parents of X_i removed:

$$p(x_i | do(x_i)) = \prod_{j \neq i} p(x_j | pa_j) = \frac{p(x)}{p(x_i | pa_i)}. \quad (7)$$

⁸ It might be preferable to use a more precise term “predictability” instead of influence.

⁹ DAG = Directed Acyclic Graph. The word ‘graph’ refers to the mapping from the set of (factorized) joint probability densities over X and the actual directed acyclic graph that represents the set of conditional independencies implicit in the factorization of the joint pdf $p(x)$.

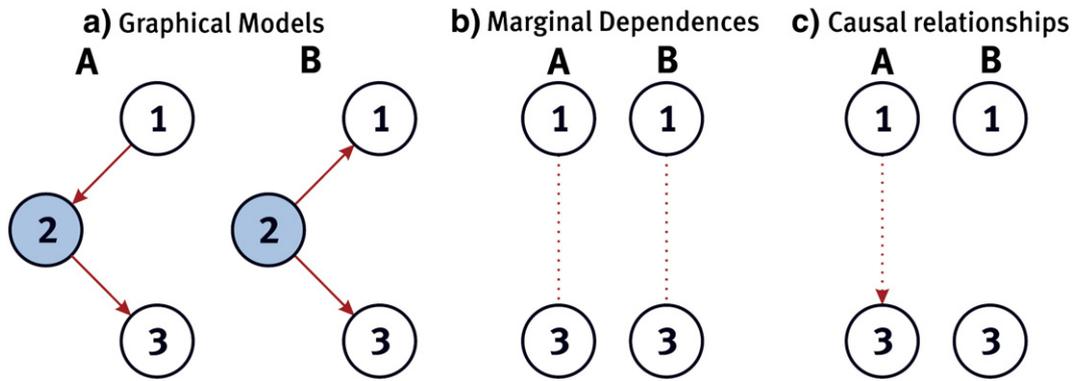


Fig. 5. The missing region problem. a) Two typical graphical models including a hidden node (node 2). b) Marginal dependence relationships implied by the causal structure depicted in (a), after marginalizing over the hidden node 2; the same moral graph can be derived from directed (causal) graphs A and B. c) Causal relationships implied by the causal structure depicted in (a), after marginalizing over the hidden node 2. Note that these are perfectly consistent with the moral graph in (b), depicting (non causal) statistical dependencies between nodes 1 and 3, which are the same for both A and B.

Here, pa_j denotes the set of all the parents of the j th node in the graph and $p(x)$ is the full joint distribution. Such interventional probabilities exhibit two properties:

$$\begin{cases} P1 : p(x_i | do(pa_i)) = p(x_i | pa_i) \\ P2 : p(x_i | do(pa_i), do(s)) = p(x_i | do(pa_i)) \end{cases} \quad (8)$$

for all i and for every subset S of variables disjoint of $\{X_i, PA_i\}$. Property 1 renders every parent set PA_i exogenous relative to its child X_i , ensuring that the conditional $p(x_i | pa_i)$ probability coincides with the effect (on X_i) of setting PA_i to pa_i by external control. Property 2 expresses the notion of invariance: once we control its direct causes PA_i , no other interventions will affect the probability of X_i . These properties allow us to evaluate the (probabilistic) effect of interventions from the definition of the joint density $p(x)$ associated with the pre-intervention graph.

This treatment of interventions provides a semantics for notions such as “causal effects” or “causal influence”. For example, to see whether a variable X_i has a causal influence on X_j , we compute (using the truncated factorization in Eq. (7)) the marginal distribution of X_j under the actions $do(X_i = x_i)$ and check whether that distribution is sensitive to x_i . It is easy to see that only descendants of X_i can be influenced by X_i ; deleting the factor $p(x_i | pa_i)$ from the joint distribution turns X_i into a root node¹⁰ in the mutilated graph. This can be contrasted with (undirected) probabilistic dependencies that can be deduced from the factorization of the joint distribution *per se*. These dependencies can be thought as (non-causal and non-directed) correlations among measured variables that can be predicted on the basis of the structure of the network.

In the context of brain connectivity, the measures of interventional and conditional probabilities map onto the notions of effective connectivity and functional connectivity respectively. Let us consider two typical situations that arise in the context of the missing region problem. These are summarized in Fig. 5.

Consider Fig. 5a. In situation A, node 1 influences node 2, which influences node 3. That is, the causal effect of 1 on 3 is mediated by 2. The joint distribution of the graphical causal model can be factorized as $p_A(x) = p(x_3 | x_2) p(x_2 | x_1) p(x_1)$. In situation B, both 1 and 3 have a common cause: node 2 influences both 1 and 3. The joint distribution of this graphical causal model can then be factorized as: $p_B(x) = p(x_1 | x_2) p(x_3 | x_2) p(x_2)$. It is easy to prove that in both cases (A and B), 1 and 3 are conditionally independent given 2; i.e., $p(x_1, x_3 | x_2) = p(x_1 | x_2) p(x_3 | x_2)$. This means that observing node 1 (respectively 3) does not convey additional information about 3 (respectively 1), once we know

¹⁰ A root node is a node without parents. It is marginally independent of all other variables in a DAG, except its descendants.

2. Furthermore, note that 1 and 3 are actually *marginally dependent*; i.e., $p(x_1, x_3) = \int p(x) dx_2 \neq p(x_1) p(x_3)$. This means that whatever value X_2 might take, X_1 and X_3 will be correlated. Deriving the marginal independencies from the DAG produces an undirected graph (see, e.g., Fig. 5b). This undirected graph is called a *moral graph* and its derivation is called the *moralization* of the DAG. For example, moralizing the DAG A produces a fully connected moral graph.

In brief, both situations (A and B) are similar in terms of their statistical dependencies. In both situations, functional connectivity methods would recover the conditional independence of nodes 1 and 3 if node 2 was observed, and their marginal dependence if it is not (see Fig. 5b). However, the situations in A and B are actually very different in terms of the causal relations between 1 and 3. This can be seen using the interventional probabilities defined above: let us derive the interventional probabilities expressing the causal influence of node 1 onto node 3 (and reciprocally) in situation A:

$$\begin{aligned} p_A(x_3 | do(\tilde{x}_1)) &= \int p_A(x_2, x_3 | do(\tilde{x}_1)) dx_2 \\ &= \int p(x_3 | x_2) p(x_2 | \tilde{x}_1) dx_2 \\ &= p(x_3 | \tilde{x}_1) \end{aligned} \quad (9)$$

$$\begin{aligned} p_A(x_1 | do(\tilde{x}_3)) &= \int p_A(x_1, x_2 | do(\tilde{x}_3)) dx_2 \\ &= p(x_1) \int p(x_2 | x_1) dx_2 \\ &= p(x_1). \end{aligned} \quad (10)$$

Eq. (8) simply says that the likelihood of any value that x_3 might take is dependent upon the value \tilde{x}_1 that we have fixed for x_1 (by intervention). In contradistinction, Eq. (9) says that the likelihood of any value that x_1 might take is independent of x_3 . This means that node 1 has a causal influence on node 3, i.e. there is a directed (mediated through 2) causal link from 1 to 3. The situation is quite different in B:

$$\begin{aligned} p_B(x_3 | do(\tilde{x}_1)) &= \int p_B(x_2, x_3 | do(\tilde{x}_1)) dx_2 \\ &= \int p(x_3 | x_2) p(x_2) dx_2 \\ &= p(x_3) \\ p_B(x_1 | do(\tilde{x}_3)) &= \int p_B(x_1, x_2 | do(\tilde{x}_3)) dx_2 \\ &= \int p(x_1 | x_2) p(x_2) dx_2 \\ &= p(x_1). \end{aligned} \quad (11)$$

This shows that nodes 1 and 3 are not influenced by intervention on the other. This means that here, there is no causal link between 1 and 3. This is summarized in Fig. 5c, which depicts the corresponding ‘effective’ causal graphs, having marginalized over node 2.

Causal calculus provides a simple but principled perspective on the “missing region” problem. It shows that effective connectivity analysis

can, in certain cases, address a subset of brain regions (subgraph), leaving aside potential variables (e.g., brain regions) that might influence the system of interest. The example above makes the precise confines of this statement clear: one must be able to perform interventional actions on source and target variables. Given that the principal ‘value-setting’ interventions available to us in cognitive neuroscience are experimental stimulus manipulations, our capacity for such interventions are generally limited to the primary sensory cortices. Intervention beyond sensorimotor cortex is much more difficult; although one could employ techniques such as transcranial magnetic stimulation (TMS) to perturb activity in superficial cortical areas. However, the perturbation in TMS is unnatural and known to induce compensatory changes throughout the brain rather than well-defined effects in down-stream areas.

The same undirected graph can be derived from the moralization of a set of DAGs (c.f. from Figs. 5a and b). This set contains a (potentially infinite) number of elements, and is referred to as the *equivalent class*. As stated by Pearl, the identification of causal (i.e., interventional) probabilities from observational data requires additional assumptions or constraints (see also Ramsey et al., 2010). Pearl mentions two such critical assumptions: (i) *minimality* and (ii) *structural stability*. Minimality appeals to complexity minimization, when maximizing model evidence (c.f., Occam's razor). In brief, among a set of causal models that would explain the observed data, one must choose the simplest (e.g., the one with the fewest parameters). Structural stability (also coined ‘faithfulness’) is a related requirement that is motivated from the fact that an absence of causal relationships is inferred from an observed absence of correlation. Therefore, if no association is observed, it is unlikely to be due to the particular instantiation of a given model for which this independence would be predicted (see below). Rather, it is more likely to be explained in terms of a model that would predict, for any parameter setting, the observed absence of correlation. This clearly speaks to the convergent application, mentioned above, of data driven exploratory approaches that scan the largest model space possible for correlations to be explained and a model driven confirmatory approach that appeal to structural stability: Within a Bayesian setting, we usually specify a prior distribution $p(\theta|m)$ over model parameters, which are usually assumed to be independent. This is justified when the parameters represent mechanisms that are free to change independently of one another—that is, when the system is structurally stable. In other terms, the use of such prior favors structurally stable models. In most cases, stability and minimality are sufficient conditions for solving the structure discovery inverse problem in the context of observational data. If this is not sufficient to reduce the cardinality of the equivalent class, one has to resort to experimental interventions.¹¹ Within the context of Neuroimaging, this would involve controlling the system by optimizing the experimental design in terms of the psychophysical properties of the stimuli and/or through direct biophysical stimulation (e.g., transcranial magnetic stimulation – TMS – or deep brain stimulation—DBS).

Summary

The causal calculus based on graphical models has some important connections to the distinction between functional and effective connectivity and provides an elegant framework in which one can deal with interventions. However, it is limited in two respects. First, it is restricted to discovering conditional independencies in *directed acyclic graphs*. This is a problem because the brain is a directed *cyclic* graph—every brain region is reciprocally connected (at least polysynaptically) and every computational theory of brain function rests

on some form of reciprocal or reentrant message passing. Second, the calculus ignores time: Pearl argues that what he calls a ‘causal model’ should rest upon *functional relationships* between variables, an example of which is structural equation modeling (SEM). However, these functional relationships cannot deal with (cyclic) feedback loops. In fact, DCM was invented to address these limitations, after evaluating structural causal modeling for fMRI time-series. This is why it was called *dynamic* causal modeling to distinguish it from *structural* causal modeling (Friston et al., 2003). Indeed, Pearl (2000) argues in favor of dynamic causal models, when attempting to identify what physicists call hysteresis effects, whereby the causal influence depends upon the history of the system. Interestingly, the DAG limitation can be finessed by considering dynamics and temporal precedence within structural causal modeling. This is because the arrow of time turns directed cyclic graphs into directed acyclic graphs, when the nodes are deployed over successive time points. This leads us to an examination of prediction-based measures of functional relations.

WAGS influence

The second stream of statistical causal modeling is based on the premise that a cause must precede and increase the predictability of its consequence. This type of reasoning can be traced back at least to Hume (Triacca, 2007) and is particularly popular in time series analysis. Formally, it was originally proposed (in an abstract form) by Wiener (1956) (see Appendix A) and introduced into data analysis by Granger (1963). Granger emphasized that increased predictability is a necessary but not sufficient condition for a causal relation to exist. In fact, Granger distinguished between true causal relations and “prima facie” causal relations (Granger, 1988); the former only to be inferred in the presence of “knowledge of the state of the whole universe”. When discussing “prima facie causes” we recommend the use of the neutral term “influence” in agreement with other authors (Commenges & Gégout-Petit, 2009; Gégout-Petit & Commenges, 2010). Additionally, it should be pointed out that around the same time as Grangers work, Akaike (1968), and Schweder (1970) introduced similar concepts of influence, prompting us to refer to “WAGS influence modeling” (for Wiener–Akaike–Granger–Schweder). This is a generalization of a proposal by Aalen (1987) and Aalen and Frigessi (2007) who were among the first to point out the connections between the Granger and Shweder concepts.

An unfortunate misconception in Neuroimaging identifies WAGS influence modeling (WAGS for short) with just one of the specific proposals (among others) dealt with by Granger; namely, the discrete-time linear Vector Autoregressive Model (VAR). This simple model has proven to be a useful tool in many fields, including Neuroimaging—the latter work well documented in Bressler and Seth (2010). However, this restricted viewpoint overlooks the fact that WAGS has dealt with a much broader class of systems:

1. Classical textbooks, such as Lutkepohl (2005), show how WAGS can applied VAR models, infinite order VAR, impulse response functions, Vector Autoregressive Moving Average models (VARMA), etc.
2. There are a number of nonlinear WAGS methods that have been proposed for analyzing directed effective connectivity (Freiwald et al., 1999, Solo, 2008; Gouieroux et al., 1987; Marinazzo et al., 2011; Kalitzin et al., 2007)
3. Early in the econometrics literature, causal modeling was extended to linear and nonlinear random differential equations in continuous time (Bergstrom, 1988). These initial efforts have been successively generalized (Aalen, 1987; Commenges & Gégout-Petit, 2009; Comte & Renault, 1996; Florens & Fougere, 1996; Gill & Petrović, 1987; Gégout-Petit & Commenges, 2010; Mykland, 1986; Petrović & Stanojević, 2010) to more inclusive types of dynamical systems.
4. Schweder (1970) describes WAGS concepts for counting processes in continuous, time which has enjoyed applications in Survival

¹¹ For example, the back- and front-door criteria (Pearl, 2000) can be used to optimize the intervention.

Analysis—a formalism that could well be used to model interactions expressed in neural spike train data.

We now give an intuitive explanation of some of these definitions (the interested reader can refer to the technical literature for more rigorous treatments). Let us again consider triples of (possibly vector) time series $X_1(t), X_2(t), X_3(t)$, where we want to know if time series $X_1(t)$ is influenced by time series $X_2(t)$ conditional on $X_3(t)$. This last variable can be considered as any time series to be controlled for (if we were omniscient, the “entire universe”!). Let $X[a, b] = \{X(t) | t \in [a, b]\}$ denote the history of a time series in the discrete or continuous time interval $[a, b]$. There are several types of influence. One distinction is based on what part of the present or future of $X_1(t)$ can be predicted by the past or present of $X_2(\tau) \tau < t$. This leads to the following classification:

- If $X_2(\tau): \tau < t$, can influence any future value of $X_1(s)$ for $s > t$, then it is a global influence.
- If $X_2(\tau) \tau < t$, can influence $X_1(t)$ it is a local influence.
- If $X_2(\tau) \tau = t$ can influence $X_1(t)$ it is a contemporaneous influence.

Another distinction is whether one predicts the whole probability distribution (**strong** influence) or only given moments (**weak** influence). These two classifications give rise to six types of influence as schematized in Fig. 6 and Table 4 and 5. Briefly, the formal definitions are as follows.

$X_1(t)$ is strongly, conditionally, and globally independent of $X_2(t)$ given $X_3(t)$ (not SCGi), if

$$P(X_1(\infty, t] | X_1(t, -\infty], X_2(t, -\infty], X_3(t, -\infty]) = P(X_1(\infty, t] | X_1(t, -\infty], X_3(t, -\infty]). \tag{12}$$

When this condition does not hold we say $X_2(t)$ strongly, conditionally, and globally influences (SCGi) $X_1(t)$ given $X_3(t)$. Note that the whole future of X_1 is included (hence the term “global”). And the whole past of all time series is considered. This means these definitions accommodate non-Markovian processes (for Markovian processes, we only consider the previous time point). Furthermore, these definitions do not depend on an assumption of linearity or any given functional form (and are therefore applicable to any of the state equations in Table 2). Note also that this definition is appropriate for point processes, discrete and continuous time series, even for categorical (qualitative valued) time series. The only problem with this formulation is that it calls on the whole probability distribution and therefore its practical assessment requires the use of measures such as mutual information.

$X_1(t)$ is weakly, conditionally and globally independent of $X_2(t)$ given $X_3(t)$ (not WCGi), if

$$E[X_1(\infty, t] | X_1(\infty, t], X_2(t, -\infty], X_3(t, -\infty)] = E[X_1(\infty, t] | X_1(t, -\infty], X_3(t, -\infty)]. \tag{13}$$

If this condition does not hold we say $X_2(t)$ weakly, conditionally and globally influences (WCGi) $X_1(t)$ given $X_3(t)$. This concept extends to any number of moments (such as the variance of the process). There are a number of relations between these concepts: not SCGi implies not WCGi for all its moments and the converse is true for influences (WCGi implies SCGi), but we shall not go into details here; see Florens and Mouchart (1985), Florens (2003), Florens and Fougere (1996), and Florens and Mouchart (1982).

Global influence refers to influence at any time in the future. If we want to capture the idea of immediate influence we use the local

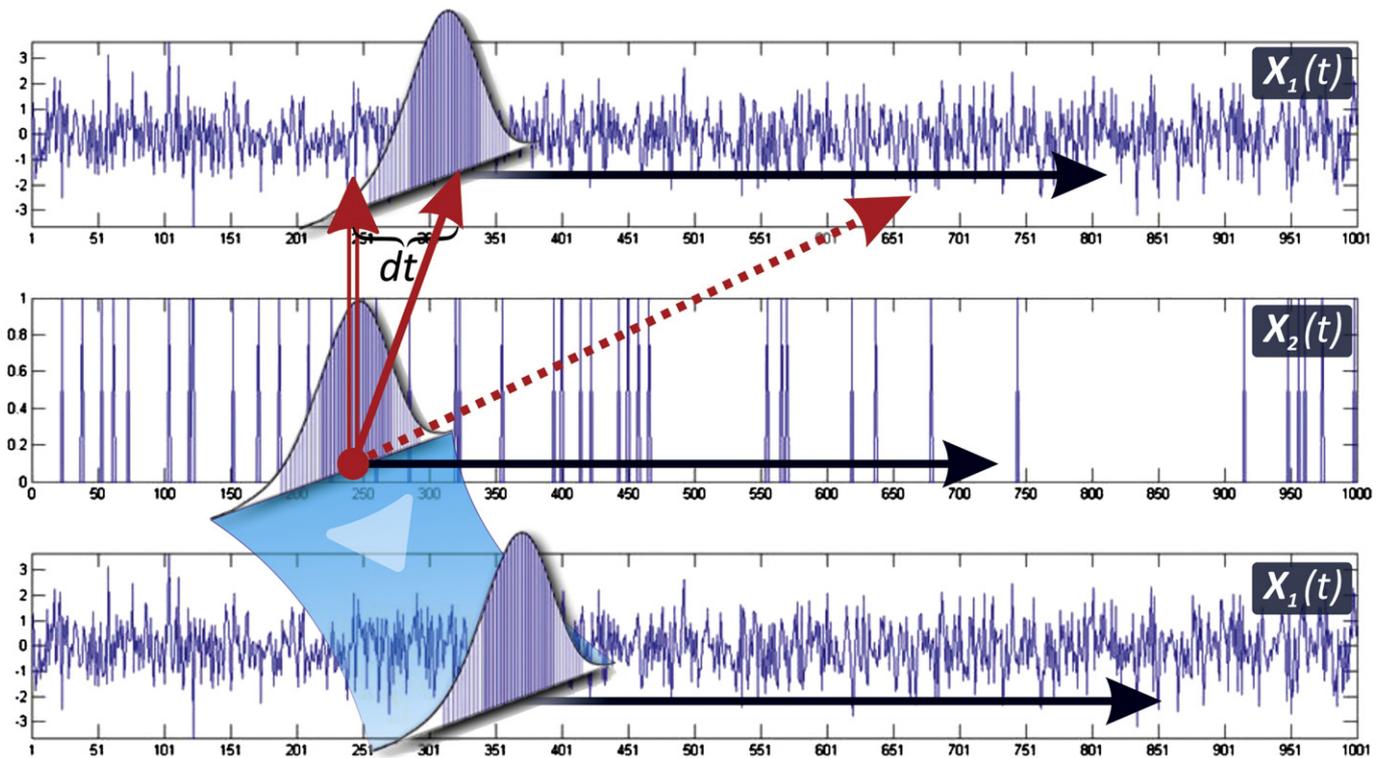


Fig. 6. Wiener-Akaike-Granger-Schweder (WAGS) Influences. This figure illustrates the different types of WAGS influence measures. In the middle $X_2(t)$ a continuous time point process, which may be influencing the differentiable continuous time process $X_1(t)$ (top and bottom). This process may have local influence (full arrows), which indicate predictability in the immediate future (dt), or global influence (dashed arrow) at any set of future times. If predictability pertains to the whole probability distribution, this is a strong influence (bottom), and a weak influence (top) if predictability is limited to the moments (e.g., expectation) of this distribution.

Table 4
Conditional Independence relations.

	Strong (Probability Distribution)	Weak (Expectation)
Global (for all horizons)	Strongly, Conditionally, Globally, independence (not SCGi)	Weakly, Conditionally, Globally, independence (not WCGi)
Local (Immediate future)	Strongly, Conditionally, Locally, independence (not SCLi)	Weakly, Conditionally, Globally, independence (not WCLi)
Contemporaneous	Strongly, Conditionally, Contemporaneously, independence (not SCCi)	Weakly, Conditionally, Contemporaneously, independence (not WCCi)

Table 5
Types of Influence defined by absence of the corresponding independences in Table 4.

	Strong (Probability Distribution)	Weak (Expectation)
Global (for all horizons)	Strongly, Conditionally, Globally, influence (SCGi) - Strong Granger or Sims influence	Weakly, Conditionally, Globally, influence (WCGi) - Weak Granger or Sims influence
Local (Immediate future)	Strongly, Conditionally, Locally, influence (SCLi) - Influence (Possibly indirect)	Weakly, Conditionally, Globally, influence (WCLi) - Direct Influence
Contemporaneous	Strongly, Conditionally, Contemporaneously, influence (SCCi)	Weakly, Conditionally, Contemporaneously, influence (WCCi)

concepts defined above. The concepts of strong and weak local influence have very simple interpretations if we are modeling in discrete time and events occur every Δt . To see this, consider the expectation based weak conditionally local independence (not WCLi) in discrete time:

$$E[X_1(t + \Delta t) | X_1[t, -\infty], X_2[t, -\infty], X_3[t, -\infty]] = E[X_1(t + \Delta t) | X_1[t, -\infty], X_3[t, -\infty]]. \tag{14}$$

If this condition does not hold we have that $X_2(t)$ weakly, conditionally and locally influences (WCLi) $X_1(t)$ given $X_3(t)$. Strong local concepts are defined similarly by considering conditional independences. For the usual discrete time, real valued time series of Neuroimaging, all these concepts are equivalent as shown by Florens and Mouchart (1982) and Solo (2007). As an example, consider the multivariate autoregressive model of the previous section

$$X(t + \Delta t) = \sum_{k=1}^p A_k X(t - (k-1)\Delta t) + e(t + \Delta t) \tag{15}$$

with the innovation term $e_{t+\Delta t}$ being GWN with covariance matrix $\Sigma := \Sigma_{discrete}$. For this familiar case $E[X[t + \Delta t] | X[t, -\infty]] = \sum_k A_k X(t - (k-1)\Delta t)$, and analyzing influence reduces to finding which coefficients of the autoregressive coefficients are zero. However, in continuous time there is a problem when $\Delta t \rightarrow 0$, since the stochastic processes we are dealing with are at least almost surely continuous and $\lim_{\Delta t \rightarrow 0} E[X_1(t + \Delta t) | X_1[t, -\infty], X_2[t, -\infty], X_3[t, -\infty]] = \lim_{\Delta t \rightarrow 0} E[X_1(t + \Delta t)]$ is trivially satisfied (limits are now taken in the sense of a quadratic mean) because the $X_1(t)$ process is path continuous—it will only depend on itself. To accommodate this situation instead we shall use the following definition for not WCLi (Commenges & Gégout-Petit, 2009; Comte & Renault, 1996; Florens & Fougere, 1996; Gégout-Petit & Commenges, 2010; Renault, Sekkat, & Szafarz, 1998):

$$\lim_{\Delta t \rightarrow 0} E \left[\frac{X_1(t + \Delta t) - X_1(t)}{\Delta t} \middle| X_1(t, -\infty), X_2(t, -\infty), X_3(t, -\infty) \right] = \lim_{\Delta t \rightarrow 0} E \left[\frac{X_1(t + \Delta t) - X_1(t + \Delta t)}{\Delta t} \middle| X_1(t, -\infty), X_3(t, -\infty) \right]. \tag{16}$$

As noted by Renault et al. (1998) (whom we follow closely here), for finite Δt this is equivalent to the usual definitions. Now how does this definition relate to the linear SDE in Eq. (3)?

For three time series:

$$\begin{pmatrix} dX_1(t) \\ dX_2(t) \\ dX_3(t) \end{pmatrix} = A \begin{bmatrix} X_1(t) \\ X_2(t) \\ X_3(t) \end{bmatrix} dt + dB(t). \tag{17}$$

Integrating from t to Δt , we have

$$X_1(t + \Delta t) - X_1(t) = \int_t^{t+\Delta t} [a(1,1)X_1(\tau) + a(1,2)X_2(\tau) + a(1,3)X_3(\tau) + \sigma_{bb}(B_1(t + \Delta t), -B_2(t))] d\tau$$

$$\Rightarrow \lim_{\Delta t \rightarrow 0} E \left[\frac{X_1(t + \Delta t) - X_1(t)}{\Delta t} \middle| X_1(t), X_2(t), X_3(t) \right] = a(1,1)X_2(t) + a(1,2)X_2(t) + a(1,3)X_3(t).$$

This shows that, in effect, the detection of an influence will depend on whether the coefficients of the matrix A are zero or not. For nonlinear systems this holds with the local linear approximation. This treatment highlights the goal of WAGS, like structural causal modeling, is to detect conditional independencies; in this (AR) example, weak and local.

The issue of contemporaneous influence measures is quite problematic. In discrete time, it is clear that the covariance matrix of two or more time series may have cross-covariances that are due to an “environmental” or missing variable $Z(t)$. This was discussed by Akaike and a nice example of this effect is described in Wong and Ozaki (2007), which also explains the relation of the Akaike measures of influence to others used in the literature. For continuous time (Comte and Renault, 1996) define strong (second order) conditional contemporaneous independence (not SCCi) if:

$$\text{cov}[X_1(\infty, t], X_2(\infty, t) | X_1[t, -\infty], X_2[t, -\infty], X_3[t, -\infty]] = 0. \tag{18}$$

Note that this is the same definition for continuous time as for the discrete AR example (Eq. (15)) and is equivalent to requiring that the elements of the corresponding innovation covariance matrix Σ be zero. These authors then went on to define weak contemporaneous conditional independence (not WCCi) if:

$$\lim_{\Delta t \rightarrow 0} \{ \text{cov}[X_1(t + \Delta t), X_2(t + \Delta t) | X_1[t, -\infty], X_2[t, -\infty], X_3[t, -\infty]] \} = 0. \tag{19}$$

In the absence of these conditions we have strong (weak) contemporaneous conditional influences which are clearly non-

directional. In his initial paper (Granger, 1963) defined a contemporaneous version of his influence measure in discrete time. Much later, (Geweke, 1984) decomposed his own WAGS measure into a sum of parts, some depending on lag information and others reflecting contemporaneous (undirected) influences, see in these C&C (Bressler and Seth, 2010). However, Granger (in later discussions) felt that if the system included all relevant time series this concept would not be valid, unless these influences were assigned a directionality (see Granger, 1988, pp. 204–208). In this sense, he was proposing a Structural Equation Modeling approach to the covariance structure of the autoregressive model innovations. As will be mentioned below (WAGS influence section) this is something that has been explored in the econometrics literature by Demiralp and Hoover (2008), Moneta and Spirtes (2006), but not to our knowledge in Neuroimaging.

More general models

As we have seen, strong global measures of independence are equivalent to conditional independence and are therefore applicable to very general stochastic processes. For weak local conditional independence, the situation is a little more difficult and we have given examples, which involve a limit in the mean of a derivative-type operator expression. The more general theory, too technical to include here, entails successive generalizations by Mykland (1986), Aalen (1987), Commenges and Gégout-Petit (2009), and Gégout-Petit and Commenges (2010). The basic concept can be stated briefly as follows (we drop conditioning on a third time series for convenience). Suppose we have stochastic processes that are semi-martingales of the form, $X(t) = P^X(t) + M^X(t)$. Here $P^X(t)$ is a predictable stochastic process¹² of bounded variation, which is known as the “compensator” of the semi-martingale, and M_t^X is a martingale.¹³ Predictability is the key property that generalizes Wiener’s intuition. The martingale component is the unpredictable part of the stochastic process we are interested in.¹⁴ Now suppose we have two stochastic processes $X(t)$ and $W(t)$. If:

1. The martingales M^{X_1} and M^{X_2} are orthogonal (no contemporaneous interactions).
2. $P^{X_1}(t)$ is measurable¹⁵ with respect to $X_1[t, -\infty]$ only (without considering $X_2(t)$).

then $X_1(t)$ is said to be weakly locally independent of $X_2(t)$. In Gégout-Petit and Commenges (2010) the concept of \sim WLCI is generalized to a general class of random phenomena that include random measures, marked point processes, diffusions, and diffusions with jumps, covering many of the models in Table 2. In fact, this theory may allow unification of the analysis of random behavioral events, LFP, spike recordings, and EEG, just to give a few examples.

¹² Roughly speaking, if $P^X(t)$ is a predictable process, then it is “known” just ahead of time t . For a rigorous definition and some discussion see <http://myyn.org/m/article/predictable-process/>.

¹³ For a martingale $M(t)$, $E(M(t+s)|X[t, -\infty]) = M(t)$ for all t and s . This states that the expected value of $M(t+s)$ is that at time t , there is no “knowledge” (in the sense of expected value) for the future from the past, hence this type of process is taken as a representation of unpredictability.

¹⁴ This is a form of the famous Doob–Meyer decomposition of a stochastic process (Medvegyev, 2007).

¹⁵ Roughly speaking $P^X(t)$ is measurable with respect to the process $X_1[t, -\infty]$ and not $X_2[t, -\infty]$ if all expected values of $P^X(t)$ can be obtained by integrating $X_1[t, -\infty]$ without reference to $X_2[t, -\infty]$. The technical definition can be found in Medvegyev (2007). Basically this definition is based on the concept of a “measurable function” extended to the sets of random variables that comprise the stochastic processes.

Direct influence

Weak local independence might be considered an unnecessarily technical condition for declaring the absence of an influence; in that strong (local or global) influence measures should be sufficient. An early counterexample of this was provided by Renault et al. (1998), where they considered a model where $X(t)$ is \sim WLCI of $W(t)$, given $Z(t)$. See Fig. 7 for an illustration of this divergence between local and global influences. This has led Commenges and Gégout-Petit (2009) to define WLCI as the central concept for “direct influence” whereas SCGI is an influence that can be mediated directly or indirectly through other time series.

An important point here is the degree to which the definition of WAGS influence depends on the martingale concept or, indeed, on that of a stochastic process. As discussed in The observation equation section, there are a number of instances in which Markovian models developed for financial time series may not apply for Neuroimaging data. However, the concepts are probably generally valid, as we shall illustrate with some examples:

- The analytical random processes used in generalized coordinates are quite different from those usually studied in classical SDE theory but have been known for a long time (Belyaev, 1959). In fact, there has been quite a lot of work on their predictability (Lyman et al., 2000) and indeed there is even work on VARMA modeling of this type of process (Pollock, 2010).
- We have already seen that the definitions of influence do not depend on Markovian assumptions as noted by Aalen (1987).
- The use of deterministic bilinear systems in DCM (Penny et al., 2005) suggests that (non-stochastic) ODEs may be incorporated into the WAGS framework. This sort of assimilation has in fact been proposed by Commenges and Gégout-Petit (2009) as a limiting case of the definition based on semi-martingales above. Extensions of the definition might be required when dealing with chaotic dynamics but, even here, measure theoretic definitions are probably valid.¹⁶ An interesting discussion of determinism versus stochastics can be found in Ozaki (1990).

The use or development of WAGS theory for systems that were not initially considered by the aforementioned papers may well be a fruitful area of mathematical research. In particular, WAGS may be especially powerful when applied to processes defined on continuous spatial manifolds (Valdes-Sosa, 2004; Valdés-Sosa et al., 2006). To our knowledge, WAGS has yet to be developed for the case of continuous time and space models; for example, those expressed as stochastic or random Partial Differential Equations.

Testing and measuring WAGS influence

Above, we have covered different types of WAGS influence. With these definitions in place we now distinguish between testing for the presence of an influence (inference on models) and estimating the strength of the influence (inference on parameters). There is an extensive literature on this, which we shall not go into here. Examples of testing versus measuring for discrete time VAR models include the Dickey–Fuller test and the Geweke measure of influence. In the electrophysiological literature, there are a number of measures proposed. A review and a toolbox for these measures can be found in Seth (2009). From the point of view of effective connectivity, many of these measures have an uncertain status. This is because effective connectivity is only defined in relation to a generative model. In turn, this means there are only two quantities of interest (that permit

¹⁶ In particular the Sinai–Ruelle–Bowen measure for hyperbolic dynamical systems (Chueshov, 2002).

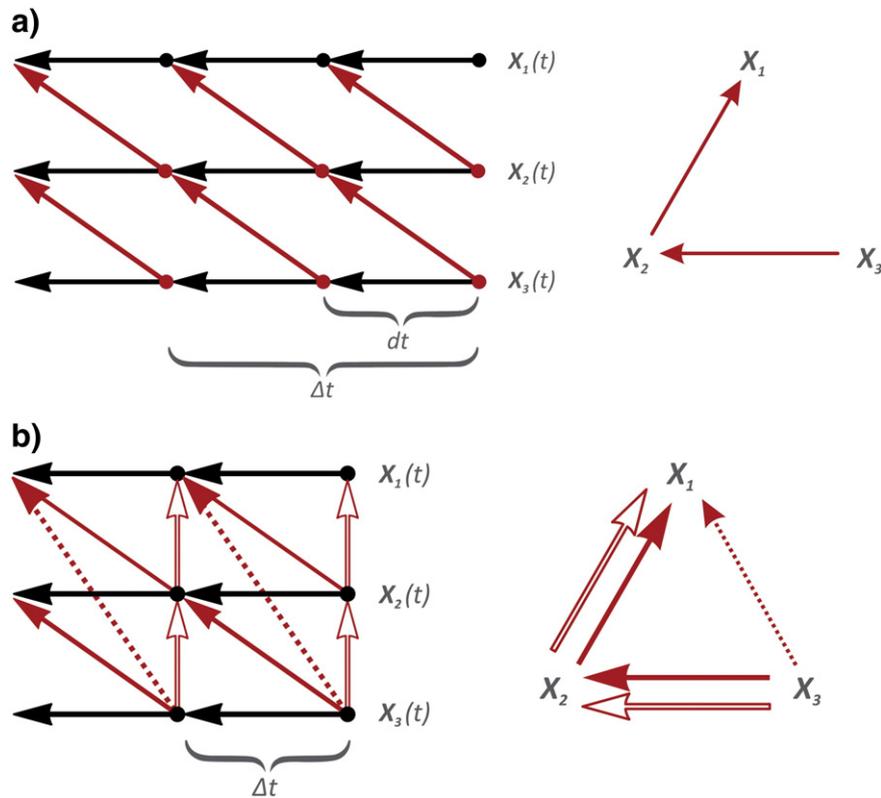


Fig. 7. The missing time problem. This figure provides a schematic representation of spurious causality produced by sub-sampling. a) Three time series $X_1(t)$, $X_2(t)$, and $X_3(t)$ are shown changing at an “infinitesimal” time scale with steps dt , as well as at a coarser sampled time scale with set Δt . Each time series, influences itself at later moments. In the example $X_3(t)$ directly influences $X_2(t)$, with no direct influence on $X_1(t)$. In turn $X_2(t)$ directly influences $X_1(t)$, with no direct influence on $X_3(t)$. Finally $X_1(t)$ does not influence either $X_3(t)$ nor $X_2(t)$. There are no contemporaneous influences. b) When only observing at the coarser time scale Δt , spurious contemporaneous influences (mediated by intermediate nodes) appear between $X_2(t)$ and $X_1(t)$ and between $X_3(t)$ and $X_2(t)$. In addition a spurious direct influence appears between $X_3(t)$ and $X_1(t)$. The graphical representations of the true and spurious causal relations are to the right of each figure where an arrow represents direct influence and a double arrow represents contemporaneous influence. Estimating these spurious influences can only be avoided by explicitly modeling their effect from continuous models or using models such as VARMA models which are resistant to this phenomena.

inference on models and parameters respectively): the relative evidence for a model with and without a connection and the estimate (conditional density over) the connection parameter. For DCM the first quantity is the Bayes factor and for GCM it is the equivalent likelihood ratio (Granger causal F-statistics). In DCM, the conditional expectation of the parameter (effective connectivity) measures the strength, while for GCM this is the conditional estimate of the corresponding autoregression coefficient. Other measures (e.g., partial directed coherence) are simply different ways of reporting these conditional estimates. The next section explores the use of WAGS measures of direct and indirect effects within the Structural Causal modeling framework, thus bringing together the two major strands of statistical causal modeling.

Dynamic structural causal modeling

There have been recent theoretical efforts to embed WAGS into Structural Causal Modeling, which one could conceive of (in the language of Granger) as providing a means to find out which “prima facie causes” are actual “causes”. One of the first people to use the methods from Structural Causal Modeling was Granger himself: Swanson and Granger (1997) used Bayes-Net methods described in Spirtes et al. (2000) in combination with autoregressive modeling. Similar approaches have been adopted by Demiralp and Hoover (2008) and Moneta and Spirtes (2006), which address the search for directed contemporaneous influences mentioned above.

However, we should mention three current attempts to combine Structural Causal Modeling with WAGS influence analysis. We shall

follow White in calling models that can be described by both theoretical frameworks Dynamic Structural Systems:

1. Eichler has been developing graphical time series models that are based on discrete time WAGS. Recently, in work with Didelez the formalization of interventions has been introduced and equivalents for the backdoor and front-door criteria of Structural Causality have been defined. Thus, for discrete systems, this work could result in practical criteria for defining when it is possible to infer causal structure from WAGS in discrete time.
2. White has created a general formalism for Dynamical Structural Systems (White and Lu, 2010) based on the concept of settable systems (White and Chalak, 2009), which supports model optimization, equilibrium and learning. The effects of intervention are also dealt with explicitly.
3. Commenges and Gégout-Petit (2009) have also proposed a general framework for causal inference that combines elements of Bayes-Nets and WAGS influence and has been applied to epidemiology. Specifically, as mentioned above, they introduce a very general definition of WAGS that is valid for continuous/discrete time processes. This definition can be applied to a mixture of SDEs and point processes and distinguishes between direct influences and indirect influences. They then relate the definition to graphical models, with nodes connected by direct influences only and place their work in the context of General Systems Theory. Interestingly, they stress the need for an observation equation to assure causal explanatory power.

The common theme of all these efforts is to supplement predictability with additional criteria to extend WAGS influence to inference on

causal mechanisms. In the words of [Gégout-Petit and Commenges \(2010\)](#): “A causal interpretation needs an epistemological act to link the mathematical model to a physical reality.” We will illustrate these ideas with a particular type of SSM, known as a (stochastic) dynamic causal model (DCM):

$$\begin{cases} \dot{x} = f(x, \theta, u) + \omega \\ y = g(x, \theta) + \varepsilon \end{cases} \quad (20)$$

where x are (hidden) states of the system, θ are evolution parameters, u are the experimental control variables, ω are random fluctuations and ε is observation noise. Inverting this model involves estimating the evolution parameters θ , which is equivalent to characterizing the structural transition density $p(\dot{x}|do(x))$, having accounted for observational processes.¹⁷ Here, time matters because it prevents instantaneous cyclic causation, but still allows for dynamics. This is because identifying the structural transition density $p(\dot{x}|do(x))$ effectively decouples the children of $X(t)$ (in the future) from its parents (in the past). Let us now examine a bilinear form of this model

$$f(x) = Ax + \sum_i u_i B^{(i)} x + Cu + \sum_j x_j D^{(j)} x. \quad (21)$$

Then we have:

$$\begin{aligned} A &= \lim_{x,u \rightarrow 0} \frac{\partial}{\partial x} E[\dot{x}|do(x)] \\ B^{(i)} &= \frac{\partial^2}{\partial x \partial u_i} E[\dot{x}|do(x)] \\ C &= \lim_{x \rightarrow 0} \frac{\partial}{\partial u_i} E[\dot{x}|do(x)] \\ D^{(j)} &= \frac{\partial^2}{\partial x \partial x_j} E[\dot{x}|do(x)]. \end{aligned} \quad (22)$$

The meaning of A ; i.e. the effective connectivity is the rate of change (relative to x) of the expected motion $E[\dot{X}]$ where X is held at $x \approx 0$.¹⁸ It measures the *direct effect* of connections. Importantly, *indirect effects* can be derived from the effective connectivity. To make things simple, consider the following 3-region DCM depicted in [Fig. 8](#):

$$\begin{aligned} \dot{x}_1 &= A_{11}x_1 + \omega_1 \\ \dot{x}_2 &= A_{21}x_1 + A_{22}x_2 + \omega_2 \\ \dot{x}_3 &= A_{31}x_1 + A_{32}x_2 + A_{33}x_3 + \omega_3. \end{aligned} \quad (23)$$

The effect of node 1 on node 3 is derived from the calculus of the intervention $do(X_1 = x_1)$, where X_1 is held constant at x_1 but X_2 is permitted to run its natural course. This intervention confirms that node 1 has both a direct and an indirect effect on node 3 (through node 2).¹⁹ Interestingly, indirect effects can also be derived by

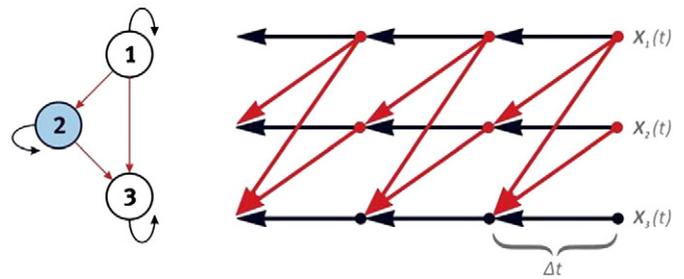


Fig. 8. Direct and indirect effects. Causal relationships implied by the DCM given in Eq. (23). On the left the apparent graph, that includes feedback which precludes causal analysis. Note that the causal links are actually expressed through implicit delays, which makes this graph a DAG, which is seen more clearly on the right where each node is expanded at several time instants.

projecting Eq. (20) onto generalized coordinates; i.e. by deriving the evolution function of the augmented state space $\tilde{x} = (x, \dot{x}, \ddot{x}, \dots)^T$ (see Friston et al., 2008a,b for a variational treatment of stochastic dynamical systems in generalized coordinates). For example, deriving the left and the right hand side of the last equation in Eq. (23) with respect to time yields:

$$\begin{aligned} \ddot{x}_3 &= \tilde{A}_{31}x_1 + \tilde{A}_{32}x_2 + \tilde{A}_{33}x_3 + \tilde{\omega}_3 \\ \tilde{A}_{31} &= \underbrace{A_{31}(A_{11} + A_{33})}_{\text{direct effect}} + \underbrace{A_{32}A_{21}}_{\text{indirect effect}} \\ \tilde{A}_{32} &= A_{32}(A_{22} + A_{33}) \\ \tilde{A}_{33} &= A_{33}A_{33} \end{aligned} \quad (24)$$

where $\tilde{\omega}_3$ lumps all stochastic inputs (and their time derivatives) together. The total effect of node 1 onto node 3 is thus simply decomposed through the above second order ODE (Eq. (24)), as the sum of direct and indirect effects. One can see that the indirect causal effect of node 1 on node 3 is proportional to the product $A_{32}A_{21}$ of the path coefficients of the links $[1 \rightarrow 2]$ and $[2 \rightarrow 3]$. This speaks to a partial equivalence of the *do* calculus and the use of generalized coordinates, when modeling both direct and mediated (indirect) effects. This is because embedding the evolution equation into a generalized coordinates of motion naturally accommodates dynamics and the respective contributions of direct/indirect connections (and correlations induced by non-Markovian state noise ω). However, the embedding (truncation) order has to be at least as great as the number of intermediary links to capture indirect effects.

This type of reasoning is very similar to the treatment of direct and indirect influences under WAGS influence and exemplifies a convergence of Structural Causal (Bayes-Net) Modeling and WAGS influence. One could summarize this ambition by noting the “arrow of time” converts realistic (cyclic) graphical models – that include feedback and cyclic connections – into a DAG formalism, to allow full causal inference. So what are the limits of this approach in Neuroimaging?

Challenges for causal modeling in Neuroimaging

The papers in this C&C highlight challenges that face methods for detecting effective connectivity. These challenges arise mainly in the analysis of BOLD signals. To date, the only experimental examination of these issues is reported in the paper that originated this series ([David et al., 2008](#)). The main message from the ensuing exchanges is the need to account for the effect of the HRF; that is, to include an appropriate observation model in the analysis, along with careful evaluation of form, priors and Identifiability.

Another approach to testing the validity and limits of the methods discussed above has been through computer simulations. The results of these simulations have been mixed. A number of papers have supported the use of GCM in fMRI ([Deshpande et al., 2009](#); [Stevenson](#)

¹⁷ Note that the interventional interpretation of DCM is motivated by the (temporal) asymmetry between the left- and the right-hand terms in Eq. (24). Its right-hand term gives us the expected rate of change $E[\dot{X}(t)]$ of $X(t)$ if we fix $X(t)$ to be x (i.e. if we perform the action $do(x)$), but does not provide any information about what $X(t)$ is likely to be if we fix its rate of change $\dot{X}(t)$. This is best seen by noting that the system’s motion $\dot{X}(t)$ is a proxy for the system’s *future* state $X(t + \Delta t)$, which cannot influence its own past $X(t)$. Interestingly, this shows how interventional and prediction-over-time oriented (i.e. WAGS) interpretations of DCM are related.

¹⁸ The original motivation for the neural evolution equation of DCM for fMRI data considered the system’s states x as being perturbations around the steady-state activity x_0 . Thus, $x = 0$ actually corresponds to steady (background) activity within the network (x_0).

¹⁹ Interventional probabilities in a dynamical setting have recently been derived in, e.g., [Eichler and Didelez \(2010\)](#).

and Körding, 2010; Witt and Meyerand, 2009). Others have shown advantages for Bayes-Net methods in short time series and for GCM for longer time series (Zou et al., 2009).

An extensive set of simulations (NETSIM) has been carried out by Smith et al. (2010b) using non-stationary (Poisson-type) neural innovations in several configurations of nodes and simulating hemodynamics using the fMRI version of DCM. Many different methods were compared (apart from DCM), distinguishing between those that estimate undirected association (functional connectivity) from those that estimate “lagged” dependence (essentially a form of effective connectivity). The main conclusion was that a few undirected association methods that only used the information in the zero lag covariance matrixes perform well in identifying functional connectivity from fMRI. However, lag-based methods “perform worse”. We speculate that lag information is lost by filtering with a (regionally variable) HRF and sub-sampling. Thus one could expect that (stochastic) DCM might perform better, as supported by a comparison of SEM and DCM (Penny et al., 2004).

Interesting as these results are, several points remain unresolved. In the first place, more biophysically realistic simulations are called for, especially in the simulation of neurodynamics. The neurodynamics model in DCM for fMRI is intentionally generic, to ensure identifiability when deconvolving fMRI time-series. There is work suggesting that discrete time Vector Autoregressive Moving Average models are immune to sub-sampling and noise relative to VAR models (Amendola et al., 2010; Solo, 1986; 2007). Considering that WAGS influence modeling with VARMA models is in the standard time series textbooks (Lutkepohl, 2005), it is surprising that this model has not been used in Neuroimaging, with the notable exception of (Victor Solo, 2008).

NETSIM has not yet been tested using continuous time models. The problem, as pointed out by the creators of NETSIM and (Roebroeck et al., 2005), is not only sub-sampling but the combined effect of sub-sampling and the low pass filtering of the HRF. However, these problems only pertain to AR models. Continuous time DCMs have an explicit forward model of (fast) hidden states and are not confounded by sub-sampling or the HRF, provided both are modeled properly in the DCM. The key issue is whether DCM can infer hidden states in the absence of priors (i.e., stimulus functions) that are unavailable for design-free (resting state) fMRI studies of the sort generated by NETSIM. This is an unsettled issue that will surely be followed up in the near future, with the use of biophysically more informed models and new DCM developments; e.g., DCM in generalized coordinates, stochastic DCMs and the DCM–GCM combinations that are being tested at the moment.

It should further be noted that the effect of sub-sampling (and hemodynamic convolution) are only a problem at certain spatial and temporal scales. Undoubtedly it must be a concern, when inferring the dynamics of fast neural phenomena. However, it is clear that brain activity spans many different spatial (Michael Breakspear and Stam, 2005) and temporal (Vanhatalo et al., 2005) scales. Multi-scale time series methods (including WAGS influence measures) have already been used in econometrics (Gencay et al., 2002) and could be applied in neuroscience.

One example of events that occur at a time scale that is probably sufficiently slow to allow simple (AR) WAGS influence analysis are resting state fluctuations observed in concurrent EEG/fMRI recordings. The analysis of causal relations between EEG and BOLD have been studied by several authors (Eichler, 2005; Jiao et al., 2010; Valdés-Sosa et al., 2006) and is illustrated in Fig. 4: The autoregressive coefficients of this first order sparse VAR model suggest that:

1. There are hardly any lag 0 (or contemporaneous) interactions between ROIs.
2. The only coefficients that survive the FDR threshold in the fMRI are those that link each ROI to its own past.

3. There is no influence of the fMRI on the EEG.
4. There are many, interesting interactions, among the EEG sources.
5. There are a number of influences of the EEG sources on the fMRI.

This is a consistent causal model of EEG induced fMRI modulation—valid only for the slow phenomena that survive convolution with the HRF and for the alpha band EEG activity that was investigated here. Of course there are neural phenomena that might show up at as contemporaneous at this sampling rate—but we have filtered them out. An interesting analysis of information recoverable at each scale can be found in Deneux and Faugeras (2010).

Conclusion and suggestions for further work

1. We believe that the simulation efforts that are being carried currently out are very useful and should be extended to cover a greater realism in the neurodynamics, as well as to systematically test new proposals.
2. It will be also be important to have standardized experimental data from animals as a resource for model testing. Ideally this data set should provide intracranial recordings of possible neural drivers, BOLD-fMRI, surface EEG, diffusion MRI based structural connectivity and histological based connectivity matrices.²⁰
3. There is a clear need for tools that can assess model evidence (and establish their Identifiability) when dealing with large model spaces of biophysically informed SSMs. These should be brought to bear on the issue of bounds on model complexity, imposed by the HRF convolution and sub-sampling in fMRI.
4. We foresee the following theoretical developments in Causal modeling for effective connectivity:
 - a. The fusion of Bayes-Net and WAGS methods.
 - b. The WAGS tools developed for combined point and continuous time stochastic processes may play an important role in the connectivity analysis of EEG/fMRI, LFP and spike train data.
 - c. WAGS methods must be extended to non-standard models, among others: non-Markovian, RDE, and delay differential equations.
5. The development of exploratory (nonparametric), large scale state-space methods that are biophysically constrained and contain modality specific observation equations. This objective will depend critically on the exploration of large model spaces and is in consistent with the recent surge of methods analyzing “Ultra-High” dimensional data.
6. The explicit decomposition of multiple spatial and frequency scales.
7. Effective connectivity in the setting of Neural Field Modeling

We hope to have focused attention on these issues, within a unifying framework that integrates apparently disparate and important approaches. We are not saying that DCM and GCM are equivalent, but rather that an integration is possible within a Bayesian SSM framework and the use of model comparison methods. Our review of the field has been based on the use of state space models (SSM). While we are aware that SSMs are not the only possible framework for analyzing effective connectivity, this formulation allowed us to present a particular view that we feel will stimulate further work.

Besides reviewing current work we have discussed a number of new mathematical tools: Random Differential Equations, non-Markovian models, infinitely differentiable sample path processes, as well as the use of graphical causality models. We also considered the use of

²⁰ Such a data set in an animal model including EEG, EcoG, DWI tractography and fMRI is being gathered by Jorge Riera (Tohoku University), within a collaboration including F. H. Lopes da Silva, Thomas Knoesche, Olivier David, and the authors of this paper. This data set will be made publicly available in the near future.

continuous-time AR and ARMA models. It may well be that some of these techniques will not live up to expectations, but we feel our field will benefit from these and other new tools that confront some of the particular challenges addressed in this discussion series.

Acknowledgments

We dedicate this paper to Rolf Kötter for the many insights and the promotion of the Brain Connectivity Workshops that influenced this work. We also wish to thank Steve Smith for stimulating discussions, as well as to Daniel Commenges, Rolando Biscay, Juan Carlos Jimenez, Guido Nolte, Tohru Ozaki, Victor Solo, Nelson Trujillo, and Kamil Uludag for helpful input to the contents of this paper. An important part of the work described here was conceived and executed during the “The Keith Worsley workshop on Computational Modeling of Brain Dynamics: from stochastic models to Neuroimages (09w5092)” organized by the Banff International Research Station.

Appendix A. Wiener's original definition of causality

This approach was first formalized by Wiener (1956) as follows.²¹ Consider a strictly stationary (possibly complex) stochastic processes²² $X_1(t, \omega)$ defined as a collection of random variable for all integer time instants t and realizations ω . Wiener showed how to construct its “innovation”—the unit variance white noise time series $E_1(t, \omega)$ which is uncorrelated with the past of $X_1(t, \omega)$. The innovation $E_2(t, \omega)$ can also be constructed for a second time series $X_2(t, \omega)$. Now consider the random variable $K_1(\omega)$, that part of $E_1(t, \omega)$ uncorrelated with its own past and that of $E_2(t, \omega)$. The variance of this random variable lies between 0 and 1 and is the degree to which the time series $X_1(t, \omega)$ does not depend on the past of $X_2(t, \omega)$. One minus this variance is the Wiener measure C of the causal effect of $X_2(t, \omega)$ on $X_1(t, \omega)$. This measure of influence was in fact expressed by Wiener as an infinite sum:

$$I_{2 \rightarrow 1}^W = \sum_{m=1}^{\infty} |\rho(t, t-m)|^2 + \sum_{m=1}^{\infty} \left| \sum_{n=1}^{\infty} \rho(t, t-m) \overline{\rho(t-n, t-m)} \right|^2 + \dots$$

$$\rho(t, s) = E[X_1(t) \overline{X_2(s)}] \quad (25)$$

where $\overline{X(s)}$ indicates the complex conjugate of a time series.

As pointed out in Bressler and Seth (2010) this definition is not practical. We elaborate on why: First, it is limited to strictly stationary processes and involves an infinite series of moments without specification of how to perform the requisite calculations. More seriously, it only involves a finite number of series and ignores the potential confounding effect of unobserved (or latent) causes. More importantly, it adopts the “functional formulation” of von Mises that lost out to the currently predominant “stochastic formulation” of Kolmogorov and Doob (Von Mises and Doob, 1941). Nevertheless Wiener's definition has several points that deserve to be highlighted:

1. It was not limited to autoregressive models but was based on the more general Moving Average Representation (MAR).
2. Although defined explicitly for discrete time stochastic processes, the extension to continuous time was mentioned explicitly.
3. Applications in neuroscience were anticipated. In fact, Wiener elaborated on its possible use: “Or again, in the study of brain waves we may be able to obtain electroencephalograms more or

less corresponding to electrical activity in different parts of the brain. Here the study of the coefficients of causality running both ways and of their analogs for sets of more than two functions f may be useful in determining what part of the brain is driving what other part of the brain *in its normal activity*”.

4. It is instructive to compare this initial definition with modern accounts of direct influence.

References

- Aalen, O.O., 1987. Dynamic modeling and causality. *Scand. Actuarial J.* 13, 177–190.
- Aalen, O.O., Frigessi, A., 2007. What can statistics contribute to a causal understanding. *Scand. J. Stat.* 34 (1), 155–168. doi:10.1111/j.1467-9469.2006.00549.x
- Akaike, H., 1968. On the use of a linear model for the identification of feedback systems. *Annals of the Institute of Statistical Mathematics*, 20(1). Springer, pp. 425–439. Retrieved from <http://www.springerlink.com/index/MP5748216213R74Q.pdf>
- Amendola, A., Niglio, M., Vitale, C., 2010. Temporal aggregation and closure of VARMA models: some new results. In: Palumbo, F., Lauro, C.N., Greenacre, M.J. (Eds.), *Data Analysis and Classification*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 435–443. doi:10.1007/978-3-642-03739-9
- Angelova, M., Wennberg, B., 2010. On analytic and algebraic observability of nonlinear delay systems. *Automatica*, 46(4). Elsevier Ltd., pp. 682–686. doi:10.1016/j.automatica.2010.01.031
- Astrom, K.J., 1969. On the choice of sampling rates in parametric identification of time series. *Inf. Sci.* 1, 273–278.
- August, E., Papachristodoulou, A., 2009. A new computational tool for establishing model parameter identifiability. *J. Comput. Biol.* 16 (6), 875–885. doi:10.1089/cmb.2008.0211
- Belyaev, Y.K., 1959. Analytic random processes. *Theory Probab. Appl.* 4 (4), 402. doi:10.1137/1104040
- Bergstrom, A.R., 1966. Nonrecursive models as discrete approximations to systems of stochastic differential equations. *Econometrica* 34 (1), 173–182.
- Bergstrom, A.R., 1984. Continuous time stochastic models and issues of aggregation. In: Griliches, Z., Lnrilligato, M.D. (Eds.), *Handbook of Econometrics*, Volume II. Elsevier Science Publishers B.V.
- Bergstrom, A.R., 1988. Continuous-time models, realized volatilities, and testable distributional implications for daily stock returns. *Econometric Theory* 4 (3), 365–383. doi:10.1002/jae.1105
- Bojak, Ingo, Liley, D.T.J., 2010. Axonal velocity distributions in neural field equations. *PLoS Comput. Biol.* 6 (1), 1–25. doi:10.1371/journal.pcbi.1000653
- Bosch-Bayard, J., Valdés-Sosa, P., Virues-Alba, T., Aubert-Vázquez, E., John, E.R., Harmony, T., et al., 2001. 3D statistical parametric mapping of EEG source spectra by means of variable resolution electromagnetic tomography (VARETA). *Clinical EEG (Electroencephalography)*, 32(2). ECNS, pp. 47–61. Retrieved September 13, 2010, from <http://www.ncbi.nlm.nih.gov/pubmed/11360721>
- Brandt, S.F., Pelster, A., Wessel, R., 2007. Synchronization in a neuronal feedback loop through asymmetric temporal delays. *Europhys. Lett.* 79 (3), 38001. doi:10.1209/0295-5075/79/38001
- Breakspear, Michael, Stam, C.J., 2005. Dynamics of a neural system with a multiscale architecture. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360 (1457), 1051–1074. doi:10.1098/rstb.2005.1643
- Breakspear, M., Roberts, J.A., Terry, J.R., Rodrigues, S., Mahant, N., Robinson, P.A., 2006. A unifying explanation of primary generalized seizures through nonlinear brain modeling and bifurcation analysis. *Cereb. Cortex* 16 (9), 1296–1313. doi:10.1093/cercor/bhj072
- Bressler, S.L., Seth, A.K., 2011. Wiener–Granger causality: a well established methodology. *Neuroimage* 58, 323–329 (this issue).
- Bunge, M., 2009. *Causality and Modern Science*. Book, Fourth. Dover Publications, New York.
- Calbo, G., Cortés, J.-C., Jódar, L., 2010. Mean square power series solution of random linear differential equations. *Computers & Mathematics with Applications*, 59(1). Elsevier Ltd., pp. 559–572. doi:10.1016/j.camwa.2009.06.007
- Candy, J.V., 2006. *Model Based Signal Processing*. Book. IEEE Press, 701 pp.
- Carbonell, F., Biscay, R.J., Jimenez, J.C., de la Cruz, H., 2007. Numerical simulation of nonlinear dynamical systems driven by commutative noise. *J. Comput. Phys.* 226 (2), 1219–1233. doi:10.1016/j.jcp.2007.05.024
- Cartwright, N., 2007. *Hunting Causes and Using Them: Approaches in Philosophy and Economics*. Politics. Cambridge University Press, Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, Sao Paulo.
- Chambers, Marcus J., Thornton, M.A., 2009. Discrete time representation of continuous time ARMA processes. *Por Clasificar*, pp. 1–19.
- Chen, C.C., Kiebel, S.J., Friston, K.J., 2008. Dynamic causal modelling of induced responses. *Neuroimage* 41 (4), 1293–1312. doi:10.1016/j.neuroimage.2008.03.026
- Chen, B., Hu, J., Zhu, Y., Sun, Z., 2009. Parameter identifiability with Kullback–Leibler information divergence criterion. *Int. J. Adapt. Control Signal Process.* 23 (10), 940–960. doi:10.1002/acs.1078
- Chueshov, I.D., 2002. Introduction to the Theory of Infinite-Dimensional Dissipative Systems. Book. ACTA Scientific Publishing House, Kharkov, pp. 1–419. Retrieved from <http://www.emis.de/monographs/Chueshov/>
- Commenges, D., Gégout-Petit, A., 2009. A general dynamical statistical model with causal interpretation. *J. R. Stat. Soc. B Stat. Methodol.* 71 (3), 719–736. doi:10.1111/j.1467-9868.2009.00703.x

²¹ With some loss of rigor we have simplified the definitions, making our notation consistent with current time series analysis. For greater detail please consult the original references.

²² That is $\Pr(X_1(t_1, \omega), \dots, X_1(t_n, \omega)) = \Pr(X_1(t_1 + \tau, \omega), \dots, X_1(t_n + \tau, \omega))$ for all for all n and τ .

- Comte, F., Renault, E., 1996. Noncausality in continuous time models. *Econometric Theory* 12, 215–256.
- Coombes, S., 2010. Large-scale neural dynamics: simple and complex. *NeuroImage* 52 (3), 731–739. doi:10.1016/j.neuroimage.2010.01.045
- Coombes, S., Venkov, N., Shiao, L., Bojak, I., Liley, D., Laing, C., 2007. Modeling electrocortical activity through improved local approximations of integral neural field equations. *Phys. Rev. E* 76 (5), 1–8. doi:10.1103/PhysRevE.76.051901
- Cox, D.R., Wermuth, N., 2004. Causality: a statistical view. *Int. Stat. Rev.* 72, 285–305.
- Daunizeau, J., David, O., Stephan, K.E., 2011a. Dynamic causal modelling: a critical review of the biophysical and statistical foundations. *NeuroImage* 58, 312–322 (this issue).
- Daunizeau, J., Friston, K.J., Kiebel, S.J., 2009b. Variational Bayesian identification and prediction of stochastic nonlinear dynamic causal models. *Physica D* 238 (21), 2089–2118. doi:10.1016/j.physd.2009.08.002
- Daunizeau, Jean, Kiebel, Stefan J., Friston, Karl J., 2009c. Dynamic causal modelling of distributed electromagnetic responses. *NeuroImage*, 47(2). Elsevier Inc., pp. 590–601. doi:10.1016/j.neuroimage.2009.04.062
- David, O., Kilner, J.M., Friston, K.J., 2006. Mechanisms of evoked and induced responses in MEG/EEG. *NeuroImage* 31 (4), 1580–1591. doi:10.1016/j.neuroimage.2006.02.034
- David, Olivier, 2011. fMRI connectivity, meaning and empiricism Comments on: Roebroeck et al. The identification of interacting networks in the brain using fMRI: Model selection, causality and deconvolution. *NeuroImage* 58, 306–309 (this issue).
- David, Olivier, Guillemain, I., Saitlet, S., Rey, S., Deransart, C., Segebarth, C., et al., 2008. Identifying neural drivers with functional MRI: an electrophysiological validation. *PLoS Biol.* 6 (12), 2683–2697. doi:10.1371/journal.pbio.0060315
- Deco, G., Jirsa, V.K., Robinson, Peter A., Breakspear, Michael, Friston, Karl, 2008. The dynamic brain: from spiking neurons to neural masses and cortical fields. *PLoS Comput. Biol.* 4 (8), e1000092. doi:10.1371/journal.pcbi.1000092
- Demiralp, S., Hoover, K., 2008. A bootstrap method for identifying and evaluating a structural vector autoregression*. *Oxford Bulletin of Economics and Statistics*. Retrieved August 27, 2010, from <http://www3.interscience.wiley.com/journal/120174048/abstract>
- Deneux, T., Faugeras, O., 2010. EEG-fMRI fusion of paradigm-free activity using Kalman filtering. *Neural Comput.* 22 (4), 906–948. doi:10.1162/neco.2009.05-08-793.
- Deshpande, G., Sathian, K., Hu, X., 2009. Effect of hemodynamic variability on Granger causality analysis of fMRI. *NeuroImage*, 52(3). Elsevier Inc., pp. 884–896. doi:10.1016/j.neuroimage.2009.11.060
- Dhamala, M., Rangarajan, G., Ding, M., 2008. Analyzing information flow in brain networks with nonparametric Granger causality. *NeuroImage* 41 (2), 354–362. doi:10.1016/j.neuroimage.2008.02.020
- Eichler, M., 2005. A graphical approach for evaluating effective connectivity in neural systems. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360 (1457), 953–967. doi:10.1098/rstb.2005.1641
- Eichler, M., Didelez, V., 2010. On Granger causality and the effect of interventions in time series. *Lifetime data analysis* 16 (1), 3–32. doi:10.1007/s10985-009-9143-3
- Faugeras, O., Touboul, J., Cessac, B., 2009. A constructive mean-field analysis of multi-population neural networks with random synaptic weights and stochastic inputs. *Front. Comput. Neurosci.* 3, 1. doi:10.3389/fncom.10.001.2009 (February).
- Florens, J.-P., 2003. Some technical issues in defining causality. *J. Econometrics* 112, 127–128.
- Florens, J.-P., Fougere, D., 1996. Noncausality in continuous time. *Econometrica* 64 (5), 1195–1212 Retrieved from <http://www.jstor.org/stable/2171962>
- Florens, J.-P., Mouchart, M., 1982. A note on noncausality. *Econometrica* 50 (3), 583–591 Retrieved from <http://www.jstor.org/stable/1912602>
- Florens, A.J., Mouchart, M., 1985. A linear theory for noncausality. *Econometrica* 53 (1), 157–176.
- Freiwald, W., Valdes-Sosa, P.A., Bosch-Bayard, Jorge, Biscay-Lirio, R., Jimenez, Juan Carlos, Rodríguez, L.M., et al., 1999. Testing non-linearity and directedness of interactions between neural groups in the macaque inferotemporal cortex. *J. Neurosci. Methods* 94 (1), 105–119. doi:10.1016/S0165-0270(99)00129-6
- Friston, K.J., 2008a. Variational filtering. *NeuroImage* 41, 747–766.
- Friston, Karl J., 2008b. Hierarchical models in the brain. *PLoS Comput. Biol.* 4 (11), e1000211. doi:10.1371/journal.pcbi.1000211
- Friston, Karl, 2009a. Causal modelling and brain connectivity in functional magnetic resonance imaging. *PLoS Biol.* 7 (2), e33. doi:10.1371/journal.pbio.1000033
- Friston, Karl, 2011. Dynamic causal modeling and Granger causality comments on: the identification of interacting networks in the brain using fMRI: model selection, causality and deconvolution. *NeuroImage* 58, 303–305 (this issue).
- Friston, K.J., Daunizeau, J., 2008. DEM: a variational treatment of dynamic systems. *NeuroImage* 41, 849–885. doi:10.1016/j.neuroimage.2008.02.054
- Friston, K.J., Harrison, L., Penny, W., 2003. Dynamic causal modelling. *NeuroImage* 19 (4), 1273–1302.
- Friston, K.J., Mechelli, A., Turner, R., Price, C.J., 2000. Nonlinear responses in fMRI: the Balloon model, Volterra kernels, and other hemodynamics. *NeuroImage* 12 (4), 466–477. doi:10.1006/nimg.2000.0630
- Frosini, B.V., 2006. Causality and causal models: a conceptual. *Int. Stat. Rev.* 305–334 (June 2004).
- Galka, A., Yamashita, O., Ozaki, Tohru, Biscay, Rolando, Valdés-Sosa, Pedro, 2004. A solution to the dynamical inverse problem of EEG generation using spatiotemporal Kalman filtering. *NeuroImage* 23 (2), 435–453. doi:10.1016/j.neuroimage.2004.02.022
- Galka, A., Ozaki, Tohru, Muhle, H., Stephani, U., Siniatchkin, M., 2008. A data-driven model of the generation of human EEG based on a spatially distributed stochastic wave equation. *Cogn. Neurodynamics* 2 (2), 101–113. doi:10.1007/s11571-008-9049-x
- Garnier, H., Wang, L., 2008. Identification of continuous time models from sampled data. In: Garnier, H., Wang, L. (Eds.), *Engineering*. Springer Verlag, London Limited.
- Ge, T., Kendrick, K.M., Feng, J., 2009. A novel extended Granger Causal Model approach demonstrates brain hemispheric differences during face recognition learning. *PLoS Comput. Biol.* 5 (11), e1000570. doi:10.1371/journal.pcbi.1000570
- Gégout-Petit, A., Commenges, D., 2010. A general definition of influence between stochastic processes. *Lifetime Data Anal.* 16 (1), 33–44. doi:10.1007/s10985-009-9131-7
- Gencay, R., Selcuk, F., Whitcher, B., 2002. An introduction to wavelets and other filtering methods in finance and economics. : Statistics, Vol. 12. Academic Press.
- Geweke, J., 1984. Measures of conditional linear dependence and feedback between time series. *J. Am. Stat. Assoc.* 79 (388), 907–915.
- Gill, J.B., Petrović, L., 1987. Causality and stochastic dynamical systems. *SIAM J. Appl. Math.* 47 (6), 1361–1366.
- Glover, G.H., 1999. Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage* 9 (4), 416–429 Retrieved August 21, 2010, from <http://www.ncbi.nlm.nih.gov/pubmed/10191170>
- Glymour, C., 2009. What Is Right with 'Bayes Net Methods' and What Is Wrong with 'Hunting Causes and Using Them'? *The British Journal for the Philosophy of Science* 61 (1), 161–211. doi:10.1093/bjps/axp039
- Gourieroux, C., Monfort, A., Renault, Eric, 1987. Kullback Causality Measures. *Granger, C.W.J.*, 1963. Economic processes involving feedback. *Inf. Control* 48, 28–48.
- Granger, C.W.J., 1988. Some recent developments in a concept of causality. *J. Econometrics* 39, 199–211.
- Hansen, L.P., Sargent, T.J., 1983. The dimensionality of the aliasing problem in models with rational spectral densities. *Econometrica* 51 (2), 377–387.
- Havlicek, M., Jan, J., Calhoun, V.M., 2009. Extended time-frequency Granger causality for evaluation of functional network connectivity in event-related fMRI data. Conference of the IEEE, pp. 4440–4443. Retrieved August 27, 2010, from <http://www.ncbi.nlm.nih.gov/pubmed/19963833>
- Havlicek, Martin, Jan, Jiri, Brazdil, M., Calhoun, V.D., 2010. Dynamic Granger causality based on Kalman filter for evaluation of functional network connectivity in fMRI data. *NeuroImage*, 53(1). Elsevier Inc., pp. 65–77. doi:10.1016/j.neuroimage.2010.05.063
- Havlicek, M., Friston, K.J., Jan, J., Brazdil, M., Calhoun, V.D., 2011. Dynamic modeling of neuronal responses in fMRI using cubature Kalman filtering. *NeuroImage*. Elsevier Inc. doi:10.1016/j.neuroimage.2011.03.005
- Holden, H., Oksendal, B., Ub, J., Zhang, T., 1996. Holden, Oksendal et al 1996—Stochastic Partial Differential Equations. Birkhauser.
- Jansen, B.H., Rit, V.G., 1995. Biological Cybernetics in a mathematical model of coupled cortical columns. *Biol. Cybern.* 366, 357–366.
- Jentzen, A., Kloeden, P.E., 2009. Pathwise Taylor schemes for random ordinary differential equations. *BIT Numer. Math.* 49 (1), 113–140. doi:10.1007/s10543-009-0211-6
- Jiao, Q., Lu, G., Zhang, Z., Zhong, Y., Wang, Z., Guo, Y., et al., 2010. Granger causal influence predicts BOLD activity levels in the default mode network. *Hum. Brain Mapp.* 1–8. doi:10.1002/hbm.21065
- Jirsa, V.K., et al., 2002. Spatiotemporal Forward Solution of the EEG and MEG Using Network Modeling. *IEEE Transactions on Medical Imaging* 21 (5), 493–504.
- Kailath, T., 1980. *Linear Systems*. Book. Prentice Hall, New Jersey.
- Kalitzin, S.N., Parra, J., Velis, D.N., Lopes da Silva, F.H., 2007. Quantification of unidirectional nonlinear associations between multidimensional signals. *IEEE Trans. Biomed. Eng.* 54 (3), 454–461. doi:10.1109/TBME.2006.888828
- Larsson, E.K., Mossberg, M., Soderstrom, T., 2006. An overview of important practical aspects of continuous-time ARMA system identification. *Circuits Syst. Signal Process.* 25 (1), 17–46. doi:10.1007/s00034-004-0423-6
- Lauritzen, S., 1996. Graphical Models. Oxford University Press.
- Ljung, L., Glad, T., 1994. On global identifiability for arbitrary model parametrizations. *Automatica* 30 (2), 265–276.
- Łuczka, J., 2005. Non-Markovian stochastic processes: colored noise. *Chaos (Woodbury, N.Y.)* 15 (2), 26107. doi:10.1063/1.1860471
- Lutkepohl, H., 2005. *New Introduction to Multiple Time Series Analysis*. Book. Springer, pp. 1–764.
- Lyman, R.J., Edmonson, W.W., McCullough, S., Rao, M., 2000. The predictability of continuous-time, bandlimited processes. *IEEE Trans. Signal Process.* 48 (2), 311–316.
- Machamer, P., Darden, L., Craver, C.F., Machamer, P., 2000. Thinking about mechanisms. *Philos. Sci.* 67 (1), 1–25.
- Maiwald, Thomas, Timmer, Jens, 2008. Dynamical modeling and multi-experiment fitting with Potters-Wheel. *Bioinformatics (Oxford, England)* 24 (18), 2037–2043. doi:10.1093/bioinformatics/btn350
- Marinazzo, D., Liao, W., Chen, H., Stramaglia, S., 2011. Nonlinear connectivity by Granger causality. *NeuroImage* 58, 330–338 (this issue).
- Marreiros, A.C., Kiebel, Stefan J., Daunizeau, Jean, Harrison, L.M., Friston, Karl J., 2009. Population dynamics under the Laplace assumption. *NeuroImage*, 44(3). Elsevier Inc., pp. 701–714. doi:10.1016/j.neuroimage.2008.10.008
- Marrelec, G., Benali, H., Ciuciu, P., Pélégriani-Issac, M., Poline, J.-B., 2003. Robust Bayesian estimation of the hemodynamic response function in event-related BOLD fMRI using basic physiological information. *Hum. Brain Mapp.* 19 (1), 1–17. doi:10.1002/hbm.10100
- Martínez-Montes, E., Valdés-Sosa, P.A., Miwakeichi, F., Goldman, R.I., Cohen, M.S., 2004. Concurrent EEG/fMRI analysis by multiway Partial Least Squares. *NeuroImage* 22 (3), 1023–1034. doi:10.1016/j.neuroimage.2004.03.038
- Marzetti, L., Del Gratta, C., Nolte, G., 2008. Understanding brain connectivity from EEG data by identifying systems composed of interacting sources. *NeuroImage* 42 (1), 87–98. doi:10.1016/j.neuroimage.2008.04.250
- Mccrorie, J. Roderick, 2003. The problem of aliasing in identifying finite parameter continuous time stochastic models. *Acta Applicandae Mathematicae* 79, 9–16.
- Mccrorie, J.R., Chambers, M.J., 2006. Granger causality and the sampling of economic. *J. Econometrics* 132, 311–326.
- Medvedev, P., 2007. *Stochastic Integration Theory*. Oxford University Press, USA. Retrieved July 8, 2010, from <http://books.google.com/books?hl=en&lr=&id=>

- pZGKC_PVvBsC&oi=fnd&pg=PR13&dq=Stochastic+Integration+Theory&ots=IogSzUxwK&sig=j29s0LLApDxAueHhDp5qNylcOQ
- Minchev, B., Wright, W., 2005. A review of exponential integrators for first order semi-linear problems. Preprint Numerics. Trondheim, Norway. Retrieved August 29, 2010, from <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:A+review+of+exponential+integrators+for+first+order+semi-linear+problems#0>
- Moneta, A., Spirtes, Peter, 2006. Graphical models for the identification of causal structures in multivariate time series models. Proceedings of the 9th Joint Conference on Information Sciences (JCIS), 1. Atlantis Press, Paris, France, pp. 1–4. doi:10.2991/jcis.2006.171
- Moran, R.J., Stephan, K.E., Kiebel, S.J., Rombach, N., O'Connor, W.T., Murphy, K.J., et al., 2008. Bayesian estimation of synaptic physiology from the spectral responses of neural masses. *NeuroImage* 42 (1), 272–284. doi:10.1016/j.neuroimage.2008.01.025
- Mykland, P., 1986. Statistical Causality.
- Nalatore, H., Ding, M., Rangarajan, G., 2007. Mitigating the effects of measurement noise on Granger causality. *Phys. Rev. E* 75 (3). doi:10.1103/PhysRevE.75.031123
- Nolte, G., Meinecke, F., Ziehe, A., Müller, K.-R., 2006. Identifying interactions in mixed and noisy complex systems. *Phys. Rev. E* 73 (5), 1–6. doi:10.1103/PhysRevE.73.051913
- Nolte, G., Ziehe, A., Nikulin, V., Schlögl, A., Krämer, N., Brismar, T., et al., 2008. Robustly estimating the flow direction of information in complex physical systems. *Phys. Rev. Lett.* 100 (23), 1–4. doi:10.1103/PhysRevLett.100.234101
- Nolte, G., Marzetti, L., Valdes Sosa, P., 2009. Minimum Overlap Component Analysis (MOCA) of EEG/MEG data for more than two sources. *J. Neurosci. Methods* 183 (1), 72–76. doi:10.1016/j.jneumeth.2009.07.006
- Ozaki, T., 1990. Contribution to the discussion of M.S. Bartlett's paper, 'Chance and chaos'. *J. R. Stat. Soc. Ser. A* 153, 330–346.
- Ozaki, Tohru, 1992. A bridge between nonlinear time series models and nonlinear stochastic dynamical systems: a local linearization approach. *Stat. Sin.* 2, 113–135.
- Ozaki, Tohru, 2011. Statistical Time Series Modelling Approach to Signal Decomposition, Inverse Problems and Causality Analysis for Neuroscience Data. CRC Press.
- Pearl, J., 2000. *Causality: Models, Reasoning and Inference*. Cambridge University Press.
- Pearl, J., 2003. Statistics and causal inference: a review. *Test* 12 (2), 281–345.
- Penny, W.D., Stephan, K.E., Mechelli, A., Friston, K.J., 2004. Modelling functional integration: a comparison of structural equation and dynamic causal models. *NeuroImage* 23 (Suppl 1), S264–S274. doi:10.1016/j.neuroimage.2004.07.041
- Penny, W., Ghahramani, Z., Friston, K., 2005. Bilinear dynamical systems. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360 (1457), 983–993. doi:10.1098/rstb.2005.1642
- Petrović, L., Stanojević, D., 2010. Statistical causality, extremal measures and weak solutions of stochastic differential equations with driving semimartingales. *J. Math. Modell. Algorithms* 9 (1), 113–128. doi:10.1007/s10852-009-9121-5
- Phillips, P.C.B., 1973. The problem of identification in finite parameter continuous time models. *J. Econometrics* 1, 351–362.
- Phillips, P.C.B., 1974. The estimation of some continuous time models. *Econometrica* 42 (5), 803–823.
- Pollock, D., 2010. Oversampling of stochastic processes. Working Papers, 2. Retrieved August 27, 2010, from <http://ideas.repec.org/p/wse/wpaper/44.html>
- Ramsey, J.D., Hanson, S.J., Hanson, C., Halchenko, Y.O., Poldrack, R.A., Glymour, C., 2010. Six problems for causal inference from fMRI. *NeuroImage*, 49(2). Elsevier Inc., pp. 1545–1558. doi:10.1016/j.neuroimage.2009.08.065
- Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., et al., 2009. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics Oxford England* 25 (15), 1923–1929. doi:10.1093/bioinformatics/btp358
- Renault, Eric, Sekkat, K., Szafarz, A., 1998. Testing for spurious causality in exchange rates. *J. Empir. Finance* 5, 47–66.
- Riera, J.J., Jimenez, J.C., Wan, X., Kawashima, R., Ozaki, T., 2007a. Nonlinear local electrovascular coupling. II: from data to neuronal masses. *Hum. Brain Mapp.* 354 (August 2006), 335–354. doi:10.1002/hbm.20278
- Riera, J.J., Jimenez, J.C., Wan, X., Kawashima, R., Ozaki, T., 2007b. Nonlinear local electrovascular coupling. II: from data to neuronal masses. *Hum. Brain Mapp.* 354 (August 2006), 335–354. doi:10.1002/hbm.20278
- Riera, Jorge J., Wan, Xiaohong, Jimenez, Juan Carlos, Kawashima, Ryuta, 2006. Nonlinear local electrovascular coupling. I: a theoretical model. *Hum. Brain Mapp.* 27 (11), 896–914. doi:10.1002/hbm.20230
- Robinson, P.M., 1991. Automatic frequency domain inference on semiparametric and nonparametric models. *Econometrica* 59 (5), 1329. doi:10.2307/2938370
- Robinson, P.A., Chen, P.-chia, Yang, L., 2008. Physiologically based calculation of steady-state evoked potentials and cortical wave velocities. *Biol. Cybern.* 98 (1), 1–10. doi:10.1007/s00422-007-0191-z
- Roebrock, A., Formisano, E., Goebel, R., 2005. Mapping directed influence over the brain using Granger causality and fMRI. *NeuroImage* 25 (1), 230–242 Retrieved from <Go to ISI>: <http://WOS:000227369600021>
- Roebrock, Alard, Formisano, Elia, Goebel, Rainer, 2011a. Reply to Friston and David fMRI: model selection, causality and deconvolution. *NeuroImage* 58, 310–311 (this issue).
- Roebrock, Alard, Formisano, Elia, Goebel, Rainer, 2011b. The identification of interacting networks in the brain using fMRI: model selection, causality and deconvolution. *NeuroImage* 58, 296–302 (this issue).
- Saccomani, M.P., Audoly, S., Bellu, G., D'Angiò, L., 2010. Examples of testing global identifiability of biological and biomedical models with the DAISY software. *Comput. Biol. Med.* 40 (4), 402–407. doi:10.1016/j.compbiomed.2010.02.004
- Sanchez-Bornot, J., Martinez-Montes, E., Lage-Castellanos, Agustín, Vega-Hernández, M., Valdes-Sosa, P.A., 2008. Uncovering sparse brain effective connectivity: a voxel-based approach using penalized regression. *Stat. Sin.* 18, 1501–1518 Retrieved from <http://www3.stat.sinica.edu.tw/statistica/password.asp?vol=18&num=4&art=14>
- Sargan, J.D., 1974. Some discrete approximations to continuous time stochastic models. *J. R. Stat. Soc. B* 36 (1), 74–90.
- Schwartz, E.L., 1977. Spatial mapping in the primate sensory projection: analytic structure and relevance to perception. *Biol. Cybern.* 25, 181–194.
- Schweder, T., 1970. Composable Markov processes. *J. Appl. Probab* 7 (2), 400. doi:10.2307/3211973
- Seth, A.K., 2009. Granger Causal Connectivity Analysis: A MATLAB Toolbox.
- Shampine, L.F., Gahinet, P., 2006. Delay-differential-algebraic equations in control theory. *Appl. Numer. Math.* 56, 574–588. doi:10.1016/j.apnum.2005.04.025
- Shardlow, T., 2003. Numerical simulation of stochastic PDEs for excitable media. Analysis University of Manchester. Numerical Analysis Report 437.
- Smith, J.F., Pillai, A., Chen, K., Horwitz, B., 2010a. Identification and validation of effective connectivity networks in functional magnetic resonance imaging using switching linear dynamic systems. Manuscript. *NeuroImage* 52 (3), 1027–1040. doi:10.1016/j.neuroimage.2009.11.081
- Smith, S.M., Miller, K.L., Salimi-Khorshidi, G., Webster, M., Beckmann, C.F., Nichols, T.E., et al., 2010b. Network modelling methods for fMRI. *NeuroImage*. doi:10.1016/j.neuroimage.2010.08.063
- Solo, V., 1986. Topics in advanced time series analysis. In: Pino, G., Rebollo, R. (Eds.), *Lectures in Probability and Statistics*, Vol. 1215. Springer, Berlin Heidelberg. doi:10.1007/BFb0075871
- Solo, V., 2007. On causality I: sampling and noise. 46th IEEE Conference on Decision and Control. IEEE, pp. 3634–3639. Retrieved June 29, 2010, from <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:On+Causality+I:+Sampling+and+Noise#0>
- Solo, V., 2008. On causality and mutual information. Proceedings of the 47th IEEE Conference on Decision and Control, pp. 4939–4944.
- Spirtes, P., Glymour, C., Scheines, R., 2000. *Causation, Prediction and Search*. Cambridge University Press.
- Stephan, Klaas Enno, Kasper, L., Harrison, L.M., Daunizeau, Jean, den Ouden, H.E.M., Breakspear, Michael, et al., 2008. Nonlinear dynamic causal models for fMRI. *NeuroImage* 42 (2), 649–662. doi:10.1016/j.neuroimage.2008.04.262
- Stevenson, I.H., Kording, K.P., 2010. On the similarity of functional connectivity between neurons estimated across timescales. *PLoS One* 5 (2), e9206. doi:10.1371/journal.pone.0009206
- Supp, G.G., Schlögl, A., Trujillo-Barreto, Nelson, Müller, M.M., Gruber, T., 2007. Directed cortical information flow during human object recognition: analyzing induced EEG gamma-band responses in brain's source space. *PLoS One* 2 (1), e684. doi:10.1371/journal.pone.0000684
- Sussmann, H., 1977. An interpretation of stochastic differential equations as ordinary differential equations which depend on the sample point. *Am. Math. Soc.* 83 (2), 296–298 Retrieved August 21, 2010, from <http://www.ams.org/journals/bull/1977-83-02/S0002-9904-1977-14312-7/S0002-9904-1977-14312-7.pdf>
- Swanson, N.R., Granger, C.W.J., 1997. Impulse response functions based on a causal approach to residual orthogonalization in vector autoregressions. *J. Am. Stat. Assoc.* 92 (437), 357. doi:10.2307/2291481
- Triacca, U., 2007. Granger causality and contiguity between stochastic processes. *Phys. Lett. A* 362, 252–255. doi:10.1016/j.physleta.2006.10.024
- Valdes-Sosa, P.A., Jimenez, J.C., Riera, J., Biscay, R., Ozaki, T., 1999. Nonlinear EEG analysis based on a neural mass model. *Biol. Cybern.* 81 (5–6), 415–424 Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10592017>
- Valdés-Sosa, P.A., Bosch, J., Jiménez, J., Trujillo, N., Biscay, R., Morales, F., et al., 1999. The statistical identification of nonlinear brain dynamics: a progress report. *Non linear Dynamic and Brain Functioning*, pp. 1–22. Retrieved August 28, 2010, from http://www.ism.ac.jp/~ozaki/publications/paper/1999_Valdes_Bosch_nov_sci.pdf
- Valdes-Sosa, P.A., 2004. Spatio-temporal autoregressive models defined over brain manifolds. *Neuroinformatics* 2 (2), 239–250. doi:10.1385/Nl:2:2:239
- Valdes-Sosa, P.A., Riera, J., Casanova, R., 1996. Spatio temporal distributed inverse solutions. In: Aine, C.J., Okada, Y., Stroink, G., Switthenby, S.J., Wood, C.C. (Eds.), *Biomag 96: Proceedings of the Tenth International Conference on Biomagnetism*, Volume 1. Springer, pp. 377–380.
- Valdés-Sosa, P.A., Hernández, J., Vila, P., 1996. EEG spike and wave modelled by a stochastic limit cycle. *Neuroreport* Retrieved August 21, 2010, from http://journals.lww.com/neuroreport/Abstract/1996/09020/EEG_spike_and_wave_modelled_by_a_stochastic_limit.37.aspx
- Valdés-Sosa, P.A., Sánchez-Bornot, J.M., Lage-Castellanos, A., Vega-Hernández, M., Bosch-Bayard, J., Melie-García, L., et al., 2005. Estimating brain functional connectivity with sparse multivariate autoregression. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 360 (1457), 969–981. doi:10.1098/rstb.2005.1654
- Valdés-Sosa, P.A., Sánchez-Bornot, J., Vega-Hernández, M., Melie-García, L., Lage-Castellanos, A., Canales-Rodríguez, E., 2006. Granger causality on spatial manifolds: applications to neuroimaging. *Handbook of Time Series Analysis: Recent Theoretical Developments and Applications*, pp. 1–53.
- Valdes-Sosa, P.A., Sanchez-Bornot, J.M., Sotero, R.C., Iturria-Medina, Y., Aleman-Gomez, Y., Bosch-Bayard, Jorge, et al., 2009a. Model driven EEG/fMRI fusion of brain oscillations. *Hum. Brain Mapp.* 30 (9), 2701–2721. doi:10.1002/hbm.20704
- Valdés-Sosa, P.A., Vega-Hernández, Mayrim, Sánchez-Bornot, J.M., Martínez-Montes, E., Bobes, M.A., 2009b. EEG source imaging with spatio-temporal tomographic nonnegative independent component analysis. *Hum. Brain Mapp.* 30 (6), 1898–1910. doi:10.1002/hbm.20784
- Vanhatalo, S., Voipio, J., Kaila, K., 2005. Full-band EEG (FbEEG): an emerging standard in electroencephalography. *Clinical Neurophysiology*, 116(1). Elsevier, pp. 1–8.

- Retrieved August 29, 2010, from <http://linkinghub.elsevier.com/retrieve/pii/S1388245704003748>
- Victor Solo, 2008. Spurious causality and noise with fMRI and MEG. Organization for Human Brain Mapping annual Meeting.
- Von Mises, R., Doob, J.L., 1941. Discussion of papers on probability theory. *Ann. Math. Stat.* 12 (2), 215–217.
- White, H., Chalak, K., 2009. Settable systems: an extension of Pearl's causal model with optimization, equilibrium, and learning. *J. Mach. Learn. Res.* 10, 1–49.
- White, H., Lu, X., 2010. Granger causality and dynamic structural systems. *J. Financ. Econometrics* 8 (2), 193–243. doi:10.1093/jfifin/nbq006
- Wiener, N., 1956. The theory of prediction. In: Beckenbach, E. (Ed.), *Modern Mathematics for Engineers*. McGraw-Hill, New York.
- Witt, S.T., Meyerand, M.E., 2009. The effects of computational method, data modeling, and TR on effective connectivity results. *Brain Imaging Behav.* 3 (2), 220–231. doi:10.1007/s11682-009-9064-5
- Wong, K.F.K., Ozaki, Tohru, 2007. Akaike causality in state space. Instantaneous causality between visual cortex in fMRI time series. *Biol. Cybern.* 97 (2), 151–157. doi:10.1007/s00422-007-0165-1
- Woodward, J., 2003. *Making Things Happen: A Theory of Causal Explanations*. Oxford University Press, Oxford, New York, Bangkok, Buenos Aires, Cape Town.
- Wright, S., 1921. Correlation and causation. *J. Agric. Res.* 20, 557–585.
- Zou, C., Denby, K.J., Feng, J., 2009. Granger causality vs. dynamic Bayesian network inference: a comparative study. *BMC Bioinformatics* 10, 122. doi:10.1186/1471-2105-10-122