

Dynamic causal modelling for fMRI: A two-state model

A.C. Marreiros,* S.J. Kiebel, and K.J. Friston

Wellcome Trust Centre for Neuroimaging, Institute of Neurology, UCL, 12 Queen Square, London WC1N 3BG, UK

Received 14 March 2007; revised 10 July 2007; accepted 6 August 2007
Available online 25 August 2007

Dynamical causal modelling (DCM) for functional magnetic resonance imaging (fMRI) is a technique to infer directed connectivity among brain regions. These models distinguish between a neuronal level, which models neuronal interactions among regions, and an observation level, which models the hemodynamic responses each region. The original DCM formulation considered only one neuronal state per region. In this work, we adopt a more plausible and less constrained neuronal model, using two neuronal states (populations) per region. Critically, this gives us an explicit model of intrinsic (between-population) connectivity within a region. In addition, by using positivity constraints, the model conforms to the organization of real cortical hierarchies, whose extrinsic connections are excitatory (glutamatergic). By incorporating two populations within each region we can model selective changes in both extrinsic and intrinsic connectivity.

Using synthetic data, we show that the two-state model is internal consistent and identifiable. We then apply the model to real data, explicitly modelling intrinsic connections. Using model comparison, we found that the two-state model is better than the single-state model. Furthermore, using the two-state model we find that it is possible to disambiguate between subtle changes in coupling; we were able to show that attentional gain, in the context of visual motion processing, is accounted for sufficiently by an increased sensitivity of excitatory populations of neurons in V5, to forward afferents from earlier visual areas.

© 2007 Elsevier Inc. All rights reserved.

Keywords: Functional magnetic resonance imaging; Dynamic causal modelling; Neural mass model

Introduction

Dynamic causal modelling (DCM) for fMRI is a natural extension of the convolution models used in the standard analysis of fMRI (Friston et al., 2003). This extension involves the explicit modelling of activity within and among regions of a

hypothesized network, at the neuronal level. The general idea behind DCM is to construct a reasonably realistic neuronal model of interacting cortical regions with neurophysiologically inspired parameters. These parameters are estimated such that the predicted blood oxygenation level dependent (BOLD) series, which results from converting the neural dynamics into hemodynamics, correspond as closely as possible to the observed BOLD series.

Standard DCMs for fMRI are based upon a bilinear approximation to neuronal dynamics with one state per region. The neuronal dynamics are described by the differential equations describing the dynamics of a single state that summarizes the neuronal or synaptic activity of each area; this activity then induces a hemodynamic response as described by an extended Balloon model (Buxton et al., 1998). Examples of DCM for fMRI can be found in Mechelli et al. (2004), Noppeney et al. (2006), Stephan et al. (2005), and Griffiths et al. (2007) (for a review on the conceptual basis of DCM and its implementation for functional magnetic resonance imaging data and event-related potentials, see Stephan et al., 2007).

Dynamical causal modelling differs from established methods for estimating effective connectivity from neurophysiological time series, which include structural equation modelling and models based on multivariate autoregressive processes (Harrison et al., 2003; McIntosh and Gonzalez-Lima, 1994; Roebroeck et al., 2005). In these models, there is no designed perturbation and the inputs are treated as unknown and stochastic. DCM assumes the input to be known, which seems appropriate for designed experiments. Further, DCM is based on a parameterized set of differential equations which can be extended to better describe the system.

Here, we extend the original model to cover two states per region. These states model the activity of inhibitory and excitatory populations. This has a number of key advantages. First, we can relax the *shrinkage priors* used to enforce stability in single-state DCMs because the interaction of excitatory–inhibitory pairs confers dynamical stability on the system. Second, we can model both extrinsic and intrinsic connections. Third, we can enforce positivity constraints on the extrinsic connections (i.e., interregional influences of excitatory popula-

* Corresponding author. Fax: +44 20 7813 1420.

E-mail address: amarreiros@fil.ion.ucl.ac.uk (A.C. Marreiros).

Available online on ScienceDirect (www.sciencedirect.com).

tions). Finally, this re-parameterization enables one to model context-dependent changes in coupling as a proportional increase or decrease in connection strength (cf., the additive effects used previously; Friston et al., 2003).

Shrinkage priors are simply priors or constraints on the parameters that shrink their conditional estimates towards zero (i.e., their prior expectation is zero and the prior variance determines the degree of shrinkage, in relation to observation noise). They were employed in early formulations of DCM to ensure coupling strengths did not attain very high weights, which generate exponentially diverging neuronal activity. However, this motivation for shrinkage priors is rather *ad hoc* and, as we will discuss later, confounds model specification and comparison.

This paper is structured as follows. In the first section, we present the two-state DCM, with two states per region. In the subsequent section, we provide a stability analysis of the two-state DCM. In the third section, we describe model inversion; i.e., prior distributions, Bayesian estimation, conditional inference and model comparison. In Simulations—model comparisons, we compare the single- and two-state DCM using synthetic and real data to establish its face validity. Finally, an empirical section then demonstrates the use of the two-state DCM by looking at attentional modulation of connections during visual motion processing. From these analyses, we conclude that the two-state DCM is a better model for fMRI data than the single-state DCM.

Theory

Dynamic causal modelling for fMRI—single-state models

In this section we review briefly dynamic causal models of fMRI data (Friston et al., 2003). In the next section, we extend this model to accommodate two neuronal sources per region. A dynamic causal model is, like the general linear model, an equation which expresses predicted responses in terms of some parameters and explanatory variables. In our case, the explanatory variables are the experimental inputs u , which correspond to stimulus functions in conventional models. The causal aspect comes from control theory, in which the response of causal models can be predicted from the current and past input. Critically, dynamic models do not predict the response *per se* but its rate of change. This rate of change can be a complicated nonlinear function of the models unknown parameters and known inputs. The form and parameterization of this function is entailed by the specific DCM used. In fMRI, people generally use a simple bilinear form with coupling parameter matrices A , $B^{(j)}$ and C . Critically, the $B^{(j)}$ matrices parameterize interactions between inputs and states; hence, bilinear. The DCMs considered below are multiple-input multiple-output systems that comprise N_u inputs and N_r outputs with one output per region. The N_u inputs correspond to designed causes (e.g., boxcar or impulse stimulus functions used in conventional analyses). Each region produces a measured output that corresponds to the observed BOLD signal. These time-series would normally be the averages or first eigenvariates of N_r selected regions.

Interactions among regions are modelled at the neuronal level. In single-state models each region has one state variable. This state is a simple summary of neuronal (i.e., synaptic) activity $x^{(i)}$,

in a region. Friston et al. (2003) used a bilinear form to describe their dynamics:

$$\dot{z} = F(x, u, \theta) \approx \mathfrak{I}x + Cu$$

$$\mathfrak{I} = A + \sum_j u_j B^{(j)}$$

$$A = \frac{\partial F}{\partial x} = \frac{\partial \dot{x}}{\partial z} \Big|_{u=0}$$

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

$$C = \frac{\partial F}{\partial u} \Big|_{x=0} \quad (1)$$

This model is used to generate neuronal activity; later we will add hemodynamics and noise to furnish a probabilistic model of fMRI measurements. In this model, the state vector $x(t)$ contains one scalar per region. The changes in neuronal (i.e., synaptic) activity are described by the sum of three effects. First, the matrix A encodes directed connectivity between pairs of regions. The elements of this connectivity matrix are not a function of the input, and can be considered as an endogenous or condition-invariant. Second, the elements of $B^{(j)}$ represent the changes of connectivity induced by the inputs, u_j . These condition-specific modulations or bilinear terms $B^{(j)}$ are usually the interesting parameters. The endogenous and condition-specific matrices are mixed to form the total connectivity or Jacobian matrix \mathfrak{I} . Third, there is a direct exogenous influence of each input u_j on each area, encoded by the matrix C . The parameters of this system, at the neuronal level, are given by $\theta^n \supseteq A, B^1, \dots, B^{N_u}, C$. At this level, one can specify which connections one wants to include in the model. Connections (i.e., elements of the matrices) are removed by setting their prior mean and variance to zero. We will illustrate this later.

The bilinear form in Eq. (1) can be regarded as an approximation to any function, $F(z, u, \theta)$, because it is simply a Taylor expansion around $z=0$ and $u=0$; retaining only terms that are first-order in the states or input. In this sense, the bilinear model can be regarded as a generic approximation, to any [unknown] function describing neuronal dynamics, in the vicinity of its fixed-point; i.e., when the neuronal states are at equilibrium or zero.

At the observation level, for each region, the neuronal state forms an input to a hemodynamic model that generates the BOLD signal. Region-specific hemodynamics are modelled by four extra hemodynamic state-variables. The corresponding first-order ordinary differential equations are parameterized by five region-specific parameters, θ^h (see Friston et al., 2003 for a complete description). Here, we summarize the integration of the neuronal and hemodynamic states by the generalized convolution

$$h(t) = h(u(t), \theta). \quad (2)$$

By integrating the ordinary differential equations of both levels, we can compute the systems predicted hemodynamic response $h(t)$ as a continuous function of time, given the neuronal and hemodynamic parameters $\theta \cong \theta^n, \theta^h$, input $u(t)$, and some initial states. Finally, this response is sampled appropriately to form predictions for the BOLD time-series (Kiebel et al., 2007). By assuming the observation error ε is Gaussian, we implicitly specify a likelihood model for the i th observation in the time-series

$$y_i = h(t_i) + \varepsilon_i \Leftrightarrow p(y|\theta) = N(h(\theta), \Sigma(\lambda)). \quad (3)$$

Where $\Sigma(\lambda)$ is the error covariance that is parameterized by an unknown hyperparameter. To complete the model specification, we need only the priors, $p(\theta)$. However, we will first review the extension of conventional likelihood models and their implicit re-parameterization.

Dynamic causal modelling for fMRI—two-state models

We now extend the standard DCM above to incorporate two state variables per region. These model the activity of an inhibitory and excitatory population respectively. Schematics of the single- and two-state models are shown in Fig. 1.

The Jacobian matrix, \mathfrak{S} represents the effective connectivity within and between regions. *Intrinsic* or within-region coupling is encoded by the leading diagonal blocks (see Fig. 1), and *extrinsic* or between-region coupling is encoded by the off-diagonal blocks. Each within-region block has four entries, $\mathfrak{S}_{ii}^{**} = \{\mathfrak{S}_{ii}^{EE}, \mathfrak{S}_{ii}^{II}, \mathfrak{S}_{ii}^{EI}, \mathfrak{S}_{ii}^{IE}\}$. These correspond to all possible intrinsic connections between the

excitatory and inhibitory states, $\{x_i^E, x_i^I\}$ of the i th region. These comprise self-connections, $E \rightarrow E, I \rightarrow I$ and interstate connections $E \rightarrow I, I \rightarrow E$. We enforce the connections, $E \rightarrow E, I \rightarrow E, I \rightarrow I$ to be negative (i.e., $\mathfrak{S}_{ii}^{EE}, \mathfrak{S}_{ii}^{EI}, \mathfrak{S}_{ii}^{II} \leq 0$), which means they mediate a dampening effect on population responses. This negativity is imposed by using log-normal priors; we use the negative exponential of an underlying coupling parameter with a normal prior (see below). Although the excitatory self-connections are negative, we do not mean to suggest that there are direct inhibitory connections among excitatory units, rather the multitude of mechanisms that self-organize neuronal activity (e.g., adaptation, gain-control, refractoriness, polysynaptic input from recurrent axonal collaterals, etc.) will conspire to make the effective self-connection negative. The extrinsic connections among areas are assumed to be positive (i.e., $\mathfrak{S}_{ij}^{EE} \geq 0$) and are mediated exclusively by coupling among excitatory populations (cf., glutamatergic projections in the real brain). In accord with known anatomy, we disallow long-range coupling among inhibitory populations.

The two-state DCM has some significant advantages over the standard DCM. First, intrinsic coupling consists of excitatory and inhibitory influences, which is biologically more plausible. Also, the interactions between inhibitory and excitatory subpopulations confer more stability on the overall system. This means we can relax the shrinkage priors used to enforce stability in single-state DCMs. Furthermore, we can now enforce positivity constraints on the extrinsic connections (i.e., interregional influences among excitatory populations) using log-normal priors and scale parameters as above for the intrinsic connections. This means changes in connectivity are now expressed as a proportional increase or

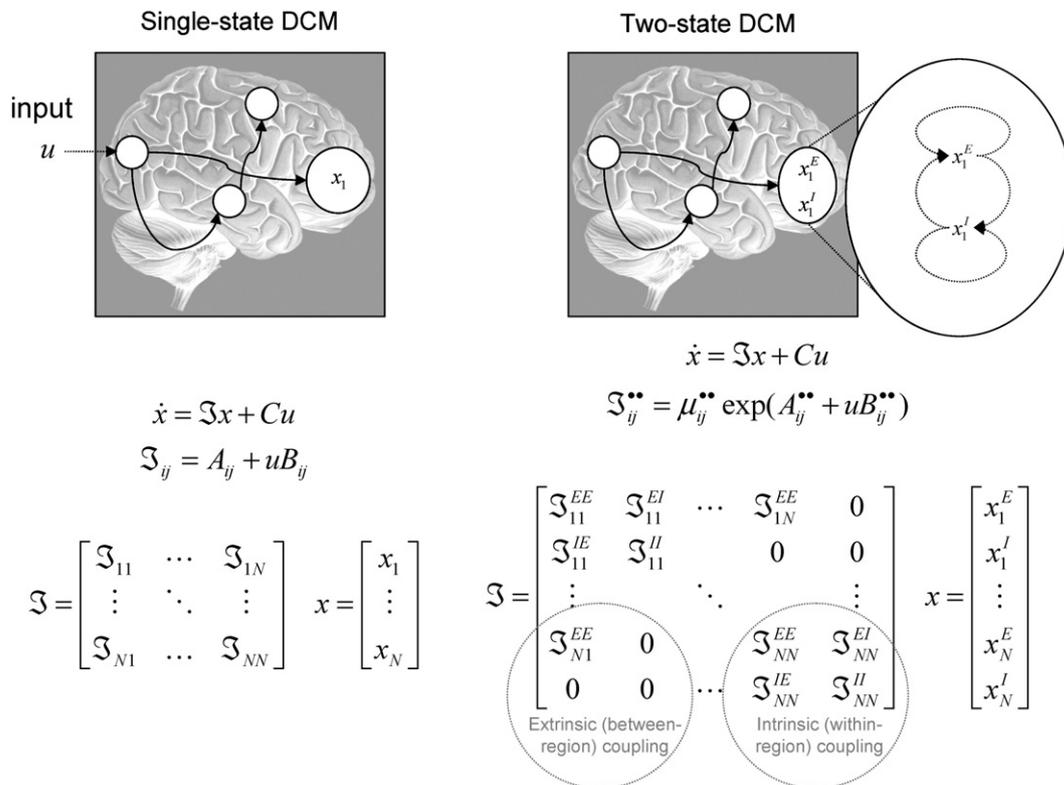


Fig. 1. Schematic of the single-state DCM (left) and the current two-state DCM (right). The two-state model has an inhibitory and an excitatory subpopulation. The positivity constraints are explicitly represented in the two-state connectivity matrix by exponentiation of underlying scale parameters (bottom right).

decrease in connection strength. In what follows, we address each of these issues, starting with the structural stability of two-state systems and the implications for priors on their parameters.

Stability and priors

In this section, we will describe a stability analysis of the two-state system, which informs the specification of the prior distributions of the parameters. Network models like ours can display a variety of different behaviours (e.g., Wilson and Cowan, 1973). This is what makes them so useful, but there are parameterizations which make the system unstable. By this we mean that the system response increases exponentially. In real brains, such behaviour is not possible and this domain of parameter space is highly unlikely to be populated by neuronal systems. The prior distributions should reflect this by assigning a prior probability of zero to unstable domains. However, this is not possible because we have to use normal priors to keep the model inversion analytically tractable. Instead, we specify priors that are centered on stable regions of parameter space.

In the original single-state DCM, we had a single state per region and a self-decay for each state (see Fig. 2). This kind of system allows for only an exponential decay of activity in each region, following a perturbation of the state by exogenous input or incoming connections. For this model, Friston et al. (2003) chose shrinkage priors, which were used to initialize the inversion scheme in a stable regime of parameter space, in which neuronal activity decayed rapidly. The conditional parameter estimates were then guaranteed to remain in a stable regime through suitable checks during iterative optimization of the parameters.

Two-state models (Fig. 3) can exhibit much richer dynamics compared to single-state models. Indeed, one can determine analytically the different kinds of periodic and harmonic oscillatory network modes these systems exhibit. This is important because it enables us to establish stability for any prior mean on the parameters. This entails performing a linear stability analysis by examining the eigenvalues of the Jacobian, \mathfrak{J} under the prior expectation of the parameters. The system is asymptotically stable if these eigenvalues (cf., Lyapunov spectrum) have only negative real parts (Dayan and Abbott, 2001). This is the procedure we adopt below.

It should be noted that, in generic coupled nonlinear systems, instability of a linearly stable fixed point does not always lead to exponential growth, but may lead to the appearance of a stable nonlinear regime. In the case of a Hopf bifurcation (as in Wilson and Cowan, 1973), a limit cycle appears near the unstable fixed point, which can model alpha rhythms and other oscillatory phenomena. Indeed, a system close to a linear instability exhibits longer and more complex nonlinear transients on perturbation (e.g.,

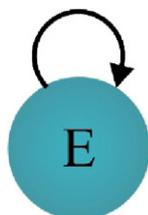


Fig. 2. Schematic of single-state DCM (one region).

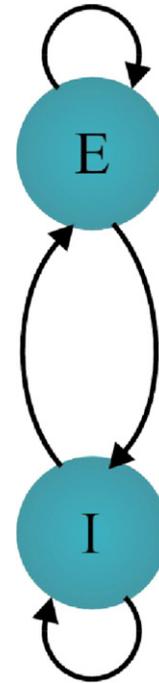


Fig. 3. Schematic of two-state DCM (one region).

Friston, 1997). This is a further reason to avoid using shrinkage priors (that preclude systems close to instability). However, because the bilinear model is linear in its states, its unstable fixed points are necessarily associated with exponential growth.

Priors

We now describe how we specify the priors and enforce positivity or negativity constraints on the connections. We seek priors that are specified easily and are not a function of connectivity structure; because this can confound model comparison (Penny et al., 2004). The strategy we use is to determine a stable parameterization for a single area, use this for all areas and allow only moderate extrinsic connections. In this way, the system remains stable for all plausible network structures.

Priors have a dramatic impact on the landscape of the objective function that is optimized: precise prior distributions ensure that the objective function has a global minimum that can be attained robustly. Under Gaussian assumptions, the prior distribution $p(\theta)$ is defined by its mean and covariance Σ . In our expectation–maximization inversion scheme, the prior expectation is also the starting estimate. If we chose a stable prior, we are guaranteed to start in a stable domain of parameter space. After this initialization, the algorithm could, of course, update to an unstable parameterization because we are dealing with a dynamic generative model. However, these updates will be rejected because they cannot increase the objective function: in the rare cases an update to an unstable regime actually occurs (and the objective function decreases), the algorithm returns to the previous estimate and halves its step-size, using a Levenberg–Marquardt scheme (Press et al., 1999). This is repeated iteratively, until the objective function increases, at which point the update is accepted and the optimization proceeds. Therefore, it is sufficient to select priors whose mean lies in a stable domain of parameter space.

Stability is conferred by enforcing connectivity parameters to be strictly positive or negative. In particular, the intrinsic, E→E, I→I, I→E connections are negative while the E→I and all extrinsic E→E connections are positive. We use

$$\begin{bmatrix} -1 & -0.5 \\ 0.5 & -1 \end{bmatrix}.$$

as the prior mode (most likely *a priori*) for a single region's Jacobian, where its states, $x = [x_1^E, x_1^I]^T$ summarize the activity of its constituent excitatory and inhibitory populations. This Jacobian has eigenvalues of $-1 \pm 0.5i$ and is guaranteed to be stable. We then replicate these priors over regions, assuming weak positive excitatory extrinsic E→E connections, with a prior of 0.5 Hz. For example, a three-region model, with hierarchical reciprocal extrinsic connections and states, $x = [x_1^E, x_1^I, x_2^E, x_2^I, x_3^E, x_3^I]^T$ would have Jacobian with a prior mode of

$$\mu = \begin{bmatrix} -1 & -0.5 & 0.5 & 0 & 0 & 0 \\ 0.5 & -1 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & -1 & -0.5 & 0.5 & 0 \\ 0 & 0 & 0.5 & -1 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & -1 & -0.5 \\ 0 & 0 & 0 & 0 & 0.5 & -1 \end{bmatrix}$$

The eigenvalue spectrum of this Jacobian is show in Fig. 4 (left panel), along with the associated impulse response functions for an input to the first subpopulation (right panels); evaluated with $x(t) = \exp(\mu t)x(0)$. It can be seen for this architecture we expect neuronal dynamics to play out over a time-scale of about 1 s. Note that these dynamics are not enforced; they are simply the most likely *a priori*.

Positivity constraints and scale-parameters

To ensure positivity or negativity, we scale these prior modes, μ with scale-parameters, which have log-normal priors. This is implemented using underlying coupling parameters with Gaussian or normal priors; for example, the extrinsic connections are parameterized as $\mathfrak{J}_{ij}^{**} = \mu_{ij}^{**} \exp(A_{ij}^{**} + uB_{ij}^{**})$, where $p(A_{ij}^{**}) = N(0, \nu)$ and we have assumed one input. A mildly informative log-normal prior obtains when the prior variance $\nu \approx 1/16$. This allows for a scaling around the prior mode, μ_{ij}^{**} of up to a factor of two, where

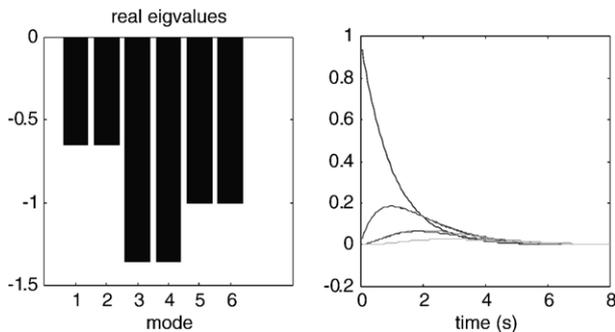


Fig. 4. Stability analysis for a two-state DCM (three regions).

Table 1

Log-evidences for three different models using synthetic data generated by the backward, forward and intrinsic models (see text)

Models	Synthetic data		
	Backward	Forward	Intrinsic
Backward	523.93 (99.9%)	494.19 (0.0%)	477.50 (0.7%)
Forward	382.07 (0.0%)	538.09 (99.9%)	439.22 (0.0%)
Intrinsic	497.67 (0.0%)	503.55 (0.0%)	482.47 (99.3%)

The diagonal values show higher log evidences, which indicate that the two-state DCM has internal consistency. The percentages correspond to the conditional probability of each model, assuming uniform priors over the three models examined under each data set.

the sign of the mode determines whether the connection is positive or negative. In what follows, we use a prior variance for the endogenous and condition-specific coupling parameters, A_{ij}^{ik} and B_{ij}^{ik} of $\nu = 1/16$.

Re-parameterizing the system in terms of scale-parameters entails a new state equation (see Fig. 1), which replaces the Bilinear model in Eq. (1)

$$\dot{x} = \mathfrak{J}x + Cu$$

$$\mathfrak{J}_{ij}^{**} = \mu_{ij}^{**} \exp(A_{ij}^{**}) + \sum_k u_k B_{ij}^{** (k)} = \mu_{ij}^{**} \exp(A_{ij}^{**}) \prod_k \exp(u_k B_{ij}^{** (k)})$$

$$\mathfrak{J} = \begin{bmatrix} \mathfrak{J}_{11}^{EE} & \mathfrak{J}_{11}^{EI} & \cdots & \mathfrak{J}_{1N}^{EE} & 0 \\ \mathfrak{J}_{11}^{IE} & \mathfrak{J}_{11}^{II} & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ \mathfrak{J}_{N1}^{EE} & 0 & & \mathfrak{J}_{NN}^{EE} & \mathfrak{J}_{NN}^{EI} \\ 0 & 0 & \cdots & \mathfrak{J}_{NN}^{IE} & \mathfrak{J}_{NN}^{II} \end{bmatrix} x = \begin{bmatrix} x_1^E \\ x_1^I \\ \vdots \\ x_N^E \\ x_N^I \end{bmatrix} \quad (4)$$

In this form, it can be seen that condition-specific effects u_k act to scale the connections by $\exp(u_k B_{ij}^{** (k)}) = \exp(B_{ij}^{** (k)} u_k)$. When $B_{ij}^{** (k)} = 0$, this scaling is $\exp(u_k B_{ij}^{** (k)}) = 1$ and there is no effect of input on the connection strength. The hemodynamic priors are based on those used in Friston (2002).

Having specified the form of the DCM in terms of its likelihood and priors, we can now estimate its unknown parameters, which represent a summary of the coupling among brain regions and how they change under different experimental conditions.

Bayesian estimation, inference and model comparison

For a given DCM, say model m , parameter estimation corresponds to approximating the moments of the posterior distribution given by Bayes rule

$$p(\theta|y, m) = \frac{p(y|\theta, m)p(\theta|m)}{p(y|m)}. \quad (5)$$

The estimation procedure employed in DCM is described in Friston et al. (2003) and Kiebel et al. (2006). The posterior moments (conditional mean η and covariance Ω) are updated iteratively using Variational Bayes under a fixed-form Laplace (i.e.,

Gaussian) approximation to the conditional density $q(\theta)=(\eta,\Omega)$. This is formally equivalent to expectation–maximization (EM) that employs a local linear approximation of Eq. (2) and Eq. (4) about the current conditional expectation.

Often, one wants to compare different models for a given data set. We use Bayesian model comparison, using the model evidence (Penny et al., 2004), which is

$$p(y|m) = \int p(y|\theta, m)p(\theta|m)d\theta. \tag{6}$$

Note that the model evidence is simply the normalization constant in Eq. (5). The evidence can be decomposed into two components: an accuracy term, which quantifies the data fit, and a complexity term, which penalizes models with redundant parameters. In the following, we approximate the model evidence for model m , under the Laplace approximation, with

$$\ln p(y|m) \approx \ln p(y|\lambda, m). \tag{7}$$

This is simply the maximum value of the objective function attained by EM. The most likely model is the one with the largest log-evidence. This enables Bayesian model selection. Model

comparison rests on the likelihood ratio of the evidence for two models. This ratio is the Bayes factor B_{ij} . For models i and j

$$\ln B_{ij} = \ln p(y|m = i) - \ln p(y|m = j). \tag{8}$$

Conventionally, strong evidence in favour of one model requires the difference in log-evidence to be about three or more (Penny et al., 2004). Under the assumption that all models are equally likely *a priori*, the marginal densities $p(y|m)$ can be converted into the probability of the model given the data $p(m|y)$ (by normalizing so that they sum to one over models). We will use this probability to quantify model comparisons below (see Tables).

Simulations—model comparisons

Here we establish the face validity of the DCM described in the previous section. This was addressed by integrating DCMs with known parameters, adding observation noise to simulate responses and inverting the models. Crucially, we used different models during both generation and inversion and evaluated all combinations to ensure that model selection identified the correct model.

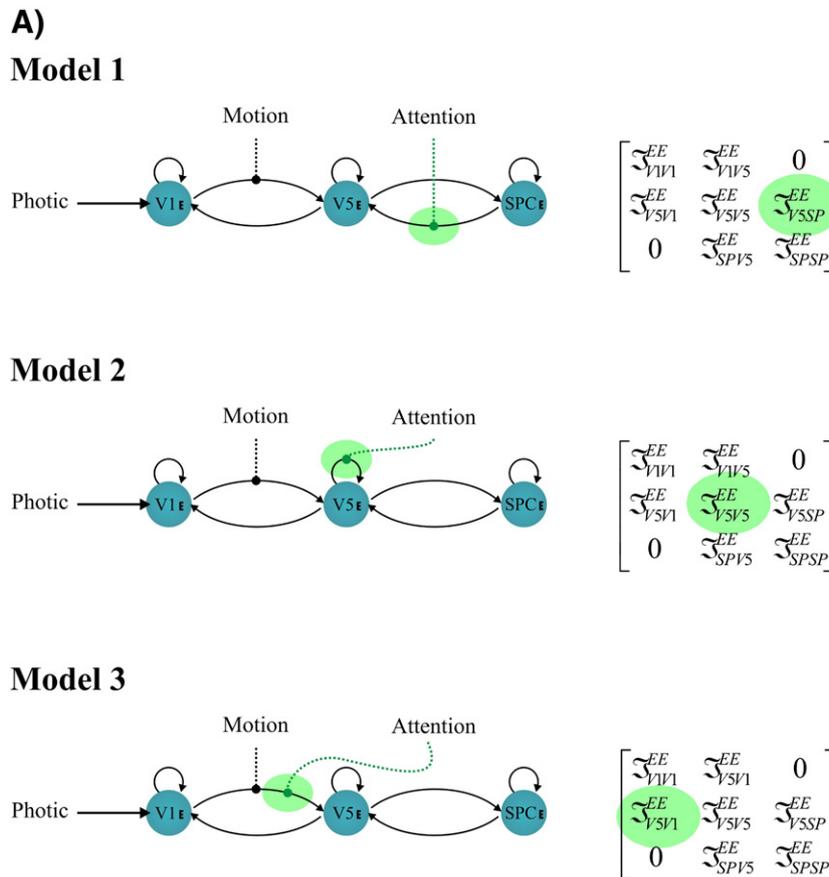
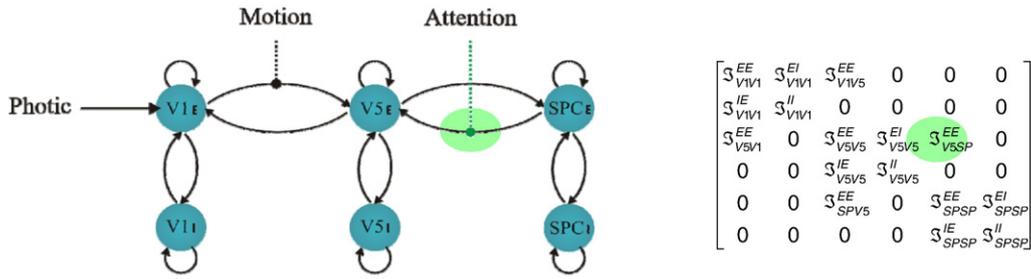
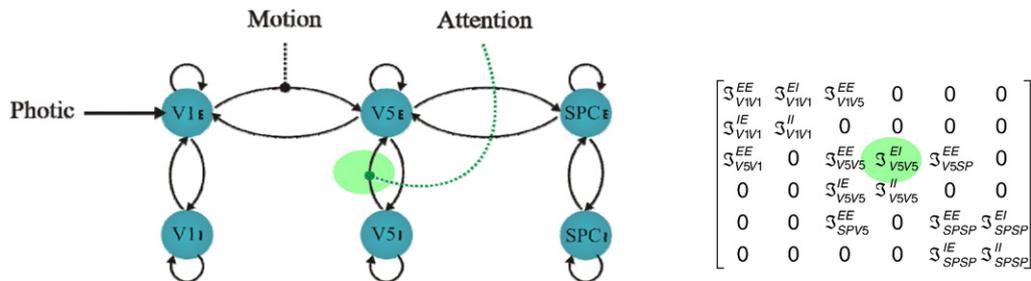


Fig. 5. In all models photic stimulation enters V1 and the motion variable modulates the connection from V1 to V5. Models 1, 2 and 3 all assume reciprocally and hierarchically organized connections. They differ in how attention modulates the influences on V5; Model 1 assumes modulation of the backward extrinsic connection, Model 2 assumes modulation of intrinsic connections in V5 and Model 3 assumes modulation of the forward connection. A: single-state DCMs. B: two-state DCMs.

B) –Two-State DCM
Model 1



Model 2



Model 3

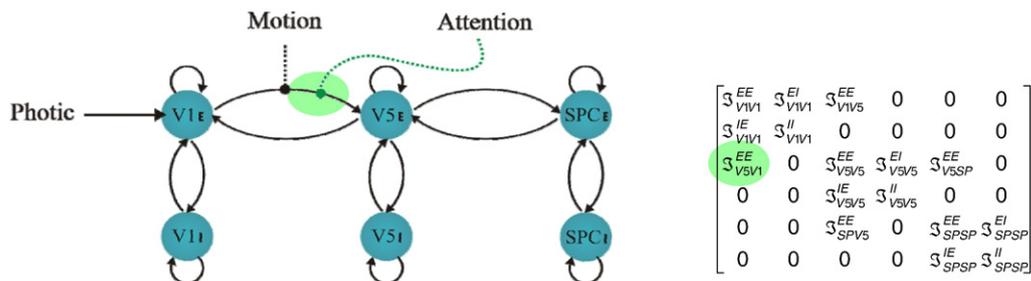


Fig. 5 (continued).

The DCMs used the posterior or conditional means from three different models estimated using real data (see next section). We added random noise such that the final data had a signal-to-noise ratio of three, which corresponds to typical DCM data¹. We created three different synthetic data sets corresponding to a forward, backward and intrinsic model of attentional modulation of connections in the visual processing stream. We used a hierarchical three-region model where stimulus-bound visual input entered at the first or lowest region. In the forward model, attention increased coupling in the extrinsic forward connection to the middle region; in the backward model it changed backward influences on the middle region and in the intrinsic model attention changed the intrinsic I→E connection. In all models, attention increased the sensitivity of the same excitatory population to different sorts of afferents.

¹ Note that a DCM time-series of a single region is the first eigenvariate of a cluster of voxels and is relatively denoised.

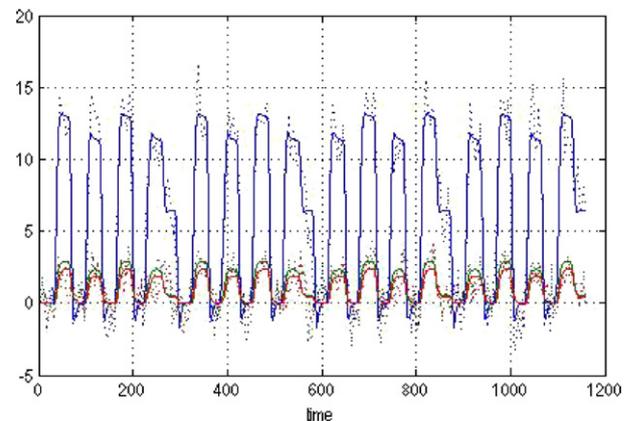


Fig. 6. Plot of the DCM fit to visual attention fMRI data, using the two-state Model 3. Solid: Prediction; Dotted: Data.

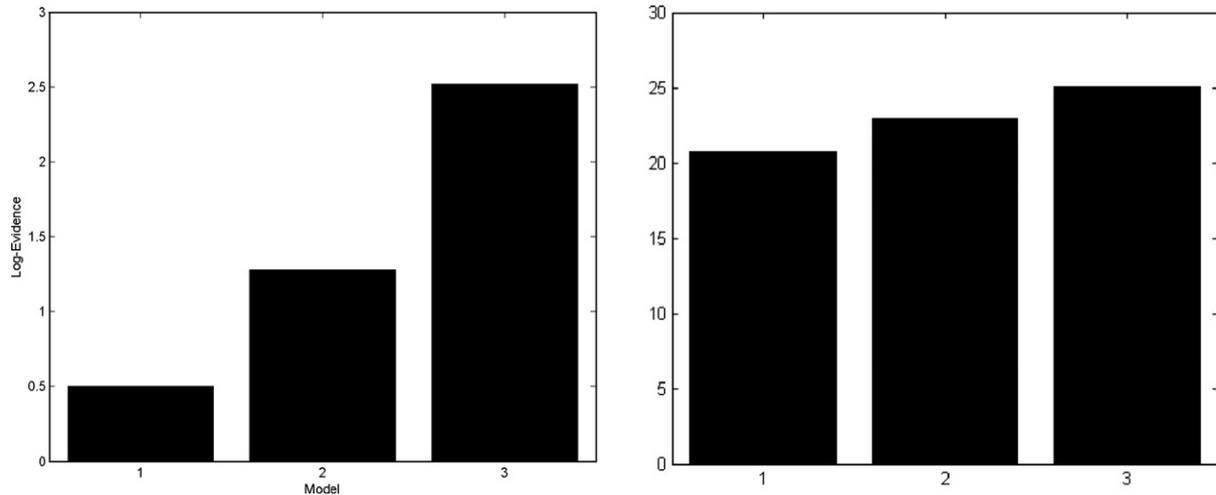


Fig. 7. Results of the Bayesian model comparisons among DCMs for single-state (left) and two-state (right) formulations. The graphs show the log-evidences for each model: Model 3 (modulation of the forward connections by attention) is superior to the other two models. The two-state model log-evidences are better than any single-state model (note the difference in scale).

We then used the three models to fit each of these three synthetic data sets, giving nine model inversions. Table 1 presents the log-evidences for each inversion. The highest evidences were obtained for models that were used to generate the synthetic data: these correspond to the diagonal entries. These results show that model comparison can identify reliably the correct model, among competing and subtly different two-state models.

Empirical analyses—model comparisons

In this section we ask whether the two-state extension described in this paper is warranted, in terms of providing a better explanation of real data. This was addressed by inverting the single- and two-state models using the same empirical data. These data have been used previously to validate DCM and are available from <http://www.fil.ion.ucl.ac.uk/spm>. We analyzed data from a study of attentional modulation during *visual motion processing* (Büchel and Friston, 1997). The experimental manipulations were encoded as three exogenous inputs: A ‘**photic stimulation**’ input indicated when dots were presented on a screen, a ‘**motion**’ variable indicated that the dots were moving and the ‘**attention**’ variable indicated that the subject was attending to possible velocity changes. The activity was modelled in three regions V1, V5 and superior parietal cortex (SPC).

We compared the single- and two-state DCM over the following three model variants. Model 1 assumed that attention modulates the backward extrinsic connection from SPC to V5.

Model 2 assumed that attention modulates the intrinsic connection in V5 and Model 3 assumed attention modulates the forward connection from V1 to V5. All models assumed that the effect of motion was to modulate the connection from V1 to V5. In Fig. 5 we show each of these three variants for the single- and two-state DCM.

We inverted all models using the variational EM scheme above and compared all six DCMs using Bayesian model comparison. As a representative example of the accuracy of the DCM predictions, we show the predicted and observed BOLD series for Model 3 (two-state) in Fig. 6. The results of the Bayesian model comparison are shown in Fig. 7, in terms of the log-evidences (in relation to a baseline model with no attentional modulation). These results show two things. First, both models find strong evidence in favour of Model 3, i.e., attention modulates the forward connection from V1 to V5. Second, there is strong evidence that the two-state Models 2 and 3 are better than any single-state model. The respective log-evidences for this Bayesian model comparison among DCMs (Büchel and Friston data) are shown in Table 2. Again, the table shows that the forward model is the best model, among either the single- or two-state DCMs. Moreover, there is very strong evidence in favour of the two-state model over the single-state model, because the differences in log-evidences are all greater than five. For reference; the log-evidence for the baseline model with no attentional modulation was -1649.9 .

These results represent an inference on model space. To illustrate inference on parameter space, Fig. 8 shows the conditional

Table 2

This table shows the log-evidences for the two models, single and two-state DCMs, plotted in the previous figure

	Backward	Forward	Intrinsic
Single-state DCM	-1649.38 (0.00%)	-1647.36 (0.00%)	-1648.60 (0.00%)
Two-state DCM	-1629.20 (1.08%)	-1624.80 (88.12%)	-1626.90 (10.79%)
Difference in log-evidence	20.180	22.560	21.700

Forward modulation is the best for both models. We can also see that there is very strong evidence in favour of the two-state model over the single-state model. The percentages in bold correspond to the conditional probability of each model, given the data and assuming uniform priors over the six models examined.

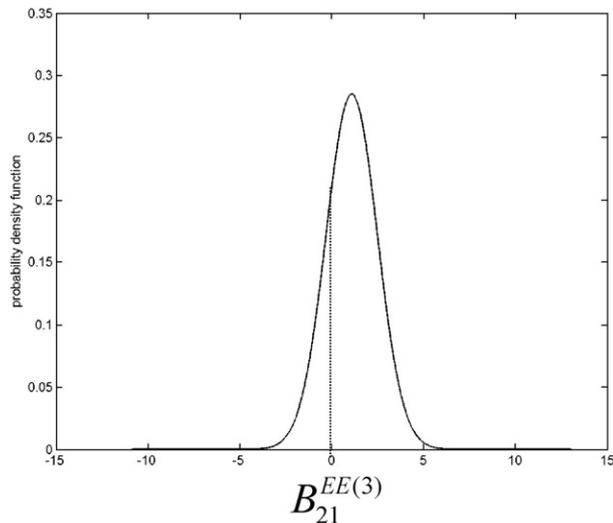


Fig. 8. Posterior probability density functions for the Gaussian parameter, $B_{ij}^{EE(3)}$ associated with attentional modulation of the forward connection in the best model. There is an 88% confidence that this gain is greater than one (area under the Gaussian to the right of the dashed line). The dashed line indicates $B_{ij}^{EE(3)}=0 \Rightarrow \exp(B_{ij}^{EE(3)})=1$.

density of the parameters representing attentional gain of the forward connection in the best model. We show this conditional density on the Gaussian parameter, $B_{ij}^{EE(3)}$ (with an implicit gain or scale-parameter $\exp(B_{ij}^{EE(3)})$) associated with attention (i.e., when $u_3=1$). It can be seen that we can be 88% confident that this gain is greater than one.

Discussion

In this paper, we have described a new DCM for fMRI, which has two states per region instead of one. With the two-state DCM, it is possible to relax shrinkage priors used to guarantee stability in single-state DCMs. Moreover, we can model both extrinsic and intrinsic connections, as well as enforce positivity constraints on the extrinsic connections.

Using synthetic data, we have shown that the two-state model has internal consistency. We have also applied the model to real data, explicitly modelling intrinsic connections. Using model comparison, we found that the two-state model is better than the single-state model and that it is possible to disambiguate between subtle changes in coupling; in the example presented here, we were able to show that attentional gain, in the context of visual motion processing, is accounted for sufficiently by an increased sensitivity of excitatory populations of neurons in V5 to forward afferents from earlier visual areas.

These results suggest that the parameterization of the standard single-state DCM is possibly too constrained. With a two-state model, the data can be explained by richer dynamics at the neuronal level. This might be seen as surprising because it generally is thought that the hemodynamic response function removes a lot of information and a reconstruction of neuronal processes is not possible. However, our results challenge this assumption, i.e., DCMs with richer dynamics (and more parameters) are clearly supported by the data.

In the following, we discuss some potential extensions to current DCMs that may allow useful questions to be addressed to

fMRI data: currently, we model excitatory (glutamatergic) and inhibitory (GABAergic) connections. As a natural extension we can include further states per region, accounting for other neurotransmitter effects. Important examples here would be adaptation phenomena and activity-dependent effects of the sort mediated by NMDA receptors. This is interesting because NMDA receptors are thought to be targeted preferentially by backward connections. This could be tested empirically using a suitable multistate DCM based on an explicit neural mass model.

Another important point is that the hemodynamics in the current DCM are a function of the excitatory states only. The contributions to the BOLD signal from the inhibitory states are expressed indirectly, through dynamic interactions between the two states, at the neuronal level. One possible extension would be to model directly separate contributions of these two states, at the hemodynamic level. Hypotheses about the influence of excitatory and inhibitory populations on the BOLD signal could then be tested using model comparison.

Another extension is to generalize the interactions between the two subpopulations, i.e., to use nonlinear functions of the states in the DCM. Currently, this is purely linear in the states, but one could use sigmoidal functions. This would take our model into the class described by Wilson and Cowan (1973). In this fashion, one can construct more biologically constrained response functions and bring DCMs for fMRI closer to those being developed for EEG and MEG. Again, the question of whether fMRI data can inform such neural mass models can be answered simply by model comparison. As noted above, the bilinear approximation used in the original formulation of DCM for fMRI represents a global linearization over the whole of state-space; the current extension uses the same bilinear approximation in the states (although it is nonlinear in the parameters). Further refinements to the model, such as applying a sigmoid nonlinearity (cf., Wilson and Cowan, 1973) would give a state equation that is nonlinear in the states. In this instance, we can adopt a local linearization, when integrating the system to generate predictions. In fact, our inversion scheme already uses a local linearization because the hemodynamic part of DCM for fMRI is nonlinear in the hemodynamic states (Friston, 2002). However, this approach does not account for noise on the states (i.e., random fluctuations in neuronal activity). There has already been much progress in the solution of stochastic differential equations entailed by stochastic DCMs, particularly in the context of neural mass models (see Valdes et al., 1999; Sotero et al., 2007).

Finally, in the next development of DCM for fMRI, we will evaluate DCMs based on density-dynamics. Current DCMs consider only the mean neuronal state for each population. In future work we will replace the implicit neural mass model with full-density dynamics, using the Fokker–Planck formalism. This would allow one to model the interactions between mean neuronal states (e.g., firing rates) and their dispersion or variance over each population of neurons modelled.

Conclusion

Our results indicate that one can estimate intrinsic connection strengths within network models using fMRI. Using real data, we find that a two-state DCM is better than the conventional single-state DCM. The present study demonstrates the potential of adopting generative models for fMRI time-series that are informed by anatomical and physiological principles.

Acknowledgments

This work was supported by the Portuguese Foundation for Science and Technology and the Wellcome Trust.

References

- 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.
- 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.
- Dayan, P., Abbott, L.F., 2001. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. MIT Press.
- Friston, K.J., 1997. Another neural code? *NeuroImage* 5 (3), 213–220.
- Friston, K.J., 2002. Bayesian estimation of dynamical systems: an application to fMRI. *NeuroImage* 16 (2), 513–530.
- Friston, K.J., Harrison, L., Penny, W., 2003. Dynamic causal modelling. *NeuroImage* 19, 1273–1302.
- Griffiths, T.D., Kumar, S., Warren, J.D., Stewart, L., Stephan, K.E., Friston, K.J., 2007. Approaches to the cortical analysis of auditory objects. *Hear. Res.*, doi:10.1016/j.heares.2007.01.010.
- Harrison, L.M., Penny, W., Friston, K.J., 2003. Multivariate autoregressive modelling of fMRI time series. *NeuroImage* 19 (4), 1477–1491.
- Kiebel, S.J., David, O., Friston, K.J., 2006. Dynamic causal modelling of evoked responses in EEG/MEG with lead field parameterization. *NeuroImage* 30, 1273–1284.
- Kiebel, S.J., Klöppel, S., Weiskopf, N., Friston, K.J., 2007. Dynamic causal modeling: a generative model of slice timing in fMRI. *NeuroImage* 34 (4), 1487–1496 (Feb 15).
- McIntosh, A.R., Gonzalez-Lima, F., 1994. Structural equation modeling and its application to network analysis in functional brain imaging. *Hum. Brain Mapp.* 2, 2–22.
- Mechelli, A., Price, C.J., Friston, K.J., Ishai, A., 2004. Where bottom-up meets top-down: neuronal interactions during perception and imagery. *Cereb. Cortex* 14, 1256–1265.
- Noppeney, U., Price, C.J., Penny, W.D., Friston, K.J., 2006. Two distinct neural mechanisms for category-selective responses. *Cereb. Cortex* 16 (3), 437–445.
- Penny, W.D., Stephan, K.E., Mechelli, A., Friston, K.J., 2004. Comparing dynamic causal models. *NeuroImage* 22 (3), 1157–1172.
- Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T., 1999. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press.
- Roebroeck, A., Formisano, E., Goebel, R., 2005. Mapping directed influence over the brain using Granger causality and fMRI. *NeuroImage* 25 (1), 230–242.
- Stephan, K.E., Penny, W.D., Marshall, J.C., Fink, G.R., Friston, K.J., 2005. Investigating the functional role of callosal connections with dynamic causal models. *Ann. N. Y. Acad. Sci.* 1064, 16–36.
- Stephan, K.E., Harrison, L.M., Kiebel, S.J., David, O., Penny, W.D., Friston, K.J., 2007. Dynamic causal models of neural system dynamics: current state and future extensions. *J. Biosci.* 32, 129–144.
- Sotero, R.C., Trujillo-Barreto, N.J., Iturria-Medina, Y., Carbonell, F., Jimenez, J.C., 2007. Realistically coupled neural mass models can generate EEG rhythms. *Neural Comput.* 19 (2), 478–512.
- Valdes, P.A., Jimenez, J.C., Riera, J., Biscay, R., Ozaki, T., 1999. Nonlinear EEG analysis based on a neural mass model. *Biol. Cybern.* 81 (5–6), 415–424.
- Wilson, H.R., Cowan, J.D., 1973. A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik* 13, 55–80.