ACADEMIC PRESS

Available online at www.sciencedirect.com
SCIENCE DIRECT®

NeuroImage

NeuroImage 0 (2003) 000–000

www.elsevier.com/locate/ynimg

Dynamic causal modelling

K.J. Friston,* L. Harrison, and W. Penny

The Wellcome Department of Imaging Neuroscience, Institute of Neurology, Queen Square, London WC1N 3BG, UK

Received 18 October 2002; revised 7 March 2003; accepted 2 April 2003

Abstract

In this paper we present an approach to the identification of nonlinear input–state–output systems. By using a bilinear approximation to the dynamics of interactions among states, the parameters of the implicit causal model reduce to three sets. These comprise (1) parameters that mediate the influence of extrinsic inputs on the states, (2) parameters that mediate intrinsic coupling among the states, and (3) [bilinear] parameters that allow the inputs to modulate that coupling. Identification proceeds in a Bayesian framework given known, deterministic inputs and the observed responses of the system. We developed this approach for the analysis of effective connectivity using experimentally designed inputs and fMRI responses. In this context, the coupling parameters correspond to effective connectivity and the bilinear parameters reflect the changes in connectivity induced by inputs. The ensuing framework allows one to characterise fMRI experiments, conceptually, as an experimental manipulation of integration among brain regions (by contextual or trial-free inputs, like time or attentional set) that is revealed using evoked responses (to perturbations or trial-bound inputs, like stimuli). As with previous analyses of effective connectivity, the focus is on experimentally induced changes in coupling (*cf.,* psychophysiologic interactions). However, unlike previous approaches in neuroimaging, the causal model ascribes responses to designed deterministic inputs, as opposed to treating inputs as unknown and stochastic.

Keywords: Nonlinear system identification; Functional neuroimaging; fMRI; Hemodynamic response function; Effective connectivity; Bilinear model

1. Introduction

This paper is about modelling interactions among neuronal populations, at a cortical level, using neuroimaging (hemodynamic or electromagnetic) time series. It presents the motivation and procedures for dynamic causal modelling of evoked brain responses. The aim of this modelling is to estimate, and make inferences about, the coupling among brain areas and how that coupling is influenced by changes in experimental context. Dynamic causal modelling represents a fundamental departure from existing approaches to effective connectivity because it employs a more plausible generative model of measured brain responses that embraces their nonlinear and dynamic nature.

The basic idea is to construct a reasonably realistic neuronal model of interacting cortical regions. This model is then supplemented with a forward model of how neuronal or synaptic activity is transformed into a measured response. This enables the parameters of the neuronal model (i.e., effective connectivity) to be estimated from observed data. These supplementary models may be forward models of electromagnetic measurements or hemodynamic models of fMRI measurements. In this paper we will focus on fMRI. Responses are evoked by known deterministic inputs that embody designed changes in stimulation or context. This is accomplished by using a dynamic input–state–output model with multiple inputs and outputs. The inputs correspond to conventional stimulus functions that encode experimental manipulations. The state variables cover both the neuronal activities and other neurophysiological or biophysical variables needed to form the outputs. The outputs are measured electromagnetic or hemodynamic responses over the brain regions considered.

Intuitively, this scheme regards an experiment as a designed perturbation of neuronal dynamics that are promulgated and distributed throughout a system of coupled anatomical nodes to change region-specific neuronal activity.

* Corresponding author. Fax: +44-020-7813-1445.
E-mail address: k.friston@fil.ion.ucl.ac.uk (K.J. Friston).

1053-8119/03/$ – see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S1053-8119(03)00202-7

These changes engender, through a measurement-specific forward model, responses that are used to identify the architecture and time constants of the system at the neuronal level. This represents a departure from conventional approaches (e.g., structural equation modelling and autoregression models; McIntosh and Gonzalez-Lima, 1994; Büchel and Friston, 1997; Harrison et al., in press), in which one assumes the observed responses are driven by endogenous or intrinsic noise (i.e., innovations). In contradistinction, dynamic causal models assume the responses are driven by designed changes in inputs. An important conceptual aspect of dynamic causal models, for neuroimaging, pertains to how the experimental inputs enter the model and cause neuronal responses. Experimental variables can elicit responses in one of two ways. First, they can elicit responses through direct influences on specific anatomical nodes. This would be appropriate, for example, in modelling sensory-evoked responses in early visual cortices. The second class of input exerts its effect vicariously, through a modulation of the coupling among nodes. These sorts of experimental variables would normally be more enduring; for example, attention to a particular attribute or the maintenance of some perceptual set. These distinctions are seen most clearly in relation to existing analyses and experimental designs.

1.1. DCM and existing approaches

The central idea behind dynamic causal modelling (DCM) is to treat the brain as a deterministic nonlinear dynamic system that is subject to inputs and produces outputs. Effective connectivity is parameterised in terms of coupling among unobserved brain states (e.g., neuronal activity in different regions). The objective is to estimate these parameters by perturbing the system and measuring the response. This is in contradistinction to established methods for estimating effective connectivity from neurophysiological time series, which include structural equation modelling and models based on multivariate autoregressive processes. In these models, there is no designed perturbation and the inputs are treated as unknown and stochastic. Multivariate autoregression models and their spectral equivalents like coherence analysis not only assume the system is driven by stochastic innovations, but are restricted to linear interactions. Structural equation modelling assumes the interactions are linear and, furthermore, instantaneous in the sense that structural equation models are not time-series models. In short, DCM is distinguished from alternative approaches not just by accommodating the nonlinear and dynamic aspects of neuronal interactions, but by framing the estimation problem in terms of perturbations that accommodate experimentally designed inputs. This is a critical departure from conventional approaches to causal modelling in neuroimaging and brings the analysis of effective connectivity much closer to the conventional analysis of region-specific effects. DCM calls upon the same experimental design principles to elicit region-specific interactions that we use in experiments to elicit region-specific activations. In fact, as shown later, the convolution model, used in the standard analysis of fMRI time series, is a special and simple case of DCM that ensues when the coupling among regions is discounted. In DCM the causal or explanatory variables that compose the conventional design matrix become the inputs and the parameters become measures of effective connectivity. Although DCM can be framed as a generalisation of the linear models used in conventional analyses to cover bilinear models (see below), it also represents an attempt to embed more plausible forward models of how neuronal dynamics respond to inputs and produces measured responses. This reflects the growing appreciation of the role that neuronal models may have to play in understanding measured brain responses (see Horwitz et al., 2001, for a discussion).

This paper can be regarded as an extension of our previous work on the Bayesian identification of hemodynamic models (Friston, 2002) to cover multiple regions. In Friston (2002) we focussed on the biophysical parameters of a hemodynamic response in a single region. The most important parameter was the efficacy with which experimental inputs could elicit an activity-dependent vasodilatory signal. In this paper neuronal activity is modelled explicitly, allowing for interactions among the neuronal states of multiple regions in generating the observed hemodynamic response. The estimation procedure employed for DCM is formally identical to that described in Friston (2002).

1.2. DCM and experimental design

DCM is used to test the specific hypothesis that motivated the experimental design. It is not an exploratory technique; as with all analyses of effective connectivity the results are specific to the tasks and stimuli employed during the experiment. In DCM designed inputs can produce responses in one of two ways. Inputs can elicit changes in the state variables (i.e., neuronal activity) directly. For example, sensory input could be modelled as causing direct responses in primary visual or auditory areas. The second way in which inputs affect the system is through changing the effective connectivity or interactions. Useful examples of this sort of effect would be the attentional modulation of connections between parietal and extrastriate areas. Another ubiquitous example of this second sort of contextual input would be time. Time-dependent changes in connectivity correspond to plasticity. It is useful to regard experimental factors as inputs that belong to the class that produces evoked responses or to the class of contextual factors that induces changes in coupling (although, in principle, all inputs could do both). The first class comprises trial- or stimulus-bound perturbations whereas the second establishes a context in which effects of the first sort evoke responses. This second class is typically trial-free and established by task instructions or other contextual changes.

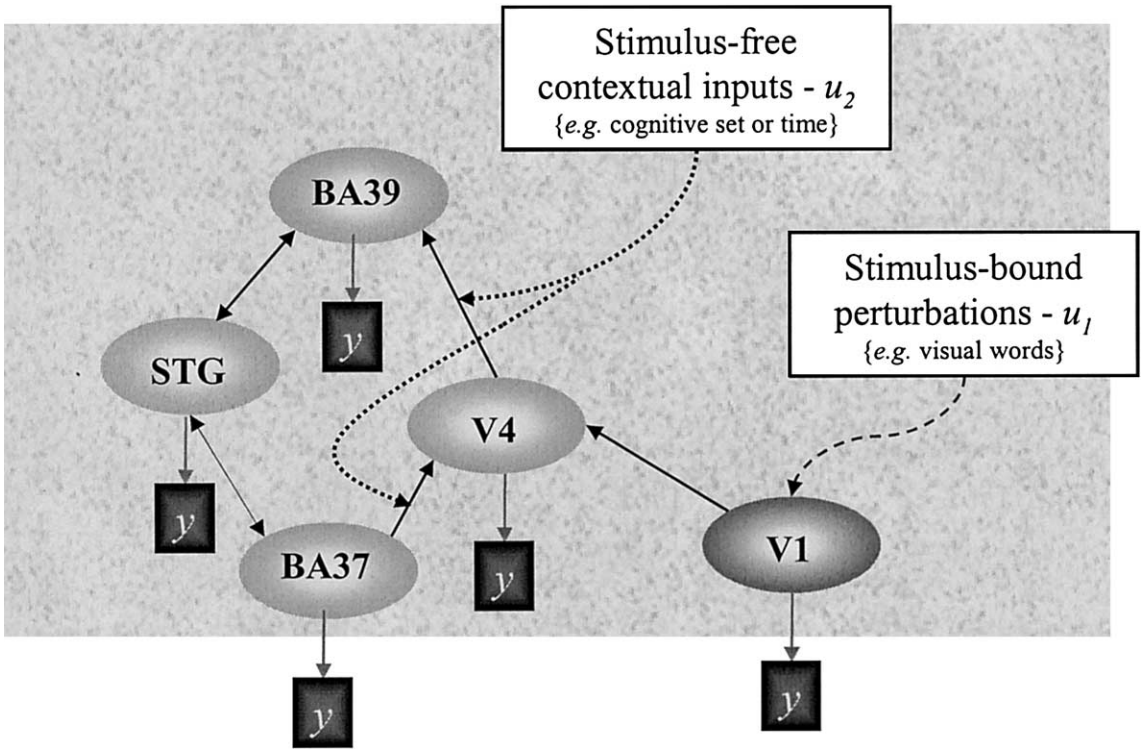


Fig. 1. This is a schematic illustrating the concepts underlying dynamic causal modelling. In particular it highlights the two distinct ways in which inputs or perturbations can elicit responses in the regions or nodes that compose the model. In this example there are five nodes, including visual areas **V1** and **V4** in the fusiform gyrus, areas 39 and 37, and the superior temporal gyrus **STG**. Stimulus-bound perturbations designated u_1 act as extrinsic inputs to the primary visual area **V1**. Stimulus-free or contextual inputs u_2 mediate their effects by modulating the coupling between **V4** and **BA39** and between **BA37** and **V4**. For example, the responses in the angular gyrus **(BA39)** are caused by inputs to **V1** that are transformed by **V4,** where the influences exerted by **V4** are sensitive to the second input. The dark square boxes represent the components of the **DCM** that transform the state variables z_i in each region (neuronal activity) into a measured (hemodynamic) response y_i.

Measured responses in high-order cortical areas are mediated by interactions among brain areas elicited by trial-bound perturbations. These interactions can be modulated by other set-related or contextual factors that modulate the latent or intrinsic coupling among areas. Fig. 1 illustrates this schematically. The important implication here for experimental design in DCM is that it should be multifactorial, with at least one factor controlling sensory perturbation and another factor manipulating the context in which the sensory-evoked responses are promulgated throughout the system (cf., psychophysiological interaction studies; Friston et al., 1997).

In this paper we use bilinear approximations to any DCM. The bilinear approximation reduces the parameters to three sets that control three distinct things: first, the direct or extrinsic influence of inputs on brain states in any particular area; second, the intrinsic or latent connections that couple responses in one area to the state of others; and, finally, changes in this intrinsic coupling induced by inputs. Although, in some instances, the relative strengths of intrinsic connections may be of interest, most analyses of DCMs focus on the changes in connectivity embodied in the bilinear parameters. The first set of parameters is generally of little interest in the context of DCM but is the primary focus in classical analyses of regionally specific effects. In classical analyses the only way experimental effects can be expressed is through a direct or extrinsic influence on each voxel because mass-univariate models (e.g., SPM) preclude connections and their modulation.

We envisage that DCM will be used primarily to answer questions about the modulation of effective connectivity through inferences about the third set of parameters described above. These will be referred to as bilinear in the sense that an input-dependent change in connectivity can be construed as a second-order interaction between the input and activity in a source region when causing a response in a target region. The key role of bilinear terms reflects the fact that the more interesting applications of effective connectivity address changes in connectivity induced by cognitive set or time. In short, DCM with a bilinear approximation allows one to claim that an experimental manipulation has “activated a pathway” as opposed to a cortical region. Bilinear terms correspond to psychophysiologic interaction terms in classical regression analyses of effective connectivity (Friston et al., 1997) and those formed by moderator variables (Kenny and Judd, 1984) in structural equation modelling (Büchel and Friston, 1997). This bilinear aspect speaks again to the importance of mul-

tifactorial designs that allow these interactions to be measured and the central role of the context in which region-specific responses are formed (see McIntosh, 2000).

1.3. DCM and inference

Because DCMs are not restricted to linear or instantaneous systems they are necessarily complicated and, potentially, need a large number of free parameters. This is why they have greater biological plausibility in relation to alternative approaches. However, this makes the estimation of the parameters more dependent upon constraints. A natural way to embody the requisite constraints is within a Bayesian framework. Consequently, dynamic causal models are estimated using Bayesian or conditional estimators and inferences about particular connections are made using the posterior or conditional density. In other words, the estimation procedure provides the probability distribution of a coupling parameter in terms of its mean and standard deviation. Having established this posterior density, the probability that the connection exceeds some specified threshold is easily computed. Bayesian inferences like this are more straightforward and interpretable than corresponding classical inferences and furthermore eschew the multiple comparison problem. The posterior density is computed using the likelihood and prior densitites. The likelihood of the data, given some parameters, is specified by the DCM (in one sense all models are simply ways of specifying the likelihood of an observation). The prior densities on the connectivity parameters offer suitable constraints to ensure robust and efficient estimation. These priors harness some natural constraints about the dynamics of coupled systems (see below) but also allow the user to specify which connections are likely to be present and which are not. An important use of prior constraints of this sort is the restriction of where inputs can elicit extrinsic responses. It is interesting to reflect that conventional analyses suppose that all inputs have unconstrained access to all brain regions. This is because classical models assume activations are caused directly by experimental factors, as opposed to being mediated by afferents from other brain areas.

Additional constraints on the intrinsic connections and their modulation by contextual inputs can also be specified but they are not necessary. These additional constraints can be used to finesse a model by making it more parsimonious, allowing one to focus on a particular connection. We will provide examples of this below. Unlike structural equation modelling, there are no limits on the number of connections that can be modelled because the assumptions and estimations schemes used by dynamic causal modelling are completely different, relying upon known inputs.

1.4. Overview

This paper comprises a theoretical section and three validation sections. In the theoretical section we present the conceptual and mathematical fundaments that are used in the remaining sections. The later sections address the face, predictive, and construct validity of DCM, respectively. Face validity ensures that the estimation and inference procedure identifies what it is supposed to. We have tried to establish face validity, using model systems and simulated data, to explore the performance of DCM over a range of hyperparameters (e.g., error variance, serial correlations among errors, etc). Some of these manipulations deliberately violate the assumptions of the model, embedded in priors, to establish that the estimation procedure remains robust in these circumstances. The subsequent section on predictive validity uses empirical data from an fMRI study of single word processing at different rates. These data were obtained consecutively in a series of contiguous sessions. This allowed us to repeat the DCM using independent realisations of the same paradigm. Predictive validity, over the multiple sessions, was assessed in terms of the consistency of the effective connectivity estimates and their posterior densities. The final section on construct validity revisits changes in connection strengths among parietal and extrastriate areas induced by attention to optic flow stimuli. We have established previously attentionally mediated increases in effective connectivity using both structural equation modelling and a Volterra formulation of effective connectivity (Büchel and Friston, 1997; Friston and Büchel, 2000). Our aim here is to show that DCM leads to the same conclusions. This paper ends with a brief discussion of DCM, its limitations and potential applications. This paper is primarily theoretical and hopes to introduce the concepts of DCM and establish its validity, at least provisionally.

2. Theory

In this section we present the theoretical motivation and operational details upon which DCM rests. In brief, DCM is a fairly standard nonlinear system identification procedure using Bayesian estimation of the parameters of deterministic input–state–output dynamic systems. In this paper the system can be construed as a number of interacting brain regions. We will focus on a particular form for the dynamics that corresponds to a bilinear approximation to any analytic system. However, the idea behind DCM is not restricted to bilinear forms.

The identification scheme described below conforms to a posterior density analysis under Gaussian assumptions. The details of this approach have already been described in relation to biophysical models of hemodynamic responses in a single brain region (Friston, 2002). That paper can be seen as a prelude to the current paper where we extend the model to cover multiple interacting regions. In the previous paper we were primarily concerned with estimating the efficacy with which input elicits a vasodilatory signal, presumably mediated by neuronal responses to the input. The

causal models in this paper can be regarded as a collection of hemodynamic models, one for each area, in which the experimental inputs are supplemented with neural activity from other areas. The parameters of interest now embrace not only the direct efficacy of experimental inputs but also the efficacy of neuronal input from distal regions, i.e., effective connectivity (see Fig. 1).

This section is divided into four parts. First, we describe the DCM itself and then summarise the estimation procedure used to find the posterior distribution of its parameters. This procedure requires priors on the parameters, which are considered in the third part. Finally, we describe conditional inferences about the parameters. Posterior density analyses find the maximum or mode of the posterior density of the parameters (i.e., the most likely coupling parameters given the data) by performing a gradient assent on the log posterior. The log posterior requires both likelihood and prior terms. The likelihood obtains from Gaussian assumptions about the errors in the observation model implied by the DCM. This likelihood or forward model is described next.

2.1. Dynamic causal models

The dynamic causal model here is a multiple-input multiple-output system that comprises m inputs and l outputs with one output per region. The m inputs correspond to designed causes (e.g., boxcar or stick stimulus functions). The inputs are exactly the same as those used to form design matrices in conventional analyses of fMRI and can be expanded in the usual way when necessary (e.g., using polynomials or temporal basis functions). In principle, each input could have direct access to every region. However, in practice the extrinsic effects of inputs are usually restricted to a single input region. Each of the l regions produces a measured output that corresponds to the observed BOLD signal. These l time series would normally be taken as the average or first eigenvariate of key regions, selected on the basis of a conventional analysis. Each region has five state variables. Four of these are of secondary importance and correspond to the state variables of the hemodynamic model presented in Friston et al. (2000). These hemodynamic states comprise a vasodilatory signal, normalised flow, normalised venous volume, and normalised deoxyhemoglobin content. These variables are required to compute the observed BOLD response and are not influenced by the states of other regions.

Central to the estimation of effective connectivity or coupling parameters are the first state variables of each region. These correspond to neuronal or synaptic activity and are a function of the neuronal states of other brain regions. We will deal first with the equations for the neuronal states and then briefly reprise the differential equations that constitute the hemodynamic model for each region.

2.1.1. Neuronal state equations

Restricting ourselves to the neuronal states $z = [z_1, \ldots, z_1]^T$, one can posit any arbitrary form or model for effective connectivity

$$\dot{z} = F(z, u, \theta), \tag{1}$$

where F is some nonlinear function describing the neurophysiological influences that activity z in all l brain regions and inputs u exert upon changes in the others. θ are the parameters of the model whose posterior density we require for inference. Some readers will note that Eq. (1) is a departure from the usual form of casual models in neuroimaging, in which the states are a static function of themselves $z = F(z, u, \theta)$ (e.g., $z = \theta z + u$, where u plays the role of an error process or innovation). These static models, such as those used by structural equation modelling, are a limiting case of the dynamic causal models considered in this paper that obtain when the inputs vary slowly in relation to neuronal dynamics (see Appendix A.1 for details).

It is not necessary to specify the form of Eq. (1) because its bilinear approximation provides a natural and useful reparameterisation in terms of effective connectivity. The bilinear form of Eq. (1) is:

$$\begin{aligned}
\dot{z} &\approx Az + \sum u_j B^j z + Cu \\
&= (A + \sum u_j B^j) z + Cu \\
A &= \frac{\partial F}{\partial z} = \frac{\partial \dot{z}}{\partial z} \\
B^j &= \frac{\partial^2 F}{\partial z \partial u_j} = \frac{\partial}{\partial u_j} \frac{\partial \dot{z}}{\partial z} \\
C &= \frac{\partial F}{\partial u}.
\end{aligned} \tag{2}$$

The Jacobian or connectivity matrix A represents the first-order connectivity among the regions in the absence of input. Effective connectivity is the influence that one neuronal system exerts over another in terms of inducing a response $\partial \dot{z}/\partial z$. In DCM a response is defined in terms of a change in activity with time $\dot{z}$. This latent connectivity can be thought of as the intrinsic coupling in the absence of experimental perturbations. Notice that the state, which is perturbed, depends on the experimental design (e.g., baseline or control state) and therefore the intrinsic coupling is specific to each experiment. The matrices B^j are effectively the change in coupling induced by the jth input. They encode the input-sensitive changes in $\partial \dot{z}/\partial z$ or, equivalently, the modulation of effective connectivity by experimental manipulations. Because B^j are second-order derivatives these terms are referred to as bilinear. Finally, the matrix C embodies the extrinsic influences of inputs on neuronal activity. The parameters $\theta^c = \{A, B^j, C\}$ are the connectivity or coupling matrices that we wish to identify and define the functional architecture and interactions among brain

regions at a neuronal level. Fig. 2 shows an example of a specific architecture to demonstrate the relationship between the matrix form of the bilinear model and the underlying state equations for each region. Notice that the units of connections are per unit time and therefore correspond to rates. Because we are in a dynamic setting a strong connection means an influence that is expressed quickly or with a small time constant. It is useful to appreciate this when interpreting estimates and thresholds quantitatively. This is illustrated below.

The neuronal activities in each region cause changes in volume and deoxyhemoglobin to engender the observed BOLD response y as described next. The ensuing hemodynamic component of the model is specific to BOLD-fMRI and would be replaced by appropriate forward models for other modalities; for example, models based on classical electromagnetics for EEG signals, caused by postsynaptic currents measured at the scalp. In principle, it would be possible to solve the inverse problem for any imaging modality to estimate the underlying neuronal processes z and then use Eq. (2) as the basis of a generic DCM that could be identified post hoc. However, augmenting the neuronal DCM with a modality-specific forward model is a mathematically equivalent but more graceful approach that subsumes all the identification issues into a single estimation procedure.

2.1.2. Hemodynamic state equations

The remaining state variables of each region are biophysical states engendering the BOLD signal and mediate the translation of neuronal activity into hemodynamic responses. Hemodynamic states are a function of, and only of, the neuronal state of each region. These equations have been described elsewhere (Friston et al., 2000) and constitute a hemodynamic model that embeds the Balloon–Windkessel model (Buxton et al., 1998; Mandeville et al., 1999). In brief, for the ith region, neuronal activity z_i causes an increase in a vasodilatory signal s_i that is subject to autoregulatory feedback. Inflow f_i responds in proportion to this signal with concomitant changes in blood volume v_i and deoxyhemoglobin content q_i.

$$\begin{aligned}
\dot{s}_i &= z_i - \kappa_i s_i - \gamma_i(f_i - 1)\\
\dot{f}_i &= s_i\\
\tau_i \dot{v}_i &= f_i - v_i^{1/\alpha}\\
\tau_i \dot{q}_i &= f_i E(f_i, \rho_i)/\rho_i - v_i^{1/\alpha} q_i / v_i.
\end{aligned} \tag{3}$$

Outflow is related to volume $f_{out}(v) = v^{1/\alpha}$ through Grubb's exponent α (Grubb et al., 1974). The oxygen extraction is a function of flow $E(f, \rho) = 1 - (1 - \rho)^{1/f}$ where ρ is resting oxygen extraction fraction. The BOLD signal is taken to be a static nonlinear function of volume and deoxyhemoglobin that comprises a volume-weighted sum of extra- and intravascular signals

$$\begin{aligned}
y_i &= g(q_i, v_i)\\
&= V_0(k_1(1 - q_i) + k_2(1 - q_i/v_i) + k_3(1 - v_i))\\
k_1 &= 7\rho_i\\
k_2 &= 2\\
k_3 &= 2\rho_i - 0.2,
\end{aligned} \tag{4}$$

where $V_0 = 0.02$ is resting blood volume fraction. Again it should be noted that the particular forms of Eqs. (3) and (4) are specific to BOLD-fMRI and should, obviously, be replaced by appropriate state and output equations for other modalities. A list of the biophysical parameters $\theta^h = \{\kappa, \gamma, \tau, \alpha, \rho\}$ is provided in Table 1 and a schematic of the hemodynamic model is shown in Fig. 3.

2.2. Estimation

In this subsection we describe the expectation maximization (EM) procedure for estimating the DCM above. More details are provided in Appendix A.2. Combining the neuronal and hemodynamic states $x = \{z, s, f, v, q\}$ gives us a full forward model specified by the neuronal state Eq. (2) and the hemodynamic Eqs. (3) and (4):

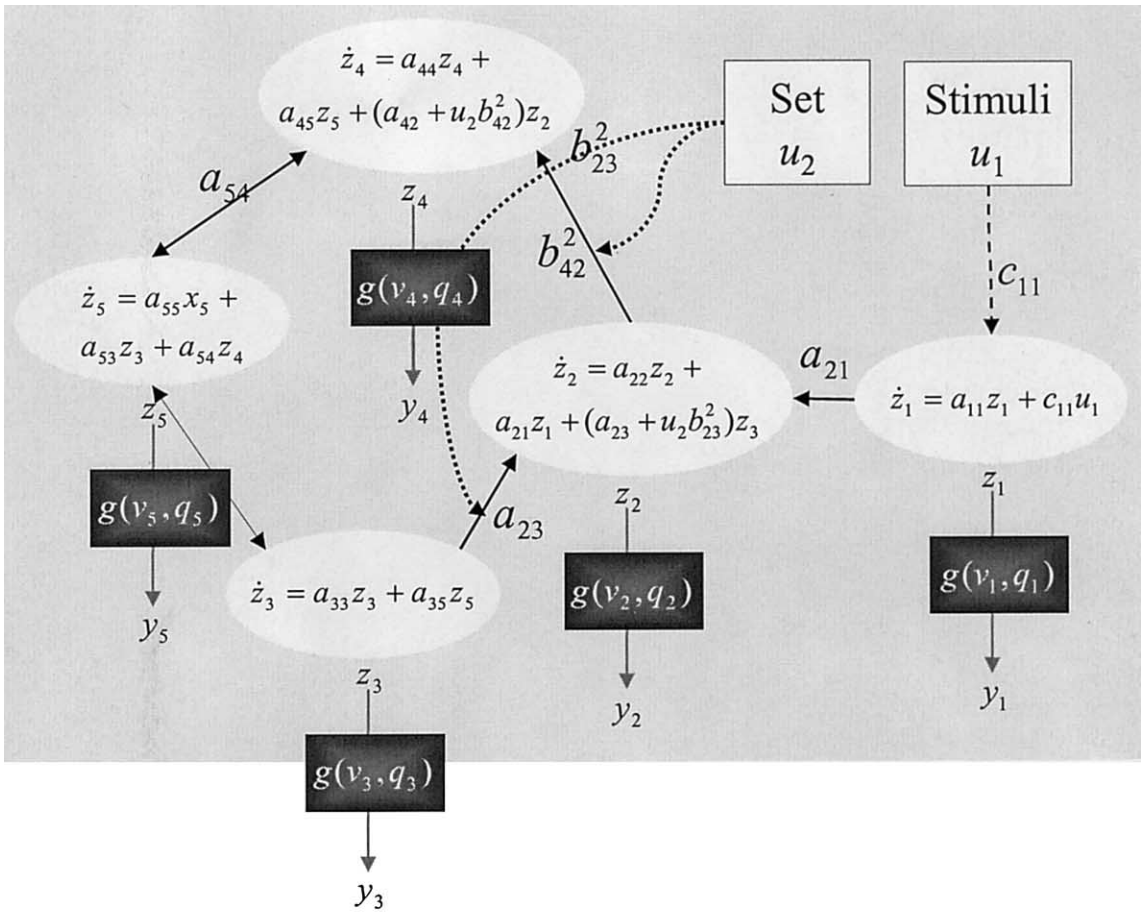
$$\begin{aligned}
\dot{x} &= f(x, u, \theta)\\
y &= \lambda(x),
\end{aligned} \tag{5}$$

with parameters $\theta = \{\theta^c, \theta^h\}$. For any set of parameters and inputs, the state equation can be integrated and passed though the output nonlinearity [Eq. (4)] to give a predicted response $h(u, \theta)$. This integration can be made quite expedient by capitalising on the sparsity of stimulus functions commonly employed in fMRI designs (see Friston, 2002, Section 3.4). Integrating Eq. (5) is equivalent to a generalised convolution of the inputs with the systems Volterra kernels. These kernels are easily derived from the Volterra expansion of Eq. (5) (Bendat, 1990),

$$\begin{aligned}
h_i(u, \theta) &= \sum_k \int_0^t \cdots \int_0^t \kappa_i^k(\sigma_1, \ldots, \sigma_k) u(t - \sigma_l),\\
&\qquad \ldots, u(t - \sigma_k) d\sigma_1, \ldots, d\sigma_k\\
\kappa_i^k(\sigma_1, \ldots, \sigma_k) &= \frac{\partial^k y_i(t)}{\partial u(t - \sigma_1), \ldots, \partial u(t - \sigma_k)},
\end{aligned} \tag{6}$$

either by numerical differentiation or analytically through bilinear approximations (see Friston, 2000, Appendix). κ_i^k is the kth order kernel for region i. For simplicity, Eq. (6) has been written for a single input. The kernels are simply a reparameterisation of the model. We will use these kernels to characterise regional impulse responses at neuronal and hemodynamic levels later.

The forward model can be made into an observation model by adding error and confounding or nuisance effects



latent connectivity induced connectivity

$$\begin{bmatrix} \dot{z}_1 \\ \vdots \\ \dot{z}_5 \end{bmatrix} = \left\{ \begin{bmatrix} a_{11} & \cdots & 0 \\ a_{21} & a_{22} & a_{23} \\ \vdots & a_{33} & a_{35} \\ a_{42} & a_{44} & a_{45} \\ 0 & \cdots & a_{53} \quad a_{54} \quad a_{55} \end{bmatrix} + u_2 \begin{bmatrix} 0 & \cdots & 0 \\ & b_{23}^2 & \\ \vdots & \ddots & \vdots \\ b_{42}^2 & & \\ 0 & \cdots & 0 \end{bmatrix} \right\} \begin{bmatrix} z_1 \\ \vdots \\ z_5 \end{bmatrix} + \begin{bmatrix} c_{11} & 0 \\ \vdots & \vdots \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

Forward, backward & self

$$\dot{z} = (A + \sum_j u_j B^j)z + Cu$$

The bilinear model

Fig. 2. This schematic (upper panel) recapitulates the architecture in Fig. 1 in terms of the differential equations implied by a bilinear approximation. The equations in each of the white areas describe the change neuronal activity z_i in terms of linearly separable components that reflect the influence of other regional state variables. Note particularly how the second contextual inputs enter these equations. They effectively increase the intrinsic coupling parameters (a_{ij}) in proportion to the bilinear coupling parameters (b_{ij}^k). In this diagram the hemodynamic component of the DCM illustrates how the neuronal states enter a region-specific hemodynamic model to produce the outputs y_i that are a function of the region's biophysical states reflecting deoxyhemoglobin content and venous volume (q_i and v_i). The lower panel reformulates the differential equations in the upper panel into a matrix format. These equations can be summarised more compactly in terms of coupling parameter matrices A, B^j, and C. This form of expression is used in the main text and shows how it relates to the underlying differential equations that describe the state dynamics.

Table 1
Priors on biophysical parameters

Parameter	Description	Prior Mean η_θ	Prior Variance C_θ
κ	Rate of signal decay	0.65 per s	0.015
γ	Rate of flow-dependent elimination	0.41 per s	0.002
τ	Hemodynamic transit time	0.98 s	0.0568
α	Grubb's exponent	0.32	0.0015
ρ	Resting oxygen extraction fraction	0.34	0.0024

$X(t)$ to give $y = h(u, \theta) + X\beta + \varepsilon$. Here β are the unknown coefficients of the confounds. In the examples used below, $X(t)$ comprised a low-order discrete cosine set, modelling low-frequency drifts and a constant term. Following the approach described in Friston (2002), we note

$$\begin{aligned} y - h(u, \eta_{\theta|y}) &\approx J\Delta\theta + X\beta + \varepsilon \\ &= [J, X]\begin{bmatrix} \Delta\theta \\ \beta \end{bmatrix} + \varepsilon \\ \Delta\theta &= \theta - \eta_{\theta|y} \\ J &= \frac{\partial h(u, \eta_{\theta|y})}{\partial \theta}. \end{aligned} \tag{7}$$

This local linear approximation then enters an iterative EM scheme, described previously (Friston, 2002) and in Appendix A.2, to give the conditional expectation $\eta_{\theta|y}$ and covariance $C_{\theta|y}$ of the parameters and restricted maximum likelihood (ReML) estimates of hyperparameters λ for the error covariance.

Until convergence {

E-step

$$J = \frac{\partial h(\eta_{\theta|y})}{\partial \theta}$$

$$\bar{y} = \begin{bmatrix} y - h(\eta_{\theta|y}) \\ \eta_\theta - \eta_{\theta|y} \end{bmatrix}, \bar{J} = \begin{bmatrix} J & X \\ 1 & 0 \end{bmatrix},$$

$$\bar{C}_\varepsilon = \begin{bmatrix} \sum \lambda_i Q_i & 0 \\ 0 & C_\theta \end{bmatrix}$$

$$C_{\theta|y} = (\bar{J}^T \bar{C}_\varepsilon^{-1} \bar{J})^{-1}$$

$$\begin{bmatrix} \Delta\eta_{\theta|y} \\ \eta_{\beta|y} \end{bmatrix} = C_{\theta|y}(\bar{J}^T \bar{C}_\varepsilon^{-1} \bar{y})$$

$$\eta_{\theta|y} \leftarrow \eta_{\theta|y} + \Delta\eta_{\theta|y}$$

M-step

$$P = \bar{C}_\varepsilon^{-1} - \bar{C}_\varepsilon^{-1} \bar{J} C_{\theta|y} \bar{J}^T \bar{C}_\varepsilon^{-1}$$

$$\frac{\partial F}{\partial \lambda_i} = -\frac{1}{2} tr\{PQ_i\} + \frac{1}{2} \bar{y}^T P^T Q_i P \bar{y}$$

$$\left\langle \frac{\partial^2 F}{\partial \lambda_{ij}^2} \right\rangle = -\frac{1}{2} tr\{PQ_i PQ_j\}$$

$$\lambda \leftarrow \lambda - \left\langle \frac{\partial^2 F}{\partial \lambda^2} \right\rangle^{-1} \frac{\partial F}{\partial \lambda} \tag{8}$$

These expressions are formally the same as Eq. (15) in Friston (2002) but for the addition of confounding effects in X. These confounds are treated as fixed effects with infinite prior variance, which does not need to appear explicitly in Eq. (8).

Note that the prediction and observations encompass the entire experiment. They are therefore large $ln\ x\ 1$ vectors whose elements run over regions and time. Although the response variable could be viewed as a multivariate time series, it is treated as a single observation vector, whose error covariance embodies both temporal and interregional correlations $C_\varepsilon = V \otimes \Sigma(\lambda) = \Sigma\lambda_i Q_i$. This covariance is parameterised by some covariance hyperparameters λ_i that scale the contribution of covariance components Q_i. The form of the covariance matrix conforms to a Kronecker tensor product of the $n \times n$ matrix V encoding temporal correlations among the errors and an unknown $l \times l$ regional error covariance Σ. In the examples below Q_i corresponds to region-specific error variances assuming the same temporal correlations $Q_i = V \otimes \Sigma_i$ in which Σ_i is a $l \times l$ sparse matrix with the ith leading diagonal element equal to one.

Eq. (8) enables us to estimate the conditional moments of the coupling parameters (and the hemodynamics parameters) plus the hyperparameters controlling observation error. However, to proceed we need to specify the prior expectation η_θ and covariance C_θ.

2.3. The priors

To form the posterior density one needs to combine the likelihood with a prior density on the parameters (see Appendix A.2). In this paper we use a fully Bayesian approach because (1) there are clear and necessary constraints on neuronal dynamics that can be used to motivate priors on the coupling parameters and (2) empirically determined priors on the biophysical hemodynamic parameters are relatively easy to specify. We will deal first with priors on the coupling parameters.

2.3.1. Priors on the coupling parameters

It is self-evident that neuronal activity cannot diverge exponentially to infinite values. Therefore, we know that, in the absence of input, the neuronal state must return to a stable mode. Mathematically, this means the largest real eigenvalue of the intrinsic coupling matrix, also known as the principal Lyapunov exponent, must be negative. We will use this constraint to establish a prior density on the coupling parameters a_{ij} that ensures the system is dissipative.

The specification of priors on the connections can be finessed by a reparameterisation of the coupling matrices

$$A \rightarrow \sigma A = \sigma \begin{bmatrix} -1 & a_{12} & \cdots \\ a_{21} & -1 & \\ \vdots & & \ddots \end{bmatrix}$$

$$B^j \rightarrow \sigma B^j = \sigma \begin{bmatrix} b^j_{11} & b^j_{12} & \cdots \\ b^j_{21} & \ddots & \\ \vdots & & \end{bmatrix}. \quad (9)$$

This factorisation into a scalar and normalised coupling matrix renders the normalised couplings adimensional, such that strengths of connections among regions are relative to the self-connections. From this point on, we will deal with normalised parameters. This particular factorisation enforces the same self-connection or temporal scaling σ in all regions. Although there is evidence that hemodynamics can vary from region to region, there is less reason to suppose that the neuronal dynamics, intrinsic to each region, will differ markedly. Having said this, it is perfectly possible to employ different factorisations (for example, factorisations

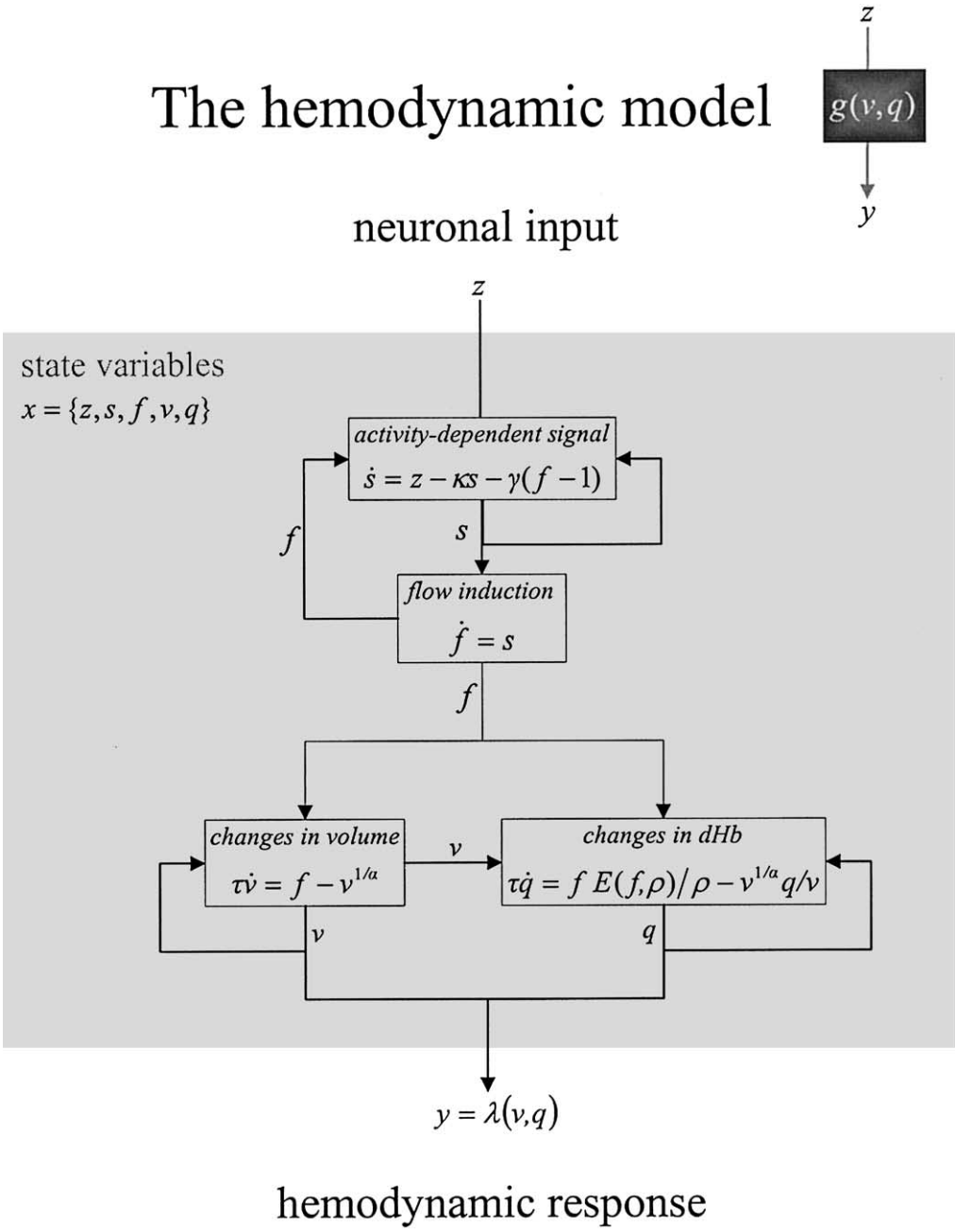


Fig. 3. This schematic shows the architecture of the hemodynamic model for a single region (regional subscripts have been dropped for clarity). Neuronal activity induces a vasodilatory and activity-dependent signal s that increases the flow f. Flow causes changes in volume and deoxyhemoglobin (v and q). These two hemodynamic states enter the output nonlinearity Eq. (4) to give the observed BOLD response y. This transformation from neuronal states z_i to hemodynamic response y_i is encoded graphically by the dark grey boxes in the previous figure and in the insert above.

that allow for region-specific self-connections, using diagonal matrices). We will use the factorisation in Eq. (9) because it simplifies the derivation and specification of priors.

If the principal Lyapunov exponent of A is less than zero, the system will converge to a point attractor. If it is zero, the system will exhibit oscillatory dynamics. It is therefore sufficient to constrain the connection strengths, with Gaussian priors on a_{ij}, to ensure that the principal exponent is negative. The problem here is that different, but equally probable, configurations of a_{ij} can have different principal exponents. One solution is to use the probability that the *biggest* possible exponent exceeds zero and choose priors that render this probability suitably small. For normalised connections, the biggest principal exponent is easily derived and the appropriate variance of a_{ij} can be computed (see Appendix A.3). This represents a simple way to place an upper bound on the principal exponent. The derivation of the prior variance $v_a = \mathrm{Var}(a_{ij})$ and moments of the scaling parameter (η_σ and v_σ) are provided in Appendix A.3.

In brief, priors on the connectivity parameters ensure that the system remains stable. Coupling matrices can be decomposed into a scaling parameter σ that corresponds to the intrinsic decay or self-inhibition of each region and a normalised coupling matrix. The spectrum of eigenvalues of the intrinsic coupling matrix determines the time constants of modes or patterns of neuronal activity expressed in response to perturbation. These are scaled by σ, whose prior expectation controls the characteristic neuronal time constants (i.e., those observed in the absence of coupling). In this work we have assumed a value of 1 s ($\eta_\sigma = 1$), motivated by the time constants of evoked neuronal transients observed using single-unit electrode recordings and EEG. The prior variance v_σ is chosen to make the probability that it is negative suitably small (in this paper 10^{-3}). The resulting prior can be expressed as a function of the implicit half-life $\tau_z(\sigma) = \ln 2/\sigma$ by noting $p(\tau_z) = p(\sigma)\partial\sigma/\partial\tau_z$. This transformation (Fig. 4) shows that we expect regional transients with time constants in the range of a few hundred milliseconds to several seconds.

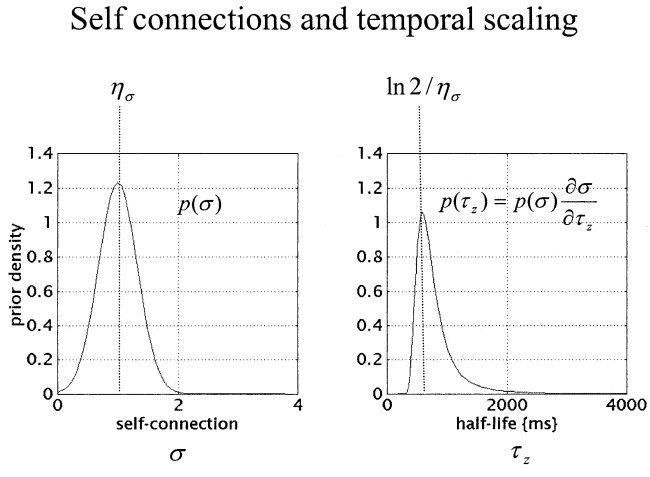


Fig. 4. Prior probability density functions for the temporal scaling parameter or self-connection σ. This has a Gaussian form (left panel) that translates into a skewed distribution, when expressed in terms of the characteristic half-life of neural transients τ_z in any particular region (right panel). This prior distribution implies that neuronal activity will decay with a half-life of roughly 500 ms, falling in the range of 300 ms to 2 s.

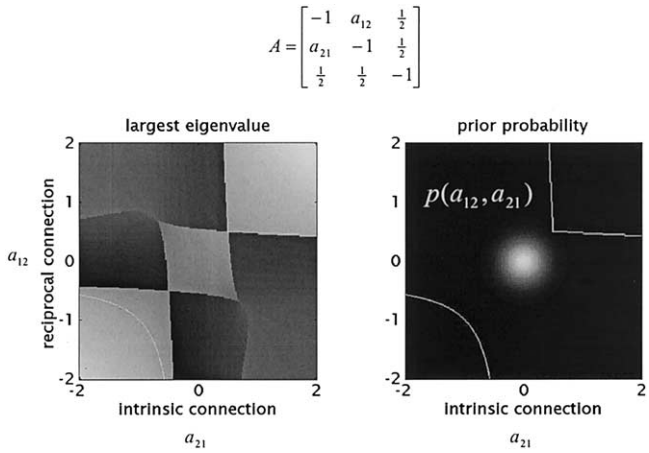


Fig. 5. Prior probability density on the intrinsic coupling parameters for a specific intrinsic coupling matrix A. The left panel shows the real value of the largest eigenvalue of A (the principal Lyapunov exponent) as a function of the connection from the first to the second region and the reciprocal connection from the second to the first. The remaining connections were held constant at 0.5. This density can be thought of as a slice through a multidimensional spherical distribution over all connections. The right panel shows the prior probability density function and the boundaries at which the largest real eigenvalue exceeds zero (dotted lines). The variance or dispersion of this probability distribution is chosen to ensure that the probability of excursion into unstable domains of parameter space is suitably small. These domains are the upper right and lower left bounded regions.

The prior distribution of individual connection strengths a_{ij} is assumed to be identically and independently distributed with a prior expectation $\eta_a = 0$ and a variance v_a that ensures the principal exponent has a small probability of being positive (here 10^{-3}). This variance decreases with the number of regions. To provide an intuition about how these priors keep the system stable, a quantitative example is shown in Fig. 5. Fig. 5 shows the prior density of two connections that renders the probability of a positive exponent less than 10^{-2}. It can be seen that this density lies in a domain of parameter space encircled by regions in which the principal exponent exceeds zero (bounded by dotted lines). See the figure legend for more details.

Priors on the bilinear coupling parameters b_{ij}^k are the same as those for the intrinsic coupling parameters. Because these represent the input-dependent component of a_{ij} they are also normalised by σ and are consequently adimensional [See Eq. (9)]. Conversely, priors on the influences of extrinsic input c_{ik} are relatively uninformative with zero expectation and unit variance.

It is important to interpret the bilinear estimators and their priors in relation to the scaling of the inputs. This is because the bilinear parameters are not invariant to trans-

formations of the inputs. For example, if we doubled the size of the inputs we would have to halve the values of b_{ij}^k to conserve their modulatory effects on interactions among neuronal states. Consequently, we always scale inputs such that their time integral equals the number of events that have occurred or the number of seconds a particular experimental context prevails (see next section). Furthermore, the inputs should be specified in a way that conforms to the prior assumption that the inputs modulate connections independently of each other. This means that reparameterising the inputs, by taking linear combinations of the original inputs, will lead to slightly different results because of the change in implicit priors. As noted in the introduction, additional constraints can be implemented by precluding certain connections. This is achieved by setting their variance to zero.

The simple but useful prior based on the principal exponent can be applied to any DCM. In contradistinction, the priors on hemodynamic parameters are specific to the fMRI application considered here.

2.3.2. Hemodynamic priors

The hemodynamic priors are based on those used in Friston (2002). In brief, the mean and variance of posterior estimates of the five biophysical parameters were computed over 128 voxels using the single word presentation data presented in the next section. These means and variances (see Table 1) were used to specify Gaussian priors on the hemodynamic parameters.

This would be quite sufficient for general purposes. However, reducing the rank of the prior covariance of the biophysical parameters can finesse the computational load on estimation. This is effectively the same as allowing only two linear mixtures of the hemodynamic parameters to change from region to region. In the examples below these mixtures were those controlling the expression of the first two eigenvectors or principal components ε_1^h and ε_2^h of the prior covariance of the response κ^1 in measurement space. For any region

$$\begin{aligned} Cov\{\kappa^1\} &= \frac{\partial \kappa^1}{\partial \theta^h} C_\theta^h \frac{\partial \kappa^{1^T}}{\partial \theta^h} \\ &= \varepsilon^h \lambda^h \varepsilon^{hT} \\ C_\theta^h &= \frac{\partial \kappa^{1^+}}{\partial \theta^h} \varepsilon^h \lambda^h \varepsilon^{hT} \frac{\partial \kappa^{1^{+T}}}{\partial \theta^h}, \end{aligned} \tag{10}$$

where + denotes pseudoinverse. This response is simply the first order kernel from Eq. (6) and depends on the prior covariance of the biophysical parameters C_θ^h. The motivation for this is based on the fact that although the biophysical parameters may vary independently, their influence on the observed hemodynamic response may be indistinguishable. The eigenvalue spectrum in the leading diagonal of λ^h suggests that there are only two modes of substantial hemodynamic variation (see Fig. 6, left panel). After setting the remaining eigenvalues to zero, the last line of Eq. (10)

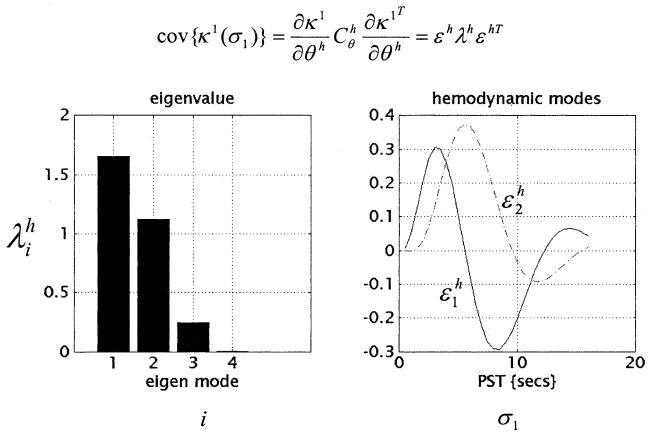


Fig. 6. An analysis of the prior probability distributions of the impulse or hemodynamic responses induced by neuronal activity in a single region. This prior distribution in measurement space has been characterised in terms of the covariance of the first-order Volterra kernel κ^1 and its eigen decomposition. The decomposition is used to show that there are only two substantial modes of hemodynamic variation caused by the prior variation in biophysical parameters. The eigenvalues λ_i^h reflect the variance expressed by each of these modes. The two modes depicted on the right side correspond to the eigenvectors ε_i^h, which are a function of peri-stimulus time (σ_1). See the main text for a full description of the variables used in this figure.

specifies the adjusted prior covariances in parameter space. Restricting the prior density of the biophysical parameters to a two-dimensional subspace is not a terribly important component of constraining the parameters because we are not interested in making inferences about them. However, restricting the search space in this fashion makes the estimation procedure more efficient and reduces computation time. It is interesting to note that the first eigenvector is almost exactly the first temporal derivative of the second, which itself looks very much like a canonical hemodynamic response (see Fig. 6, right panel). This will be important later.

Combining the prior densities on the coupling and hemodynamic parameters allows us to express the prior probability of the parameters in terms of their prior expectation η_θ and covariance C_θ

$$\theta = \begin{bmatrix} \sigma \\ a_{ij} \\ b_{ij}^k \\ c_{ik} \\ \theta^h \end{bmatrix}, \; \eta_\theta = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ \eta_\theta^h \end{bmatrix},$$

$$C_\theta = \begin{bmatrix} v_\sigma & & & & \\ & C_A & & & \\ & & C_B & & \\ & & & 1 & \\ & & & & C_h \end{bmatrix}, \tag{11}$$

where the prior covariances C_A and C_B contain leading diagonal elements v_a for all connections that are allowed to vary. Having specified the priors, we are now in a position

to form the posterior and proceed with estimation using Eq. (8).

2.4. Estimation

As noted above, the estimation scheme is a posterior density analysis under Gaussian assumptions. This is described in detail in Friston (2002). In short, the estimation scheme provides the approximating Gaussian posterior density of the parameters $q(\theta)$ in terms of its expectation $\eta_{\theta|y}$ and covariance $C_{\theta|y}$. The expectation is also known as the posterior mode or maximum a posteriori (MAP) estimator. The marginal posterior probabilities are then used for the inference that any particular parameter or contrast of parameters $c^T\eta_{\theta|y}$ (e.g., average) exceeds a specified threshold γ.

$$p = \phi_N\left(\frac{c^T\eta_{\theta|y} - \gamma}{\sqrt{c^T C_{\theta|y} c}}\right), \tag{12}$$

where ϕ_N is the cumulative normal distribution. In this paper, we are primarily concerned with the coupling parameters θ^c and, among these, the bilinear terms. The units of these parameters are Hz or per second (or adimensional if normalised) and the thresholds are specified as such. In dynamical modelling strength corresponds to a fast response with a small time constant.

2.5. Relationship to conventional analyses

It is interesting to note that conventional analyses of fMRI data, using linear convolution models, are a special case of dynamic causal models using a bilinear approximation. This is important because it provides a direct connection between DCM and classical models. If we allow inputs to be connected to all regions and discount interactions among regions by setting the prior variances on a_{ij} and b_{ij}^k to zero, we produce a set of disconnected brain regions or voxels that respond to, and only to, extrinsic input. The free parameters of interest reduce to the elements of C, which reflect the ability of input to excite neural activity in each voxel. By further setting the prior variances on the self connections (i.e., scaling parameter) and those on the hemodynamic parameters to zero we end up with a single-input single-output model at each and every brain region that can be reformulated as a convolution model as described in Friston (2002). For voxel i and input j, c_{ij} can be estimated by simply convolving the input with $\partial\kappa_i^1/\partial c_{ij}$ where κ_i^1 is the first order kernel meditating the influence of input j on output i. The convolved inputs are then used to form a general linear model that can be estimated using least squares in the usual way. This is precisely the approach adopted in classical analyses, in which $\partial\kappa_i^1/\partial c_{ij}$ is usually referred to as the hemodynamic response function. The key point here is that the general linear models used in typical data analyses are special cases of bilinear models *that embody more assumptions*. These assumptions enter through the use of highly precise priors that discount interactions among regions and prevent any variation in biophysical responses.

Having described the theoretical aspects of DCM, we now turn to applications and assessing its validity.

3. Face validity—simulations

3.1. Introduction

In this section we use simulated data to establish the utility of the bilinear approximation and the robustness of the estimation scheme described in the previous section. We used the same functional architecture to generate simulated data, under a variety of different conditions, to ensure that the Bayesian estimates are reasonable, even when the underlying assumptions are deliberately violated. Furthermore, we chose a simulated architecture that would be impossible to characterise using existing methods based on regression models (e.g., structural equation modelling). This architecture embodies loops and reciprocal connections and poses the problem of vicarious input: the ambiguity between the direct influences of one area and influences that are mediated through others.

3.1.1. The simulated system

The architecture is depicted in Fig. 7 and has been labelled so that it is consistent with the DCM characterised empirically in the next section. The model comprises three regions; a primary (**A1**) and secondary (**A2**) auditory area and a higher-level region (**A3**). There are two inputs. The first is a sensory input u_1 encoding the presentation of epochs of words at different frequencies. The second input u_2 is contextual in nature and is simply an exponential function of the time elapsed since the start of each epoch (with a time constant of 8 s). These inputs were based on a real experiment and are the same as those used in the empirical analyses of the next section. The scaling of the inputs is important for the quantitative evaluation of the bilinear and extrinsic coupling parameters. The convention adopted here is that inputs encoding events approximate delta functions such that their integral over time corresponds to the number of events that have occurred. For event-free inputs, like the maintenance of a particular instructional set, the input is scaled to a maximum of unity, so that the integral reflects the number of seconds over which the input was prevalent. The inputs were specified in time bins that were a sixteenth of the interval between scans (repetition time: TR = 1.7 s).

The auditory input is connected to the primary area; the second input has no direct effect on activity but modulates the forward connections from **A1** to **A2** so that its influence shows *adaptation* during the epoch. The second auditory area receives input from the first and sends signals to the higher area (**A3**). In addition to reciprocal backward con-

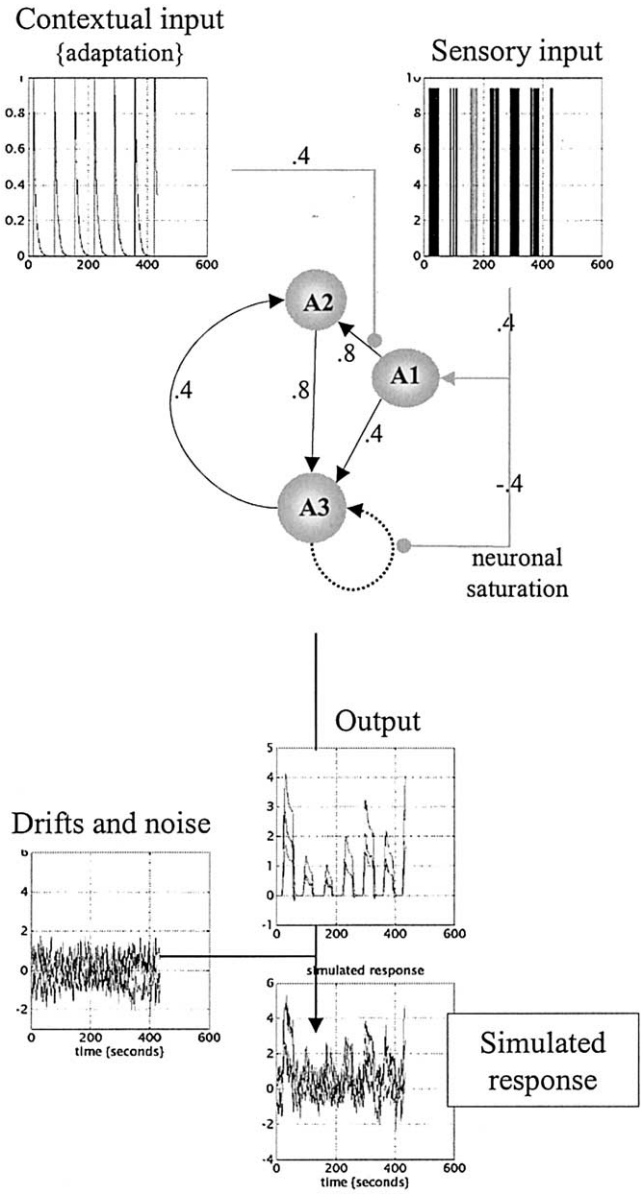


Fig. 7. This is a schematic of the architecture used to generate simulated data. Nonzero intrinsic connections are shown as directed black arrows with the strength or true parameter alongside. Here, the perturbing input is the presentation of words (sensory inputs) and acts as an intrinsic influence on **A1**. In addition, this input modulates the self-connection of **A3** to emulate saturation like-effects (see text and Fig. 8). The contextual input is a decaying exponential of within-epoch time and positively modulates the forward connection from **A1** to **A2**. The lower panel shows how responses were simulated by mixing the output of the system described above with drifts and noise as described in the text.

nection, in this simple auditory hierarchy, a connection from the lowest to the highest area has been included. Finally, the first input (word presentation) modulates the self-connections of the third region. This influence has been included to show how bilinear effects can emulate nonlinear responses. A bilinear modulation of the self-connection can augment or attenuate decay of synaptic activity, rendering the average response to streams of stimuli rate-dependent. This is because the bilinear effect will only be expressed if sufficient synaptic activity persists after the previous stimulus. This, in turn, depends on a sufficiently fast presentation rate. The resulting response emulates a saturation at high presentation rates or small stimulus onset asynchronies that have been observed empirically. Critically, we are in a position to disambiguate between neuronal saturation, modelled by this bilinear term, and hemodynamic saturation, modelled by nonlinearities in the hemodynamic component of this DCM. A significant bilinear self-connection implies neuronal saturation above and beyond that attributable to hemodynamics. Fig. 8 illustrates this neuronal saturation by plotting the simulated response of **A3** in the absence of saturation $B^1 = 0$ against the simulated response with $b^1_{3,3} = -0.4$. It is evident that there is a nonlinear subadditive effect at high response levels. It should be noted that true neuronal saturation of this sort is mediated by second order interactions among the states (i.e., neuronal activity). However, as shown in Fig. 8, we can emulate these effects by using the first extrinsic input as a surrogate for neuronal inputs from other areas in the bilinear component of the model.

Using this model we simulated responses using the values for A, B^1, B^2, and C given in Fig. 7 and the prior expectations for the biophysical parameters given in Table 1. The values of the coupling parameters were chosen to emulate those seen typically in practice. This ensured the simulated responses were realistic in relation to simulated noise. After downsampling these deterministic responses every 1.7 s (the TR of the empirical data used in the next section), we added known noise to produce simulated data. These data composed time series of 256 observations with independent or serially correlated Gaussian noise based on

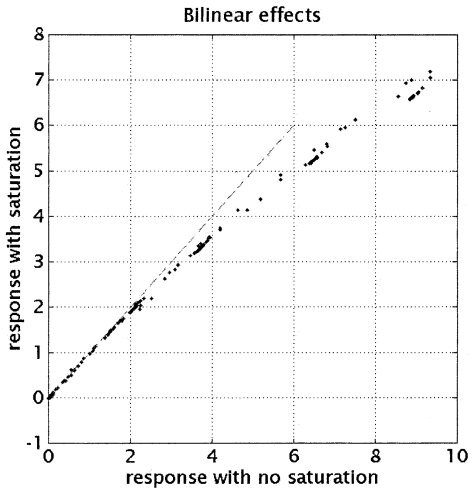


Fig. 8. This is a plot of the simulated response with saturation against the equivalent response with no saturation. These simulated responses were obtained by setting the bilinear coupling parameter b^1_{33} labelled "neuronal saturation" in the previous figure to -0.4 and zero, respectively. The key thing to observe is a saturation of responses at high levels. The broken line depicts the response expected in the absence of saturation. This illustrates how bilinear effects can introduce nonlinearities into the response.

an AR(1) process. Unless otherwise stated, the noise had 0.5 standard deviation and was i.i.d. (independently and identically distributed). The drift terms were formed from the first six components of a discrete cosine set mixed linearly with normal random coefficients, scaled by one over the order. This emulates a $1/f^2$ plus white noise spectrum for the noise and drifts. See the lower panel of Fig. 7 for an exemplar data simulation with i.i.d. noise of unit variance.

3.1.2. Exemplar analysis

The analysis described in the previous section was applied to the data shown in Fig. 7. The priors on coupling parameters were augmented by setting the variance of the off-diagonal elements of B^1 (saturation) and all but two connections in B^2 (adaptation) to zero. These two connections were the first and second forward connections of this cortical hierarchy. The first had simulated adaptation, whereas the second did not. Extrinsic input was restricted to the primary area **A1** by setting the variances of all but c_{11} to zero. We placed no further constraints on the intrinsic coupling parameters. This is equivalent to allowing full connectivity. This would be impossible with structural equation modelling. The results are presented in Fig. 9 in terms of the MAP or conditional expectations of the coupling parameters (upper panels) and the associated posterior probabilities (lower panels) using Eq. (12). It can be seen that the intrinsic coupling parameters are estimated reasonably accurately with a slight overestimation of the backward connection from **A3** to **A2**. The bilinear coupling parameters modelling adaptation are shown in the lower panels and the estimators have correctly identified the first forward connection as the locus of greatest adaptation. The posterior probabilities suggest that inferences about the coupling parameters would lead us to the veridical architecture if we considered only connections whose half-life exceeded 4 s with 90% confidence or more.

The MAP estimates allow us to compute the MAP kernels associated with each region in terms of both neuronal output and hemodynamics response using Eq. (6). The neuronal and hemodynamic kernels for the three regions are shown in Fig. 10 (upper panels). It is interesting to note that the regional variation in the form of the neuronal kernels is sufficient to induce differential onset and peak latencies, in the order of a second or so, in the hemodynamic kernels despite the fact that neuronal onset latencies are the same. This difference in form is due to the network dynamics as activity is promulgated up the system and is recursively reentered into lower levels.

The neuronal kernels are simply a way of summarising the input–output behaviour of the model in terms of neuronal states. They can be regarded as a reparameterisation of the coupling parameters. They should not be taken as estimates of neuronal responses per se because the DCM is not really specified to that level of neurobiological finesse.

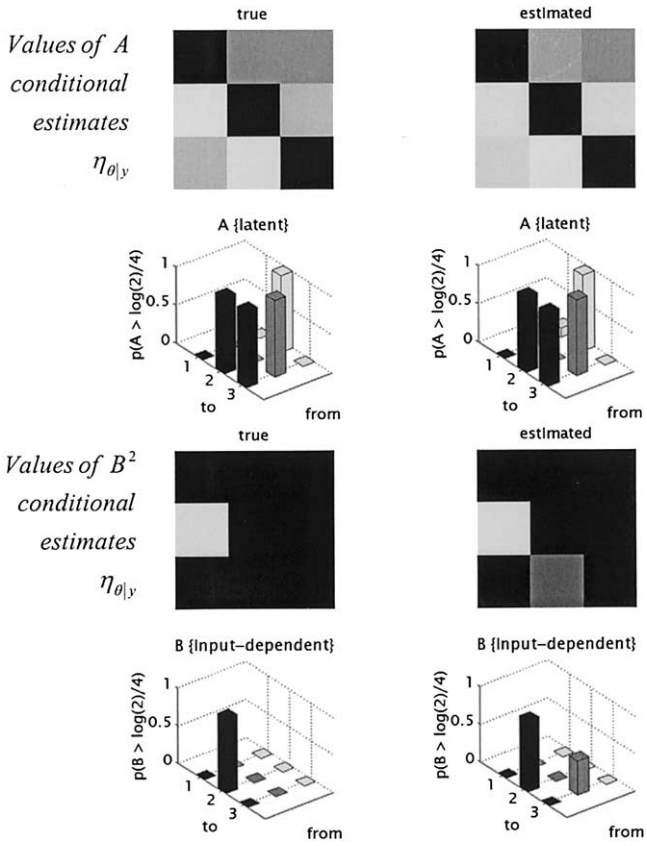


Fig. 9. Results summarising the conditional estimation based upon the simulated data of Fig. 7. The upper panels show the conditional estimates and posterior probabilities pertaining to the intrinsic coupling parameters. The lower panels show the equivalent results for bilinear coupling parameters mediating the effect of within-epoch time. Conditional or MAP estimates of the parameters are shown in image format with arbitrary scaling. The posterior probabilities that these parameters exceeded a threshold of In(2)/4 per second are shown as three-dimensional bar charts. True values and probabilities are shown on the left whereas the estimated values and posterior probabilities are shown on the right. This illustrates that the conditional estimates are a reasonable approximation to the true values and, in particular, the posterior probabilities conform to the true probabilities, if we consider values of 90% or more.

Notice also that the neuronal kernels are very protracted in relation to what one might expect to see using electrical recordings. This enduring activity, particularly in the higher two areas, is a product of the recurrent network dynamics and the rather slow time constant used in the simulations 1 s. The MAP estimates also enable us to compute the predicted response (Fig. 10, lower left panel) in each region and compare it to the true response without observation noise (Fig. 10, lower right panel). This comparison shows that the actual and predicted responses are very similar.

This estimation procedure was repeated for several series of simulated data sequences, described below, to obtain the conditional densities of the coupling parameters. By comparing the known values to these densities we were able to explore the face validity of the estimation scheme over a range of hyperparameters.

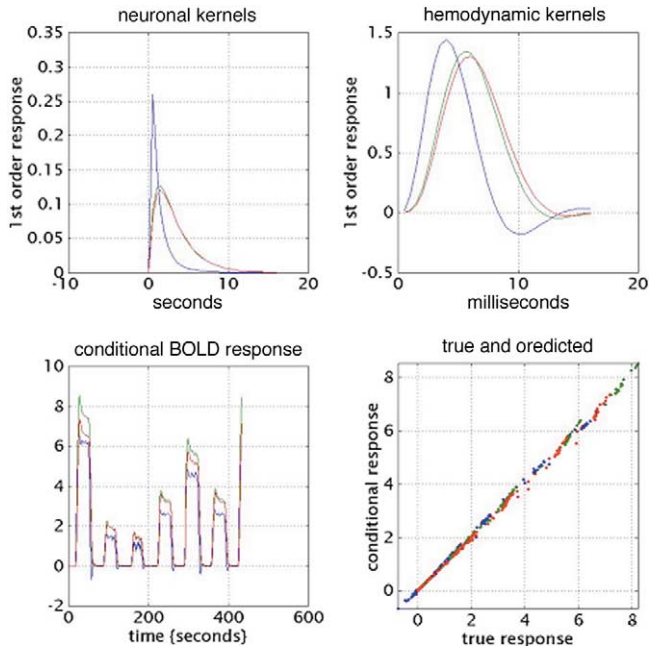


Fig. 10. These results are based upon the conditional or MAP estimates of the previous figure. The upper panels show the implied first-order kernels for neuronal responses (upper left) and equivalent hemodynamic responses (upper right) as a function of peristimulus time for each of the three regions. The lower panels show the predicted response based upon the MAP estimators and a comparison of this response to the true response. The agreement is self-evident.

3.2. Simulations

The aim of the work presented in this section was to explore the range of various hyperparameters, such as error variance, over which the estimation procedure gave useful results. The idea was to present the analysis with difficult situations to see when, and if, the procedure failed. We considered four important areas that, although not exhaustive, cover the most likely ways in which the validity of the connectivity estimates could be compromised. These four areas were, first, the level of noise or observation error and serial correlations among the errors. The second critical area was imprecise specification of the inputs in terms of their timing. This is particularly important in fMRI where multislice acquisition means that the timing of inputs, in relation to acquisition, will vary with brain region. This induces a necessary, region-specific, misspecification of input times (if they are assumed to be the same for all brain regions). The third area addressed deviations from our prior assumptions. The key parameters here are the biophysical parameters and the temporal scaling parameter or self-inhibition. These are the only parameters that have a non-zero prior expectation and real values that deviate substantially from prior expectations may affect veridical estimation of the normalised coupling parameters. Finally, we address the assumption, implicit in Eq. (4), that BOLD signals are detected with equal sensitivity throughout the brain. This is clearly not the case and calls for a simulation of region-specific dropout.

Each of these areas was investigated by simulating data over a range of hyperparameters. A valid estimation requires that the true value falls within appropriate (e.g., 90%) confidence intervals of the posterior density. Consequently, posterior expectations and confidence intervals were plotted as a function of each hyperparameter, with the true value superposed for comparison.

3.2.1. The effects of noise

In this subsection we investigate the sensitivity and specificity of posterior density estimates to the level and nature of observation noise. Data were simulated as described above and mixed with various levels of white noise. For each noise level the posterior densities of the coupling parameters were estimated and plotted against the noise hyperparameter (expressed as its standard deviation) in terms of the posterior mean and 90% confidence intervals. Fig. 11 shows some key coupling parameters that include both zero and nonzero connection strengths. The solid lines represent the posterior expectation or MAP estimator and the dashed lines indicate the true value. The grey areas encompass the 90% confidence regions. Characteristic behaviours of the estimation are apparent from these results. As one might intuit, increasing the level of noise increases the uncertainty in the posterior estimates as reflected by an increase in the conditional variance and a widening of the confidence intervals. This widening is, however, bounded by the prior variances to which the conditional variances asymptote, at very high levels of noise. Concomitant with this effect is “shrinkage” of some posterior means to their prior expectation of zero. Put simply, when the data become very noisy the estimation relies more heavily upon priors and the prior expectation is given more weight. This is why priors of the sort used here are referred to as “shrinkage priors.” These simulations suggest that for this level of evoked response, noise levels between 0 and 2 permit the connection strengths to be identified with a fair degree of precision and accuracy. Noise levels in typical fMRI experiments are about 0.5–1.5 (see next section). The units of signal and noise are adimensional and correspond to percentage whole brain mean. Pleasingly, noise did not lead to false inferences in the sense that the posterior densities always encompassed the true values even at high levels of noise (Fig. 11).

The same results are shown in a more compact form in Fig. 12a. Instead of showing the posterior densities of the coupling parameters, this figure shows the posterior densities of linear compounds or contrasts of coupling parameters. We have taken the average of coupling parameters that were zero and similarly the average of parameters that were nonzero. These two contrasts allow one to ascertain the specificity and sensitivity of the Bayesian inference, respectively (Fig. 12a, right and left panels). This was done for the intrinsic and bilinear parameters separately (Fig. 12a, upper

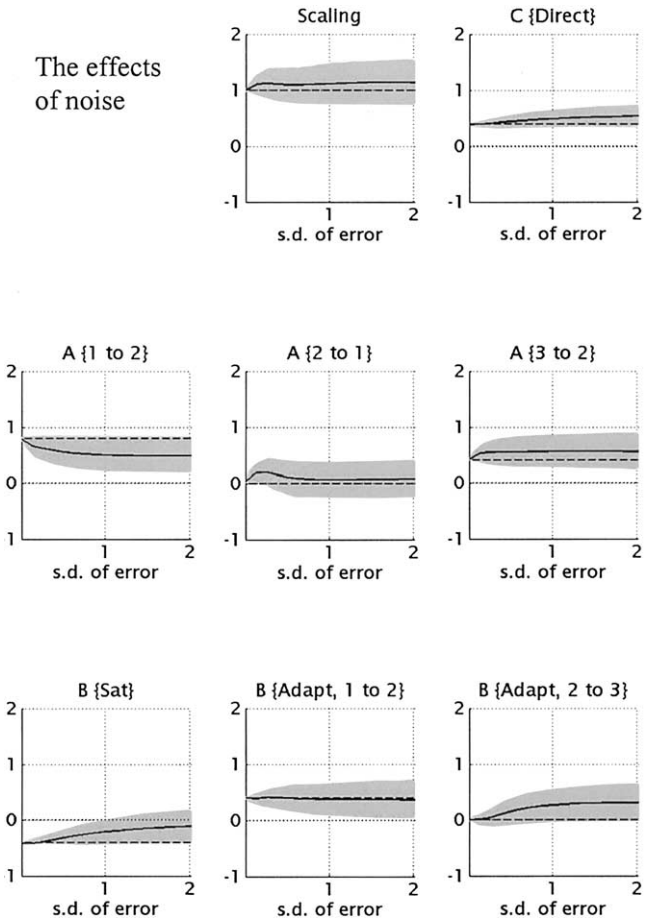


Fig. 11. Posterior densities as a function of noise levels: the analysis, summarised in the previous two figures, was repeated for simulated data sequences at different levels of noise ranging from 0 to 2 units of standard deviation. Each graph shows the conditional expectation or MAP estimate of a coupling parameter (solid line) and the 90% confidence region (grey region). The true value for each parameter is also shown (broken line). The top row shows the temporal scaling parameter and the extrinsic connection between the first input and the first area. The middle row shows some intrinsic coupling parameters and the bottom row bilinear parameters. As anticipated the conditional variance of these estimators increases with noise, as reflected by a divergence of the confidence region with increasing standard deviation of the error.

and lower panels), including only allowed connections. For zero connection strengths one would hope that the confidence intervals always contain the zero level. For connections that are present, one hopes to see the confidence interval either above or below the zero line and preferably encompassing the true value. Fig. 12a shows this to be the general case, although the 90% confidence region falls below the true values of the contrast for intrinsic connections at high levels of noise (Fig. 12a, upper left panel). To illustrate the role of priors in placing an upper bound on the conditional variance, the dark grey areas represent the confidence region in the absence of empirical evidence (i.e., prior variance) and the light grey areas represent the confidence region based upon the conditional variance, given the data. It can be seen that, at low levels of noise, the data are sufficiently precise to render our confidence about the estimates greater than that due to our prior assumptions. Conversely, at high levels of noise, observing the data does little to increase the precision of the estimates and the conditional variance approaches the prior variance. We will use this display format in subsequent sections because it is a parsimonious way of summarising the results.

The ReML hyperparameter estimates of the noise levels themselves are shown in Fig. 12b, over the 32 simulations comprising these results. The estimates are shown sepa-

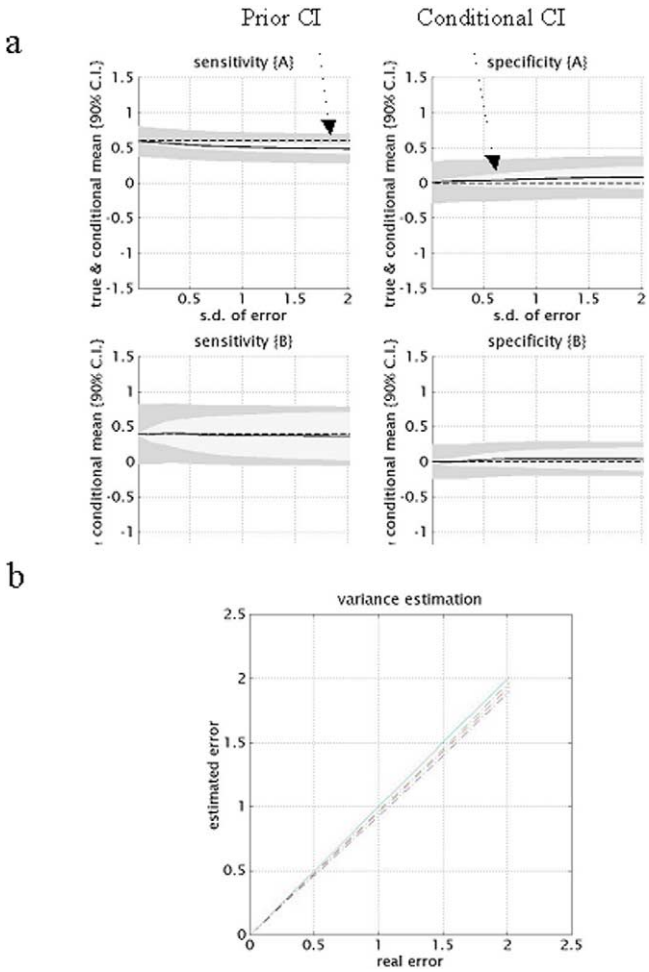


Fig. 12. Parameter and hyperparameter estimates: (a) The same results shown in Fig. 11 are shown here in a more compact form. In this format the conditional means and 90% confidence regions of parameter contrasts are displayed for the intrinsic parameters (upper panels) and the bilinear parameters (lower panels). The left panels show contrasts testing for the average of all nonzero connections and can be construed as an evaluation of sensitivity to detecting connections that are present. Conversely, the right panels show results for contrasts testing for zero connections (where they are allowed) and can be considered as an evaluation of specificity. The light grey regions correspond to confidence regions based upon the conditional variance for each contrast. This should be compared with the darker grey regions that are based upon the prior variance. The prior confidence bounds the conditional confidence as would be expected. (b) This shows the accuracy of ReML variance hyperparameter estimates by plotting them against their true values. The three broken lines correspond to the error variances in the three regions.

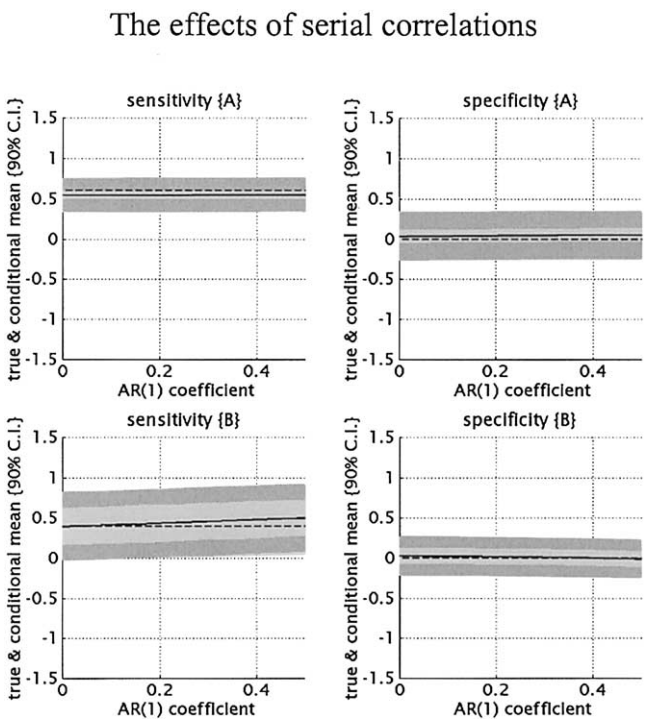


Fig. 13. The format of this figure is identical to Fig. 12a. The hyperparameter in this instance was the AR coefficient inducing serial correlations in the noise or observation error. It can be seen that the serial correlations have a minimal impact on the posterior density and ensuing conditional expectations.

rately for each of the three regions and correspond to the true values. Note that the hyperparameterisation of the error covariance accommodates different error variances in each region.

We next investigated the impact of serial correlations on the posterior density estimates. Although it is perfectly possible to estimate error variance hyperparameters controlling serial correlations of a specified form (see Friston et al., 2002), we deliberately assumed that there were no serial correlations to see how violation of this assumption corrupted the posterior estimates. Serial correlations were introduced according to an autoregressive process of first order AR(1) over a range of AR coefficients [0 to 0.5]. The results of these simulations and analyses are presented in Fig. 13 using the same format as Fig. 12. Perhaps surprisingly, serial correlations have a trivial effect on the posterior density estimates, suggesting that hyperparameters controlling the off-diagonal elements in the error variance covariance matrix do not need to be estimated. Theoretically, this is not important because the EM algorithm described in Section 2.2 can easily accommodate any number of hyperparameters. However, computationally, the presence of off-diagonal terms in the error covariance structure destroys the sparsity of the matrices rendering the computational times substantially greater (approximately by a factor of 2). The remaining simulations and empirical analyses therefore assumed the observation error was uncorrelated. For empirical data we generally use whitened data to ensure this assumption is not violated.

3.2.2. Misspecification of timing

In this subsection we explore robustness to misspecification of the inputs in terms of their timing. It might be thought that dynamic modelling of this sort would make the estimation very sensitive to timing errors; however, this is not necessarily the case. The information in the response variable is contained largely in the relative amplitudes and shapes of the hemodynamic responses and not their timings (compare the neuronal and hemodynamic kernels in Fig. 10). The utility of dynamic causal modelling is that this information can be used to estimate parameters of the model that implicitly specify timing relationships not otherwise observable in the data. The reason dynamic causal models can do this is because they have constraints on their architecture. In short, building knowledge into our estimation model allows the characterisation of data to be finessed in ways that may seem counterintuitive.

In these simulations we varied the time the response was downsampled to produce different sets of simulated data. This is equivalent to advancing or delaying the inputs in relation to the response. The results of this analysis are displayed in Fig. 14. The results show that the estimation procedure is robust in the range of $\pm$ a second. An intuition into this behaviour obtains by examining the MAP estimates of the temporal scaling parameter $\hat{\sigma}$ (Fig. 14, lower left). As the delay increases the responses appear to be premature, in relation to the specified input. This effect can be "absorbed" by the model by increasing the temporal scaling parameter and accelerating the network dynamics. Consequently as the delay hyperparameter increases so does the scaling parameter. However, the compensatory changes in temporal scaling are constrained by the priors so that the estimates of σ adopt a sigmoid shape to avoid extreme values. In concert with accelerated neuronal dynamics, the transit time decreases for all three areas. This hemodynamic compensation for mistiming the inputs should be compared with that in Fig. 15 in which the timing error were restricted to one region, as described next.

In the simulations above the delay was the same for all regions and was framed as a misspecification of input times. Even if the inputs are specified correctly, relative delays in sampling the response can still arise in multislice acquisition. To mimic the effect of sequential slice acquisition, we repeated the simulations using region-specific delays. We applied a delay of -1.5 to 1.5 s to, and only to, the response of the second region to produce the results in Fig. 15. Again the estimation seems relatively insensitive to delays of this sort. Here, the reason that timing errors do not produce inaccurate results is because the effects can be emulated by region-specific variations in delay of the hemodynamic response. Slight delays or advances in the sampling of the response are easily accommodated by changes in the biophysical parameters, to render the hemodynamic response

more delayed or acute. This can be seen in the lower right panel of Fig. 15. Here the temporal scaling parameter is less sensitive to delay, whereas the transit time for **A2** decreases to accommodate the apparently accelerated responses. In contradistinction the transit times for **A1** and **A3** increase to balance increases in temporal scaling (Fig. 15, lower left).

In summary, timing errors induced by sequential slice acquisition, or improper model specification, can be tolerated to within a second or so. This is acceptable for most fMRI studies with a short TR. In studies with a longer TR it might be necessary to "temporally realign" the data or restrict the system to proximate regions.

3.2.3. Violations of priors

In this subsection, we look at the impact of deviations from prior expectations in the hemodynamic or biophysical

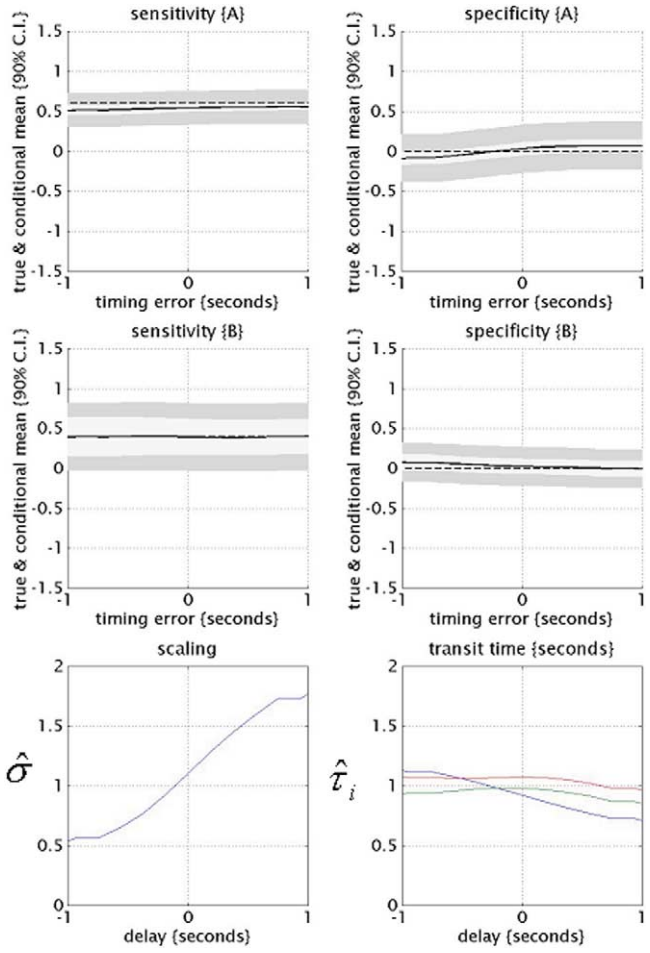


Fig. 14. As for Fig. 13. In this figure the hyperparameter varied was the timing error in input specification. In addition to the conditional estimates and confidence regions the lower two panels show the MAP estimates of the scaling parameter (lower left) and transit time for each region (lower right). These results are shown to illustrate that the estimation procedure accommodates timing errors by "speeding up" the dynamics through the temporal scaling parameter. This parameter shows an enormous variation over the delay hyperparameter from 0.5 to 1.5 per second. In contradistinction to the next figure, the transit times for each region behaved in a roughly similar fashion.

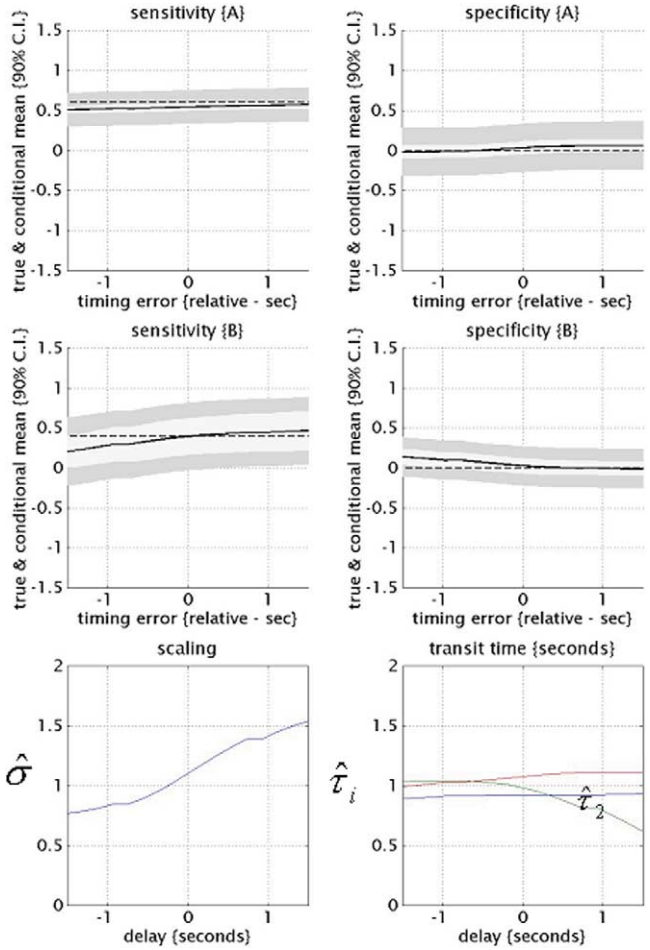


Fig. 15. As for Fig. 14, but in these simulations the timing error or delay was introduced at the level of the response in, and only in, the second region. The key difference is seen in the lower panels in which the temporal scaling parameter is relatively less affected whereas the transit time for **A2** falls from about 1 s to 0.6 s. These results, and those of the previous figure, illustrate how errors and delays in timing of the responses, in relation to the inputs, are accommodated by changes in the temporal scaling and biophysical parameters. This renders the estimates of the coupling parameters relatively unchanged (upper panels).

parameters and scaling parameter. The DCM parameters can be divided into those we are interested in, namely the normalised coupling parameters and those of less interest (the temporal scaling and biophysical parameters). It is important to establish that unexpected values of the latter do not compromise estimates of the former. We addressed this issue using two sets of simulations. First, we introduced a systematic region-specific variation in the hemodynamic parameters, to examine the impact of different hemodynamics over the brain. Second, we increase the temporal scaling well beyond its prior bounds.

In the first set of simulations biophysical deviates $\Delta\theta^h = \sqrt{C_\theta^h}Z$ were added to the prior expectation, where Z was a random normal variate. These deviates were scaled by a hyperparameter corresponding to the number of prior standard deviations. The regional variations in hemodynamics

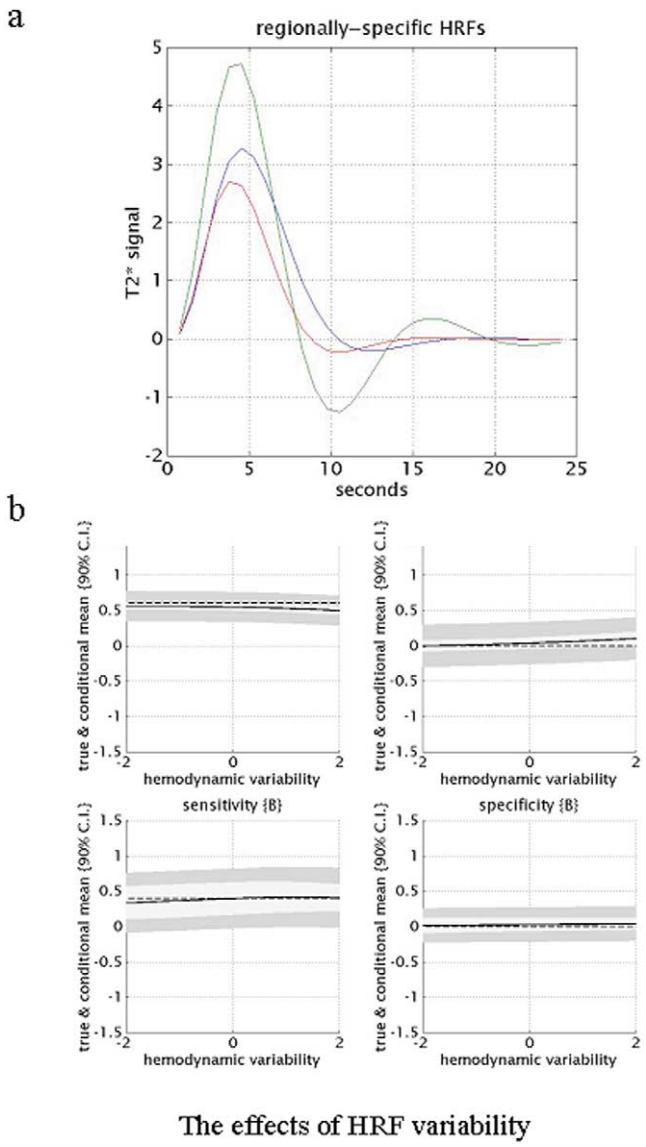


Fig. 16. Robustness to region-specific variation in the biophysical parameters. (a) Region-specific hemodynamic response functions (first-order kernels) when the deviations of the biophysical parameters are two standard deviations away from their prior expectation. (b) Conditional means and confidence regions adopting the same format as in the previous figure. The hyperparameter in this instance is the standard deviation from the prior expectation, experienced by region-specific hemodynamic or biophysical parameters. It is evident that the estimation procedure is robust, even with large degrees of regional variation in hemodynamic responses.

are characterised in Fig. 16a. These responses are the first-order hemodynamic kernels of the three regions, at the maximum of two prior standard deviations. The results of estimation are shown in Fig. 16b and demonstrate that regional variation in hemodynamics has a very small impact on estimates of the normalised coupling parameters. This is reassuring but expected, given that we allow for region-specific variability in the biophysical parameters.

In the second set of simulations we examined the effect of manipulating the scaling parameter σ that controls the characteristic time constants of neuronal transients. The prior expectation of this value was held constant at 1 per second and simulated data were generated using values between 1 and 3. A value of 3 per second corresponds to a half-life of evoked synaptic ensemble activity of $\ln 2/3 \approx 230$ ms. This is short, even for single-unit responses (Gerstein and Perkel, 1969).

The results of these analyses are shown in Fig. 17 and speak to an interesting phenomenon. Although the scaling parameter is estimated very poorly, the normalised couplings are estimated reasonably well, even with quite profound deviations from the prior expectation. This is a product of the way that we have parameterised the DCM and have specified the priors. Briefly, if there is a departure from the prior assumptions about the absolute coupling strengths, this deviation is expressed in the conditional estimates of the scaling parameter and not the normalised intrinsic coupling parameters. This is deliberate because we are inter-

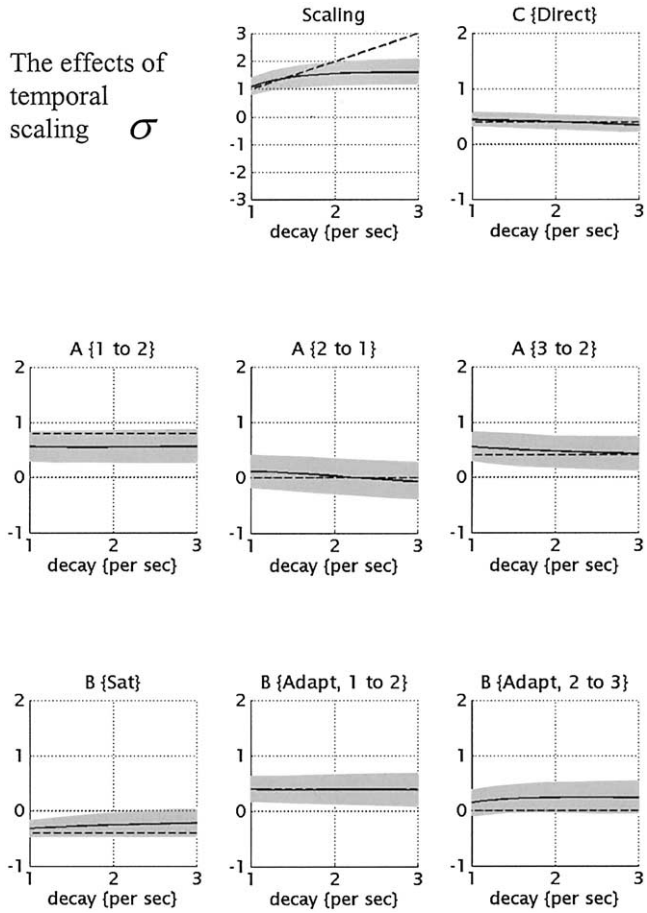


Fig. 17. The effects of variation in the temporal scaling parameter. The format of this figure follows that of Fig. 11. The hyperparameter in these simulations was the temporal scaling parameter σ that was varied between 1 and 3 per second. The key thing to take from these results is that unlikely excursions of the temporal scaling parameter from its prior expectation render its estimate unreliable (first panel). In contradistinction, the conditional estimates of the normalised coupling parameters are relatively unaffected and remain robust to this particular violation of prior assumptions.

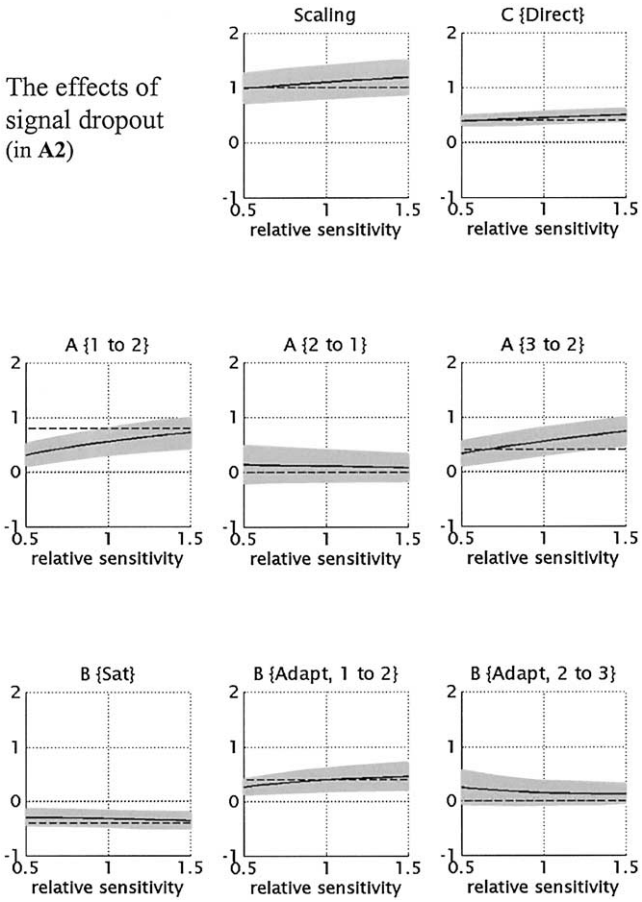


Fig. 18. These results are presented in the same format as Fig. 17. The hyperparameter in this instance was the relative sensitivity to BOLD changes in the second region. As anticipated, an increase in sensitivity to signal in **A2** causes an increase in the normalised connections to this area and a decrease in connections from it.

ested primarily in the relative strengths of effective connectivity among regions and not in their values per se. A conclusion from these analyses is that inferences about the scaling parameter are precluded unless we can be very sure the underlying neuronal dynamics conform closely to our prior beliefs. Estimates and inferences about the scaling parameter will be avoided in the rest of this paper.

3.2.4. Simulations of signal dropout

To simulate regionally specific variations in sensitivity to BOLD changes we simply scaled the response from the second area by a hyperparameter that ranged from 50 to 150%. Although this affects the data in exactly the same way as changing the amplitude of the hemodynamic response function (see Fig. 16), the results suggest a degree of caution should be exercised when assessing the conditional estimates quantitatively, in the context of substantial dropout. This is because the afferent coupling parameter (to the affected area) increases, roughly in proportion to the sensitivity with which its response is measured. Conversely, efferents (from the area) decrease. This effect is mitigated somewhat by compensatory changes in the hemodynamic parameters but these are constrained by priors and will not accommodate dropouts of 50%. The solution to this problem is to make the fixed parameters $V_0 = 0.02$ in the output Eq. (4) region-specific free parameters, with relatively uninformative priors. Because our analyses do not involve any regions subject to profound dropout we have treated V_0 as a fixed parameter in this paper.

In this section we hoped to establish the domains in which the estimates described in Section 2 can be usefully interpreted. We now turn to real data and address issues of reproducibility and predictive validity.

4. Predictive validity—an analysis of single word processing

4.1. Introduction

In this section we try to establish the predictive validity of DCM by showing that reproducible results can be obtained from independent data. The dataset we used was especially designed for these sorts of analyses, comprising over 1200 scans with a relatively short TR of 1.7 s. This necessitated a limited field of coverage but provided relatively high temporal acuity. The paradigm was a passive listening task, using epochs of single words presented at different rates. These data have been used previously to characterise nonlinear aspects of hemodynamics (e.g., Friston et al., 1998, 2000, 2002). Details of the experimental paradigm and acquisition parameters are provided in the legend to Fig. 19. These data were acquired in consecutive sessions of 120 scans enabling us to analyse the entire time series or each session independently. We first present the results obtained by concatenating all the sessions into a single data sequence. We then revisit the data, analysing each session independently to provide 10 independent conditional estimates of the coupling parameters to assess reproducibility and mutual predictability.

4.2. Analysis of the complete time series

Three regions were selected using maxima of the SPM{F} following a conventional SPM analysis (see Fig. 19). The three maxima were those that were closest to the primary and secondary auditory areas and Wernicke's area in accord with the anatomic designations provided in the atlas of Talairach and Tournoux (1988). Region-specific time series comprised the first eigenvariate of all voxels within a 4 mm radius sphere centred on each location. The anatomical locations are shown in Fig. 19. As in the simulations there were two inputs corresponding to a delta function for the occurrence of an aurally presented word and a parametric input modelling within-epoch adaptation. The outputs of the system were the three eigenvariate time series from each region. As in the previous section we allowed for

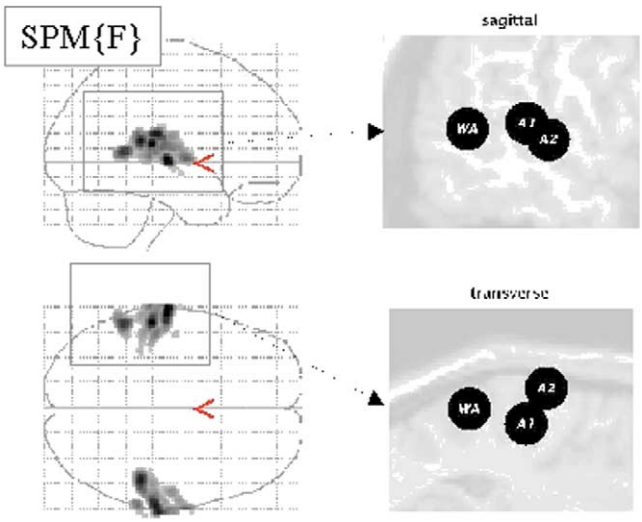


Fig. 19. Region selection for the empirical word processing example. Statistical parametric maps of the *F* ratio, based upon a conventional SPM analysis, are shown in the left panels and the spatial locations of the selected regions are shown on the right. These are superimposed on a T1-weighted reference image. The regional activities shown in Fig. 21 correspond to the first eigenvariates of a 4-mm-radius sphere centred on the following coordinates in the standard anatomical space of Talairach and Tournoux. Primary auditory area **A1,** −50, −26, 8 mm; secondary auditory area **A2,** −64, −18, 2 mm; and Wernicke's area **WA,** −56, −48, 6 mm. In brief, we obtained fMRI time series from a single subject at 2 Tesla using a Magnetom VISION (Siemens, Erlangen) whole-body MRI system, equipped with a head volume coil. Contiguous multislice T2*-weighted fMRI images were obtained with a gradient echo-planar sequence using an axial slice orientation (TE = 40 ms, TR = 1.7 s, 64 × 64 × 16 voxels). After discarding initial scans (to allow for magnetic saturation effects) each time series comprised 1200 volume images with 3 mm isotropic voxels. The subject listened to monosyllabic or bisyllabic concrete nouns (i.e., "dog," "radio," "mountain," "gate") presented at five different rates (10, 15, 30, 60, and 90 words per minute) for epochs of 34 s, intercalated with periods of rest. The five presentation rates were successively repeated according to a Latin Square design. The data were processed within SPM99 (Wellcome Department of Cognitive Neurology, http://www.fil.ion.ucl.ac.uk/spm). The time series were realigned, corrected for movement-related effects, and spatially normalised. The data were smoothed with a 5 mm isotropic Gaussian kernel. The SPM(F) above was based on a standard regression model using word presentation rate as the stimulus function and convolving it with a canonical hemodynamic response and its temporal derivative to form regressors.

a fully connected system. In other words, each region was potentially connected to every other region. Generally, one would impose constraints on highly unlikely or implausible connections by setting their prior variance to zero. However, we wanted to demonstrate that dynamic causal modelling can be applied to connectivity graphs that would be impossible to analyse with structural equation modelling. The auditory input was connected to **A1**. In addition, auditory input entered bilinearly to emulate saturation, as in the simulations. The contexual input, modelling putative adaptation, was allowed to exert influences over all intrinsic connections. From a neurobiological perspective an interesting question is whether plasticity can be demonstrated in forward connections or backward connections. Plasticity, in this instance, entails a time- dependent increase or decrease in effective connectivity and would be inferred by significant bilinear coupling parameters associated with the second input.

The inputs, outputs, and priors on the DCM parameters were entered into the Bayesian estimation procedure as described in Section 2.2. Drifts were modelled with the first 40 components of a discrete cosine set, corresponding to X in Eq. (8). The results of this analysis, in terms of the posterior densities and ensuing Bayesian inference, are presented in Fig. 20 and 21. Bayesian inferences were based upon the probability that the coupling parameters had a half-life of 8 s or less. Intuitively, this means that we only consider the influence of one region on another to be meaningfully large if this influence is expressed within a time frame of 8 s. The results show that the most probable architecture, given the inputs and data, conforms to a simple hierarchy of forward connections where **A1** influences **A2** and **WA,** whereas **A2** sends connections just to **WA** (Fig. 20). Although backward connections between **WA** and **A2** were estimated to be greater than our threshold with 82% confidence, they are not shown in Fig. 20 (which is restricted to posterior probabilities of 90% or more). Saturation could be inferred in **A1** and **WA** with a high degree of confidence with b^1_{11} and b^1_{33} being greater than 0.5. Significant plasticity or time-dependent changes were expressed predominantly in the forward connections, particularly that between **A1** and **A3**, i.e., $b^2_{13} = 0.37$. The conditional estimates are shown in more detail in Fig. 21 along with the conditional fitted responses and hemodynamic kernels. A full posterior density analysis for a particular contrast of effects is shown in Fig. 21a (lower panel). This contrast tested for the average plasticity over all forward connections and demonstrates that we can be virtually certain plasticity was greater than zero. The notion of a *contrast* of coupling parameter estimates is important because it allows one to make inferences about any linear combination of parameters. This includes differences in connection strengths, which might be important in demonstrating that one pathway is used more than another, or that backward influences are greater than forward connections. These inferences are based on the conditional density provided by Eq. (12) for any set of contrast weights c. This density affords the probability or confidence that the contrast (e.g., difference) is greater or less than zero.

This analysis illustrates three things. First, the DCM has defined a hierarchical architecture that is a sufficient explanation for the data and is indeed the most likely given the data. This hierarchical structure was not part of the prior constraints because we allowed for a fully connected system. Second, the significant bilinear effects of auditory stimulation suggest there is measurable neuronal saturation above and beyond that attributable to hemodynamic nonlinearities. This is quite important because such disambiguation is usually impossible given just hemodynamic responses. Finally, we were able to show time-dependent

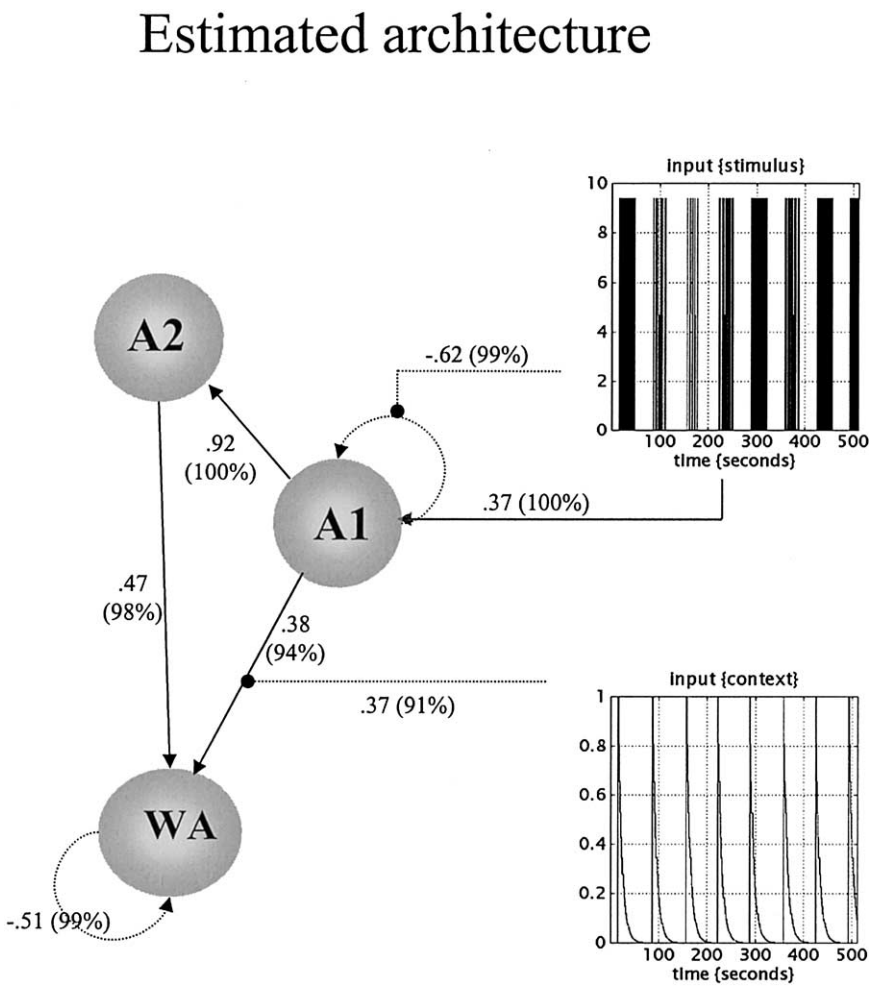


Fig. 20. Results of a DCM analysis applied to the data described in the previous figure. The display format follows that of Fig. 7. The coupling parameters are shown alongside the corresponding connections. The values in brackets are the percentage confidence that these values exceed a threshold of In(2)/8 per second.

decreases in effective connectivity in forward connections from **A1**. Although this experiment was not designed to test for plasticity, the usefulness of DCM, in studies of learning and priming, should be self-evident.

4.3. Reproducibility

The analysis above was repeated identically for each and every 120-scan session to provide 10 sets of Bayesian estimators. Drifts were modelled with the first four components of a discrete cosine set. The estimators are presented graphically in Fig. 22 and demonstrate extremely consistent results. In the upper panels the intrinsic connections are shown to be very similar in their profile, again reflecting a hierarchical connectivity architecture. The conditional means and 90% confidence regions for two connections are shown in Fig. 22a. These connections included the forward connection from **A1** and **A2** that is consistently estimated to be very strong. The backward connection from **WA** to **A2** was weaker but was certainly greater than zero in every analysis. Equivalent results were obtained for the modulatory effects or bilinear terms, although the profile was less consistent (Fig. 22b). However, the posterior density of the contrast testing for average time-dependent adaptation or plasticity is relatively consistent and again almost certainly greater than zero, in each analysis.

To illustrate the stability of hyperparameter estimates, over the 10 sessions, the standard deviations of observation error are presented for each session over the three areas in Fig. 23. Typical for studies at this field strength, the standard deviation of noise is about 0.8–1% whole-brain mean. It is pleasing to note that the session-to-session variability in hyperparameter estimates was relatively small, in relation to region-to-region differences.

In summary, independent analyses of data acquired under identical stimulus conditions, on the same subject, in the same scanning session, yield remarkably similar results. These results are biologically plausible and speak to the interesting notion that time-dependent changes, following the onset of a stream of words, are prominent in forward connections among auditory areas.

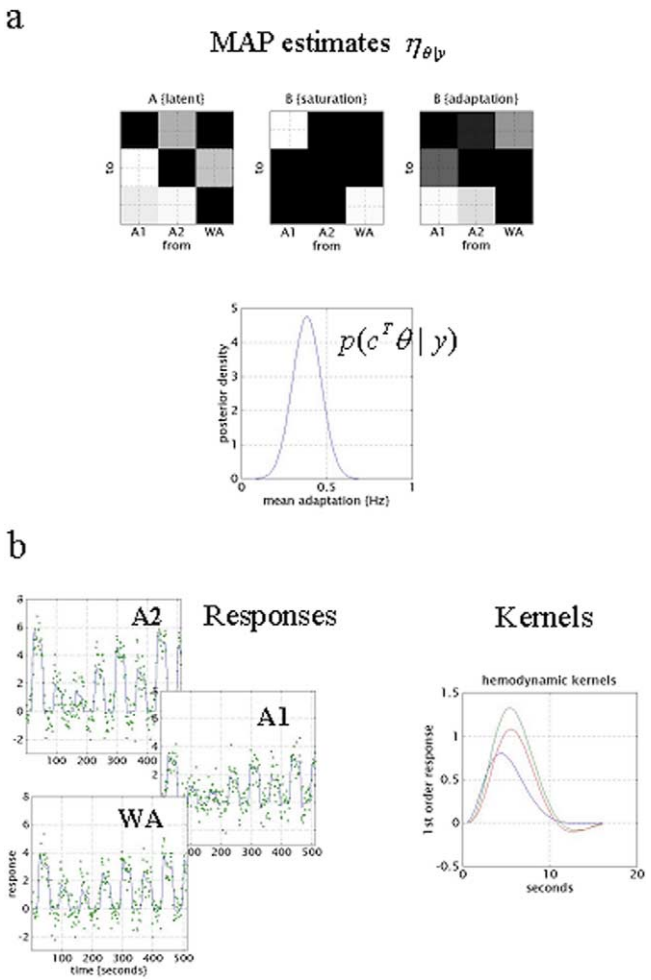


Fig. 21. This figure provides a more detailed characterisation of the conditional estimates. The images in the top row are the MAP estimates for the intrinsic and bilinear coupling parameters, pertaining to saturation and adaptation. The middle panel shows the posterior density of a contrast of all bilinear terms mediating adaptation, namely the modulation of intrinsic connections by the second time-dependent experimental effect. The predicted responses based upon the conditional estimators are shown for each of the three regions on the lower left (solid lines) with the original data (dots) after removal of confounds. A reparameterisation of the conditional estimates, in terms of the first-order hemodynamic kernels, is shown on the lower right.

5. Construct validity—an analysis of attentional effects on connections

5.1. Introduction

In this final section we take a first step towards establishing the construct validity of DCM. In a series of reports we have previously established that attention positively modulates the backward connections in a distributed system of cortical regions mediating attention to radial motion. In brief, subjects viewed optic flow stimuli comprising radially moving dots at a fixed velocity. In some epochs, subjects were asked to detect changes in velocity (that did not actually occur). This attentional manipulation was validated post hoc using psychophysics and the motion after-effect. Our previous analyses using structural equation modelling (Büchel and Friston, 1997) and a Volterra formulation of effective connectivity (Friston and Büchel, 2000) have established a hierarchical backwards modulation of effective connectivity where a higher area increases the effective connectivity among subordinate areas. These analyses have been extended using variable parameter regression and Kalman filtering (Büchel and Friston, 1998) to look at the effect of attention on interactions between **V5** and the posterior parietal complex. In this context, the Volterra formulation can be regarded as a highly finessed regression model that embodies nonlinear terms and some dynamic aspects of fMRI time series. However, even simple analyses, such as those employing psychophysiological interactions, point to the same conclusion that attention generally increases the effective connectivity among extrastriate and parietal areas. In short, we already have established that the superior posterior parietal cortex (**SPC**) exerts a modulatory role on **V5** responses using Volterra-based regression models (Friston and Büchel, 2000) and that the inferior frontal gyrus (**IFG**) exerts a similar influence on **SPC** using structural equation modelling (Büchel and Friston, 1997). The aim of this section was to show that DCM leads one to the same conclusions but starting from a completely different construct.

The experimental paradigm and data acquisition parameters are described in the legend to Fig. 24. This figure also shows the location of the regions that entered into the DCM (Fig. 24b, insert). Again, these regions were based on maxima from conventional SPMs testing for the effects of photic stimulation, motion, and attention. As in the previous section, regional time courses were taken as the first eigenvariate of spherical volumes of interest centred on the maxima shown in the figure. The inputs, in this example, comprise one sensory perturbation and two contextual inputs. The sensory input was simply the presence of photic stimulation and the first contextual one was presence of motion in the visual field. The second contextual input, encoding attentional set, was unity during attention to speed changes and zero otherwise. The outputs corresponded to the four regional eigenvariates in Fig. 24b. The intrinsic connections were constrained to conform to a hierarchical pattern in which each area was reciprocally connected to its supraordinate area. Photic stimulation entered at, and only at, **V1**. The effect of motion in the visual field was modelled as a bilinear modulation of the **V1** to **V5** connectivity and attention was allowed to modulate the backward connections from **IFG** and **SPC**.

The results of the DCM are shown in Fig. 24a. Of primary interest here is the modulatory effect of attention that is expressed in terms of the bilinear coupling parameters for this third input. As hoped, we can be highly confident that attention modulates the backward connections from **IFG** to **SPC** and from **SPC** to **V5**. Indeed, the influ-

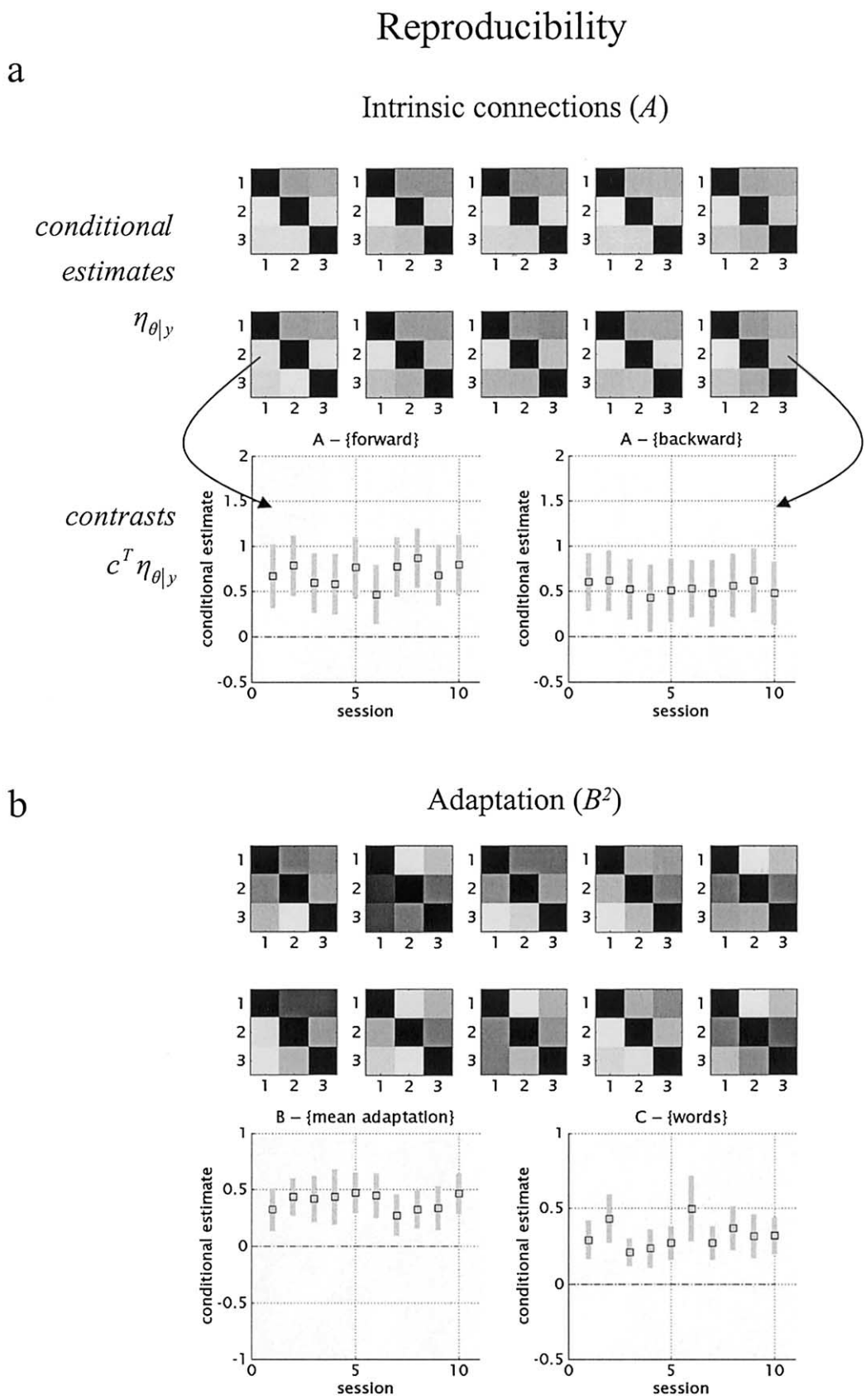


Fig. 22. Results of the reproducibility analyses. (a) Results for the intrinsic parameters. The profile of conditional estimates for the 10 independent analyses described in the text are shown in image format, all scaled to the maximum. The posterior densities, upon which these estimates are based, are shown for two selected connections in the lower two graphs. These densities are displayed in terms of their expectation and 90% confidence intervals (grey bars) for

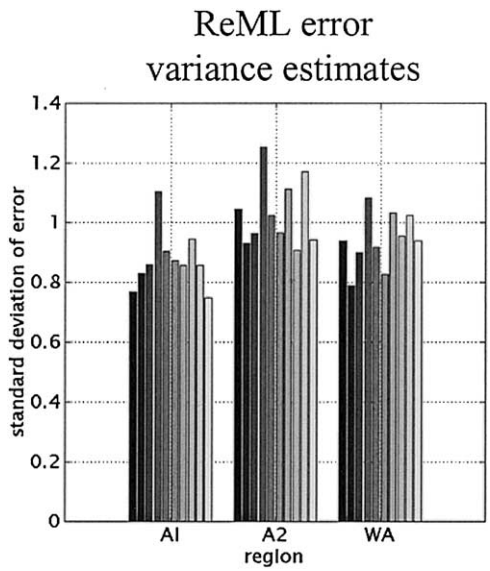


Fig. 23. ReML hyperparameter variance estimates for each region and analysis. These estimates provide an anecdotal characterisation of the within and between-area variability, in hyperparameter estimates, and show that they generally lie between 0.8 and 1 (adimensional units corresponding to percentage whole-brain mean).

ences of **IFG** on **SPC** are negligible in the absence of attention (dotted connection in Fig. 24a). It is important to note that the only way that attentional manipulation can affect brain responses was through this bilinear effect. Attention-related responses are seen throughout the system (attention epochs are marked with arrows in the plot of **IFG** responses in Fig. 24b). This attentional modulation is accounted for, sufficiently, by changing just two connections. This change is, presumably, instantiated by an instructional set at the beginning of each epoch. The second thing this analysis illustrates is how the functional segregation is modelled in DCM. Here one can regard **V1** as a "segregating" motion from other visual information and distributing it to the motion-sensitive area **V5**. This segregation is modelled as a bilinear "enabling" of **V1** to **V5** connections when, and only when, motion is present. Note that in the absence of motion the intrinsic **V1** to **V5** connection was trivially small (in fact the MAP estimate was -0.04). The key advantage of entering motion through a bilinear effect, as opposed to a direct effect on **V5,** is that we can finesse the inference that **V5** shows motion-selective responses with the assertion that these responses are mediated by afferents from **V1**.

The two bilinear effects above represent two important aspects of functional integration that DCM was designed to characterise.

6. Conclusion

In this paper we have presented dynamic causal modelling. DCM is a causal modelling procedure for dynamic systems in which causality is inherent in the differential equations that specify the model. The basic idea is to treat the system of interest, in this case the brain, as an input–state–output system. By perturbing the system with known inputs, measured responses are used to estimate various parameters that govern the evolution of brain states. Although there are no restrictions on the parameterisation of the model, a bilinear approximation affords a simple reparameterisation in terms of effective connectivity. This effective connectivity can be latent or intrinsic or, through bilinear terms, model input-dependent changes in effective connectivity. Parameter estimation proceeds using fairly standard approaches to system identification that rest upon Bayesian inference.

Dynamic causal modelling represents a fundamental departure from conventional approaches to modelling effective connectivity in neuroscience. The critical distinction between DCM and other approaches, such as structural equation modelling or multivariate autoregressive techniques, is that the input is treated as known, as opposed to stochastic. In this sense DCM is much closer to conventional analyses of neuroimaging time series because the causal or explanatory variables enter as known fixed quantities. The use of designed and known inputs in characterising neuroimaging data with the general linear model or DCM is a more natural way to analyse data from designed experiments. Given that the vast majority of imaging neuroscience relies upon designed experiments, we consider DCM a potentially useful complement to existing techniques. In the remainder of this section we consider two potential limitations of DCM and comment upon extensions.

6.1. Priors

One potential weakness of any Bayesian information procedure is its dependence upon priors. In other words, the inferences provided by DCM are only as valid as the priors used in the estimation procedure. This is not a severe limitation because the parameters about which inferences are made can be constrained by relatively uninformative priors. Although more stringent priors are applied to the hemodynamic biophysical parameters their posterior density is of no interest. In relation to the coupling parameters only the intrinsic and bilinear parameters have informative shrinkage

the forward connection from **A1** to **A2**. The equivalent densities are shown for the backward connection from **WA** to **A2**. Although the posterior probability that the latter connections exceeded the specified threshold was less than 90%, it can be seen that this connection is almost certainly greater than zero. (b) Equivalent results for the bilinear coupling matrices mediating adaptation. The lower panels here refer to the posterior densities of a contrast testing for the mean of all bilinear parameters (left) and the extrinsic connection to **A1** (right).

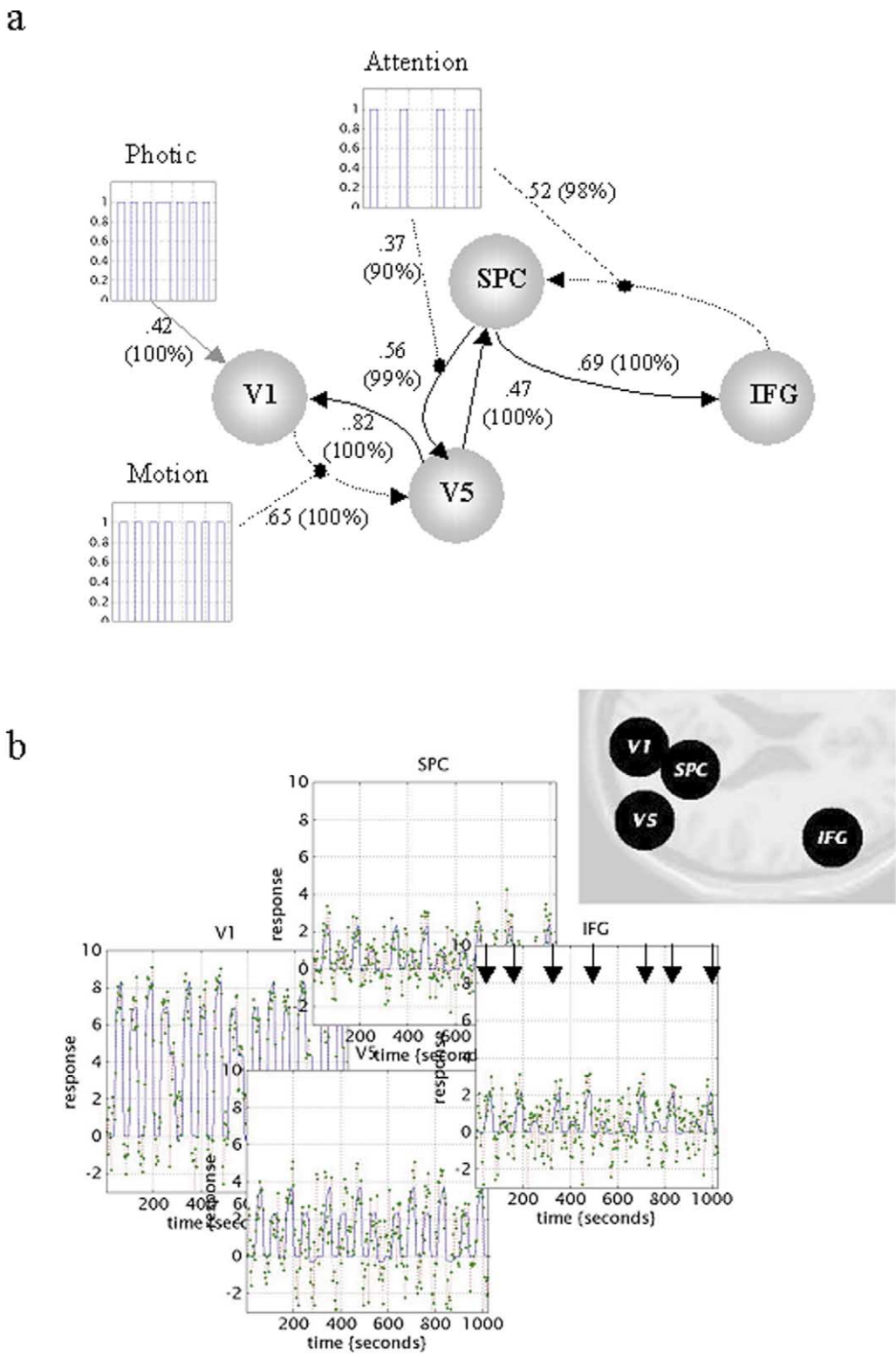


Fig. 24. Results of the empirical analysis of the attention study. (a) Functional architecture based upon the conditional estimates displayed using the same format as Fig. 20. The most interesting aspects of this architecture involved the role of motion and attention in exerting bilinear effects. Critically, the influence of motion is to enable connections from **V1** to the motion-sensitive area **V5**. The influence of attention is to enable backward connections from the inferior frontal gyrus **(IFG)** to the superior parietal cortex **(SPC)**. Furthermore, attention increases the latent influence of **SPC** on the **V5**. Dotted arrows connecting regions represent significant bilinear affects in the absence of a significant intrinsic coupling. (b) Fitted responses based upon the conditional estimates and the adjusted data are shown using the same format as in Fig. 21. The insert shows the location of the regions, again adopting the same format in previous figures. The location of these regions centred on the primary visual cortex **V1,** 6, −84, −6 mm; motion-sensitive area **V5,** 45, −81, 5 mm; superior parietal cortex, **SPC,** 18, −57, 66 mm; inferior frontal gyrus, **IFG,** 54, 18, 30 mm. The volumes from which the first eigenvariates were calculated

priors that are easily motivated by the tenable assumption that neural activity will not diverge exponentially.

6.2. Deterministic noise

The dynamic causal models considered in this paper are based upon differential equations that do not admit deterministic noise in the state equation. This means that we do not allow for noisy dynamics (or deterministic dynamics with a very high correlation dimension) to be expressed endogenously in each region. The only stochastic component of the model enters linearly as observation noise at the point the response is measured. From a neuronal perspective this is clearly a limitation because regional dynamics will not be dictated solely by designed inputs. Perhaps the best way of accommodating this, and other departures from complete biological plausibility, is to acknowledge that DCMs are only models. They do not propose to capture actual biological processes but model them in a rather abstract way that finesses the neurophysiological interpretation of the model parameters. In other words, if real neuronal dynamics could be summarised as a dynamic causal model then the parameters of that model are the best ones given the data. This does not imply that the brain actually works in the same way as the model does. For example, it is physiologically unlikely that the neural activity in every neuronal assembly within a cortical area conforms to the dynamics of a single-state variable. However, the "effective" state implied by the model may be a useful abstraction. In summary, it should be remembered that DCMs have the same status as general linear models in trying to summarise and characterise the impact of experimental perturbations on measured responses. DCMs are much closer to the underlying system in the sense that they embody dynamic aspects of neurophysiology, interactions among different brain areas, and nonlinearities, but they are still simply observation models.

6.3. Future extensions

The obvious extension of dynamic causal models is in terms of their neurophysiological plausibility. We are currently working on mean-field approximations for coupled ensembles of neurons as a more refined model of intraregional dynamics. This would enable a simple extension from a single-state variable for each region to multiple-state variables. For example, the states might include the mean activity of inhibitory and excitatory subpopulations within an area, or indeed a series of excitatory-inhibitory couples. This work is motivated by the fact that effective connectivity expressed in terms of univariate metrics, e.g., regional activity, cannot be meaningfully linked to specific neurotransmitter systems. By introducing the distinction between inhibitory and excitatory subpopulations it would be possible to model separately inhibitory (within areas) and excitatory connections (within and between areas). This would allow one to harness the empirical knowledge that most corticocortical connections are excitatory and mediated by glutamate.

A further extension would be to go beyond bilinear approximations to allow for interactions among the states. This is important when trying to model modulatory or nonlinear connections such as those mediated by backward afferents that terminate predominantly in the supragranular layers and possibly on NMDA receptors. It is evident that, as dynamic causal models become more sophisticated, they will become indistinguishable from synthetic neuronal models used in computational neurobiology. However, it is likely that the increase in the number of parameters required to define a particular DCM will necessitate the use of other imaging modalities such as EEG and MEG. The use of DCMs as forward models for the fusion or integration of multimodality data is another exciting possibility.

Acknowledgments

The Wellcome Trust supported this work. We thank Marcia Bennett for preparing the manuscript.

Software implementation note

The theory and estimation procedures described in this paper have been implemented in the SPM2 version of the statistical parametric mapping software (http://www.fil.ion.ucl.ac.uk/spm). Following a conventional analysis, a library of volumes of interest (VOI) structures can be as-

corresponded to 8-mm-radius spheres centred on these locations. Subjects were studied with fMRI under identical stimulus conditions (visual motion subtended by radially moving dots) whilst manipulating the attentional component of the task (detection of velocity changes). The data were acquired from normal subjects at 2 Tesla using a Magnetom VISION (Siemens, Erlangen) whole-body MRI system, equipped with a head volume coil. Here we analyse data from the first subject. Contiguous multi-slice T2*-weighted fMRI images were obtained with a gradient echo-planar sequence (TE = 40 ms, TR = 3.22 s, matrix size = 64 × 64 × 32, voxel size 3 × 3 × 3 mm). Each subject had four consecutive 100-scan sessions comprising a series of 10-scan blocks under five different conditions D F A F N F A F N S. The first condition (D) was a dummy condition to allow for magnetic saturation effects. F (Fixation) corresponds to a low-level baseline where the subjects viewed a fixation point at the centre of a screen. In condition A (Attention) subjects viewed 250 dots moving radially from the centre at 4.7° per second and were asked to detect changes in radial velocity. In condition N (No attention) the subjects were asked simply to view the moving dots. In condition S (Stationary) subjects viewed stationary dots. The order of A and N was swapped for the last two sessions. In all conditions subjects fixated on the centre of the screen. In a prescanning session the subjects were given five trials with five speed changes (reducing to 1%). During scanning there were no speed changes. No overt response was required in any condition.

sembled, usually based on maxima in the SPM{T} or SPM{F}. These VOI structures contain information about the original data, analysis and, critically, the region's first eigenvariate. Selecting from this list specifies regional outputs or responses. The user interface then requests constraints on the connections (in terms of which connections are allowed). Inputs are selected from the stimulus functions, originally specified (in terms of onsets and durations) to form the conventional design matrix. Estimation then proceeds automatically and the results are stored for inspection. In this implementation DCM uses exactly the same stimulus functions (inputs), confounds, and anatomical frame of reference as the conventional analysis that precedes it. This enforces a perspective on experimental design and analysis that reflects DCM as a generalisation of conventional analyses.

Appendix A.1: the relationship between dynamic and static causal models

Consider a linear DCM where we observe the states directly and there is only one state variable per region. From Eq. (2):

$$\begin{aligned}\dot{z} &= F(z, u, \theta)\\ &= Az + u\\ &= (\theta - 1)z + u.\end{aligned} \tag{A.1}$$

Here we have discounted observation error but make the inputs $u \sim N(0, Q)$ stochastic. To make the connection to structural equation models more explicit, we have expanded the intrinsic connections into off-diagonal connections and a leading diagonal matrix, modelling unit decay $A = \theta - 1$. For simplicity, we have absorbed C into the covariance structure of the inputs Q. If inputs change slowly, relative to the dynamics, the change in states will be zero at the point of observation and we obtain the regression model.

$$\begin{aligned}\dot{z} = 0 &\Rightarrow\\ (1 - \theta)z &= u\\ z &= \theta z + u.\end{aligned} \tag{A.2}$$

This special case of Eq. (2) is important because it is the basis of commonly employed methods for estimating effective connectivity in neuroimaging (e.g., SEM). Dynamic casual models do not assume the states have reach equilibrium at the point of observation. That is why they are dynamic.

Appendix A.2: parameter and hyperparameter estimation with EM

In this appendix we provide a heuristic motivation for the **E**- and **M**-steps of the estimation scheme summarised in Eq. (8). These steps can be regarded on as Fisher Scoring ascent on an objective function F that embodies the log posterior.

The E-step

The conditional expectations and covariances of the parameters are estimated in the **E**-step that performs a gradient ascent on the log posterior comprising the likelihood and prior potentials

$$\begin{aligned}l &= \ln p(\theta|y, \lambda; u)\\ &= \ln p(y|\theta, \lambda; u) + \ln p(\theta; u)\\ \ln p(y|\theta, \lambda; u) &= -\frac{1}{2}(y - h(u, \theta))^T C_\varepsilon^{-1}(y - h(u, \theta)\\ \ln p(\theta; u) &= -\frac{1}{2}(\eta_\theta - \theta)^T C_\theta^{-1}(\eta_\theta - \theta).\end{aligned} \tag{A.3}$$

On taking gradients with respect to the parameters the following Fisher scoring scheme ensues.

$$\begin{aligned}\eta_{\theta|y} &\leftarrow \eta_{\theta|y} - \left\langle \frac{\partial^2 l}{\partial \theta^2} \right\rangle^{-1} \frac{\partial l}{\partial \theta}(\eta_{\theta|y})\\ \frac{\partial l}{\partial \theta} &= J^T C_\varepsilon^{-1} r + C_\theta^{-1}(\eta_\theta - \eta_{\theta|y})\\ -\left\langle \frac{\partial^2 l}{\partial \theta^2} \right\rangle &= J^T C_\varepsilon^{-1} J + C_\theta^{-1} = C_{\theta|y}^{-1},\end{aligned} \tag{A.4}$$

where $J = \partial h(\eta_{\theta|y})/\partial\theta$, $r = y - h(u, \eta_{\theta|y})$, and $C_\varepsilon = \Sigma\lambda_i Q_i$ is the hyperparameterised error covariance. Eq. A.4 is formally the same as the **E**-step in Eq. (8), after the nuisance variables X have been included.

The M-step

The hyperparameters are estimated in the **M**-step in exactly the same way as the parameters but accounting for the fact that the log likelihood depends on the unknown parameters by integrating them out using the approximate conditional distribution $q(\theta)$. Note there are no priors on the hyperparameters. This integration motivates a lower bound on the log likelihood called the [negative] free energy in statistical physics (Neal and Hinton, 1998). By Jensen's inequality

$$\begin{aligned}\ln p(y|\lambda;u) &= \ln \int q(\theta)\frac{p(\theta, y|\lambda;u)}{q(\theta)}d\theta \geq\\ F &= \int q(\theta)\ln\frac{p(\theta, y|\lambda;u)}{q(\theta)}d\theta.\end{aligned} \tag{A.5}$$

On taking gradients with respect to the hyperparameters, the following Fisher scoring scheme can be derived.

$$\lambda \leftarrow \lambda - \left\langle \frac{\partial^2 F}{\partial \lambda^2} \right\rangle^{-1} \frac{\partial F}{\partial \lambda}(\lambda)$$

$$\frac{\partial F}{\partial \lambda_i} = \frac{1}{2} tr\{PQ_i\} - \frac{1}{2} r^T P^T Q_i P r$$

$$-\left\langle \frac{\partial^2 F}{\partial \lambda^2} \right\rangle_{ij} = \frac{1}{2} tr\{PQ_i PQ_j\}, \tag{A.6}$$

where $P = C_\varepsilon^{-1} - C_\varepsilon^{-1} J C_{\theta|y}^{-1} J^T C_\varepsilon^{-1}$. The parameter ascent on the log posterior l in the **E**-step is closely related to an ascent on the negative free energy F used for the hyperparameters in the **M**-step, with exact equivalence when $q(\theta)$ is deterministic. This can be seen if we write

$$F = \int q(\theta) \ln p(y|\theta, \lambda; u) d\theta - \int q(\theta) \ln \frac{q(\theta)}{p(\theta)} d\theta$$

$$= \langle \ln p(y|\theta, \lambda; u) \rangle_q - KL(q(\theta), p(\theta)). \tag{A.7}$$

F comprises the expected log likelihood under $q(\theta)$ and a prior term embodying the Kullback–Leibler (KL) divergence between the conditional and prior densities. $F = l$ when $q(\theta)$ shrinks to a point density over $\eta_{\theta|y}$. For completeness, it is noted that, in a linear setting, F is also the ReML (restricted maximum likelihood) objective function used in classical variance component estimation (Harville, 1977). This EM algorithm is simple and robust and has found multiple applications in our data analysis, ranging from ReML estimates of serial correlations in fMRI to hyperparameter estimation in hierarchical observation models using empirical Bayes; see Friston et al. (2002) for details. In our implementation we iterate the **E**- and **M**-steps until convergence before recomputing $J = \partial h(\eta_{\theta|y})/\partial \theta$.

Appendix A.3: priors on the coupling parameters

Consider any set of $l(l-1)$ interregional connections a_{ij}; $i \neq j$ with sum of squared values $\zeta = \Sigma a_{ij}^2$. For any given value of ζ the biggest principal Lyapunov exponent λ_a obtains when the strengths are equal $a_{ij} = a$, in which case

$$\lambda^a = (l-1)a - 1$$

$$\zeta = l(l-1)a^2. \tag{A.8}$$

This means that as the sum of squared connection strengths reaches $l/(l-1)$, the largest exponent attainable approaches zero. Consequently, if ζ is constrained to be less than this threshold, we can set an upper bound on the probability that the principal exponent exceeds zero. ζ is constrained through the priors on a_{ij}. If each connection has a prior Gaussian density with zero expectation and variance v_a, then the sum of squares has a scaled χ^2 distribution $\zeta/v_a \sim \chi_l^2(l-1)$ with degrees of freedom $l(l-1)$. v_a is chosen to make $p(\zeta > l/(l-1))$ suitably small, i.e.

$$v_a = \frac{l(l-1)}{\phi_\chi^{-1}(1-p)}, \tag{A.9}$$

where ϕ_χ is the cumulative $\chi_{l(l-1)}^2$ distribution and p is the required probability. As the number of regions increases, the prior variance decreases.

In addition to constraints on the normalised connections, the factorisation in Eq. (9) requires the temporal scaling σ to be greater than zero. This is achieved through a noncentral prior density specified in terms of its moments such that $\sigma \sim N(\eta_\sigma, v_\sigma)$ where the expectation η_σ controls the characteristic time constant of the system and the variance v_σ is chosen to ensure $p(\sigma > 0)$ is small, i.e.,

$$v_\sigma = \left(\frac{\eta_\sigma}{\phi_N^{-1}(1-p)} \right)^2, \tag{A.10}$$

where ϕ_N is the cumulative normal distribution and p the required probability.

References

Bendat, J.S., 1990. Nonlinear System Analysis and Identification from Random Data. Wiley, New York.

Büchel, C., Friston, K.J., 1997. Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modelling and fMRI. Cereb. Cortex 7, 768–778.

Büchel, C., Friston, K.J., 1998. Dynamic changes in effective connectivity characterised by variable parameter regression and Kalman filtering. Hum. Brain Mapp 6, 403–408.

Buxton, R.B., Wong, E.C., Frank, L.R., 1998. Dynamics of blood flow and oxygenation changes during brain activation: the Balloon model. MRM 39, 855–864.

Friston, K.J., Büchel, C., Fink, G.R., Morris, J., Rolls, E., Dolan, R.J., 1997. Psychophysiological and modulatory interactions in neuroimaging. NeuroImage 6, 218–229.

Friston, K.J., Josephs, O., Rees, G., Turner, R., 1998. Nonlinear event-related responses in fMRI. MRM 39, 41–52.

Friston, K.J., Büchel, C., 2000. Attentional modulation of effective connectivity from V2 to V5/MT in humans. Proc. Natl. Acad. Sci. USA 97, 7591–7596.

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, 466–477.

Friston, K.J., 2002. Bayesian estimation of dynamical systems: an application to fMRI. NeuroImage 16, 513–530.

Friston, K.J., Penny, W., Phillips, C., Kiebel, S., Hinton, G., Ashburner, J., 2002. Classical and Bayesian inference in neuroimaging: theory. NeuroImage 16, 465–483.

Gerstein, G.L., Perkel, D.H., 1969. Simultaneously recorded trains of action potentials: analysis and functional interpretation. Science 164, 828–830.

Grubb, R.L., Rachael, M.E., Euchring, J.O., Ter-Pogossian, M.M., 1974. The effects of changes in PCO2 on cerebral blood volume, blood flow and vascular mean transit time. Stroke 5, 630–639.

Harrison, L.M., Penny, W., Friston, K.J., 2003. Multivariate autoregressive modelling of fMRI time series. NeuroImage, in press.

Harville, D.A., 1977. Maximum likelihood approaches to variance component estimation and to related problems. J. Am. Stat. Assoc 72, 320–338.

Horwitz, B., Friston, K.J., Taylor, J.G., 2001. Neural modeling and functional brain imaging: an overview. Neural Networks 13, 829–846.

Kenny, D.A., Judd, C.M., 1984. Estimating nonlinear and interactive effects of latent variables. Psychol. Bull 96, 201–210.

Mandeville, J.B., Marota, J.J., Ayata, C., Zararchuk, G., Moskowitz, M.A., Rosen, B., Weisskoff, R.M., 1999. Evidence of a cerebrovascular postarteriole Windkessel with delayed compliance. J. Cereb. Blood Flow Metab 19, 679–689.

McIntosh, A.R., Gonzalez-Lima, F., 1994. Structural equation modelling and its application to network analysis in functional brain imaging. Hum. Brain Mapp 2, 2–22.

McIntosh, A.R., 2000. Towards a network theory of cognition. Neural Networks 13, 861–870.

Neal, R.M., Hinton, G.E., 1998. A view of the EM algorithm that justifies incremental, sparse and other variants, in: Jordan, M.I. (Ed.), Learning in Graphical Models. Kluwer Academic, pp. 355–368.

Talairach, J., Tournoux, P., 1988. A Co-planar Stereotaxic Atlas of a Human Brain. Thieme, Stuttgart.