Dynamic causal models of brain responses

Klaas E. Stephan, Lee M. Harrison, Will D. Penny, Karl J. Friston

The Wellcome Dept. of Imaging Neuroscience, Institute of Neurology, University College London, 12 Queen Square, London WC1N 3BG, United Kingdom

Summary

Functional magnetic resonance imaging (fMRI) is used to investigate where the neural implementation of specific cognitive processes occurs. The standard approach uses linear convolution models relating experimentally designed inputs, through a hemodynamic response function (HRF), to observed blood oxygen level dependent (BOLD) signals. However, these models are blind to the causal mechanisms that underlie observed BOLD responses, both in terms of neural dynamics and neurovascular coupling. This article reviews current developments towards causal models of how BOLD responses are generated. We focus on (i) biophysical input-state-output models with neural and hemodynamic state equations, and (ii) models of functional integration that explain local dynamics through interactions with remote areas. Forward models with parameters at the neural level, e.g. Dynamic Causal Modeling, combine both approaches, modeling the entire causal chain from external stimuli, via induced neural dynamics, to observed BOLD responses.

Introduction

Functional magnetic resonance imaging (fMRI) has become the most commonly used method to investigate human brain function. Historically, neuroimaging has been concerned predominantly with the localization of function, i.e. where in the brain neural computations mediate a cognitive process of interest. This approach rests on linear time-invariant (LTI) models that relate the time course of experimentally controlled manipulations (e.g. changes in sensory stimuli or cognitive set) to observed blood oxygen level-dependent (BOLD) signals in a voxel-specific fashion. While quite different statistical models have been suggested (see [2, 16] for recent reviews of different approaches), all treat the brain as an ensemble of isolated black boxes (i.e. voxels) whose input-output functions are characterized by BOLD responses evoked by various experimental conditions.

In this article, we briefly review current developments in causal models of how BOLD responses are generated. We focus on the following two issues: First, what are the mechanisms that translate local neural dynamics into observed BOLD signals? This question speaks to biophysical models of the neurovascular coupling. Second, how do local responses result from interactions with other brain regions? This question requires models of functional integration that consider context-dependent causal interactions among remote areas, i.e. in terms of effective connectivity.

This article is structured as follows. First, we briefly summarize standard convolution models for fMRI analysis that are blind to causal mechanisms underlying the BOLD signal. We then present current biophysical models of regionally specific responses. Finally, we
discuss progress in the field of effective connectivity. Particular emphasis will be given to Dynamic Causal Modeling (DCM, [11••]) as the first example of an emerging class of models that combine the biophysics of local responses and effective connectivity.

Convolution models and the hemodynamic response function

Most current approaches to fMRI analysis are implemented in the context of the General Linear Model (GLM)

\[ y = X\beta + e \]  

(1)

which models voxel-specific BOLD responses \( y \) in terms of a linear combination of explanatory variables in the design matrix \( X \) plus a Gaussian error term \( e \). The design matrix is based on stimulus functions that encode evoked neural responses. The relation between neural and BOLD responses is modeled by the hemodynamic impulse response function (HRF). This describes the characteristic hemodynamic response to a brief neural event and thus characterizes the input-output behavior of a given voxel. The standard convolution model for fMRI treats each voxel as an independent LTI system, convolving the stimulus functions with an HRF to give predicted hemodynamic responses that enter the design matrix as regressors [7].

The HRF may vary from voxel to voxel and subject to subject [14], and this has to be accommodated in the GLM. To allow for voxel-specific HRFs, temporal basis functions can be used to express the predicted BOLD response as the linear combination of several functions of peristimulus time [8, 16]. An alternative is to estimate the HRF directly from the data, using parametric [14] or non-parametric [6, 22] models.

In summary, regardless of whether the HRF is modeled by temporal basis functions or estimated from the data, the question addressed is where in the brain a given experimental manipulation leads to changes in BOLD signal. However, these approaches are blind to the mechanisms that underlie these changes.

Biophysical models of regional BOLD responses

By adopting a convolution model for brain responses in fMRI we are implicitly positing some underlying dynamic system that converts neuronal responses into observed hemodynamic responses. In the pioneering work by Buxton et al. (“Balloon model”, [4, 5]) and Mandeville et al. (“Windkessel model”, [21]), detailed biophysical models of the neurovascular coupling have been validated by physiological experiments. These models predict how increases in regional blood flow \( f \) dilate a venous balloon, increase its volume \( v \) and dilute venous blood to decrease deoxyhemoglobin content \( q \). The resulting BOLD signal is a non-linear function of \( v \) and \( q \) and follows the flow dynamics with a delay of about one second. This model was extended by Friston et al. [9, 10•] to include the effects of external inputs \( u \) on an auto-regulated vasodialatory signal \( s \) assuming that the relation between evoked neural activity and blood flow is linear. This linear relation had been demonstrated directly by elegant animal studies combining optical imaging, laser Doppler flowmetry and multi-electrode recordings [23•, 24] and indirectly by perfusion studies of the human brain [27].
Summary of the hemodynamic model of Friston [10•] and its adaptation for DCM [11••]. A series of experimentally controlled inputs $u$ evoke neural responses $z$ that trigger a hemodynamic cascade, modeled by 4 state equations with 5 parameters. These hemodynamic parameters comprise the rate constant of the vasodilatory signal decay ($\kappa$), the rate constant for auto-regulatory feedback by blood flow ($\gamma$), transit time ($\tau$), Grubb’s vessel stiffness exponent ($\alpha$), and capillary resting net oxygen extraction ($\rho$). Integrating the hemodynamic state equations for a given set of inputs and parameters produces a predicted BOLD response. For parameter estimation, an observation model is used that treats the observed BOLD response as a function of inputs and parameters plus some observation error.

As summarized in Fig. 1, the extended input-state-output model of Friston [10•] comprises four hemodynamic state variables combine into a vector $x = \{s, f, q, v\}$ whose interactions are described by differential equations with five hemodynamic parameters $\theta^h = \{\kappa; \gamma, \tau, \alpha, \rho\}$. These parameters have an explicit biophysical meaning (see legend of Fig. 1). The flow-inducing signal $s$ is triggered by neural responses to $m$ experimental inputs, weighted by different efficacies. These input-specific efficacies represent the neural parameters of the model: $\theta^a = \{e_1, \ldots, e_m\}$. The model represents a deterministic forward model with hidden states: for any given set of parameters $\theta = \{\theta^h, \theta^a\}$ and inputs $u$, the state equation $\dot{\theta} = f(x, u, \theta)$ can be integrated and passed through the output nonlinearity $\lambda$ to give a predicted BOLD response:

\[
\dot{x} = f(x, u, \theta)
\]

\[
y = \lambda(x)
\]
This can be extended to an observation model, representing the observed output $y$ as a function of the inputs and parameters, but without reference to the hidden states $x$:

$$y = h(u, \theta) + \epsilon$$  \hspace{1cm} (3)

This formulation is the basis for parameter estimation from measured data. For example, the distributions of the biophysical parameters $\theta^b$ were estimated from fMRI data by a Volterra series expansion [9]. The means and variances of these distributions were later used as empirical priors in a fully Bayesian scheme with an iterative EM (expectation maximization) algorithm [10(*)], using a bilinear approximation of the state equations. The latter scheme is also used in DCM, which is described below.

A limitation of the model by Friston [10(*)] is that it can only deal with measurement noise (see Eq. 2 and 3). An extended model which also considers physiological noise was suggested by Riera et al. [31(**)]. They augmented the state equation (Eq. 2) with an innovation, resulting in the stochastic differential equation

$$\dot{x} = f(x, u, \theta) + g \omega$$  \hspace{1cm} (4)

where $\omega$ is a scalar Wiener process, representing physiological noise, and $g$ is a vector defining the degree of randomness for each state variable. Eq. 4 was transformed into a non-linear state space model, from which parameters were estimated using a recursive local linearization filter [18]. This method has two advantages: (i) the parameter estimates converge to the true values, not to the values of a bilinear approximation, and (ii) the fit of the model can be evaluated easily by testing whether the distribution of the innovation terms deviates from a Gaussian distribution. Figure 2 shows an example where a BOLD signal that was generated from simulated state trajectories was reconstructed precisely from the estimated states despite the presence of both physiological and measurement noise.
Figure 2
Demonstration how a BOLD signal (blue line in A) that was generated from simulated hemodynamic state trajectories (solid lines in B) can be reconstructed (red line in A) from the estimated states (dashed lines in B) despite the presence of both physiological noise (black line in B) and observation noise (black line in A). This example used the model by Riera et al. [31••] with an additional Kalman smoothing step. The vertical dashed lines represent the onset and offset of a constant stimulus.

One limitation of the models discussed so far is that they assumed a tissue oxygen concentration of zero. Consequently, capillary oxygen extraction rate depended entirely on oxygen delivery and thus on blood flow (see Eq. 6 in [5]). While this coupling between oxygen extraction and flow was supported by earlier studies [17], recent experiments indicate that this may be an oversimplification [32]. Therefore, more sophisticated models of oxygen extraction might be useful. For example, Zheng et al. [34•] extended the Friston model [10•] by three new state variables that enable precise dynamic modeling of intra-capillary oxygen transport to tissue. Similarly, Obata et al. [28•] provided a generalized version of the original Balloon model of Buxton et al. [5] which considers both intra- and extra-vascular signal changes.

It should be noted that none of the biophysical models discussed in this section specify precisely what is meant by neural activity. Therefore, these models cannot tell us what aspect of neural information processing is reflected by the BOLD signal. Neural information processing within a given cortical unit can be described along many different dimensions, and the relation between a neurophysiological process and the resulting BOLD response can
be characterized at different scales, for example (i) local field potentials vs. spiking activity, (ii) excitatory vs. inhibitory post-synaptic potentials, or (iii) different receptor types at synapses. Sophisticated animal studies that combine multi-electrode recordings with fMRI [19, 20••] or with optical imaging techniques [23•, 24] have started to address these issues. The next step is to transform the current biophysical models of the BOLD response into comprehensive forward models with parameters at the neural level, modeling the entire causal chain from external stimuli, via induced neural dynamics, to observed BOLD responses. Such models must be parameterized in a neurophysiologically meaningful, yet parsimonious and estimable fashion. DCM [11••], which is discussed below, is a first step in that direction.

Models of functional integration

Effective connectivity
Integration within distributed neural systems is usually best understood in terms of effective connectivity: this is the influence that one neural system exerts over another, either at a synaptic or population level. It has been proposed that "effective connectivity should be understood as the experiment- and time-dependent, simplest possible circuit diagram that would replicate the observed timing relationships between the recorded neurons" [1]. This speaks to two important points: (i) effective connectivity is dynamic, i.e. activity- and time-dependent, and (ii) it depends upon a causal model of the interactions. Classical estimation procedures, employed in functional neuroimaging, were based initially on linear regression models, e.g. Structural Equation Modeling [3, 25]. Here, we briefly review current developments with a focus on multivariate autoregressive models and nonlinear dynamic causal models.

Multivariate Autoregressive Models (MAR)
Autoregressive models of fMRI data are usually not concerned with causality in a biophysical sense, i.e. how observed BOLD series result from underlying neural processes. Instead, they address the temporal aspect of causality in BOLD time series, focusing on the causal dependence of the present on the past: each data point of a time series is explained as a linear combination of past data points. This is in contrast to regression-based models of effective connectivity where the time series can be permuted without changing the results. MAR models extend the autoregressive approach to n brain regions, modeling the n-vector of regional BOLD signals at time t (y_i) as a linear combination of p past data vectors whose contributions are weighted by the parameter matrices A_i:

\[ y_t = \sum_{i=1}^{p} y_{t-i} A_i + e_t \]  

(5)

MAR models directed influences among a set of regions whose causal interactions, expressed at the BOLD level, are inferred via their mutual predictability from past time points. Although MAR is an established statistical technique, specific implementations for fMRI were suggested only very recently. Harrison et al. [15••] presented a MAR version that (i) allowed for bilinear variables representing modulatory effects on connections and (ii) used a Bayesian parameter estimation scheme suggested by Penny & Roberts [29]. This Bayesian
scheme also determined the optimal model order, i.e. the number of past time points ($p$) to be considered by the model. A complementary MAR approach, based on the idea of “Granger causality” [13], was proposed by Goebel et al. [12•]. They computed whole-brain connectivity maps by evaluating voxel-specific two-dimensional MAR models of the interactions between the current voxel and a reference voxel and introduced a statistical framework for distinguishing different types of interactions.

A logical extension of MAR models is to augment them with a biophysical forward model to enable inferences about neural parameters. So far, this type of model has only been introduced for EEG data [33]. For fMRI data, DCM is the only approach to date which marries biophysical and functional integration models.

**Dynamic Causal Modeling (DCM)**

The general idea behind DCM is to construct a reasonably realistic neuronal model of interacting cortical regions with neurophysiologically meaningful parameters. These parameters are estimated such that the predicted BOLD series, which results from converting the neural dynamics into hemodynamics, correspond as closely as possible to the observed BOLD series [11••]. In DCM, neural dynamics in several regions (represented by a neural state vector $z$ with one state per region) are driven by experimentally designed inputs that enter the model in two distinct ways: they can elicit responses through direct influences on specific anatomical nodes (e.g. evoked responses in early sensory cortices) or they can modulate the coupling among nodes (e.g. during learning or attention). DCM models the change in neural states as a non-linear function of the states, the inputs $u$ and neural parameters $\theta^a$:

$$\dot{z} = F(z,u,\theta^a)$$  \hspace{1cm} (6)

The parameters are the connectivity matrices ($\theta^a = \{A,B,C\}$) that define the functional architecture and interactions among brain regions at a neuronal level. The bilinear approximation of Eq. 6 is given by

$$\dot{z} = Az + \sum_j u_j B^j z + Cu$$

$$= (A + \sum_j u_j B^j) z + Cu$$ \hspace{1cm} (7)

in which coupling parameters correspond to partial derivatives of $F$:

$$A = \frac{\partial F}{\partial z} = \frac{\partial \dot{z}}{\partial z}$$

$$B^j = \frac{\partial^2 F}{\partial z \partial u_j} = \frac{\partial}{\partial u_j} \frac{\partial \dot{z}}{\partial z}$$

$$C = \frac{\partial F}{\partial u}$$ \hspace{1cm} (8)

The matrix $A$ represents the effective connectivity among the regions in the absence of modulatory input, the matrices $B^j$ encode the change in effective connectivity induced by the $j^{th}$ input $u_j$, and $C$ embodies the strength of direct influences of inputs on neuronal activity.
DCM combines this neural model with the biophysical forward model of Friston [10] which describes how neuronal activity translates into a BOLD response (Fig. 1). This enables the parameters and time constants of the neuronal model to be estimated from measured data, using a fully Bayesian approach with empirical priors for the biophysical parameters and conservative shrinkage priors for the coupling parameters. The posterior distributions of the parameter estimates can then be used to test hypotheses about the size and nature of modeled effects. Usually, these hypotheses concern context-dependent changes in coupling which are represented by the bilinear terms of the model. For example, applications of DCM have addressed the modulatory effects of object category [26] and attention to motion [11] (see also Fig. 3) on connections in the visual system. If there is uncertainty about which connections should be included in a model, or if one would like to compare competing hypotheses (represented by different DCMs), a Bayesian model selection procedure can be used to find the DCM that shows an optimal balance between model fit and the number of parameters [30].

![DCM Diagram](image)

**Figure 3**

Results from a DCM analysis of attention to visual motion with fMRI [11]. The fMRI data were from a study in which subjects viewed identical stimuli (radially moving dots) under different attentional manipulations of the task (detection of velocity changes) [3]. Right panel: Model structure with the conditional estimates shown alongside their connections, with the percent confidence that they exceeded threshold in brackets. Motion and attention exert bilinear effects: motion modulates the connection from V1 to the motion-sensitive area V5, whereas attention modulates the backward connections from the inferior frontal gyrus (IFG) to the superior parietal cortex (SPC) and from SPC to V5. Dotted arrows connecting regions represent significant bilinear
affects in the absence of a significant intrinsic coupling. Left panel: Fitted responses based upon the conditional estimates and the adjusted data are shown for each region in the DCM. The insert (upper left) shows the location of the regions.

Conclusion
As a complement to models for identifying where evoked brain responses are expressed, current effort is invested in models of how neuronal responses are caused. In this article, we have reviewed several models that address this causality in different ways. One promising strategy is to use comprehensive forward models with meaningful neurophysiological parameters that link experimental manipulations, via induced neural dynamics, to observed BOLD responses. We expect that such models will greatly enhance our ability to investigate and understand the neural systems that mediate specific cognitive processes.

Acknowledgments
This work was supported by the Wellcome Trust (grants 056750/Z/99/Z/GRS/KM/JAT and 069468/Z/02/Z).

References and recommended reading


• This study combines an extended Balloon model with a Bayesian estimation scheme for the neural efficacies of different experimental inputs, using empirical priors from [9] for the hemodynamic parameters. It provides the starting point for Dynamic Causal Modeling (see [11]).

• This article describes Dynamic Causal Modeling in detail, including the neural state equations, the forward model for fMRI, and the Bayesian parameter estimation scheme. The validity of DCM as well as its robustness to violations of its underlying assumptions are investigated with multiple simulations.

• The authors present an implementation of Granger causality [13] that enables the computation of whole-brain connectivity maps for a chosen reference region. They also present a method for distinguishing different causal components within interactions characterized by autoregressive models.


• This article presents a multivariate autoregressive model for assessing effective connectivity among a set of brain regions. It uses bilinear variables to model modulatory effects on connections and employs a Bayesian parameter estimation scheme. A method for choosing the optimal model order is also presented.


• This review discusses the currently available experimental data on the relative roles of different neurophysiological processes in eliciting a hemodynamic response.


• This study used a combination of optical imaging, laser Doppler flowmetry and multielectrode recordings with current source density reconstruction to investigate the relation between stimulus-evoked neural transients and the associated hemodynamics in rat cortex.


• The authors present an updated version of the classical Balloon model (see [5]) with a more precise description of the oxygen extraction rate.


• An important extension to [10] that allows for representation of physiological noise within the state equations, using a non-linear state space model and an extended Kalman filtering technique.


• This article provides an extended version of the hemodynamic model suggested in [10]. This extension incorporates a more precise model of intracapillary oxygen transport to tissue.