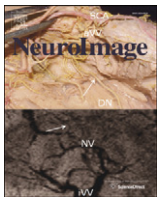
The identification of interacting networks in the brain using fMRI: model selection, causality and deconvolution

NeuroImage 58 (2011) 296–302

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg



Comments and Controversies

The identification of interacting networks in the brain using fMRI: Model selection, causality and deconvolution

Alard Roebroeck *, Elia Formisano, Rainer Goebel

Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, Postbus 616, 6200MD Maastricht, The Netherlands

ARTICLE INFO

Article history:
Received 8 June 2009
Revised 24 August 2009
Accepted 17 September 2009
Available online 25 September 2009

ABSTRACT

Functional magnetic resonance imaging (fMRI) is increasingly used to study functional connectivity in large-scale brain networks that support cognitive and perceptual processes. We face serious conceptual, statistical and data analysis challenges when addressing the combinatorial explosion of possible interactions within high-dimensional fMRI data. Moreover, we need to know, and account for, the physiological mechanisms underlying our signals. We argue here that (i) model selection procedures for connectivity should include consideration of more than just a few brain structures, (ii) temporal precedence – and causality concepts based on it – are essential in dynamic models of connectivity and (iii) undoing the effect of hemodynamics on fMRI data (by deconvolution) can be an important tool. However, it is crucially dependent upon assumptions that need to be verified.

Introduction

Understanding how interactions between brain structures ('functional and effective connectivity') support the performance of specific cognitive tasks or perceptual processes is a prominent goal in cognitive neuroscience. Neuroimaging methods, such as electroencephalography (EEG), magnetoencephalography (MEG) and functional magnetic resonance imaging (fMRI), are employed more and more to address questions of functional connectivity, inter-region coupling and networked computation that go beyond the 'where' and 'when' of task-related activity (McIntosh, 2004; Valdes-Sosa et al., 2005a; Salmelin and Kujala, 2006; Horwitz and Smith, 2008). A network perspective onto the parallel and distributed processing in the brain – even on the large scale accessible by neuroimaging methods – is a promising approach to enlarge our understanding of perceptual, cognitive and motor functions. However, we face serious conceptual, statistical and data analysis challenges when addressing the combinatorial explosion of possible interactions within high-dimensional neuroimaging data sets. Moreover, we need to know, and take account of, the actual physiological mechanisms underlying our signals (e.g., Logothetis, 2008).

Functional magnetic resonance imaging (fMRI) in particular is increasingly used not only to localize structures involved in cognitive and perceptual processes but also to study the connectivity in large-scale brain networks that support these functions. Two fMRI-based connectivity methods have gained increasing popularity in recent years: Granger causality analysis (GCA; Goebel et al., 2003; Valdes-Sosa, 2004; Roebroeck et al., 2005) and dynamic causal modeling (DCM; Friston et al., 2003). Both techniques aim to estimate directed influences between brain structures making use of the temporal dynamics in the fMRI signal. Despite the common goal, there are also differences between the two methods. Whereas GCA explicitly models temporal precedence and uses the concept of Granger causality (or G-causality), DCM employs a biophysically motivated generative model of neuronal population dynamics and hemodynamic processes. A recent article (David et al., 2008) has compared the two techniques in a rat model of absence epilepsy. Simultaneous fMRI and EEG and separate intracranial EEG (iEEG) were measured in six rats during epileptic episodes in which spike-and-wave discharges (SWDs) spread through the brain. These authors and a related commentary (Friston, 2009) concluded that (i) the concepts of temporal precedence and G-causality should not be used in fMRI connectivity analysis and (ii) explicit biophysically motivated models, such as DCM, model true causality in fMRI data, because they account for the hemodynamic processes that intervene between neural activity and fMRI signals.

We show here that these conclusions are not unequivocally supported by the actual results of David et al. (2008) and that they give only a partial view onto the important considerations in modeling brain connectivity. More specifically, we argue that the results of David et al., along with general considerations in system identification theory and neuroscience, lead to three crucial points about brain connectivity modeling:

(i) model selection procedures for connectivity should include consideration of more than just a few brain structures,

* Corresponding author. Fax: +31 43 3884125.
E-mail address: a.roebroeck@maastrichtuniversity.nl (A. Roebroeck).

1053-8119/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2009.09.036

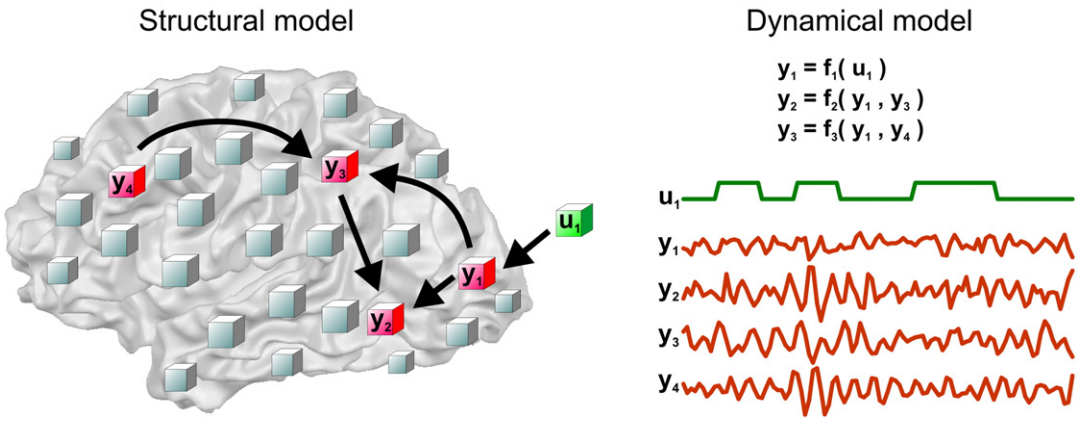


Fig. 1. A partitioning of brain connectivity models. Models to estimate connectivity from data (e.g. fMRI) can partitioned into a structural (or anatomical) model and a dynamical (or mathematical) model (Buchel and Friston, 2000). The structural model contains a selection of the structures in the brain that are assumed to be of importance in the cognitive process or task under investigation. Specifically, it specifies which regions of interest (ROIs) in the spatially rich high-dimensional fMRI data set will be considered for further analysis, as illustrated by the selection of the red boxes $y_1 \ldots y_4$. The structural model can also define the possible interactions between the ROIs in the form of one or more directed graph models that might be compared in a later model selection step. Finally the structural model also defines where exogenous inputs (that may be under control of the experimenter) can exert effects onto the network. The dynamical model embeds the structural model assumptions into parameterized equations that relate the selected measurements and inputs to each other. Connectivity modeling involves the estimation of the parameters in the dynamical model from actual measurements y_j and, possibly, inputs u_k.

(ii) temporal precedence – and causality concepts based on it – are essential in dynamic models of brain connectivity and
(iii) undoing the effect of hemodynamics on fMRI data (by deconvolution) can be an important tool. However, it is crucially dependent upon assumptions that need to be verified.

Structural model selection for brain connectivity

Brain connectivity modeling of neuroimaging data entails the estimation of multivariate mathematical models and inference on parameters that quantify the directed influence between brain structures. The estimation mathematical models from time series data generally has two important aspects: model selection and model identification (Ljung, 1999). In the model selection stage a class of models is chosen by the researcher that is deemed suitable for the problem at hand. In the model identification stage the parameters in the chosen model class are estimated from the observed data record. In practice, model selection and identification often occur in a somewhat interactive fashion where, for instance, model selection can be informed by the fit of different models to the data achieved in an identification step. The important point is that model selection involves a mixture of choices and assumptions on the part of the researcher and the information gained from the data record itself.

We can usefully partition brain connectivity models into two parts, each necessitating choices and assumptions: the structural model and the dynamical model (see Fig. 1). The structural model contains (i) a selection of the regions of interest (ROIs) in the brain that are assumed to be of importance in the cognitive process or task under investigation, (ii) the possible interactions between those structures and (iii) the possible effects of exogenous inputs onto the network. The exogenous inputs may be under control of the experimenter and often have the form of a simple indicator function that can represent, for instance, the presence or absence of a visual stimulus. The dynamical model consists of parameterized equations that relate the signals of the selected structures and exogenous inputs to each other. The functional form of these equations can embed assumptions on signal dynamics, temporal precedence or physiological processes from which signals originate. Connectivity modeling involves the estimation of (and inference on) the parameters in the dynamical model from actual measurements and possibly exogenous inputs. The number of parameters to be estimated (i.e., the total model complexity) is directly dependent on the complexity of the structural model (i.e., how many ROIs are included) and the complexity of the dynamical model. The bias/variance trade-off in model fitting dictates that overfitting a finite data set with too many parameters will lead to poor generalization of model fit to other data sets. Therefore, clear justifiable choices must be made both in the structural model and in the dynamical model to keep the number of estimated parameters in a suitable range. Applications of DCM invariably use very simple structural models (typically employing three to six ROIs) in combination with a complex parameter-rich dynamical model that we discuss below. The clear danger with overly simple structural models is that of spurious influence: an erroneous influence found between two selected regions that in reality is due to interactions with additional regions which have been ignored. Prototypical examples of spurious influence, of relevance in brain connectivity, are those between unconnected structures A and B that receive common input from, or are intervened by, an unmodeled region C.

Early applications of G-causality to fMRI data were aimed at counteracting the problems with overly restrictive structural models by employing more permissive structural models in combination with a simple dynamical model (Goebel et al., 2003; Valdes-Sosa, 2004; Roebroeck et al., 2005). We developed the technique of Granger causality mapping (GCM[1]) to explore all regions in the brain that interact with a single selected reference region using GCA of fMRI time series. By employing a simple bivariate model containing the reference region and, in turn, every other voxel in the brain, the sources and targets of influence for the reference region can be mapped. We showed that such an 'exploratory' mapping approach can form an important tool in structural model selection (Roebroeck et al., 2005). Although a bivariate model does not discern direct from indirect influences, the mapping approach locates potential sources of common input and areas that could act as intervening network nodes. Other applications of GCA to fMRI data have considered full multivariate models on large sets of selected brain regions that can model indirect influences within those sets. Valdes-Sosa et al. (2004, 2005b) applied these models to parcellations of the entire cortex in

[1] It is unfortunate and confusing that our original definition of the acronym GCM as Granger causality mapping (Goebel et al., 2003; Roebroeck et al., 2005) is used in the discussed comment (Friston, 2009) as Granger causality modeling, since 'mapping' expresses a fundamental and distinguishing characteristic of the way we apply Granger causality without employing a restrictive structural model.

conjunction with sparse regression approaches that enforce an implicit structural model selection within the set of parcels. In Deshpande et al. (2008) a full multivariate model was estimated over 25 ROIs (that were found to be activated in the investigated task) together with an explicit reduction procedure to prune regions from the full model as a structural model selection procedure.

The need for a more permissive structural model selection approach is illustrated by the work of David et al. In their study, fMRI was used to map the hemodynamic response throughout the brain to seizure activity, where ictal and inter-ictal states were quantified by the simultaneously recorded EEG. This showed widespread changes throughout the brain, including seven structures with an increased cerebral blood volume (CBV) signal during seizures (first somatosensory cortex (S1BF), thalamus, striatum, cerebellum, medulla oblongata, pons and retrosplenial cortex) and five deactivated structures. Only three of these structures were selected by the authors as the crucial nodes in the network that generates and sustains seizure activity, and thus worthy of further analysis: S1BF, thalamus and striatum. One cannot stop but wonder whether such a greatly simplified structural model is a justifiable decision given both the rich data set at hand and indications in the existing literature that generation and maintenance of seizure activity in the employed rat model involves other brain regions, such as frontoparietal cortex (e.g., Danober et al., 1998). It would be interesting to see whether a preliminary structural model selection step using a technique like GCM (on a small part of the data, not to be reused) would lead to a better justified set of selected regions. More generally, one of the strengths of the fMRI technique (and its analysis by statistical parametric mapping) is that it captures the large number of brain areas involved in many perceptual, cognitive or motor tasks. Therefore, it seems appropriate that connectivity models and structural model selection procedures consider those large number of areas, rather than fitting and comparing very simple structural models.

These considerations indicate that an important distinction must be made between exploratory and confirmatory approaches, especially in structural model selection procedures for brain connectivity. Neither approach is fundamentally wrong or right; rather they have a different but complementary goal. Exploratory techniques, like GCM, use information in the data to investigate the relative applicability of many models. As such, they have the potential to detect 'missing' regions in structural models. Confirmatory approaches, like DCM, test hypotheses about connectivity within a small set of models assumed to be applicable. Sources of common input or intervening causes are taken into account in a multivariate confirmatory model, but only if the employed structural model allows it. A confirmatory connectivity model can no more detect a missing region than a general linear model can detect a missing regressor.

Dynamical models in brain connectivity: G-causality and DCM

In addition to their different approach to structural model selection, there are subtle but important differences in the dynamical model employed by GCM and DCM. It has been claimed (Friston, 2009) that DCM employs a state-space model that embeds 'true' causality (in the form of a generative model of how the data are caused), whereas the GCM employs a statistical model of correlations in the data. However, both dynamical models can be given a state-space formulation; in both cases the inference on parameters employs statistics – be they confirmatory or exploratory, Bayesian or classical – and in both cases estimation is dependent on variance and correlations in the data. We will argue that, rather, the important distinctions between DCM and GCM are in a deterministic versus a stochastic dynamical model and in the physical interpretation of its variables.

GCM derives its name from Granger causality or G-causality, proposed by Clive Granger (Granger, 1969, 1980) and partially based upon earlier ideas of Norbert Wiener (Wiener, 1956). The aim is to give an operational definition of what 'causality' or 'influence' could mean for observations, structured in time, for multiple variables of interest. In economics, the variables of interest might be interest rates, employment numbers and the federal budget deficit. In neuroscience the variables could be invasive electrode recordings, intracranial EEG, non-invasive EEG, MEG or fMRI time series from different parts of the brain. The general idea of G-causality is that variable *A* G-causes another variable *B* if the prediction of *B*'s values improves when we use past values of *A*, given that all other relevant information is taken into account. Two more things need to be specified when we want to apply this idea to our data: (i) which model we use to make predictions and (ii) what 'all other relevant information' is. The second point is dealt with in the structural model selection process that entails the selection of a reasonable set of relevant variables (e.g., voxels, channels or ROIs), as we discussed above. The most common answer to the first point is the linear autoregressive (AR) model for discretely sampled data. The AR model is a simple model that can flexibly represent a wide range of signal dynamics, auto- and cross-correlation patterns and spectral characteristics, and is easy to estimate from data records. However, G-causality is definitely not tied exclusively to the standard linear AR model. It can be equally well instantiated in non-linear models (Freiwald et al., 1999) and time-varying models for non-stationary data (Hesse et al., 2003), and it can be framed in terms of non-parametric spectral factorization (Dhamala et al., 2008). In addition, G-causality has been extended to Markov processes and more general stochastic processes, based on Martingale theory (Aalen and Frigessi, 2007) and to continuous-time signal models (Florens and Fougere, 1996). Nonetheless, it will be informative to compare the class of linear stochastic models (LSM), of which the AR model used in GCM is a special case, with the DCM signal model to see that their crucial differences are actually subtle.

Both the LSM and DCM can be given a state-space formulation (Fig. 2). In a state-space representation the relations between measured variables y_j (e.g., fMRI data) and exogenous input variables u_k (e.g., stimulus functions) are modeled through unobservable state variables z_i. State-space representations generally consist of two sets of equations. The transition equations or state equations describe the evolution of the dynamic system over time, capturing relations among the hidden state variables z_i themselves and the influence of exogenous inputs u_k. The observation equations or measurement equations describe how the measurement variables y_j are obtained from the hidden state variables z_i and the inputs u_k. The LSM model accommodates equivalent representation of the general class of autoregressive moving average models with exogenous inputs (ARMAX models, Reinsel, 1997). Connectivity modeling of neuroimaging data involves the estimation of the elements in the coefficient matrices ($\mathbf{A}$, $\mathbf{B}^j$ and $\mathbf{C}$ in Fig. 2) from measurements $\mathbf{y}[t]$ and, possibly, the inputs $\mathbf{u}[t]$. The state-space representation makes the subtle but important differences between GCM/LSM and DCM insightful.

The first important difference in modeling signal dynamics is that LSMs employ linear stochastic transition equations, whereas those in DCMs are bi-linear and deterministic. The stochastic term in the LSM transition equation allows for variation in the state variables that cannot be explained by the inputs $\mathbf{u}[t]$. In fact, in the case of a purely autoregressive model, exogenous inputs are absent and all signal variation is modeled as driven by uncorrelated stochastic processes (called 'innovations'). This forces all dynamic and spectral complexity in the observed signals to be represented in the model parameters. It is exactly this property of comprehensive and flexible representation of signal dynamics and spectral properties that has made autoregressive models a popular tool in analyzing complex biophysical signals (Bernasconi and Konig, 1999; Ding et al., 2000; Kaminski et al., 2001; Harrison et al., 2003; Brovelli et al., 2004). In contrast, the

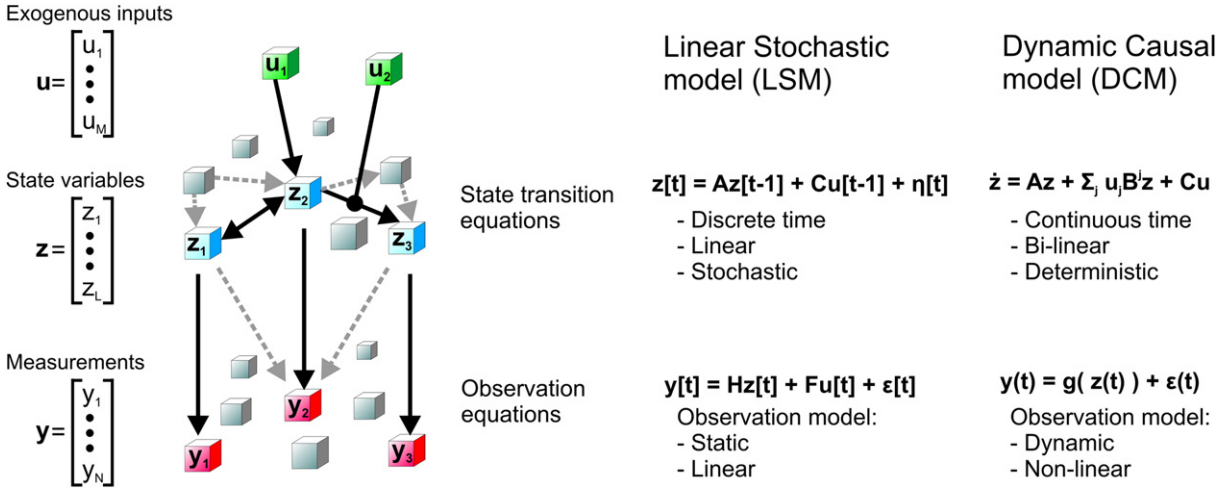


Fig. 2. State-space representations of dynamic connectivity models. The state-space representations for a linear stochastic model (LSM, often employed in Granger causality analysis and in GCM) and a dynamic causal model are shown and compared with respect to their mathematical properties. In a state-space representation the relations between measured variables y_j and, possibly, exogenous input variables u_k are modeled through unobservable state variables z_i. The individual variables vary over time and are summarized into vectors: $\mathbf{u}=(u_1, \ldots, u_M)$, $\mathbf{z}=(z_1, \ldots, z_L)$, $\mathbf{y}=(y_1, \ldots, y_N)$. State-space equations generally consist of two sets of equations. The transition equations or state equations describe the evolution of the dynamic system over time, capturing relations among state variables z_i themselves and the influence of exogenous inputs u_k. The observation equations or measurement equations relate the measurement variables y_j to the state variables z_i and inputs u_k. Connectivity modeling of neuroimaging data involves the estimation of the elements in the coefficient matrices $\mathbf{A}$, $\mathbf{B}^j$ and $\mathbf{C}$ from measurements $\mathbf{y}[t]$ and, possibly, inputs $\mathbf{u}[t]$. Whereas a linear stochastic model employs linear stochastic transition equations, those in dynamic causal modeling are bi-linear and deterministic.

transition equation in DCM for fMRI, as it is used widely to date, does not have a stochastic term. As a consequence, any and all signal dynamics that it can capture is limited to the signal subspace spanned by the assumed inputs. In other words, it assumes that all neural population dynamics can be captured without error from the chosen inputs and the transformation of that input in its 'flow' through the DCM network. The exogenous inputs mostly have a very simple form, such as a stimulus function that represents the presence or absence of a visual stimulus or level of experimental manipulation, such as attention left vs. right. The incapability of DCM to model signal variations beyond those implied by the exogenous inputs makes its connectivity estimation highly dependent on the exact number and form of the assumed inputs and the form of the structural model. Although the particular instantiation of DCM widely used to date (and used by David et al.) is indeed deterministic, stochastic extensions to DCM have been in development very recently (Friston et al., 2008; Daunizeau et al., 2009). These developments clearly have the potential to eliminate one of the differentiating aspects of LSMs and deterministic DCMs and bring the models even closer together. Interestingly, the inclusion of noise in the state equations makes inference on stochastic DCMs usefully interpretable in the stochastic framework of G-causality, reinforcing our point of the importance of this framework. In addition, a stochastic version of DCM could potentially provide an increased robustness to certain kinds of structural model misspecification that we have discussed above, such as unmodeled (or poorly modeled) sources of input to the system. However, this robustness is likely to be very limited, especially when the misspecification of structural models is more comprehensive than the omission of additional exogenous inputs. The inclusion and estimation of state noise is not a viable proxy to the actual inclusion of the right nodes in a structural model (or the consideration thereof in an exploratory or model-comparison framework).

The second important difference in modeling signal dynamics is that in DCM the state variables are given a definite physical interpretation within a generative model of the data. For every selected region a single state variable represents the neuronal or synaptic activity of a local population of neurons and (in DCM for BOLD fMRI) four or five more (Stephan et al., 2007) represent hemodynamic quantities such as capillary blood volume, blood flow and deoxy-hemoglobin content. All state variables (and the equations governing their dynamics) that serve the mapping of neuronal activity to the fMRI measurements $\mathbf{y}[t]$ (including the observation equation) can be called the observation model. Most of the physiologically motivated generative model in DCM for fMRI is therefore concerned with an observation model encapsulating hemodynamics. In contrast, in LSM/GCM the state variables may or may not have a definite physical interpretation, depending on the particular representation chosen. However, in the most straightforward representations LSM state variables are very simple functions of the measurements $\mathbf{y}[t]$. In its standard formulation, LSM/GCM does not use a biophysical model of hemodynamics. In short, the LSM observation model amounts to a linear combination of the state variables at the same moment in time (hence: static), whereas the observation model in DCM is non-linear and dynamic. Thus, investigating the observation model in DCM for fMRI and what it affords in terms of connectivity modeling will be important in the comparison of the dynamical models in DCM and GCM.

The observation model in DCM for fMRI is a biophysical model of hemodynamic coupling largely based on the Balloon model (Buxton et al., 1998) and Windkessel model (Mandeville et al., 1999). The parameters in this model, such as transit time and autoregulation, are estimated conjointly with the parameters quantifying neuronal connectivity. Thus, the forward biophysical model of hemodynamics is 'inverted' in the estimation procedure to achieve a deconvolution of fMRI time series and obtain estimates of the underlying neuronal states. The inversion of the observation model to achieve hemodynamic deconvolution of fMRI time series is an important aspect of DCM that we will discuss further below. It is important to note that the specific biophysical model for the interactions between neuronal states (neurodynamics) on one hand and the model for the hemodynamics (the observation model for fMRI) on the other hand largely dictate which of these models will absorb given aspects of the observed data. For instance, if there are delayed coherent variations between variables in the observed data and the hemodynamic model has much more affordance for delays than the neurodynamic model (as is the case in DCM), then the delay will be put into the hemodynamics in the fitting of the model. Not because it is a fact of the world that all delays are hemodynamic, but because the experimenter has implicitly assumed that to be true.

In order to compare the LSM model in GCM with the bi-linear deterministic model of neurodynamics in DCM, one would have to equate their observation model. Essentially, this is what David et al. did. They applied G-causality analysis (GCA) to deconvolved fMRI time series, obtained by the same hemodynamic deconvolution procedure that is implicit in DCM. If the deconvolution procedure succeeds (which might not always be the case, as we discuss below), it mathematically inverts the effect of hemodynamic processes on fMRI signals and uncovers the hidden neuronal population dynamics. By applying G-causality analysis to deconvolved fMRI time series, the stochastic dynamics of the LSM are augmented with the complex biophysically motivated observation model in DCM. This step is crucial if the goal is to compare the dynamic connectivity models and draw conclusions on the relative merits of linear stochastic models (estimating G-causality) and bilinear deterministic models. The results of this controlled direct comparison and comparison to the gold standard iEEG analyses are highly informative. The GCA analysis after deconvolution in particular is very convincingly in accordance with the gold standard iEEG analyses (David et al., 2008, their Figs. 4, lower right, and 7), strongly supporting the value of stochastic dynamical models and G-causality in brain connectivity analysis. In contrast, the final result of DCM analysis of the same data shows less correspondence with the gold standard, not identifying the direct influence of S1BF on the thalamus (David et al., 2008, their Figs. 5D and 7). The differences in successful capture of the direct and indirect influence, after deconvolution, are likely due to the difference between a deterministic and stochastic dynamical model, since the observation model was effectively equated. Two further notes can be made. First, David et al. did not use the bi-linear part of the standard DCM model (except for influence of the inputs on the 'self-modulation' of the neural state dynamics), thereby effectively fitting a linear version of DCM that does not allow modulation of connectivity by experimental conditions (in this case: ictal and inter-ictal states). Second, this trivariate near-linear DCM was compared to a set of bivariate tests performed with the LSM. In short, however, the results in David et al. show that the stochastic dynamics model of GCM potentially outperforms the deterministic dynamics model in DCM in a confirmatory analysis when both are given the same observation model.

Hemodynamic deconvolution

fMRI is an indirect measure of neuronal and synaptic activity. The physiological quantities directly determining signal contrast in fMRI are hemodynamic quantities such as capillary blood flow and the local ratio of oxygenated and de-oxygenated hemoglobin. The distorting effects of hemodynamic processes on the temporal structure of fMRI signals and, more importantly, the difference in hemodynamics in different parts of the brain form a severe confound for dynamic brain connectivity models. Particularly, the delay imposed upon fMRI signals with respect to the underlying neural activity is known to vary between subjects and between different brain regions of the same subject (Aguirre et al., 1998). In our work, we have acknowledged this confound and set out clear limits to the interpretation of Granger causality maps in the face of inhomogeneous hemodynamic processes over the brain. We have suggested that, at the very least, modulation of G-causality between fMRI time series by experimental context (e.g., a higher level of G-causality during attention to colour than during attention to motion) should be sought to give credibility to these analyses (Roebroeck et al., 2005). It is unfortunate that were David et al. applied GCA to original fMRI time series, they did not take note of these recommendations and did not investigate the modulation of G-causality between ictal and inter-ictal states.

The hemodynamic deconvolution approach inherent in DCM goes a step beyond our suggested approach, and tries to 'undo' the adverse effects of hemodynamic convolution. It is useful to look more closely at deconvolution operations and clarify the assumptions that go into them. A simple linear deconvolution operation will serve to illustrate the relevant points that generalize to model-based non-linear deconvolution as used in DCM. A (forward) convolution operation involves three signals: the output $y(t)$, an fMRI time series, is obtained as a convolution of the input $s(t)$, neuronal population activity, and the convolution kernel $h(t)$, often termed the hemodynamic response function (HRF) in fMRI, as illustrated in Fig. 3A. Deconvolution entails obtaining an estimate of either the input $s(t)$ or the convolution kernel $h(t)$ from knowledge of the other two factors in the convolution, among which is the output $y(t)$, possibly contaminated with additive noise. Since convolution is a commutative operation, there is no principle difference in the mathematics involved in obtaining the input $s(t)$ or the convolution kernel $h(t)$. The main

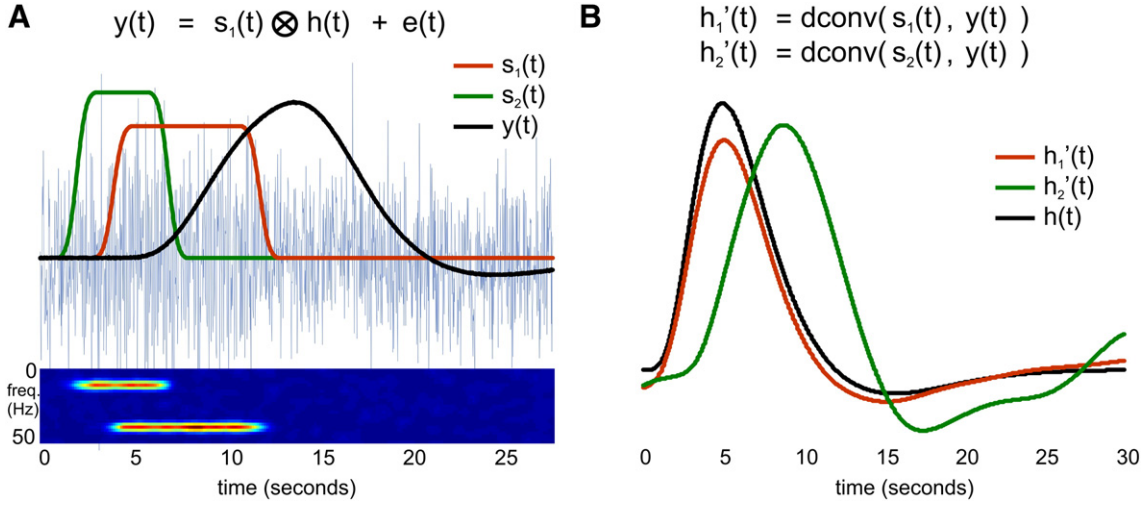


Fig. 3. Deconvolution and its dependence on assumptions on the involved signals. (A) An illustration of the coupling between an fMRI signal $y(t)$ and an underlying neuronal population activity signal $s_1(t)$, modeled as a linear convolution by an impulse response $h(t)$, the hemodynamic response function (HRF). In this simulation example, the observed fMRI signal $y(t)$ is coupled to power in the 40 Hz component $s_1(t)$ that can be obtained from a neurophysiological measurement (such as EEG, in blue in the background with its spectrogram below). An additional 12 Hz component $s_2(t)$ increases in power 2 s prior to the 40 Hz component and has a shorter duration. (B) The result of deconvolution estimates of the unobservable impulse response $h(t)$ with two different assumptions on the input. When 40 Hz power is the assumed input of hemodynamic convolution, a deconvolution estimate $h_1'(t)$ is obtained that closely approximates the 'true' impulse response $h(t)$. However, when 12 Hz power is assumed to be coupled to the fMRI signal an estimate $h_2'(t)$ is obtained that strongly deviates from the actual impulse response.

practical difference is in the type of information required: to be able to estimate h(t) one needs to know s(t) (or make strong assumptions on its form) and vice versa. It is this requirement for knowledge of – or very strong assumptions on the form of – the other convolving element that can lead to errors in a deconvolution procedure. A simulation example illustrates such errors, arising in a deconvolution estimate h(t) of the HRF when the assumptions on the input s(t) are incorrect (Fig. 3B).

David et al. performed two deconvolution operations to obtain estimates of neuronal source signals for their regions of interest to use in G-causality analysis. Using the band-limited (4–20 Hz) power of simultaneously collected EEG as the input and the recorded fMRI signal as output, the convolution kernel characterizing the HRF in each area was computed. The same input signal, derived from three EEG scalp electrodes, was used for all brain regions. This estimation was further constrained by the biophysical model of hemodynamics inherent in DCM, modified to account for the fact that an iron contrast agent was used in the fMRI imaging, giving it a stronger weighting for cerebral blood volume (CBV). The first deconvolution step showed a surprisingly extreme discrepancy of the estimated HRF in S1BF, peaking seconds later and lasting more than 30 s longer than that of other regions. In the second step, the deconvolved HRFs for S1BF, striatum and thalamus from the first step and the local CBV signal was used to deconvolve the hidden neuronal activity for each of the structures. Because of the dependency upon the HRF identified in step one, this second deconvolution step is crucially dependent upon the availability of the simultaneous EEG signal and on the decision of which aspect of that signal to use as the 'input'. Moreover, possible errors in the estimate of the local HRFs in the first step (see Fig. 3B) will propagate to the estimated neuronal source signals, in turn biasing connectivity estimates on these source signals. In DCM the two deconvolution steps are performed implicitly and conjointly with the estimation of connectivity parameters at the neuronal source level. However, the assumptions and information that go into the estimation procedures and, prominently, the need for knowledge of the input are the same.

The decision to use 4–20 Hz EEG power as input might well be justified. It characterizes the overall EEG signal power increase that accompanies seizure episodes and might be a good measure of increased neuronal firing and synaptic activity that demands increased metabolism. However, the precise coupling of hemodynamics and local metabolism with neuronal and synaptic activity is complex and partially unknown (Logothetis et al., 2001; Niessing et al., 2005). Moreover, in the great majority of applications of DCM to fMRI a simultaneously recorded EEG signal is not available. In this case the input to the hemodynamic convolution is entirely dependent on the assumed exogenous inputs **u**[t] that mostly have a simple discrete step-function form representing the presence of stimuli and the level of experimental conditions (e.g., memory load or attention condition). The neuronal population activity of brain structures that are not directly influenced by exogenous inputs is determined in DCM by the influence of other brain regions in the structural model. Thus, to some degree, successful deconvolution in DCM is also dependent on a veridical structural model, which once again highlights the importance of robust structural model selection procedures.

To our opinion, hemodynamic deconvolution might indeed improve the possibilities of fMRI-based connectivity estimates, but the assumptions that go into it deserve further investigation and scrutiny. The work of David et al. strongly supports the need for an independent measurement of neuronal and synaptic activity such as simultaneously acquired EEG. However, which part of this EEG signal should be considered to 'drive' the hemodynamic and metabolic processes that underlie BOLD and CBV signals deserves careful consideration. More generally, undoing the effects of hemodynamic convolution will require detailed knowledge of the biophysical mechanisms involved, which remain a topic of intense research.

Modeling brain connectivity: a synthesis

Any model is necessarily an abstraction; it cannot contain the full reality in all its detail. That is what makes it tractable and useful as a model. This is nicely paraphrased in a famous statement from Box and Draper (1987), p424: "Essentially, all models are wrong, but some are useful." However, that is not the end of our endeavor; it is the beginning, as another version of the statement makes clear (Box and Draper, 1987), p74: "Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful." We have argued here that useful models for brain connectivity have well justified assumptions, both in their structural model and in their dynamical model. The structural model should include all brain structures relevant to the task as informed by prior knowledge and exploratory analysis of a part of the data (not to be reused in later confirmatory steps). A biophysically motivated forward model of hemodynamics may be useful in dynamical models for connectivity analysis of fMRI data. However the hemodynamic deconvolution implied in the 'inversion' of these forward models is crucially dependent on (i) the availability of and assumptions on hidden information (the input) and (ii) the accuracy of the employed biophysical model of hemodynamics.

The brain is an immensely complex system that is neither linear, nor deterministic; neither bivariate, nor predictable. The abstractions and choices to be made in useful models of brain connectivity are therefore unlikely to be accommodated by one single 'master' model that does better than all other models on all counts. However, the considerations above do set out clear paths for development of connectivity models in cognitive neuroimaging. Development of connectivity models that can flexibly adjust the amount of parameters and prior assumptions in the structural and dynamical models could be very useful. Such models would force investigators to explicitly specify where they constrain their modeling effort by prior knowledge and assumptions and where they inform model selection by exploratory analysis of the data. Models that combine the stochastic dynamics of LSM with complex biophysical observation models as used in DCM, along with exogenous inputs, could form a useful step towards lifting the limitations of each individual model class. Finally, the discussion of hemodynamic deconvolution of fMRI data clearly points to the need for an independent measurement of neuronal and synaptic activity such as simultaneously acquired EEG. Future modeling efforts that enter such additional signals as measurement variables with their own observation model, rather than as exogenous inputs, would offer important advantages.

In any future development or evaluation of the relative advantages of current models, one should keep in mind that models are only as good as the assumptions that go into them. The choice for a confirmatory or exploratory approach or a domain specific versus a general model cannot be justified by claiming that confirmatory statistics on domain specific models are the only road to truth. Thorough review and discussion of the relative merits of different brain connectivity models can only lead to a balanced account of these issues and to developments that bring the field forward.

References

Aalen, O.O., Frigessi, A., 2007. What can statistics contribute to a causal understanding? Board of the Foundation of the Scandinavian Journal of Statistics 34, 155–168.
Aguirre, G.K., Zarahn, E., D'Esposito, M., 1998. The variability of human, BOLD hemodynamic responses. Neuroimage 8, 360–369.
Bernasconi, C., Konig, P., 1999. On the directionality of cortical interactions studied by structural analysis of electrophysiological recordings. Biol. Cybern. 81, 199–210.
Box, G.E.P., Draper, N.R., 1987. Empirical model-building and response surfaces. John Wiley, New York.
Brovelli, A., Ding, M., Ledberg, A., Chen, Y., Nakamura, R., Bressler, S.L., 2004. Beta oscillations in a large-scale sensorimotor cortical network: directional influences revealed by Granger causality. Proc. Natl. Acad. Sci. U. S. A. 101, 9849–9854.
Buchel, C., Friston, K., 2000. Assessing interactions among neuronal systems using functional neuroimaging. Neural. Netw. 13, 871–882.

Buxton, R.B., Wong, E.C., Frank, L.R., 1998. Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. Magn. Reson. Med. 39, 855–864.
Danober, L., Deransart, C., Depaulis, A., Vergnes, M., Marescaux, C., 1998. Pathophysiological mechanisms of genetic absence epilepsy in the rat. Prog. Neurobiol. 55, 27–57.
Daunizeau, J., Friston, K.J., Kiebel, S.J., 2009. Variational Bayesian identification and prediction of stochastic nonlinear dynamic causal models. Physica D: Nonlinear Phenomena 238, 2089–2118.
David, O., Guillemain, I., Saillet, S., Reyt, S., Deransart, C., Segebarth, C., Depaulis, A., 2008. Identifying neural drivers with functional MRI: an electrophysiological validation. PLoS Biol. 6, 2683–2697.
Deshpande, G., Hu, X., Stilla, R., Sathian, K., 2008. Effective connectivity during haptic perception: a study using Granger causality analysis of functional magnetic resonance imaging data. Neuroimage 40, 1807–1814.
Dhamala, M., Rangarajan, G., Ding, M., 2008. Analyzing information flow in brain networks with nonparametric Granger causality. Neuroimage 41, 354–362.
Ding, M., Bressler, S.L., Yang, W., Liang, H., 2000. Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modeling: data preprocessing, model validation, and variability assessment. Biol. Cybern. 83, 35–45.
Florens, J.P., Fougere, D., 1996. Noncausality in continuous time. Econometrica 64, 1195–1212.
Freiwald, W.A., Valdes, P., Bosch, J., Biscay, R., Jimenez, J.C., Rodriguez, L.M., Rodriguez, V., Kreiter, A.K., Singer, W., 1999. Testing non-linearity and directedness of interactions between neural groups in the macaque inferotemporal cortex. J. Neurosci. Methods 94, 105–119.
Friston, K., 2009. Causal modelling and brain connectivity in functional magnetic resonance imaging. PLoS Biol. e33, 7.
Friston, K.J., Harrison, L., Penny, W., 2003. Dynamic causal modelling. Neuroimage 19, 1273–1302.
Friston, K.J., Trujillo-Barreto, N., Daunizeau, J., 2008. DEM: a variational treatment of dynamic systems. Neuroimage 41, 849–885.
Goebel, R., Roebroeck, A., Kim, D.S., Formisano, E., 2003. Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping. Magn. Reson. Imaging 21, 1251–1261.
Granger, C.W.J., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37, 424–438.
Granger, C.W.J., 1980. Testing for causality: a personal viewpoint. J. Econ. Dyn. Control 2, 329–352.
Harrison, L., Penny, W.D., Friston, K., 2003. Multivariate autoregressive modeling of fMRI time series. Neuroimage 19, 1477–1491.
Hesse, W., Moller, E., Arnold, M., Schack, B., 2003. The use of time-variant EEG Granger causality for inspecting directed interdependencies of neural assemblies. J. Neurosci. Methods 124, 27–44.
Horwitz, B., Smith, J.F., 2008. A link between neuroscience and informatics: large-scale modeling of memory processes. Methods 44, 338–347.
Kaminski, M., Ding, M., Truccolo, W.A., Bressler, S.L., 2001. Evaluating causal relations in neural systems: granger causality, directed transfer function and statistical assessment of significance. Biol. Cybern. 85, 145–157.
Ljung, L., 1999. System identification: theory for the user2nd Edition. Prentice-Hall, New Jersey.
Logothetis, N.K., 2008. What we can do and what we cannot do with fMRI. Nature 453, 869–878.
Logothetis, N.K., Pauls, J., Augath, M., Trinath, T., Oeltermann, A., 2001. Neurophysiological investigation of the basis of the fMRI signal. Nature 412, 150–157.
Mandeville, J.B., Marota, J.J., Ayata, C., Zaharchuk, G., Moskowitz, M.A., Rosen, B.R., Weisskoff, R.M., 1999. Evidence of a cerebrovascular postarteriole windkessel with delayed compliance. J. Cereb. Blood Flow Metab. 19, 679–689.
McIntosh, A.R., 2004. Contexts and catalysts: a resolution of the localization and integration of function in the brain. Neuroinformatics 2, 175–182.
Niessing, J., Ebisch, B., Schmidt, K.E., Niessing, M., Singer, W., Galuske, R.A., 2005. Hemodynamic signals correlate tightly with synchronized gamma oscillations. Science 309, 948–951.
Reinsel, G.C., 1997. Elements of multivariate time series analysis2nd Edition. Springer-Verlag, New York.
Roebroeck, A., Formisano, E., Goebel, R., 2005. Mapping directed influence over the brain using Granger causality and fMRI. Neuroimage 25, 230–242.
Salmelin, R., Kujala, J., 2006. Neural representation of language: activation versus long-range connectivity. Trends Cogn. Sci. 10, 519–525.
Stephan, K.E., Weiskopf, N., Drysdale, P.M., Robinson, P.A., Friston, K.J., 2007. Comparing hemodynamic models with DCM. Neuroimage 38, 387–401.
Valdes-Sosa, P.A., 2004. Spatio-temporal autoregressive models defined over brain manifolds. Neuroinformatics 2, 239–250.
Valdes-Sosa, P.A., Kotter, R., Friston, K.J., 2005a. Introduction: multimodal neuroimaging of brain connectivity. Philos. Trans. R Soc. Lond. B Biol. Sci. 360, 865–867.
Valdes-Sosa, P.A., Sanchez-Bornot, J.M., Lage-Castellanos, A., Vega-Hernandez, M., Bosch-Bayard, J., Melie-Garcia, L., Canales-Rodriguez, E., 2005b. Estimating brain functional connectivity with sparse multivariate autoregression. Philos. Trans. R Soc. Lond. B Biol. Sci. 360, 969–981.
Wiener, N., 1956. The theory of prediction. In: Berkenbach, EF (Ed.), Modern mathematics for engineers. McGraw-Hill, New York.