NeuroImage 47 (2009) 590–601

Contents lists available at ScienceDirect

NeuroImage

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

Technical Note

Dynamic causal modelling of distributed electromagnetic responses

Jean Daunizeau *, Stefan J. Kiebel, Karl J. Friston

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

ARTICLE INFO

Article history:
Received 5 February 2009
Revised 6 April 2009
Accepted 14 April 2009
Available online 3 May 2009

Keywords:
Dynamic causal modelling
EEG
MEG
Neural-mass
Neural-field
Variational Bayes
Inversion
System identification
Source reconstruction

ABSTRACT

In this note, we describe a variant of dynamic causal modelling for evoked responses as measured with electroencephalography or magnetoencephalography (EEG and MEG). We depart from equivalent current dipole formulations of DCM, and extend it to provide spatiotemporal source estimates that are spatially distributed. The spatial model is based upon *neural-field* equations that model neuronal activity on the cortical manifold. We approximate this description of electrocortical activity with a set of local standing-waves that are coupled though their temporal dynamics. The ensuing distributed DCM models source as a mixture of overlapping patches on the cortical mesh. Time-varying activity in this mixture, caused by activity in other sources and exogenous inputs, is propagated through appropriate lead-field or gain-matrices to generate observed sensor data. This spatial model has three key advantages. First, it is more appropriate than equivalent current dipole models, when real source activity is distributed locally within a cortical area. Second, the spatial degrees of freedom of the model can be specified and therefore optimised using model selection. Finally, the model is linear in the spatial parameters, which finesses model inversion. Here, we describe the distributed spatial model and present a comparative evaluation with conventional equivalent current dipole (ECD) models of auditory processing, as measured with EEG.

© 2009 Elsevier Inc. All rights reserved.

Introduction

We have previously introduced a dynamic causal modelling (DCM) for event-related potentials and fields as measured with EEG and MEG (David and Friston 2003; David et al., 2005, 2006; Kiebel et al., 2006; Kiebel et al., 2007; Garrido et al., 2007). This extended the application of DCM beyond fMRI (Friston et al., 2003; Marreiros et al., 2008a) to cover EEG, MEG and local field potentials (Moran et al., 2007). However, all current DCMs model hemodynamic or electromagnetic signals as arising from a network of sources, where each source is considered to be a point process; i.e., an equivalent current dipole. In other words, the network is modelled as a *graph*, where sources correspond to *nodes* and conditional dependencies among the hidden states of each node are mediated by effective connectivity (known as *edges*). In this work, we replace the nodes with a distributed and continuous set of sources on the cortical surface. This provides a more realistic spatial model of underlying activity and, in the context of electromagnetic models, renders the DCM linear in its spatial parameters (c.f., Fuchs et al., 1999). The aim of this note is to describe this extension and compare it with models based upon point-sources or equivalent current dipoles (ECD). This model rests on the notions of mesostates (Daunizeau and Friston 2007) and anatomically informed basis functions (Phillips et al., 2002) but is motivated using neural-field theory (Amari, 1995; Jirsa and Haken, 1996; Liley et al., 2002).

DCM entails the specification of a generative model for an observed time-series and the inversion of this model to make inferences on model space and the parameters of each model. These inferences use the model evidence and posterior density of the parameters, respectively. In DCM, the underlying generative model is based on some state-equations (i.e., a state-space model) that describe the evolution of hidden states as a function of themselves and exogenous inputs (e.g., a stimulus function). The state-equations are supplemented with an observer function of the states to generate observed responses. By integrating the state-equation and applying the observer function, one obtains predicted responses. Under Gaussian assumptions about observation error these predictions furnish a likelihood model of observed responses. This likelihood model is combined with priors on the parameters to provide a full forward model of the data, which can be inverted using standard techniques (e.g., Friston et al., 2007a). These techniques generally rest on optimising a free-energy bound on the models log-evidence to approximate the posterior density of the model parameters.

In this work, we focus on the mapping from neuronal states to observed measurements at the sensors. We depart from equivalent current dipole models and employ an approximate neural-field model. Neural-field models describe electrocortical activity in terms of neuronal states (e.g. mean firing rate and post-synaptic membrane depolarisation) that are continuous over space (Amari 1995; Jirsa 1996; Liley 2002). This approach has shown how local and distal connectivity can interact to generate realistic spatiotemporal patterns of cortical activity that might underlie EEG rhythms (Nunez 1974) and

* Corresponding author. Fax: +44 207 813 1445.
E-mail address: j.daunizeau@fil.ion.ucl.ac.uk (J. Daunizeau).

1053-8119/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2009.04.062

their perceptual correlates, like visual hallucinations (Ermentrout and Cowan 1979). The patterns studied using neural-field models include *bumps* (transient clustering of activity) and *travelling waves* (which have been associated with synchronous discharges seen during epileptic seizures; Connors and Amitai 1993). These patterns are engendered by local (mesoscopic) connectivity. However, several authors have pointed out the importance of large-scale (macroscopic) connectivity in stabilizing local spatiotemporal dynamics (Jirsa and Kelso 2000; Breakspear and Stam 2005; Qubbaj and Jirsa 2007; Hutt and Atay 2005).

In this paper, we use theoretical results from neural-field theory, which combine mesoscopic and macroscopic connectivity, to model M/EEG. Specifically, we approximate the neural-field description of electrocortical activity with a set of distributed and continuous cortical sources that behave as standing-waves with compact local support. These standing-waves are coupled by temporal dynamics and follow from a truncated space-time decomposition of the solution of the underlying neural-field equations. The aims of this note are to (i) describe this neural-field DCM, (ii) compare it with established equivalent current dipole variants and (iii) to provide a framework for more realistic neural-field DCMs. We discuss these models in relation to the subtle balance between their face validity and identifiability.

In the first section of this paper, we derive a standing-wave approximation to the neural-field formulation and the ensuing parameterization of the DCM for distributed responses. In the second section, we present a comparative evaluation of DCMs based upon ECDs and distributed sources. We compare these models in terms of their relative log-evidence using a multi-subject EEG dataset, acquired during an auditory mismatch negativity paradigm. We conclude with a discussion of the benefits and potential uses of DCM for distributed responses.

DCM for distributed responses

In this section, we approximate a neural-field description of electrocortical activity with local standing-waves. We then combine the ensuing spatial model with the temporal state-space models used in previous DCMs for event-related responses (David and Friston 2003; David et al., 2005, 2006; Kiebel et al., 2006, 2007; Garrido et al., 2007). This combination provides a full spatiotemporal DCM for distributed responses, which can be fitted or inverted in the usual way.

From neural-fields to standing-waves

Neural-fields and mesoscopic modelling

We start with a description of the dynamics of a single neuron within an ensemble of neurons. These dynamics can be modelled as a temporal convolution of the average (mean-field) firing of the local population that is seen by the neuron:

$$x_j^{(i)}(t) = \int G(t-t')\left(\left\langle H\left(x_j^{(i)}(t)-\theta\right)\right\rangle_j + \gamma_{iu}u(t)\right)dt'$$
$$G(t) = \begin{cases} \gamma\dfrac{t}{\kappa}exp\left(-\dfrac{t}{\kappa}\right) & t\geq 0 \\ 0 & t<0. \end{cases} \quad (1)$$

Here, $x_j^{(i)}(t)$ is the post-synaptic membrane potential (PSP) of the j-th neuron in the i-th population; G is the alpha-kernel; H is the Heaviside function that models firing above depolarisation threshold θ; κ is a lumped rate-constant and γ controls the maximum post-synaptic potential. Eq. 1 assumes that any neuron senses all the neurons in the population it belongs to. This means endogenous input (from this population) can be written as the expected firing rate over that population (cf., Marreiros et al., 2008c). Here exogenous input (from another population or stimulus-bound subcortical input) is modelled as an injected current u scaled by the parameter γ_{iu}. Eq. 1 can be reformulated in terms of an ODE (cf., David and Friston, 2003):

$$\ddot{x}_j^{(i)} - \kappa^2\gamma\left(\left\langle H\left(x_j^{(i)}-\theta\right)\right\rangle_j + \gamma_{iu}u\right) + 2\kappa\dot{x}_j^{(i)} + \kappa^2 x_j^{(i)} = 0. \quad (2)$$

Given Eq. 2 we can now model the dynamics of the population mean PSP by taking its expectation over neurons $\mu^{(i)} = \langle x_i^{(i)}\rangle_j$ (Marreiros et al., 2008b):

$$\ddot{\mu}^{(i)} + 2\kappa\dot{\mu}^{(i)} + \kappa^2\mu^{(i)} = \kappa^2\varsigma^{(i)}$$
$$\varsigma^{(i)} = \gamma_{iu}u + \sum_j \gamma_{ij}S\left(\mu^{(j)}\right)$$
$$S\left(\mu^{(j)}\right) = \frac{1}{1+\exp(-\rho(\mu^{(j)}-\theta))} \quad (3)$$

where $\mu^{(j)}$ corresponds to the mean PSP in each population j sending exogenous input. This is a conventional neural-mass model that effectively applies a linear synaptic (alpha) kernel to input—$\varsigma^{(i)}$, from other populations. This input is a nonlinear (sigmoid) function of depolarisation (Jansen and Rit 1995), which can be thought of as the cumulative probability distribution of PSPs over the population sending afferent signals. See Fig. 1, which shows the explicit form of the state-equations for a cortical source containing three fields or populations.

In DCM, a cortical source is typically modelled using three neuronal subpopulations corresponding roughly to spiny stellate *input* cells (in the granular layer) intrinsic inhibitory interneurons (assigned to the supragranular layer) and deep pyramidal *output* cells in the infragranular layer. The connectivity within (intrinsic) and between (extrinsic) sources conforms to the laminar rules articulated in Felleman and Van Essen (1991). This is implicitly modelled in Eq. 3 through the mixture of exogenous and endogenous inputs ς that depends on the connectivity or coupling parameters—γ_{ij}. This sort of neural-mass model has been used to emulate electrophysiological recordings (e.g. Jansen and Rit 1995; Wendling et al., 2000; David et al., 2005) and as a generative model for event-related potentials in DCM (David et al., 2006).

However, these neural-mass models are not formulated to model spatially extended cortical regions (a square centimeter or so); they model the states of point processes, typically one macrocolumn (about 10,000 neurons, or a square millimeter of cortex; Breakspear and Stam, 2005). Neural-field models are important generalizations of neural-mass models, which account for the spatial spread of activity, through local connectivity between macrocolumns. In these models, states like the PSP of each cortical layer can be regarded as a continuum or *field*, which is a function of space r and time: $\mu^{(i)}(t)\rightarrow\mu^{(i)}(\mathbf{r},t)$. This allows one to formulate the dynamics of each field in terms of partial differential equations (PDE). These are essentially wave-equations that accommodate lateral interactions among neural-masses (e.g., cortical columns). Key forms for neural-field equations were proposed and analyzed by Nunez (1974) and Amari (1975). Jirsa and Haken (1996) generalized these models and also considered delays in the propagation of spikes over space. The introduction of delays leads to dynamics that are reminiscent of those observed empirically. Typically, neural-field models can be construed as a spatiotemporal convolution that can be written in terms of a Green function (see e.g. Jirsa et al., 2002):

$$\mu^{(i)}(\mathbf{r},t) = \int G(\mathbf{r}-\mathbf{r}',t-t')\varsigma^{(i)}(\mathbf{r}',t')dt'd\mathbf{r}'$$
$$G(\mathbf{r}-\mathbf{r}',t-t') = \delta\left(t-t'-\frac{1}{c}|\mathbf{r}-\mathbf{r}'|\right)\frac{1}{\gamma}\exp\left(-\frac{1}{\gamma}|\mathbf{r}-\mathbf{r}'|\right). \quad (4)$$

Here, G is a Green function (modelling mesoscopic lateral connectivity), $|\mathbf{r}-\mathbf{r}'|$ is the distance between $\mathbf{r}$ and $\mathbf{r}'$, c is the speed of spike propagation, γ controls the spatial decay of lateral

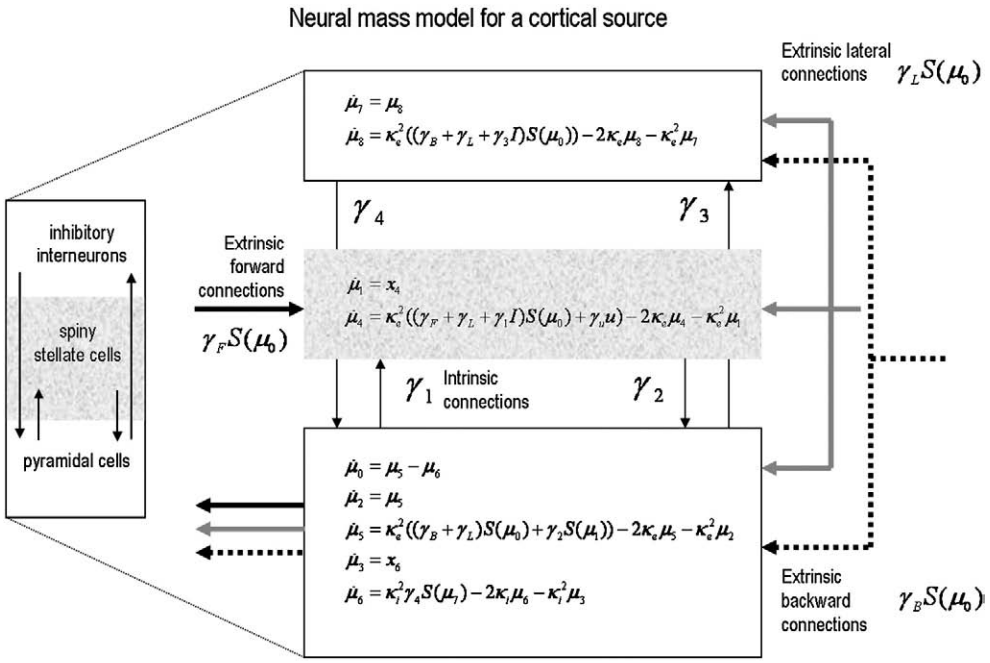

Fig. 1. Neural-mass model. This figure depicts the schematic cytoarchitectonics of a cortical source, along with the differential equations used to model the dynamics of each of the three subpopulations (pyramidal, spiny stellate and inhibitory interneurons). These subpopulations have been assigned to granular and agranular cortical layers, which receive forward and backward connections, respectively. Here we have expressed the second-order ODEs in the text with pairs of first-order ODEs. This clarifies how the coupling parameters mediate influences among and between sources. Note that the infragranular population comprises two subpopulations (one excitatory and the one inhibitory). Source or region-specific superscripts have been dropped here for clarity.

interactions (within a neural-field) and, as above, the input $\varsigma^{(i)}$ models both the effective connectivity between the neural-fields of different populations or layers. Eq. 4 is formulated as a simple convolution[1]; the corresponding second-order equations of motion are the neural wave-equations (see Appendix 1):

$$\left(\frac{\partial^2}{\partial t^2} + 2\kappa\frac{\partial}{\partial t} + \kappa^2 - \frac{3}{2}c^2\nabla^2\right)\mu^{(i)}(\mathbf{r},t) = c\kappa\varsigma^{(i)}(\mathbf{r},t) \quad (5)$$

where $\kappa = c/\gamma$ and ∇^2 is the Laplacian operator that returns the spatial curvature. Note the similarity in form of Eqs. 3 and 5. These sorts of models have been extremely useful in modelling spatiotemporally extended dynamics, which unfold on the cortical manifold (see Deco et al. (2008) for a recent review, Coombes et al. (2007) for a more informed derivation of 2D neural fields and Robinson et al. (1997) for a seminal analysis of the properties of coupled neural-fields).

Approximating the dynamics of neural-fields

In what follows, we will try to approximate the dynamics described by the partial differential equations above, with a system whose dynamics can be described with the ordinary differential equations used in neural-mass formulations. Using separation of variables, it is fairly easy to show (see Appendix 2) that the solution of the neural-field equations can be expressed as a superposition of spatiotemporal modes that can be factorised into spatial and temporal components. For the i-th field or population:

$$\mu^{(i)}(r,t) = \sum_k v_k^{(i)}(t)w_k^{(i)}(r). \quad (6)$$

Here, $w_k^{(i)}(r)$ is the k-th spatial mode or pattern and is the solution to the eigenvalue problem $\nabla^2 w_k^{(i)} + \lambda_k^{(i)} w_k^{(i)} = 0$. Note that this eigenvalue problem has to satisfy Dirichlet boundary conditions, i.e. the spatial modes are zero at the edges of the cortical region supporting the neural field. The temporal expressions of these modes are the eigenfunctions $v_k^{(i)}(t)$ of the field, which obey the following second-order ODE:

$$\ddot{v}_k^{(i)} + 2\kappa\dot{v}_k^{(i)} + \left(\kappa^2 - \frac{3}{2}\lambda_k^{(i)}c^2\right)v_k^{(i)} = c\kappa\varsigma_k^{(i)}(t) \quad (7)$$

where the scalar input seen by each mode $\varsigma_k^{(i)}(t)$ is given by projecting the input field $\varsigma^{(i)}(\mathbf{r},t)$ onto that mode

$$\varsigma_k^{(i)}(t) = \int w_k^{(i)}(\mathbf{r})\varsigma^{(i)}(\mathbf{r},t)d\mathbf{r}. \quad (8)$$

This means the solution of the partial differential equations describing the spatiotemporal dynamics of neural-fields (Eq. 6) can be decomposed into spatial modes, $w_k^{(i)}(\mathbf{r})$, weighted by the solutions of the coupled ODEs in Eq. 7, which describe the temporal dynamics of the neural-field.

We now want to simplify this description without compromising the dynamical repertoire of the model. Previous work on EEG/MEG source reconstruction suggests that most of the variance in EEG/MEG measurements can be accounted for by a set of temporally coherent and spatially extended cortical sources (see Daunizeau et al., 2006; Daunizeau and Friston, 2007; and Friston et al., 2007a,b). This coarse-grain description of electrocortical activity corresponds to a truncated spatiotemporal decomposition, in which each cortical region has just one spatial mode, whose activity is modulated over time (see also Jirsa et al. 2002; Wennekers 2008; and Robinson et al., 2001). Here, we motivate a related approximation based on equilibrium arguments. In the absence of exogenous input, each spatial mode decays at a rate that is proportional to $\kappa\left(1 + \frac{c}{2}\sqrt{\frac{3}{2}\lambda_k^{(i)}}\right)$. This is important, because high propagation velocities c, will dissipate the spatial modes quickly, with the exception of the fundamental mode w_0, which has a zero eigenvalue; $\lambda_0^{(i)} = 0$. This means that, after a short period of time, the depolarisation of the i-th

[1] When considering 2D neural fields on the cortical manifold, Eq. 3 is an approximation that is valid whenever the spatial decay of lateral interaction is short enough (typically smaller than the average distance between two cortical sulci; see Appendix 1).

population or field will become a standing-wave; corresponding to fluctuations of the fundamental mode:

$$w_0^{(i)}(\mathbf{r})\left(\ddot{v}_0^{(i)}(t) + 2\kappa\dot{v}_0^{(i)}(t) + \kappa^2 v_0^{(i)}(t)\right) = c\kappa w_0^{(i)}(\mathbf{r})\varsigma_0^{(i)}(t). \quad (9)$$

Here, $v_0^{(i)}(t)$ describes the temporal evolution of this mode. Critically, these dynamics have exactly the same form as the neural-mass model; i.e., when $\lambda_k^{(i)} = 0$, Eq. 9 is formally identical to Eq. 3. This suggests that we can model distributed responses using a single mode or pattern, whose fluctuations are coupled by the dynamics of conventional neural-mass models (see Appendix for details).

In summary, by ignoring all but the fundamental mode, we can model the spatiotemporal dynamics of each population or layer as fluctuations in a single spatial mode—$w_0^{(i)}(\mathbf{r})$. Under this approximation, the dynamics of coupled populations become a simple system of coupled standing-waves, each of which behaves like a neural-mass. The temporal dynamics $v_0^{(i)}(t)$ of these modes are exactly the same as neural-mass models, where the mean PSP is replaced by the eigenfunction:

$$\begin{aligned}\ddot{v}_0^{(i)} + 2\kappa\dot{v}_0^{(i)} + \kappa^2 v_0^{(i)} &= c\kappa\varsigma_0^{(i)}(t)\\ \varsigma_0^{(i)}(t) &= \gamma_{iu}u + \sum_j \gamma_{ij} S\left(v_0^{(j)}\right)\\ &:= \int w_0^{(i)}(r)\varsigma^{(i)}(r,t)dr.\end{aligned} \quad (10)$$

These approximations allow us to relate neural-mass DCMs to more realistic neural-field models. From the perspective of neural-fields, neural-mass models correspond to an approximation, which is valid when the system is close to equilibrium; i.e. when the interactions between the modes do not drive the system into autonomous behaviour (bumps or travelling waves) and most modes decay quickly. This is typically assumed to be the case for event-related responses (ERPs), which are generally slow damped oscillatory responses to stimulation (e.g. Kiebel et al., 2007; Garrido et al., 2008). It would be possible to increase the number of modes per population or field to provide a more complete neural-field model; however, this is beyond the scope of the present work. Our model now comprises a set of neural-masses, whose dynamics modulate the expression of some unknown but fixed spatial modes. Next, we consider how these modes are modelled.

The spatial model

Due to the Dirichlet constraints at the boundary of the cortical regions and local variations in cortical curvature, the fundamental mode $\mathbf{w}_0^{(i)}$ of the Laplacian operator can have an arbitrary spatial profile. Therefore, we model it as a mixture of spatial basis functions, derived from the gain-matrix associated the cortical region:

$$\mathbf{w}_0^{(i)} = \sum_n \boldsymbol{U}_n^{(i)}\beta_n^{(i)}. \quad (11)$$

Here, $\boldsymbol{U}_n^{(i)}$ are the spatial eigenvectors of the gain-matrix $\mathbf{L}^{(i)}$ associated with the set of vertices of the cortical mesh belonging to the i-th source or region, and $\beta_n^{(i)}$ are the unknown spatial parameters of our DCM. In addition, we assume that each cortical layer (neuronal population) within each region can contribute to the EEG/MEG signal measured at the sensors. This leads to the following DCM for distributed responses:

$$\boldsymbol{y}(t) = \sum_i \mathbf{L}^{(i)}\mathbf{w}_0^{(i)}\sum_j J_j v_0^{(ij)}(t) + \boldsymbol{\varepsilon}. \quad (12)$$

where $\boldsymbol{y}(t)$ is the column vector of instantaneous EEG/MEG scalp measurements and $\mathbf{L}^{(i)}$ are the gain-matrices for the i-th region. The unknown relative contributions J_j of the eigenfunctions $v_0^{(ij)}(t)$ of the j-th population in the i-th cortical region are assumed to be the same for all regions. Note that the fundamental mode is the same for all populations within the same region because it depends only on the geometry of the regional cortical manifold. The free-parameters of the DCM now comprise the spatial parameters $\vartheta \supset \{\beta, J\}$ (Eq. 11) and the neuronal parameters $\vartheta \supset \{\kappa, \gamma\}$ of the ODE (Eq. 10); these encode synaptic rate-constants and coupling parameters, respectively.

The decomposition of the spatial mode into the principal components of the gain-matrix (Eq. 11) suppresses redundancy in the spatial model; in the sense that spatial modes that cannot be seen by the sensors are precluded. In our implementation, the user specifies the coordinates of the sources comprising the network in canonical space (Mattout et al., 2007; Talairach and Tournoux 1988). The mesh points constituting each source are then identified automatically as those points lying within a sphere centred on the prior source location. We then take the first eight eigenvectors of $\mathbf{L}^{(i)}\mathbf{L}^{(i)T}$ to produce the spatial basis functions $\mathbf{U}^{(i)}$. The lead-fields are computed using BrainStorm (http://neuroimage.usc.edu/brainstorm/), after co-registering the channel locations to a subject-specific canonical mesh (Mattout et al., 2007). This involves warping a template mesh (in canonical space) to match the anatomy of each subject; so that individual differences in anatomy are accommodated but the mapping between subject-specific meshes and canonical space is preserved. The warping uses standard nonlinear spatial normalisation tools in SPM (http://www.fil.ion.ucl.ac.uk/spm).

Model inversion

Model inversion proceeds using standard variational techniques under the Laplace assumption as described in previous communications (e.g., Friston et al., 2007a). The products of this inversion are a free-energy approximation to the model's log-evidence $\ln p(y|m)$ and an approximating posterior density on the model parameters, $q(\vartheta) = N(\mu_\vartheta, \Sigma_\vartheta)$, where μ_ϑ is the posterior expectation and Σ_ϑ is the posterior covariance. This inversion entails the computation of the gradients and curvatures of the log-likelihood function, provided by the likelihood model (Eq. 11). This involves computing the derivatives of the predicted response with respect to model parameters; i.e., integrating the neuronal state-equations to see how they respond to stimulus-related input (a parameterized Gaussian bump-function of peristimulus time) and then repeating this under small perturbations of the parameters. Critically, the computation of the derivatives with respect to the spatial parameters can be simplified greatly if the response is linear in the parameters. This is the case for the distributed source model, under which

$$\begin{aligned}\frac{\partial \mathbf{y}}{\partial \beta_n^{(i)}} &= \mathbf{L}^{(i)}\mathbf{U}_n^{(i)}\sum_j J_j v_0^{(ij)}\\ \frac{\partial \mathbf{y}}{\partial J_j} &= \sum_i \mathbf{L}^{(i)}\sum_n \mathbf{U}_n^{(i)}\beta_n^{(i)} v_0^{(ij)}.\end{aligned} \quad (13)$$

This is not the case for the DCMs based on ECDs, which have nonlinear observer functions with six spatial parameters (encoding the location of the source and its orientation). With the present spatial model, we only have to integrate the system once, given the current estimate of the neuronal parameters, γ, κ to get $v_0^{(ij)}$. These are then used to compute the gradients in Eq. 13. This speeds up the iterative variational scheme, as compared to the conventional DCMs based on ECDs.

In what follows, we will focus on model comparison under ECD and distributed spatial models, using the same temporal model. We

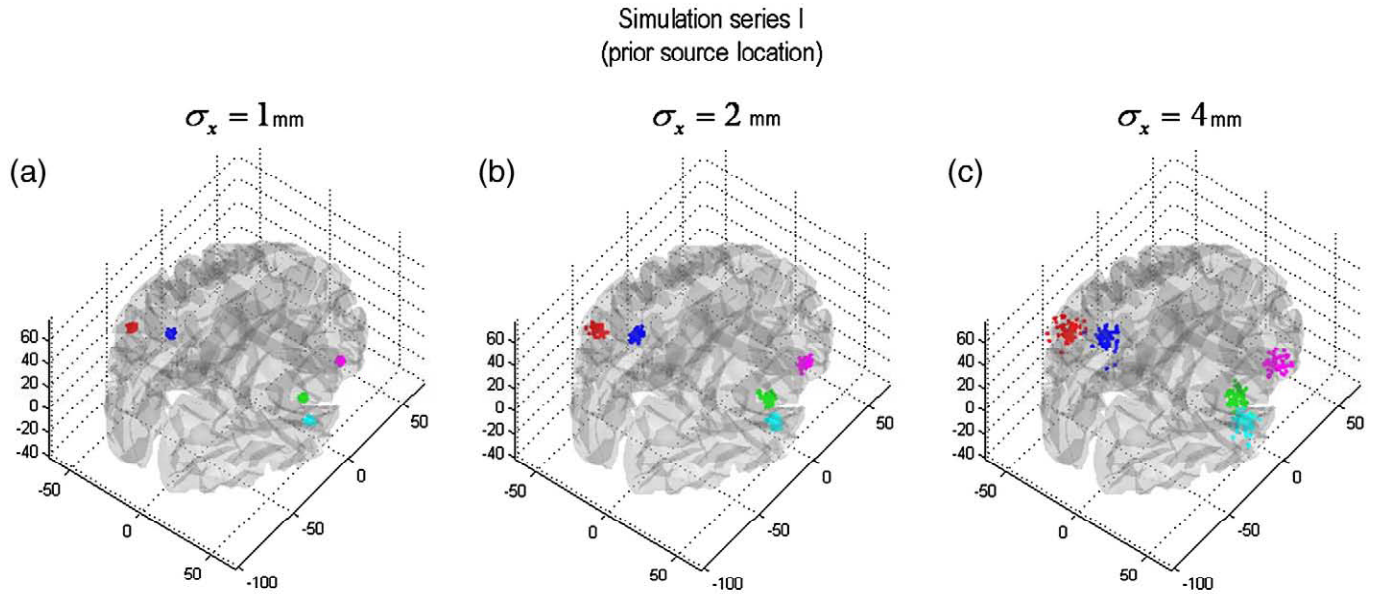

Fig. 2. Simulation series I: changing the prior locations. This figure depicts the three levels of perturbation to the prior location of the sources, as quantified by the standard deviation σ_x of the distance between the true (simulated) position of the sources and the location that has been used to specify the DCM. (a) The 50 random samples of prior location of the five sources (for $\sigma_x = 1$mm). (b) Id for $\sigma_x = 2$mm. (c) Id for $\sigma_x = 4$mm.

will use Monte Carlo simulations to assess sensitivity and real ERP data to compare the spatial models in terms of their evidence. A difference in log-evidence of three is usually considered significant; because this suggests a relative likelihood of 20:1. Under flat priors on the models, this means that one can be 95% confident that one model is better than the other.

Comparative evaluations

A sensitivity analysis

We are primarily interested in making inferences about the connectivity of the network generating data (encoded by γ_{ij}). However, the estimation of these parameters will be sensitive to the specification of the generative model (e.g. the prior position of the sources). In this section, we quantify the relative robustness (if any) of the DCM for distributed responses, relative to ECD models, to variations of the generative model. To assess robustness we compared the changes in the posterior estimates of the neuronal parameters (i.e. synaptic efficacies and rate-constants, which are common to both models) when changing the prior or likelihood of the DCM.

To equate the degrees of freedom (number of parameters) in both models, we used six spatial basis functions to model each mode (ECD models have six spatial parameters encoding the location and orientation of each dipole). First, we computed the predictions y after fitting two DCMs (ECD and distributed) to real mismatch negativity event-related potentials (ERPs) (see next section). This produced two sets of data, generated by ECD and distributed DCMs, with different but known neuronal parameters. These were then used as synthetic data for a series of DCM inversions, as follows:

We ran two sets of simulations. In series I, we perturbed the prior mean of the [five] source locations. We examined three levels of perturbation: $\sigma_x \in \{1,2,4\}$ mm, where σ_x was the standard deviation of random Gaussian perturbations to the prior mean (see Fig. 2). In series

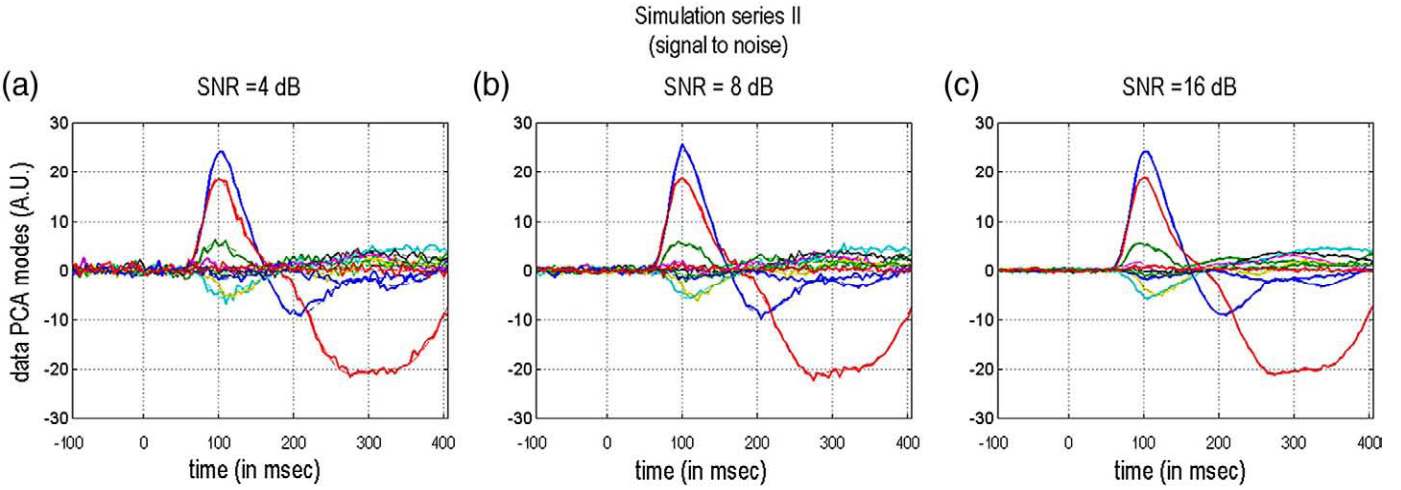

Fig. 3. Simulation series II: changing signal to noise. This figure shows the three levels of measurement noise in the synthetic data as quantified by the signal-to-noise ratio (SNR). (a) One sample of a synthetic data set (projected onto the sixteen spatial components of a PCA decomposition), at SNR = 4 dB. (b) at SNR = 8 dB. (c) at SNR = 16 dB.

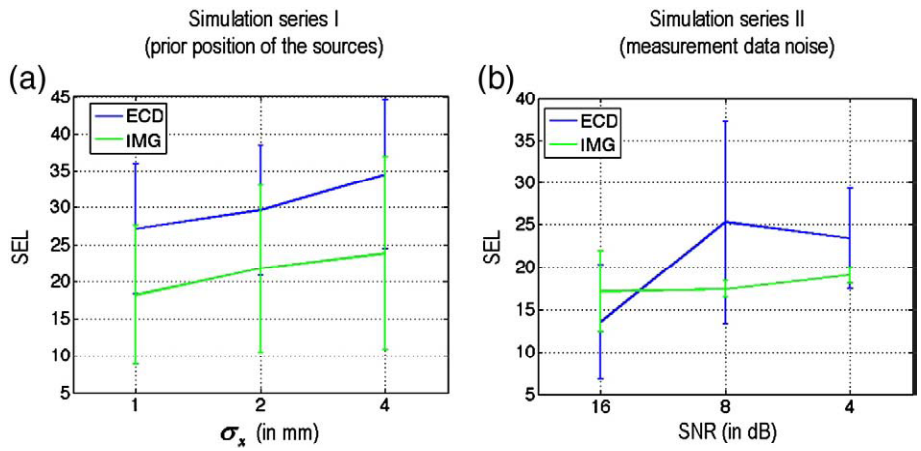

Fig. 4. Monte Carlo simulation results-squared error loss. Both graphs show the squared error loss (SEL) as a function of the level of prior dislocation and noise (error bars correspond to one standard deviation). (a) SEL as a function of perturbation on the prior location of the sources—σ_x. (b) SEL as a function of signal to noise-SNR. Except for the highest level of SNR (16 dB), where both ECD and distributed DCMs behave similarly, the spatially distributed DCM is consistently better than its ECD variant.

II, we perturbed the likelihood by adding Gaussian noise to the data; using three signal-to-noise ratios: SNR∈{4,8,16} dB. SNR is defined as: SNR = 10 ln var $(\hat{y})$/var(ε) (see Fig. 3). We used 50 Monte Carlo samples for both series.

Given the true parameters of the generative model and their posterior estimator, we can evaluate the squared error loss:

$$SEL(\vartheta) = \sum_i \left(\vartheta_i - \hat{\vartheta}_i\right)^2 \tag{14}$$

where ϑ is the i-th neuronal parameter. The SEL is a standard estimation error measure, whose posterior expectation is minimized by the mean of the posterior density. This means that using the posterior mean as an estimator $\hat{\vartheta} = \langle\vartheta\rangle_q$ of unknown ϑ is optimal with respect to squared error loss.

We investigated how the SEL changed as a function of prior location and SNR, for both the ECD and the distributed solutions (see Fig. 4). It can be seen that the distributed DCM is consistently better than its ECD homologue, except at the highest SNR (16 dB), where both models show the same squared error loss. In short, the estimation error on the neuronal parameters, as measured by the squared error loss is much smaller for the distributed DCM, which is less sensitive to noise and inaccurate priors than its ECD variant. In addition, we evaluated the quality of posterior confidence intervals: under the Laplace approximation; $q(\vartheta) = N(\mu_\vartheta, \Sigma_\vartheta)$, this reduces to

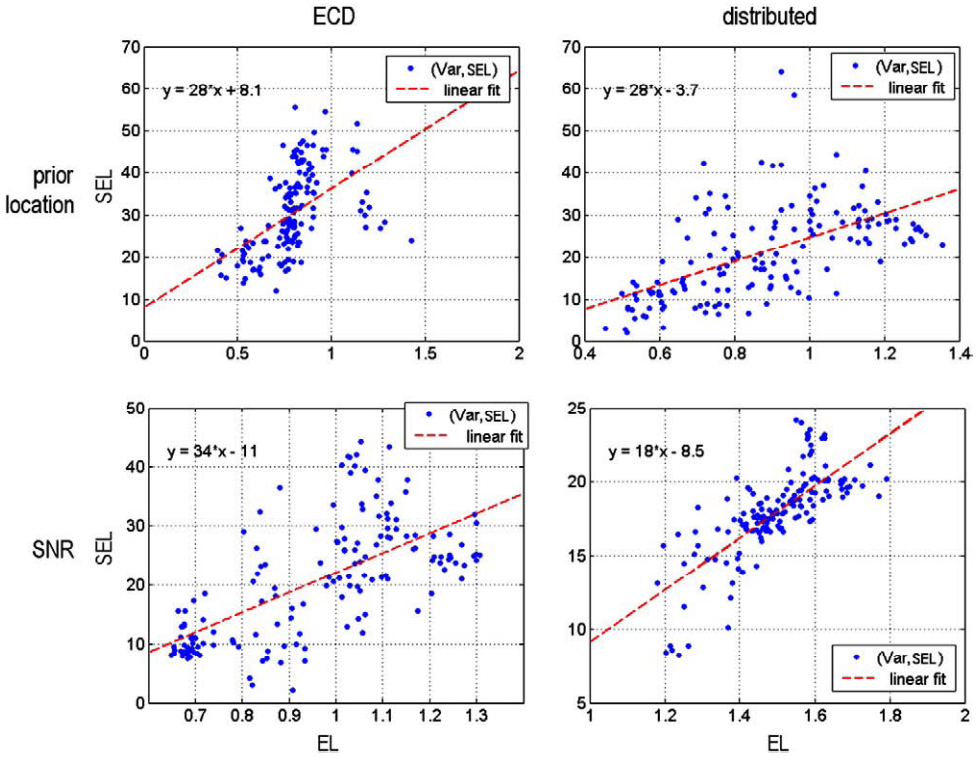

Fig. 5. Monte Carlo results—posterior confidence. Both graphs show the expected (x-axis) versus the observed (y-axis) squared error loss (SEL) for both series of simulations and spatial variants of the DCM. Top: changes in prior locations. Bottom: changes in signal to noise. Left: ECD-DCM. Right: distributed DCM. Although there is an order of magnitude difference between the predicted and observed SEL, they are strongly correlated. Note the increase in correlation for distributed DCMs over ECD-DCMs (bottom right).

assessing the accuracy of the posterior covariance in relation to the SEL, since;

$$EL(q) = \langle SEL(\vartheta)\rangle_q = tr(\Sigma_\vartheta). \quad (15)$$

where the expected loss EL(q) is the Bayesian estimator of SEL (see Robert, 1992). This equivalence means we can assess the posterior covariance in terms of the relationship between the expected and the sampled SEL; for both the ECD and distributed solutions. A good correlation between the expected EL(q) = $tr(\Sigma_\vartheta)$ and observed SEL means that the inference scheme is self-consistent; i.e., it adapts its level of confidence in proportion to the real (observed) estimation error. Fig. 5, shows the expected versus the observed SEL for both series of simulations and DCMs. Although there is an order of magnitude difference between the predicted and the observed SEL, they are strongly correlated. In addition, the correlation between EL (q) and SEL under different levels of noise is significantly higher for the distributed DCM.

In summary, the DCM of distributed responses is more robust to violations of priors and levels of noise; furthermore, it is more self-consistent in that the observed and expected estimation loss is more tightly coupled, relative to ECD models. We now turn to empirical comparisons, using the relative evidence for both models in real data.

Model comparisons using EEG data

In this section, we apply both ECD and distributed DCMs to the grand-mean responses from an eleven-subject auditory mismatch negativity study (Garrido et al., 2007). The term 'mismatch negativity' (MMN) describes an evoked response component elicited by the presentation of a rare auditory stimulus in a sequence of repetitive standard stimuli (Näätänen, 2003). The rare stimulus typically causes a more negative response. The difference between deviant and standard tone reaches a minimum at about 100 ms, and exhibits a second minimum later between 100 and 200 ms.

We first performed a conventional imaging source reconstruction to specify the underlying neuronal network, in terms of the number and prior expectations of source locations (Friston et al., 2007b). Fig. 6 shows the results of a source reconstruction for the first subject and highlights the prior source locations selected for the DCM analyses. This network is shown in Fig. 7a, which includes the extrinsic (between source) connections (cf. Garrido et al., 2007). In brief, we allowed for forward and backward connections between an early

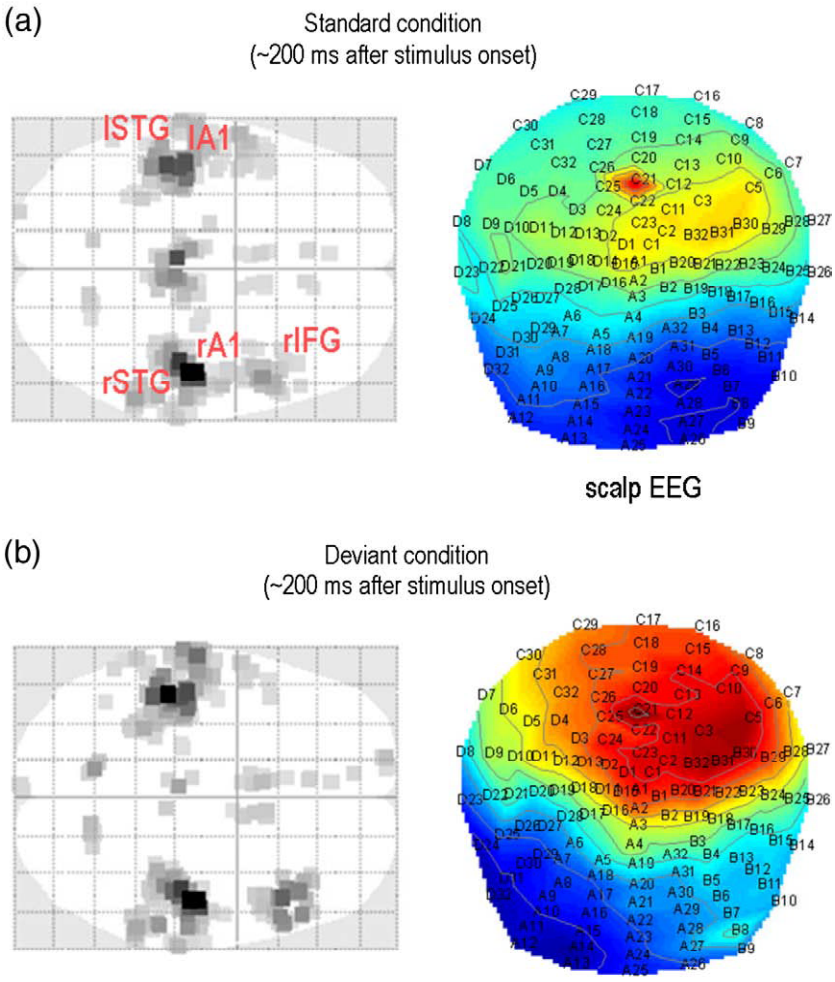

Fig. 6. Mismatch negativity study: scalp data and source reconstructions. The mismatch negativity (MMN) is the pattern elicited when contrasting a standard condition (repeated high-pitched tones) with a deviant condition (sparse low-pitched tones). This figure shows both the scalp topography and the corresponding source reconstructions at the time of the maximum difference (approx. 200 ms after onset). (a) Standard condition: the maximum intensity projection (MIP) on the source reconstruction shows five key sources: right/left primary auditory cortex (A1), right/left superior temporal gyrus (STG) and right inferior frontal gyrus (IFG). (b) Deviant condition: the MIP shows the same five sources (with different amplitudes).

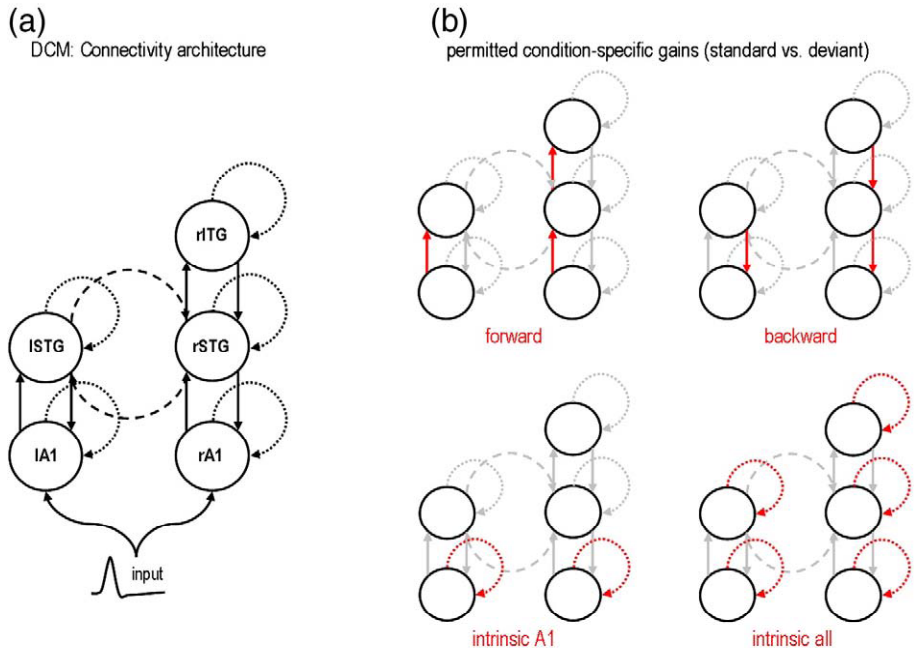

Fig. 7. Mismatch negativity study: DCM architectures. The five sources identified by source reconstruction MIP (see Fig. 6) were used to construct a DCM network as follows: (a) both primary auditory sources were coupled with forward and backward connections to ipsilateral STG sources. The latter were reciprocally connected through lateral connections, and right STG was coupled with forward and backward connections to rIFG. Within this graph, we compared eleven models, corresponding to different combinations of connectivity changes between the standard and the deviant conditions of the MMN paradigm. These condition-specific changes are depicted in (b): four sets of connections were allowed to change: forward connections, backward connections, intrinsic connections for bilateral A1, and all intrinsic connections. We then derived eleven DCMs from combinations of these four sets; "F", "B", "FB", "FI", "BI", "FBI", "FA", "BA", "FBA", and "0" (see Table 1).

bilateral auditory (rA1 and lA1) source and bilateral superior temporal gyrus (rSTG and lSTG) areas, as well as forward and backward connections between the right STG and a source located in the inferior frontal gyrus (rIFG). We also included transcallosal lateral connections between the STG sources.

The conventional understanding of the MMN rests on change-sensitive neuronal populations. In Kiebel et al., (2007), we considered two hypotheses, which explain the MMN either by adaptation or within a predictive coding framework (Friston, 2005; Garrido et al., 2008). We have shown that hypotheses like these can be formulated and tested using DCM, by allowing connections to change between the deviant and the standard conditions. In particular, we can test hypotheses about the mechanisms underlying the MMN by modelling the response evoked by a deviant using the same parameters as for the standard response, except for a gain in selected connections. Here, we repeat this analysis using both spatial variants of DCM.

Table 1 shows the different architectures we considered in terms of which connections were allowed to change (from the deviant to standard conditions). In brief, these different models correspond to different explanations for the MMN: the adaptation hypothesis (change in intrinsic connections) and the predictive coding hypothesis (change in intrinsic and extrinsic connections). We refer the interested reader to Kiebel et al., (2007):

Four sets of connections were allowed to change between the deviant and the standard conditions (see Fig. 7b); 'forward' implies permissible changes in all forward connections; 'backward', all backward connections; intrinsic A1′, changes in connectivity intrinsic to A1 and 'intrinsic all', all intrinsic connections. We then constructed eleven DCMs from combinations of these four basic differences, namely; "F", "B", "FB", "FI", "BI", "FBI", "FA", "BA", "FBA", and "0" (see Table 1). The last model precluded any changes between the two conditions and constitutes a null model.

In these comparative analyses, we also investigated the effect of changing the spatial support of the cortical regions in the distributed DCMs. This was achieved by varying the radius (1, 2, 4, 8, 16 and 32 mm) of the sphere (centred on the prior location), which defines the mesh vertices in each cortical region. We used the free-energy approximation to the log-evidence to compare the $11 \times 7 = 77$ models; eleven DCMs with seven spatial models (six distributed models with different spheres and one ECD model). The ensuing log-evidences are shown in Fig. 8. For almost all DCMs, the ECD models were significantly less likely than distributed DCMs ($\max F_{\text{distributed}} - \max F_{ECD} = 294.8$). Note that this model comparison automatically accommodates differences in model complexity. These differences were small because the ECD and the distributed and ECD-DCMs had the same number of parameters.

When the sphere radius in the distributed DCM is reduced, the DCMs have very similar log-evidences (Fig. 8a). Note also that the model evidence of the best DCM (FBA) for distributed DCM with small (1, 2, and 4 mm) spheres and the ECD model are very close to each other. This means that these spatial models converge when inverting a

Table 1
Condition-specific effects (standard versus deviant): gain in coupling strength.

	Forward	Backward	Intrinsic A1	Intrinsic all
F	X			
B		X		
FB	X	X		
FI	X		X	
BI		X	X	
FBI	X	X	X	
FA	X			X
BA		X		X
FBA	X	X		X
A				X
0				

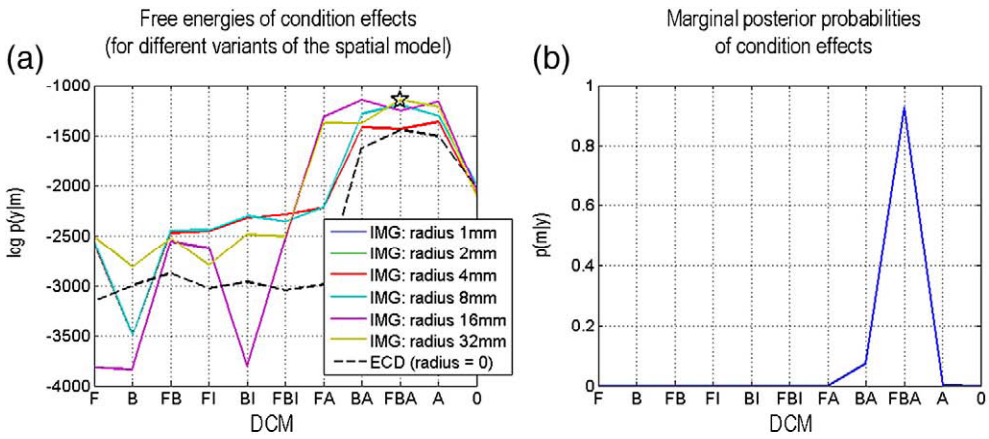

Fig. 8. Mismatch negativity study: Bayesian model comparison results. Bayesian model comparison was applied to an 11×7 factorial model space. Eleven condition effects (see Fig. 6 b) and 7 spatial variants of each DCM (6 distributed DCMs, with different cortical regions (sphere radii $\in 1, 2, 4, 8, 16$, and 32 mm) and one ECD-DCM. (a) Free-energies (log-evidences) for each of the 11×7 models. The star indicates that the best of all DCMs is a distributed FBA model (in which all connections were allowed to change) with the largest region. (b) Marginal posterior probabilities of the eleven DCMs (marginalising over spatial models).

well-specified neuronal model. In other words, there is no significant difference in model evidence between ECD and small patches, since the latter are approximated well by a single dipole. Fig. 8 also shows the marginal posterior probabilities of the eleven DCMs; marginalising over all spatial variants. This integrates out dependency on the spatial parameters and replicates the finding of Garrido et al., (2007) that the most plausible DCM seems to combine changes in forward, backward and intrinsic connections. For this DCM, there was strong evidence that the distributed DCM (all radii) was a better model than the ECD equivalent.

Discussion

We have described a variant of dynamic causal modelling for event-related potentials or fields as measured with EEG and MEG. We motivated this DCM as an approximation to a continuous neural-field model, using a mixture of overlapping patches, with compact spatial support, on the cortical surface. Time-varying activity in this mixture, caused by activity in other sources and experimental inputs, is propagated through appropriate lead-field or gain-matrices to generate observed channel data. In comparison to ECD variants of DCM, this distributed DCM has three advantages; it has greater face validity, the degrees of freedom of the spatial model can be specified (and therefore optimised using model selection) and the model is linear in the spatial parameters (which finesses computational load). Both our simulations and the application to an EEG auditory mismatch negativity dataset demonstrated the superiority of distributed DCMs, when compared to their ECD homologues.

The greater face validity of spatially distributed DCMs is similar to that of imaging source reconstruction solutions, when compared to ECD-like solutions: the spatial extent of each regional source must be modelled properly when inverting such models (see below). Furthermore, the neural-mass models we use (Jansen and Rit 1995) were designed originally to model mesoscopic electrocortical activity, at a spatial scale finer than that of EEG/MEG. Using simple approximations of neural-field models, we have proposed a simple modification of neural-mass models that render them able to emulate macroscopic spatiotemporal dynamics. Specifically, these modifications allow us to account for the spatial deployment of sources, which appears to be necessary to explain EEG/MEG data (see MMN results section).

Although not pursued here, the number of basis functions or different sizes of cortical regions could be optimised. One would repeat the inversion using different basis functions and evaluate the model evidences (as for the analysis of cortical sources in Fig. 8). This would allow one to optimise the degrees of freedom of the spatial model, in relation to the spatial information supported by the data; similarly for the size of the cortical patches used to model source-specific activity.

Note that there is a formal link between the spatially distributed DCM proposed in this work and EEG/MEG source reconstruction techniques (see e.g. Daunizeau et al., 2006; Friston et al., 2007b). The key difference between these two approaches rests on the formal constraints used by DCM. These constrain the temporal expression of source activity to conform to a biologically plausible time-course (Scherg and Von Cramon 1985). The interpretation of a DCM analysis is not usually concerned with the spatial profile of source activity but focuses on the coupling parameters and how they change with experimental manipulations. However, it is interesting to regard the DCM inversion as a biophysically and neurobiologically informed imaging source reconstruction (see Kiebel et al., 2006). In other words, one can regard the Bayesian inversion of spatially distributed DCM as a generalisation of classical forward model inversion used to reconstruct source activity from observed EEG or MEG data. The only difference between classical inversion and DCM is that the source activity has to conform to a biophysically plausible model. Generally, this model entails interactions among sources so that activity in one source is caused by activity in others. Classical forward models focus exclusively on the spatial observer function of the hidden states and ignore formal constraints on the temporal expression of source activity. The resulting spatial models are either ECD-based models or distributed source models of the sort used in image reconstruction (Baillet and Garnero 1997; Pascual-Marqui 2002; Phillips et al., 2005). Exactly the same distinction between ECD and distributed reconstructions can be applied in the context of DCM. In this note, we have described a distributed spatial model that complements existing ECD dynamic causal models.

In the future, it is possible that DCMs will be based on models that are closer to full neural-field models. These models might be more appropriate for EEG and MEG data because they account for continuous lateral interactions within each cortical region. Neural-field models can generate time-dependent dynamics that are expressed as bumps or propagating waves over the cortical surface. In this work, we truncated our space-time decomposition to the fundamental mode (a zero-order approximation). As a consequence, the neural-fields behave as interacting standing-waves; i.e. regionally specific invariant patterns of activity oscillating in response to mutual

influence. This space-time separation is a simplified variant of the sort of the spatiotemporal behaviours that could be obtained using a more realistic wave-equation (c.f. Eq. 6). Our zero-order approximation could be relaxed to increase the complexity of the neural-field model. This can be done by including more modes (see Eqs. 7 and 8 and Appendix 2). This would allow one to replace a full PDE to a set of coupled ODEs.

Two additional comments should be made: first, the derivation of the 2D neural field PDE relies on the assumption that lateral (isotropic) interactions are deployed over a small spatial scale (see Appendix 1). As a consequence, only long spatial wavelengths (relative to the spatial decay of lateral interactions) can be expressed in the 2D cortical neural field. This means that mesoscale phenomena like patchy feature maps (e.g. orientation preference or ocular dominance) in V1 might not be captured accurately (see Bressloff 2003 for a recent discussion of isotropic connectivity and Coombes et al., 2007 for an extension of the long-wavelength approximation to *patchy propagators*).

Second, we motivated our standing wave (fundamental mode) approximation to the neural field by noting that at high propagation velocity, higher harmonics will dissipate quickly. This is consistent with more realistic models (including axonal propagation), which also suggest that higher harmonics are damped more heavily (Nunez 1995). However, our standing wave approximation to experimentally manipulated (excited) neural fields is different in nature from the emergence of *global* standing-waves as proposed in Nunez and Srinivasan (2006). The latter global waves are thought to underlie global coherence of cortical activity in the absence of stimulation (e.g. eyes-closed resting alpha-band activity). Global standing-waves can be thought of as a resonance phenomenon, whose wavelength is related to the size of the brain. Nunez points out that mental tasks "enhance cell assembly activity [i.e. functional segregation], thereby reducing global field behaviour". This is in contradistinction to the present work, which postulates that local standing-waves emerge from the interaction of segregated neural ensembles. According to this view, segregation is *necessary* for the standing-waves to emerge, in the sense that it prevents activity spreading over the cortical mantle. In turn, this makes extrinsic functional integration (i.e. between region top-down and bottom-up effects, as opposed to within region lateral interactions) the principal mechanism responsible for sustained large-scale cortical activity.

Software note

All the routines and ideas described in this paper can be implemented with the academic freeware SPM8 (http://www.fil.ion.ucl.ac.uk/spm).

Acknowledgments

The Wellcome Trust funded this work. We would like to thank Marta Garrido for providing the EEG data and Marcia Bennett for invaluable help in preparing this manuscript. Also, we would like to thank the anonymous reviewers for their very helpful comments.

Appendices

Approximate 2D neural fields on the cortical manifold

Deriving the partial differential equation describing the spatiotemporal dynamics of neural fields from the underlying integrodifferential equation is a difficult problem because: (i) even on a Euclidean space, the Fourier analysis of the 2D neural field is not exact and (ii) on curved (Riemannian) manifolds, Euclidean distance measures do not apply. In this appendix, we discuss approximate solutions to these problems.

First, we consider the neural field unfolding on a planar surface tangential to the cortical manifold. Let $\mathbf{r}=(r,\theta)$ denote the 2D position (in polar coordinates) on this Euclidean space. Recall that the spatiotemporal convolution that operates on the input $\varsigma(r,t)$ is given by Eq. 4

$$\begin{aligned}\mu(\mathbf{r},t) &= G(r,t)*_{\varsigma}(\mathbf{r},t)\\ G(r,t) &= \delta\left(t-\frac{1}{c}r\right)\underbrace{\frac{1}{\gamma}\exp\left(-\frac{1}{\gamma}r\right)}_{\text{lateral interaction}}\end{aligned} \tag{A1}$$

where * denotes convolution and G is the convolution kernel, whose spatial scale is controlled by the decay of lateral interactions γ. Therefore, if the Fourier transform $\tilde{G}(k,\omega)$ of G can be represented in the rational form $\tilde{G}(k,\omega)=R(k^2,i\omega)/P(k^2,i\omega)$, we have $P(k^2,i\omega)G(k,\omega)=R(k^2,i\omega)\zeta(k,\omega)$. By identifying $k^2\leftrightarrow-\nabla^2$ and $i\omega\leftrightarrow\partial/\partial t$, an inverse Fourier transform will yield the PDE in terms of spatial and temporal derivatives (see Coombes et al., 2007). Given the functional form of the lateral interactions, Liley et al. (2002) proposed an expansion of $P(k^2,i\omega)$ near $k=0$, yielding the "long-wavelength" approximation:

$$\left(\frac{\partial^2}{\partial t^2}+2\kappa\frac{\partial}{\partial t}+\kappa^2-\frac{3}{2}c^2\nabla^2\right)\mu(\mathbf{r},t)=c\kappa\varsigma(\mathbf{r},t) \tag{A2}$$

where $\kappa=c/\gamma$. This expression is very close but not identical to that obtained more simply (and exactly) for 1D neural fields (e.g. Deco et al., 2008). The long-wavelength approximation basically implies that $k\ll 1/\gamma$, i.e. $\Delta r\gg\gamma$, where Δr is the typical spatial wavelength. This is not a critical assumption when modelling EEG scalp data; since the head volume conductor acts as a low-pass spatial filter such that scalp potentials are dominated by the long-wavelength components engendered by cortical sources (Nunez and Srinivasan 2006).

The PDE (A2) derives from the spatially invariant form of the Green function above (A1), which, on a 2D Riemannian manifold is:

$$G(\mathbf{r},\mathbf{r}',t)=\delta\left(t-\frac{d(\mathbf{r},\mathbf{r}')}{c}\right)\frac{1}{\gamma}\exp\left(-\frac{d(\mathbf{r},\mathbf{r}')}{\gamma}\right), \tag{A3}$$

where $\mathbf{r}$(resp. $\mathbf{r}'$) is the position on the cortical surface of the target (resp. source) neuron of the neural field, and $d(\mathbf{r},\mathbf{r}')$ is the distance metric on the cortical manifold. Note that the cortical manifold more precisely, each hemisphere) is homotopic to a sphere, which means that its metric is well-behaved and that local geographic (angular) coordinates can be defined on the cortical mantle (Toro and Burnod 2003). This implies that a patch of the cortical mantle is homotopic to an open set in R^2, where there is a bijective mapping from the angular coordinates to the Euclidean polar coordinates above. If the fall-off distance γ is small compared to the inverse curvature (smoothness) of the manifold Δ, the Green function can be approximated by a spatially invariant convolution kernel:

$$G(\mathbf{r},\mathbf{r}',t)\overset{\gamma\ll\Delta}{\longrightarrow}\delta\left(t-\frac{|\mathbf{r}-\mathbf{r}'|}{c}\right)\frac{1}{\gamma}\exp\left(-\frac{|\mathbf{r}-\mathbf{r}'|}{\gamma}\right)\equiv G(\mathbf{r}-\mathbf{r}',t). \tag{A4}$$

This is because contributions from the manifold that diverge from the tangent surface (e.g. neighbouring sulci) will be negligible. Note that this "short-scale lateral connectivity" is required because the assumption of isotropic lateral interaction posits that the boundary effects are negligible far from boundaries. Moreover, it justifies the above "long-wavelength" approximation, in the sense that such a system can almost only resonate at long-wavelength harmonics (relative to γ). One can think of the short-scale lateral connectivity in a 1D neural field as a string of interacting (infinitesimal) elements (e.g., under tension). The energy required to excite the string is proportional to the frequency of its harmonics, which means that higher harmonics (short wavelengths) will decay quickly. These considerations mean that under "short-scale lateral connectivity", the

PDE obtained from the lateral interaction function above (A4) is approximately valid.

Approximate solutions for neural-fields

The neuronal activity of a cortical layer or population can be modelled as a neural-field $\mu^{(i)}(r,t)$ that satisfies the following partial differential equation (PDE):

$$\left(\frac{\partial^2}{\partial t^2} + 2\kappa\frac{\partial}{\partial t} - \frac{3c^2}{2}\nabla^2\right)\mu^{(i)}(\mathbf{r},t) = f^{(i)}(\mathbf{r},t)$$
$$f^{(i)}(\mathbf{r},t) = c\kappa\varsigma^{(i)}(\mathbf{r},t) - \kappa^2\mu^{(i)}(\mathbf{r},t) \tag{A5}$$

Here, $\mu^{(i)}(\mathbf{r},t)$ is the field associated with the i-th population. In (A5), we have lumped input and decay terms into $f^{(i)}(\mathbf{r},t)$. Using separation of variables, we can express the solution of this PDE as:

$$\mu^{(i)}(\mathbf{r},t) = \sum_k v_k^{(i)}(t)w_k^{(i)}(\mathbf{r}) \tag{A6}$$

where the spatial modes $w_k^{(i)}(\mathbf{r})$ are the solutions to the eigenvalue problem:

$$\nabla^2 w_k^{(i)} + \lambda_k^{(i)} w_k^{(i)} = 0. \tag{A7}$$

Note that the self-adjoint property of the Laplacian operator ∇^2 ensures the eigenvalues $\lambda_k^{(i)}$ are real and the spatial eigenvectors or modes $w_k^{(i)}(\mathbf{r})$ are real and orthonormal

$$\int w_k^{(i)}(\mathbf{r})w_{k'}^{(i)}(\mathbf{r})d\mathbf{r} = \delta_{kk'}. \tag{A8}$$

This orthonormal property means we can express the eigenfunctions $v_k^{(i)}(t)$ as

$$v_k^{(i)}(t) = \int w_k^{(i)}(\mathbf{r})\mu^{(i)}(\mathbf{r},t)d\mathbf{r}. \tag{A9}$$

If we differentiate (A9) with respect to time and use (A5) to eliminate $\dot{\mu}^{(i)}$, we obtain

$$\begin{aligned}\dot{v}_k^{(i)}(t) &= \int w_k^{(i)}(\mathbf{r})\,\dot{\mu}^{(i)}(\mathbf{r},t)d\mathbf{r} \\ &= \frac{1}{2\kappa}\left(-\int w_k^{(i)}\,\ddot{\mu}d\mathbf{r} + \frac{3}{2}c^2\int w_k^{(i)}\nabla^2\mu^{(i)}d\mathbf{r} + \int w_k^{(i)}fd\mathbf{r}\right).\end{aligned} \tag{A10}$$

From the form of the solution (A6) and (A7), we have

$$\begin{aligned}\ddot{\mu}^{(i)}(\mathbf{r},t) &= \sum_k \ddot{v}_k(i)(t)w_k^{(i)}(\mathbf{r}) \\ \nabla^2\mu^{(i)}(\mathbf{r},t) &= \sum_k v_k^{(i)}(t)\nabla^2 w_k^{(i)}(\mathbf{r}) = \sum_k \lambda_k^{(i)}v_k^{(i)}(t)w_k^{(i)}(\mathbf{r}).\end{aligned} \tag{A11}$$

The input $f^{(i)}(\mathbf{r},t)$ and its stimulus-related component can also be expressed as a transform pair

$$\begin{aligned}f^{(i)}(\mathbf{r},t) &= \sum_k f_k^{(i)}(t)w_k^{(i)}(\mathbf{r}) \\ f_k^{(i)}(t) &= \int w_k^{(i)}(\mathbf{r})f^{(i)}(\mathbf{r},t)d\mathbf{r} \\ u(\mathbf{r},t) &= \sum_k u_k(t)w_k^{(i)}(\mathbf{r}) \\ u_k(t) &= \int w_k^{(i)}(\mathbf{r})u(\mathbf{r},t)d\mathbf{r}.\end{aligned} \tag{A12}$$

Substituting these expressions into (A10), we can use the orthogonality of the modes (A8) to eliminate terms that depend on r and express the temporal dynamics as an ODE

$$\begin{aligned}\dot{v}_k^{(i)}(t) &= \frac{1}{2\kappa}\left(-\sum_{k'}\delta_{kk'}\ddot{v}_{k'}^{(i)}(t) + \frac{3}{2}c^2\sum_{k'}\delta_{kk'}\lambda_{k'}^{(i)}v_{k'}^{(i)}(t) + \sum_{k'}\delta_{kk'}f_{k'}^{(i)}(t)\right) \\ &= \frac{1}{2\kappa}\left(-\ddot{v}_k^{(i)}(t) + \frac{3}{2}c^2\lambda v_k^{(i)}(t) + f_k^{(i)}(t)\right)\end{aligned} \tag{A13}$$

Rearranging (A13) and substituting for $f_k^{(i)}(t)$ from (A5) and (A12) gives

$$\begin{aligned}&\frac{\partial^2}{\partial t^2}v_k^{(i)}(t) + 2\kappa\frac{\partial}{\partial t}v_k^{(i)}(t) + \left(\kappa^2 - \frac{3}{2}\lambda_k^{(i)}c^2\right)v_k^{(i)}(t) \\ &\quad = c\kappa\int w_k^{(i)}(\mathbf{r})\varsigma^{(i)}(\mathbf{r},t)d\mathbf{r}.\end{aligned} \tag{A14}$$

Recall that ρ is the parameter of the sigmoid activation function $S(\mu^{(i)})$ in Eq. 3. We can further simplify the expression for the input using the first-order approximation $S(\mu^{(i)}) \approx \rho\mu^{(i)}(r,t)$ to give:

$$\begin{aligned}\varsigma^{(i)}(\mathbf{r},t) &= \gamma_{iu}u(\mathbf{r},t) + \sum_j \gamma_{ij}S\left(\mu^{(j)}(\mathbf{r},t)\right) \\ u(\mathbf{r},t) &= \sum_k u_k(t)w_k^{(i)}(\mathbf{r}) \\ S\left(\mu^{(i)}\right) &= \rho\mu^{(j)}(\mathbf{r},t) = \rho\sum_k v_k^{(j)}(t)w_k^{(j)}(\mathbf{r}).\end{aligned} \tag{A15}$$

Here, we have made the simplifying assumption that all the populations have the same spatial support and modes; and that the coupling between layers, γ_{ij} is uniform. This allows us to further approximate the input for the k-th mode with:

$$\begin{aligned}\int w_k^{(i)}(\mathbf{r})\varsigma^{(i)}(\mathbf{r},t)d\mathbf{r} &\approx \gamma_{iu}\sum_{k'}\delta_{kk'}u_{k'}(t) + \sum_j\gamma_{ij}\rho\sum_{k'}\delta_{kk'}v_{k'}^{(j)}(t) \\ &\approx \gamma_{iu}u_k(t) + \sum_j\gamma_{ij}S\left(v_k^{(j)}(t)\right)\end{aligned} \tag{A16}$$

In effect, the modes are uncoupled by their orthogonality. This means the mean-field effects are only communicated within, not between modes. Substituting (A16) into (A14) gives an approximate ODE for the temporal expression of each mode:

$$\begin{aligned}\ddot{v}_k^{(i)}(t) + 2\kappa\dot{v}_k^{(i)}(t) + \left(\kappa^2 - \frac{3}{2}\lambda_k^{(i)}c^2\right)v_k^{(i)}(t) &= c\kappa\varsigma_k^{(i)}(t) \\ \varsigma_k^{(i)}(t) &= \gamma_{iu}u_k(t) + \sum_j\gamma_{ij}S\left(v_k^{(j)}(t)\right).\end{aligned} \tag{A17}$$

If we retain only the fundamental mode, then $\lambda_k^{(i)} = \lambda_0^{(i)} = 0$ and (A17) has exactly the same form as the neural-mass model in Eq. 3 but where the mean depolarisation is replaced by the eigenfunction of each mode. This also describes the fluctuations of the standing-wave in Eq. 9.

More generally, when different populations have different spatial modes (i.e., populations in different cortical regions), one would have to replace γ_{ij} with $\Gamma_{kk'}^{(ij)}$ and sum over modes and populations; where the parameter $\Gamma_{kk'}^{(ij)}$ couples the k-th mode of the i-th population to the k'-th mode in population j. This parameter can model inhomogeneous extrinsic connections that couple spatial modes in different parts of the brain; this is the implicit meaning of $\gamma_{ij} = \Gamma_{00}^{(ij)}$ in the main text.

References

Amari, S., 1995. Homogeneous nets of neuron-like elements. Biol. Cyber. 17, 211–220.

Baillet, S., Garnero, L., 1997. A Bayesian approach to introducing anatomo-functional priors in the EEG/MEG inverse problem. IEEE Trans. Biomed. Eng. 44 (5), 374–385.

Breakspear, M., Stam, C.J., 2005. Dynamics of a neural system with a multiscale architecture. Phil. Trans. R. Soc. B 360, 1051–1074.

Bressloff, P.C., 2003. Spatially periodic modulation of cortical patterns by long-range horizontal connections. Physica D 185, 131–157.

Coombes, S., Venkov, N.A., Shiau, L., Bojak, I., Liley, D.T.J., Laing, C.R., 2007. Modeling electrocortical activity through local approximations of integral neural-field equations. Phys. Rev. E 76, 1539–1547.

Connors, B.W., Amitai, Y., 1993. Generation of epileptiform discharges by local circuits in neocortex. In: Scwartzkroin, P.A. (Ed.), Epilepsy: Models, Mechanisms and Concepts. Cambridg University Press, pp. 388–424.
Daunizeau, J., Friston, K.J., 2007. A mesostate-space model for EEG and MEG. NeuroImage 38 (1), 67–81.
Daunizeau, J., Mattout, J., Clonda, D., Goulard, B., Benali, B., Lina, J.M., 2006. Bayesian spatio-temporal approach for EEG sources reconstruction: conciliating ECD and distributed models. IEEE Trans. Biomed. Eng. 53, 503–516.
David, O., Friston, K.J., 2003. A neural-mass model for MEG/EEG: coupling and neuronal dynamics. NeuroImage 20 (3), 1743–1755.
David, O., Harrison, L., Friston, K.J., 2005. Modelling event-related responses in the brain. NeuroImage 25 (3), 756–770 Apr 15.
David, O., Kiebel, S.J., Harrison, L.M., Mattout, J., Kilner, J.M., Friston, K.J., 2006. Dynamic causal modeling of evoked responses in EEG and MEG. NeuroImage 30 (4), 1255–1272.
Deco, G., Jirsa, V.K., Robinson, P., Breakspear, M., Friston, K., 2008. The dynamic brain: from spiking neurons to neural-masses and cortical fields. Plos. Comp. Biol. 4 (8), e1000092.
Ermentrout, G.B., Cowan, J.D., 1979. A mathematical theory of visual hallucination patterns. Biol. Cyber. 34, 137–150.
Felleman, D.J., Van Essen, D.C., 1991. Distributed hierarchical processing in the primate cerebral cortex. Cereb. Cortex 1, 1–47.
Friston, K., 2005. A theory of cortical responses. Phil. Trans. R. Soc. B 360, 815–836.
Friston, K.J., Harrison, L., Penny, W., 2003. Dynamic causal modelling. NeuroImage 19 (4), 1273–1302.
Friston, K.J., Mattout, J., Trujillo-Barreto, N., Ashburner, J., Penny, W., 2007a. Variational free energy and the Laplace approximation. NeuroImage 34 (1), 220–234.
Friston, K., Harrison, L., Daunizeau, J., Kiebel, S., Phillips, C., Trujillo-Barreto, N., Henson, R., Flandin, G., Mattout, J., 2007b. Multiple sparse priors for the M/EEG inverse problem. NeuroImage 39 (3), 1104–1120.
Fuchs, M., Wagner, M., Kohler, T., Wischmann, H.A., 1999. Linear and nonlinear current density reconstructions. J. Clin. Neurophysiol. 16 (3), 267–295.
Garrido, M.I., Kilner, J.M., Kiebel, S.J., Stephan, K.E., Friston, K.J., 2007. Dynamic causal modelling of evoked potentials: a reproducibility study. NeuroImage 36 (3), 571–580.
Garrido, M.I., Friston, K.J., Kiebel, K.J., Stephan, K.E., Baldeweg, T., Kilner, J.M., 2008. The functional anatomy of the MMN: a DCM study of the roving paradigm. Neuroimage 42, 936–944.
Hutt, A., Atay, F.M., 2005. Analysis of nonlocal neural-fields for both general and gamma-distributed connectivities. Physica D 203, 30–54.
Jansen, B.H., Rit, V.G., 1995. Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. Biol Cybern 73, 357–366.
Jirsa, V., Haken, H., 1996. Field theory of electromagnetic brain activity. Phys. Rev. Letters 77 (5), 960–963.
Jirsa, V.K., Kelso, J.A.S., 2000. Spatiotemporal pattern formation in neural systems with heterogeneous connection topologies. Phys. Rev. E 62 (6), 8462–8465.
Jirsa, V.K., Jantzen, K.J., Fuchs, A., Kelso, J.A.S., 2002. Spatiotemporal forward solution of the EEG and MEG using network modeling. IEEE Trans. Med. Imag. 21 (3), 493–504.
Kiebel, S.J., David, O., Friston, K.J., 2006. Dynamic causal modelling of evoked responses in EEG/MEG with lead field parameterization. NeuroImage 30 (4), 1273–1284.
Kiebel, S.J., Garrido, M.I., Friston, K.J., 2007. Dynamic causal modelling of evoked responses: the role of intrinsic connections. NeuroImage 36 (2), 332–345.
Liley, D.T.J., Cadush, P.J., Dafilis, M.P., 2002. A spatially continuous mean field theory of electrocortical activity. Network: Comput. Neural Syst. 13, 67–113.
Mattout,, J., Henson, R.N., Friston,, K.J., 2007. Canonical source reconstruction for MEG. Computational Intelligence and Neuroscience Article ID 67613 doi:10.1155/2007/67613.
Marreiros, A.C., Kiebel, S.J., Friston, K.J., 2008a. Dynamic causal modelling for fMRI: A two-state model. NeuroImage 39 (1), 269–278.
Marreiros, A.C., Daunizeau, J., Kiebel, S.J., Friston, K.J., 2008b. Population dynamics: variance and the sigmoid activation function. Neuroimage 42, 146–157.
Marreiros, A.C., Kiebel, S., Daunizeau, J., Harrison, L., Friston, K.J., 2008c. Population dynamics under the Laplace assumption. Neuroimage 44, 701–714.
Moran, R.J., Kiebel, S.J., Stephan, K.E., Reilly, R.B., Daunizeau, J., Friston, K.J., 2007. A neural-mass model of spectral responses in electrophysiology. NeuroImage 37 (3), 706–720.
Näätänen, R., 2003. Mismatch negativity: clinical research and possible applications. Int. J. Psychophysiol. 48, 179–188.
Nunez, P.L., 1974. The brain wave-equation: a model for the EEG. Math. Biosci. 21, 279–297.
Nunez, P.L., 1995. Neocortical Dynamics and Human EEG Rhythms. Oxford University Press, New York.
Nunez, P.L., Srinivasan, R., 2006. A theoretical basis for standing and travelling brain waves measured with human EEG with implications for an integrated consciousness. Clin. Neurophysiol. 11, 2424–2435.
Pascual-Marqui, R.D., 2002. Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. Methods Find Exp. Clin. Pharmacol. 24 Suppl. D, 5–12.
Phillips, C., Rugg, M.D., Friston, K.J., 2002. Anatomically informed basis functions for EEG source localization: combining functional and anatomical constraints. NeuroImage 16 (3 Pt 1), 678–695.
Phillips, C., Mattout, J., Rugg, M.D., Maquet, P., Friston, K.J., 2005. An empirical Bayesian solution to the source reconstruction problem in EEG. NeuroImage. 24, 997–1011.
Qubbaj, M.R., Jirsa, V.K., 2007. Neural-field dynamics with heterogeneous connection topology. Phys. Rev. Letters 98 (23), 238102.
Robert, C., 1992. L'analyse statistique Bayesienne, Ed. Economica.
Robinson, P.A., Rennie, C.J., Wright, J.J., 1997. Propagation and stability of waves of electrical activity in the cerebral cortex. Phys. Rev. E 56 (1), 826.
Robinson, P.A., Loxley, P.N., O, Connor, S.C., Rennie, C.J., 2001. Modal analysis of corticothalamic dynamics, electroencephalographic spectra, and evoked potentials. Phys. Rev. E 63, 041909.
Scherg, M., Von Cramon, D., 1985. Two bilateral sources of the late AEP as identified by a spatio-temporal dipole model. Electroencephalogr. Clin. Neurophysiol. 62 (1), 32–44 Jan.
Talairach, J., Tournoux, P., 1988. Co-Planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System—An Approach to Cerebral Imaging. Thieme Medical Publishers, New York, NY.
Toro, R., Burnod, Y., 2003. Geometric atlas: modelling the cortex as an organized surface. Neuroimage 20 (3), 1468–1484.
Wendling, F., Bellanger, J.J., Bartolomei, F., Chauvel, P., 2000. Relevance of nonlinear lumped-parameter models in the analysis of depth-EEG epileptic signals. Biol. Cybern. 83, 367–378.
Wennekers, T., 2008. Tuned solutions in dynamic neural-fields as building blocks for extended EEG models. Cogn. Dynamics 2 (2), 137–146.