

# Dynamic causal modelling for EEG and MEG

Stefan J. Kiebel · Marta I. Garrido ·  
Roselyn J. Moran · Karl J. Friston

Received: 9 January 2008 / Accepted: 28 February 2008 / Published online: 23 April 2008  
© The Author(s) 2008

**Abstract** Dynamic Causal Modelling (DCM) is an approach first introduced for the analysis of functional magnetic resonance imaging (fMRI) to quantify effective connectivity between brain areas. Recently, this framework has been extended and established in the magneto/encephalography (M/EEG) domain. DCM for M/EEG entails the inversion a full spatiotemporal model of evoked responses, over multiple conditions. This model rests on a biophysical and neurobiological generative model for electrophysiological data. A generative model is a prescription of how data are generated. The inversion of a DCM provides conditional densities on the model parameters and, indeed on the model itself. These densities enable one to answer key questions about the underlying system. A DCM comprises two parts; one part describes the dynamics within and among neuronal sources, and the second describes how source dynamics generate data in the sensors, using the lead-field. The parameters of this spatiotemporal model are estimated using a single (iterative) Bayesian procedure. In this paper, we will motivate and describe the current DCM framework. Two examples show how the approach can be applied to M/EEG experiments.

**Keywords** Magnetoencephalography · Electroencephalography · Dynamic system · Connectivity · Bayesian

## Introduction

In the following, we first highlight several new methodological developments, which we believe are important for new approaches to M/EEG analysis. This introductory motivation is intended to be general. The key components will be reprised using concrete examples of Dynamic Causal Modelling (DCM) later.

The analysis of MEG and EEG data can be approached from various angles. To the un-initiated and expert researcher alike, the diversity of methods can be simply breath-taking. However, most researchers typically avoid switching among methods, because exploration of the methods landscape can incur a high cost. Rather, most M/EEG laboratories have adopted a dual strategy: Firstly, experiments are analyzed using some ‘safe’ strategy based on a kernel of robust and widely accepted methods. This is usually the method of choice for analysis and publication, for example (Picton et al. 2000). Secondly, new methods are evaluated when they look interesting and if there are enough resources for doing so. This strategy might be considered as an ideal mixture of exploitation and exploration. But why is this approach so endemic in the M/EEG field? Take fMRI analysis, after an initial decade of methods exploration, most groups nowadays agree on the main methodological issues. For example, in a recent special issue of the journal *Human Brain Mapping*, a comparison of a dozen different approaches highlighted the use of common analysis strategies (Poline et al. 2006).

In M/EEG there does not seem to be the same level of consensus. There are many analysis schemes available, which look at different components of the signal: analysis of evoked responses (Kiebel and Friston 2004; van Wassenhove et al. 2005), of single trials using multivariate techniques like independent component analysis, analysis

---

S. J. Kiebel (✉) · M. I. Garrido · R. J. Moran · K. J. Friston  
The Wellcome Trust Centre for Neuroimaging, Institute of  
Neurology, University College London, 12 Queen Square,  
London WC1N 3AR, UK  
e-mail: skiebel@fil.ion.ucl.ac.uk

of oscillations using time-frequency power or coherence analysis (Friston et al. 2006; Gross et al. 2001; Liljestrom et al. 2005; Makeig et al. 2002). These analyses can proceed in sensor or brain-space based on source reconstruction using equivalent current dipoles (ECDs) or imaging solutions (Baillet and Garnero 1997; Daunizeau et al. 2006, 2007; Jun et al. 2005; Mattout et al. 2006). Variations and mixtures of all these approaches exist. There might be a simple reason for this diversity of methods: M/EEG data contain much more information about underlying neuronal dynamics than the fMRI signal (Daunizeau et al. 2007). The underlying dynamics, while still largely a mystery, confer great potential on M/EEG. Classical M/EEG analysis methods usually try to reduce the amount of temporal detail, for example by averaging over temporal windows and channels. This is an appropriate strategy for extracting behaviourally relevant features from the data (Rugg and Curran 2007). However, averaging, in particular in sensor-space, also moves the analysis away from the brain enforcing inferences about summary measures (Makeig et al. 2002). There is not only uncertainty about how data should be analyzed but also about how these signals are generated and what they tell us about the underlying system. Therefore, a key topic in M/EEG methods research is to identify models that describe the mapping from the underlying neuronal system to the observed M/EEG response, which incorporate known or assumed constraints (David and Friston 2003; Sotero et al. 2007). Using biophysically and neuronally informed forward models means we can use the M/EEG to make explicit statements about the underlying biophysical and neuronal parameters (David et al. 2006).

In the past years, there have been three important developments in M/EEG analysis. The first is that standard computers are now powerful enough to perform sophisticated analyses in a routine fashion (David et al. 2006). These analyses would have been impractical ten years ago, even for low-density EEG measurements. Secondly, the way methods researchers describe their M/EEG models has changed dramatically in the last decade. Recent descriptions tend to specify the critical assumptions underlying the model, followed by the inversion technique. This is useful because models for M/EEG can be complex; specifying the model explicitly also makes a statement about how one believes data were generated (Daunizeau and Friston 2007). This makes model development more effective and transparent because fully specified models can be compared to other models.

The third substantial advance is the advent of Empirical or hierarchical Bayesian approaches to M/EEG model inversion. Bayesian approaches are important, because they allow for the introduction of constraints that ensure robust parameter estimation, for example (Auranen et al. 2007;

Nummenmaa et al. 2007; Penny et al. 2007; Zumer et al. 2007). This is vital once the model is complex enough to generate ambiguities (conditional dependencies) among groups of parameters. One could avoid correlations among parameter estimates by avoiding complex models. However, this would preclude further research into the mechanisms behind the M/EEG. An empirical Bayesian formulation allows the data to resolve these ambiguities and uncertainties. The traditional argument against the use of Bayesian methods is that the priors introduce ‘artificial’ or ‘biased’ information not solicited by the data. Essentially, the claim is that the priors enforce solutions, which are desired by the researcher. This argument can be discounted for three reasons: (i) In Empirical Bayes the weight afforded by the priors is determined by the data, not the analyst. (ii) Bayesian analysis provides the posterior distribution, which encodes uncertainty about the parameters, after observing the data. If the posterior is similar to the prior, then the data do not contain sufficient information to enable qualitative inference. This can be tested explicitly using the model evidence; the fact that a parameter cannot be resolved is usually informative in itself. (iii) Usually, Bayesian analysis explores a selection of models, followed by model comparison (Garrido et al. 2007). For example, one can invert a model derived from one’s favourite cognitive neuroscience theory, along with other alternative models. The best model can then be found by comparing model evidences (see below) using standard decision criteria (Penny et al. 2004).

In summary, we argue that the combination of these developments allow for models that are sophisticated enough to capture the full richness of the data. The Bayesian approach is central to this new class of models, without which is not possible to constrain complex models or deal with inherent correlations among parameter estimates. Bayesian model comparison represents the important tool of selecting the best among competing models, which is central to the scientific process.

Dynamic Causal Modelling provides a generative spatiotemporal model for M/EEG responses. The idea central to DCM is that M/EEG data are the response of a dynamic input–output system to experimental inputs. It is assumed that the sensory inputs are processed by a network of discrete but interacting neuronal sources. For each source, we use a neural mass model, which describes responses of neuronal subpopulations. Each population has its own (intrinsic) dynamics governed by the neural mass equations, but also receives extrinsic input, either directly as sensory input or from other sources. The whole set of sources and their interactions are fully specified by a set of first-order differential equations that are formally related to other neural mass models used in computational models of M/EEG (Breakspear et al. 2006; Rodrigues et al. 2006). We assume that the depolarization of pyramidal cell populations gives rise to

observed M/EEG data; one specifies how these depolarizations are expressed in the sensors through a conventional lead-field. The full, spatiotemporal model takes the form of a nonlinear state-space model with hidden states modelling (unobserved) neuronal dynamics, while the observation (lead-field) equation is instantaneous and linear in the states. In other words, the model consists of a temporal and spatial part with temporal (e.g., connectivity between two sources) and spatial parameters (e.g., lead-field parameters, like ECD locations). In the next section, we describe the DCM equations and demonstrate how the ensuing model is inverted using Bayesian techniques. We illustrate inference using evoked responses from a multi-subject data set. We also introduce a recent addition to the DCM framework, which can be used to make inferences about M/EEG steady-state responses. We conclude with a discussion about current DCM algorithms and point to some promising future developments.

### Dynamic Causal Modelling—theory

Intuitively, the DCM scheme regards an experiment as a designed perturbation of neuronal dynamics that are promulgated and distributed throughout a system of coupled anatomical sources to produce region-specific responses. This system is modelled using a dynamic input–state–output system with multiple inputs and outputs. Responses are evoked by deterministic inputs that correspond to experimental manipulations (i.e., presentation of stimuli). Experimental factors (i.e., stimulus attributes or context) can also change the parameters or causal architecture of the system producing these responses. The state variables cover both the neuronal activities and other neurophysiological or biophysical variables needed to form the outputs. Outputs are those components of neuronal responses that can be detected by MEG/EEG sensors. In our model, these components are depolarizations of a ‘neural mass’ of pyramidal cells. DCM starts with a reasonably realistic neuronal model of interacting cortical regions. This model is then supplemented with a spatial forward model of how neuronal activity is transformed into measured responses, here, M/EEG scalp-averaged responses. This enables the parameters of the neuronal model (e.g., effective connectivity) to be estimated from observed data. For M/EEG data, this spatial model is a forward model of electromagnetic measurements that accounts for volume conduction effects (Mosher et al. 1999).

#### Hierarchical MEG/EEG neural mass model

DCMs for M/EEG adopt a neural mass model (David and Friston 2003) to explain source activity in terms of the

ensemble dynamics of interacting inhibitory and excitatory subpopulations of neurons, based on the model of Jansen and Rit (1995). This model emulates the activity of a source using three neural subpopulations, each assigned to one of three cortical layers; an excitatory subpopulation in the granular layer, an inhibitory subpopulation in the supra-granular layer and a population of deep pyramidal cells in the infra-granular layer. The excitatory pyramidal cells receive excitatory and inhibitory input from local interneurons (via intrinsic connections, confined to the cortical sheet), and send excitatory outputs to remote cortical sources via extrinsic connections. See also Grimbert and Faugeras (2006) for a bifurcation analysis of this model.

In David et al. (2005), we developed a hierarchical cortical model to study the influence of forward, backward and lateral connections on evoked responses. This model embodies directed extrinsic connections among a number of sources, each based on the Jansen model (Jansen and Rit 1995), using the connectivity rules described in Felleman and Van Essen (1991). Using these rules, it is straightforward to construct any hierarchical cortico-cortical network model of cortical sources. Under simplifying assumptions, directed connections can be classified as: (i) Bottom-up or forward connections that originate in the infragranular layers and terminate in the granular layer. (ii) Top-down or backward connections that connect infragranular to agranular layers. (iii) Lateral connections that originate in infragranular layers and target all layers. These long-range or extrinsic cortico-cortical connections are excitatory and are mediated through the axonal processes of pyramidal cells. For simplicity, we do not consider thalamic connections, but model thalamic afferents as a function encoding subcortical input (see below).

The Jansen and Rit model emulates the MEG/EEG activity of a cortical source using three neuronal subpopulations. A population of excitatory pyramidal (output) cells receives inputs from inhibitory and excitatory populations of interneurons, via intrinsic connections. Within this model, excitatory interneurons can be regarded as spiny stellate cells found predominantly in layer 4 and in receipt of forward connections. Excitatory pyramidal cells and inhibitory interneurons occupy agranular layers and receive both intrinsic and extrinsic backward and lateral inputs. The ensuing DCM is specified in terms of its state-equations and an observer or output equation

$$\begin{aligned}\dot{x} &= f(x, u, \theta) \\ h &= g(x, \theta)\end{aligned}\tag{1}$$

where  $x$  are the neuronal states of cortical sources,  $u$  are exogenous inputs, and  $h$  is the system’s response.  $\theta$  are quantities that parameterize the state and observer

equations (see also below under ‘Prior assumptions’). The state-equations are ordinary first-order differential equations and are derived from the behaviour of the three neuronal subpopulations, which operate as linear damped oscillators. The integration of the differential equations pertaining to each subpopulation can be expressed as a convolution of the exogenous input to produce the response (David and Friston 2003). This convolution transforms the average density of pre-synaptic inputs into an average postsynaptic membrane potential, where the convolution kernel is given by

$$p(t)_e = \begin{cases} \frac{H_e}{\tau_e} t \exp(-t/\tau_e) & t \geq 0 \\ 0 & t < 0 \end{cases} \quad (2)$$

Here, the subscript “*e*” stands for “excitatory”. Similarly subscript “*i*” is used for inhibitory synapses.  $H$  controls the maximum postsynaptic potential, and  $\tau$  represents a lumped rate constant. An operator  $S$  transforms the potential of each subpopulation into firing rate, which is the exogenous input to other subpopulations. This operator is assumed to be an instantaneous sigmoid nonlinearity of the form

$$S(x) = \frac{1}{1 + \exp(-\rho_1(x - \rho_2))} - \frac{1}{1 + \exp(\rho_1\rho_2)} \quad (3)$$

where the free parameters  $\rho_1$  and  $\rho_2$  determine its slope and translation. Interactions, among the subpopulations, depend on internal coupling constants,  $\gamma_{1,2,3,4}$ , which control the strength of intrinsic connections and reflect the total number of synapses expressed by each subpopulation (see Fig. 1). The integration of this model, to form predicted responses, rests on formulating these two operators (Eqs. 2, 3) in terms of a set of differential equations as described in David and Friston (2003). These equations, for all sources, can be integrated using the matrix exponential of the systems Jacobian as described in the appendices of David et al. (2006). Critically, the integration scheme allows for conduction delays on the connections, which are free parameters of the model. A DCM, at the network level, obtains by coupling sources with extrinsic forward, backward and lateral connections as described above.

#### Event-related input and event-related response-specific effects

To model event-related responses, the network receives inputs from the environment via input connections. These connections are exactly the same as forward connections and deliver inputs  $u$  to the spiny stellate cells in layer 4 of specified sources. In the present context, inputs  $u$  model afferent activity relayed by subcortical structures and is modelled with two components: The first is a gamma

density function (truncated to peri-stimulus time). This models an event-related burst of input that is delayed with respect to stimulus onset and dispersed by subcortical synapses and axonal conduction. Being a density function, this component integrates to one over peri-stimulus time. The second component is a discrete cosine set modelling systematic fluctuations in input, as a function of peri-stimulus time. In our implementation, peri-stimulus time is treated as a state variable, allowing the input to be computed explicitly during integration. Critically, the event-related input is exactly the same for all ERPs. The effects of experimental factors are mediated through event-related response (ERR)-specific changes in connection strengths. See Fig. 1 for a summary of the resulting differential equations.

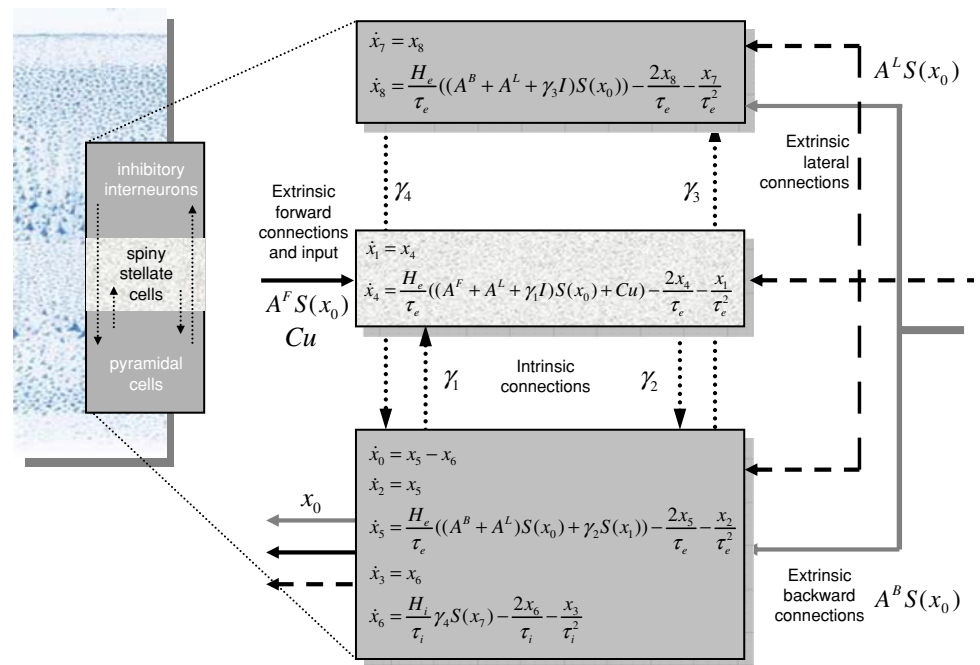
We can model differential responses to different stimuli in two ways. The first is when the effects of experimental factors are mediated through changes in extrinsic connection strengths (David et al. 2006). For example, this extrinsic mechanism can be used to explain ERP (event-related potential) differences by modulating forward (bottom-up) or backward (top-down) coupling. The second mechanism involves changing the intrinsic architecture; of the sort mediating local adaptation. Changes in connectivity are expressed as differences in intrinsic, forward, backward or lateral connections that confer a selective sensitivity on each source, in terms of its response to others. The experimental or stimulus-specific effects are modelled by coupling gains

$$\begin{aligned} A_{ijk}^F &= A_{ij}^F B_{ijk} \\ A_{ijk}^B &= A_{ij}^B B_{ijk} \\ A_{ijk}^L &= A_{ij}^L B_{ijk} \end{aligned} \quad (4)$$

Here,  $A_{ij}$  encodes the strength of a connection to the  $i$ th source from the  $j$ th and  $B_{ijk}$  encodes its gain for the  $k$ th ERP. The superscripts ( $F$ ,  $B$ , or  $L$ ) indicate the type of connection, i.e., forward, backward or lateral (see also Fig. 1). By convention, we set the gain of the first ERP to unity, so that the gains of subsequent ERPs are relative to the first. The reason we model extrinsic modulations in terms of gain (a multiplicative factor), as opposed to additive effects, is that by construction, connections should always be positive. This is assured; provided both the connection and its gain are positive. In this context, a [positive] gain of less than one represents a decrease in connection strength.

Note that if we considered the gains as elements of a gain matrix, the intrinsic gain would occupy the leading diagonal. Intrinsic modulation can explain important features of typical evoked responses, which are difficult to model with a modulation of extrinsic connections (Kiebel et al. 2007). We model the modulation of intrinsic

**Fig. 1** Neuronal state-equations. A source consists of three neuronal subpopulations, which are connected by four intrinsic connections with weights  $\gamma_{1,2,3,4}$ . Mean firing rates (Eq. 3) from other sources arrive via forward  $A^F$ , backward  $A^B$  and lateral connections  $A^L$ . Similarly, exogenous input  $Cu$  enters receiving sources. The output of each subpopulation is its trans-membrane potential (Eq. 2)



connectivity by a gain on the amplitude  $H_e$  of the synaptic kernel (Eq. 2). A gain greater than one effectively increases the maximum response that can be elicited from a source. For the  $i$ th source:

$$H_{ek}^{(i)} = H_e^{(i)} B_{iik} \tag{5}$$

The spatial forward model

The dendritic signal of the pyramidal subpopulation of the  $i$ th source  $x_0^{(i)}$  is detected remotely on the scalp surface in M/EEG. The relationship between scalp data  $h$  and source activity is assumed to be linear and instantaneous

$$h = g(x, \theta) = L(\theta^L) x_0 \tag{6}$$

where  $L$  is a lead-field matrix (i.e., spatial forward model), which accounts for passive conduction of the electromagnetic field (Mosher et al. 1999). Here, we assume that the spatial expression of each source is caused by one ECD. Of course, one can use different source models, e.g. extended patches on the cortical surface (see Section “Discussion”). The head model for the dipoles is based on four concentric spheres, each with homogeneous and isotropic conductivity. The four spheres approximate the brain, skull, cerebrospinal fluid, and scalp. The parameters of the model are the radii and conductivities for each layer. Here, we use as radii 71, 72, 79, and 85 mm, with conductivities 0.33, 1.0, 0.0042, and 0.33 S/m, respectively. The potential at the sensors requires an evaluation of an infinite series, which can be approximated using fast algorithms (Mosher et al. 1999; Zhang 1995). The lead-field of each ECD is then a function of three location and three orientation or

moment parameters  $\theta^L = \{\theta^{pos}, \theta^{mom}\}$ . For the ECD forward model, we used a Matlab (Mathworks) routine that is freely available as part of the FieldTrip package (<http://www2.ru.nl/fcdonders/fieldtrip/>), under the GNU general public license.

Dimension reduction

For computational reasons, it is expedient to reduce the dimensionality of the sensor data, while retaining the maximum amount of information. This is assured by projecting the data onto a subspace defined by its principal eigenvectors  $E$

$$\begin{aligned} y &\leftarrow Ey \\ L &\leftarrow EL \\ \varepsilon &\leftarrow E\varepsilon \end{aligned} \tag{7}$$

where  $\varepsilon$  is the observation error (see below). The eigenvectors are computed using principal component analysis or singular value decomposition (SVD). Because this projection is orthonormal, the independence of the projected errors is preserved, and the form of the error covariance components assumed by the observation model remains unchanged. In this paper, we reduce the sensor data such that the retained modes capture at least 99% of the variability of the data.

The observation or likelihood model

In summary, our DCM comprises a state-equation that is based on neurobiological heuristics and an observer

equation based on an electromagnetic forward model. By integrating the state-equation and passing the ensuing states through the observer equation, we generate a predicted measurement. This corresponds to a generalized convolution of the inputs to generate a response  $h(\theta)$  (Eq. 6). This generalized convolution gives an observation model for the vectorized data  $y$  and the associated likelihood

$$y = \text{vec}(h(\theta) + X\theta^X) + \varepsilon$$

$$p(y|\theta, \lambda) = N(\text{vec}(h(\theta) + X\theta^X), \text{diag}(\lambda) \otimes V) \quad (8)$$

Measurement noise,  $\varepsilon$  is assumed to be zero-mean Gaussian and independent over channels, i.e.,  $\text{Cov}(\text{vec}(\varepsilon)) = \text{diag}(\lambda) \otimes V$ , where  $\lambda$  is an unknown vector of channel-specific variances.  $V$  represents the error's temporal auto-correlation matrix, which we assume is the identity matrix. This is tenable because we down-sample the data to about 8 ms. Low-frequency noise or drift components are modelled by  $X$ , which is a block diagonal matrix with a low-order discrete cosine set for each evoked response and channel. The order of this set can be determined by Bayesian model selection (see below). This model is fitted to data by tuning the free parameters  $\theta$  to minimize the discrepancy between predicted and observed MEG/EEG time series under model complexity constraints (more formally, the parameters minimize the Variational Free Energy; see below). In addition to minimizing prediction error, the parameters are constrained by a prior specification of the range they are likely to lie in Friston et al. (2003). These constraints, which take the form of a prior density  $p(\theta)$ , are combined with the likelihood,  $p(y|\theta)$ , to form a posterior density  $p(\theta|y) \propto p(y|\theta) p(\theta)$  according to Bayes' rule. It is this posterior or conditional density we want to estimate. Gaussian assumptions about the errors in Eq. 8 enable us to compute the likelihood from the prediction error. The only outstanding quantities we require are the priors, which are described next.

#### Prior expectation

The connectivity architecture is constant over peri-stimulus time and defines the dynamical behaviour of the DCM. We have to specify prior assumptions about the connectivity parameters to estimate their posterior distributions. Priors have a dramatic impact on the landscape of the objective function to be optimized: precise prior distributions ensure that the objective function has a global minimum that can be attained robustly. Under Gaussian assumptions, the prior distribution  $p(\theta_i)$  of the  $i$ th parameter is defined by its mean and variance. The mean corresponds to the prior expectation. The variance reflects the amount of prior information about the

parameter. A tight distribution (small variance) corresponds to precise prior knowledge. The parameters of the state-equation can be divided into six subsets: (i) extrinsic connection parameters, which specify the coupling strengths among sources, (ii) intrinsic connection parameters, which reflect our knowledge about canonical micro-circuitry within a source, (iii) conduction delays, (iv) synaptic and sigmoid parameters controlling the dynamics within an source, (v) input parameters, which control the subcortical delay and dispersion of event-related responses, and, importantly, (vi) intrinsic and extrinsic gain parameters. Table 1 list the priors for these parameters; see also David et al. (2006) for details. Note that we fixed the values of intrinsic coupling parameters as described in Jansen and Rit (1995). Inter-laminar conduction delays are usually fixed at 2 ms and inter-regional delays have a prior expectation of 16 ms.

#### 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

$$p(\theta|y, m) = \frac{p(y|\theta, m)p(\theta, m)}{p(y|m)} \quad (9)$$

The estimation procedure employed in DCM is described in Friston et al. (2003). The posterior moments (mean  $\eta$  and covariance  $\Sigma$ ) are updated iteratively using Variational Bayes under a fixed-form Laplace (i.e., Gaussian) approximation to the conditional density  $q(\theta) = N(\eta, \Sigma)$ . This can be regarded as an Expectation-Maximization (EM) algorithm that employs a local linear approximation of Eq. 8 about the current conditional expectation. The E-step conforms to a Fisher-scoring scheme (Fahrmeir and Tutz 1994) that performs a descent on the variational free energy  $F(q, \lambda, m)$  with respect to the conditional moments. In the M-Step, the error variances  $\lambda$  are updated in exactly the same way. The estimation scheme can be summarized as follows:

Repeat until convergence

$$\mathbf{E}\text{-Step} \quad q \leftarrow \min_q F(q, \lambda, m)$$

$$\mathbf{M}\text{-Step} \quad \lambda \leftarrow \min_\lambda F(q, \lambda, m)$$

$$F(q, \lambda, m) = \langle \ln q(\theta) - \ln p(y|\theta, \lambda, m) - \ln p(\theta|m) \rangle_q$$

$$= D(q||p(\theta|y, \lambda, m)) - \ln p(y|\lambda, m) \quad (10)$$

Note that the free energy is simply a function of the log-likelihood and the log-prior for a particular DCM and  $q(\theta)$ . The expression  $\langle \cdot \rangle_q$  denotes the expectation under the density  $q$ .  $q(\theta)$  is the approximation to the posterior density

**Table 1** Prior densities of parameters (for connections to the *i*th source from the *j*th, in the *k*th evoked response)

Extrinsic coupling parameters	$A_{ijk}^F = A_{ij}^F B_{ijk}$ $A_{ij}^F = 32 \exp(\theta_{ij}^F)$ $\theta_{ij}^F \sim N(0, \frac{1}{2})$ $A_{ijk}^B = A_{ij}^B B_{ijk}$ $A_{ij}^B = 16 \exp(\theta_{ij}^B)$ $\theta_{ij}^B \sim N(0, \frac{1}{2})$ $A_{ijk}^L = A_{ij}^L B_{ijk}$ $A_{ij}^L = 4 \exp(\theta_{ij}^L)$ $\theta_{ij}^L \sim N(0, \frac{1}{2})$ $B_{ijk} = \exp(\theta_{ijk}^B)$ $\theta_{ij}^L \sim N(0, \frac{1}{2})$ $C_i = \exp(\theta_i^C)$ $\theta_i^C \sim N(0, \frac{1}{2})$
Intrinsic coupling parameters	$\gamma_1 = 128$ $\gamma_2 = \frac{4}{5}\gamma_1$ $\gamma_3 = \frac{1}{4}\gamma_1$ $\gamma_4 = \frac{1}{4}\gamma_1$
Conduction delays (ms)	$\Delta_{ii} = 2$ $\Delta_{ij} = 16 \exp(\theta_{ij}^\Delta)$ $\theta_{ij}^\Delta \sim N(0, \frac{1}{16})$
Synaptic parameters (ms)	$T_e^{(i)} = 8 \exp(\theta_i^T)$ $\theta_i^T \sim N(0, \frac{1}{8})$ $H_{e,k}^{(i)} = B_{ik} H_e^{(i)}$ $H_e^{(i)} = 4 \exp(\theta_i^H)$ $\theta_i^H \sim N(0, \frac{1}{8})$ $T_i = 16$ $H_i = 32$
Sigmoid parameters	$\rho_1^{(i)} = \frac{2}{3} \exp(\theta_i^{\rho_1})$ $\theta_i^{\rho_1} \sim N(0, \frac{1}{8})$ $\rho_2^{(i)} = \frac{1}{3} \exp(\theta_i^{\rho_2})$ $\theta_i^{\rho_2} \sim N(0, \frac{1}{8})$
Input parameters (s)	$u(t) = b(t, \eta_1, \eta_2) + \sum \theta_i^c \cos(2\pi(i-1)t)$ $\theta_i^c \sim N(0, 1)$ $\eta_1 = \exp(\theta_1^\eta)$ $\theta_1^\eta \sim N(0, \frac{1}{16})$ $\eta_2 = 16 \exp(\theta_2^\eta)$ $\theta_2^\eta \sim N(0, \frac{1}{16})$
Spatial (ECD) parameters (mm)	$\theta_i^{pos} \sim N(L_i^{pos}, 32I_3)$ $\theta_i^{mom} \sim N(0, 8I_3)$

$p(\theta|y, \lambda, m)$  we require. The **E**-step updates the moments of  $q(\theta)$  (these are the variational parameters  $\eta$  and  $\Sigma$ ) by minimizing the variational free energy. The free energy is the Kullback–Leibler divergence (denoted by  $D(\cdot||\cdot)$ ), between the real and approximate conditional density minus the log-likelihood. This means that the conditional moments or variational parameters maximize the marginal log-likelihood, while minimizing the discrepancy between the true and approximate conditional density. Because the divergence does not depend on the covariance parameters, minimizing the free energy in the **M**-step is equivalent to finding the maximum likelihood estimates of the covariance parameters. This scheme is identical to that employed by DCM for functional magnetic resonance imaging (Friston et al. 2003). Source code for this routine can be found in the Statistical Parametric Mapping software package (see Software note below), in the function ‘spm\_nlsi\_N.m’.

Bayesian inference proceeds using the conditional or posterior density estimated by iterating Eq. 10. Usually this involves specifying a parameter or compound of parameters as a contrast,  $c^T \eta$ . Inferences about this contrast are made using its conditional covariance,  $c^T \Sigma c$ . For example, one can compute the probability that any contrast is greater than zero or some meaningful threshold, given the data. This inference is conditioned on the particular model specified. In other words, given the data and model, inference is based on the probability that a particular contrast is bigger than a specified threshold. In some situations one may want to compare different models. This entails Bayesian model comparison.

Different models are compared using their evidence (Penny et al. 2004). The model evidence is

$$p(y|m) = \int p(y|\theta, m)p(\theta|m)d\theta \tag{11}$$

Note that the model evidence is simply the normalization constant in Eq. 9. The evidence can be decomposed into two components: an accuracy term, which quantifies the data fit, and a complexity term, which penalizes models with a large number of parameters. Therefore, the evidence embodies the two conflicting requirements of a good model, that it explains the data and is as simple as possible. In the following, we approximate the model evidence for model *m*, under a normal approximation (Friston et al. 2003), by

$$\ln p(y|m) \approx \ln p(y|\lambda, m) \tag{12}$$

This is simply the maximum value of the objective function attained by EM (see the **M**-Step in Eq. 10). The most likely model is the one with the largest log-evidence. This enables Bayesian model selection. Model comparison rests on the likelihood ratio  $B_{ij}$  (i.e., Bayes Factor) of the evidence or relative log-evidence for two models. For models *i* and *j*

$$\ln B_{ij} = \ln p(y|m = i) - \ln p(y|m = j) \tag{13}$$

Conventionally, strong evidence in favour of one model requires the difference in log-evidence to be three or more (Penny et al. 2004). This threshold criterion plays a similar role as a *p*-value of  $0.05 = 1/20$  in classical statistics (used to reject the null hypothesis in favour of the alternative

model). A difference in log-evidence of greater than three (i.e., a Bayes factor more than  $\exp(3) \sim 20$ ) indicates that the data provide strong evidence in favour of one model over the other. This is a standard way to assess the differences in log-evidence quantitatively.

### Models for steady-state responses

The description up to this point has assumed that we are dealing with evoked responses, which can be located precisely in time. As we have shown above, given stimulus timing information, one can model the measured M/EEG response as the system whose dynamics are prescribed by neural mass models. In theory, one can also estimate the input itself, e.g., its onset and duration. This is a hard, nonlinear problem and we acknowledge most experiments control the input by design. However, there are experiments for which one does not know the input function. For example, in sleep research, one might want to model networks that receive internally generated input functions. Similarly, experiments that measure electrophysiological data (M/EEG, local field potentials) over long times, without any designed inputs, must assume the exact input function to be unknown. In such cases, one can posit that the data have been induced by input with an assumed statistical distribution. Linear system models offer particularly amenable analysis strategies to explore state-space models. The form of the neural mass model equations above are an example of such state-space models.

For steady-state responses the system can, in essence, be understood as a filter with an accompanying transfer function

$$H(s) = \frac{(s - \zeta_1)(s - \zeta_2)(s - \zeta_3)\dots}{(s - \lambda_1)(s - \lambda_2)(s - \lambda_3)\dots} \quad (14)$$

Here  $s$  represents real and imaginary frequency components  $\zeta_i$  represent the system “zeros” where the frequency response is zero and  $\lambda_i$ , the system “poles” where the frequency goes to infinity. This function describes how any spectral input is shaped to produce spectral output. This presumes time invariance in the inputs and response and can describe neatly the dynamics in the frequency domain,  $s$ .

With DCM, we can use this strategy for steady-state paradigms by assuming a white noise (flat spectral) input. Assuming the system operates in a steady-state around its fixed point, we can linearize the nonlinear differential equations to describe the system response in the frequency domain. Note that by response we now mean the spectral output that is shaped by the system transfer function. This linearization allows us to establish a mapping from the system parameters to the predicted frequency spectrum (Moran et al. 2007). Importantly, as with multiple evoked responses,

this enables us to model differences between two or more spectra, acquired under different conditions, as consequences of specific parameter changes. These parameters might be intrinsic or extrinsic connections, but can also be, for example, the excitatory rate constants  $\tau_e$  (Eq. 2), which have a marked influence on the frequency spectrum (Moran et al. 2007). The basic idea is to manipulate the (real) system (e.g. by experimental changes in the level of a neurotransmitter), model this change in terms of changes in specific DCM parameters, and then test the implicit hypothesis using Bayesian inference, i.e., model comparison and posterior probabilities. This strategy has been applied in (Moran et al. in press) using just one source, where we showed that changes in glutamate levels as measured by microdialysis can be modelled by a change in DCM parameters. We will come back to this work below.

### Illustrative examples

#### Mismatch negativity

In this section, we illustrate the use of DCM for ERP/ERFs by analysing data acquired under a mismatch negativity (MMN) paradigm. Critically, DCM allows us to test hypotheses about the changes in connectivity between sources. In this example study, we will test a specific hypothesis (see below) about the MMN generation and compare various models over twelve subjects. The results shown here are a part of a series of papers that consider the MMN and its underlying mechanisms in detail (Garrido et al. 2007).

Novel sounds, or oddballs, embedded in a stream of repeated sounds, or standards, produce a distinct response that can be recorded non-invasively with MEG and EEG. The MMN is the negative component of the waveform obtained by subtracting the event-related response to a standard from the response to an oddball, or deviant. This response to sudden changes in the acoustic environment peaks at about 100–200 ms from change onset (Sams et al. 1985) and exhibits an enhanced negativity that is distributed over auditory and frontal areas, with prominence in frontal regions.

The MMN is believed to be an index of automatic change detection reflecting a pre-attentive sensory memory mechanism (Tiitinen et al. 1994). There have been several compelling mechanistic accounts of how the MMN might arise. The most common interpretation is that the MMN can be regarded as a marker for error detection, caused by a break in a learned regularity, or familiar auditory context. The early work by Näätänen and colleagues suggested that the MMN results from a comparison between the auditory input and a memory trace of previous sounds. In agreement



with this theory, others (Naatanen and Winkler 1999; Sussman and Winkler 2001; Winkler et al. 1996) have postulated that the MMN would reflect on-line modifications of the auditory system, or updates of the perceptual model, during incorporation of a newly encountered stimulus into the model—the *model-adjustment hypothesis*. Hence, the MMN would be a specific response to stimulus change and not to stimulus alone. This hypothesis has been supported by Escera et al. (2003) who provided evidence that the prefrontal cortex is involved in a top-down modulation of a deviance detection system in the temporal cortex. In the light of the Näätänen model, it has been claimed that the MMN is caused by two underlying functional processes, a sensory memory mechanism related to temporal generators and an automatic attention-switching process related to the frontal generators (Giard et al. 1990). Accordingly, it has been shown that the temporal and frontal MMN sources have distinct behaviours over time (Rinne et al. 2000) and that these sources interact with each other (Jemel et al. 2002). Thus the MMN could be generated by a temporofrontal network (Doeller et al. 2003; Opitz et al. 2002), as revealed by M/EEG and fMRI studies. This work has linked the early component (in the range of about 100–140 ms) to a sensorial, or non-comparator account of the MMN, elaborated in the temporal cortex, and a later component (in the range of about 140–200 ms) to a cognitive part of the MMN, involving the frontal cortex (Maess et al. 2007).

Using DCM, we modelled the MMN generators with a temporo-frontal network comprising bilateral sources over the primary and secondary auditory and frontal cortex. Following the model-adjustment hypothesis, we assume that the early and late component of the MMN can be explained by an interaction of temporal and frontal sources or network nodes. The MMN itself is defined as the difference between the responses to the oddball and the standard stimuli. Here, we modelled both evoked responses and explained the MMN, i.e., differences in the two ERPs, by a modulation of DCM parameters. There are two kinds of parameters that seem appropriate to induce the difference between oddballs and standards: (i) modulation of extrinsic connectivity between sources, and (ii) modulation of intrinsic parameters in each source. Modulation of intrinsic parameters would correspond to a mechanism that is more akin to an *adaptation hypothesis*, i.e., the MMN is generated by local adaptation of populations. This is the hypothesis considered by Jaaskelainen et al. (2004) who report evidence that the MMN is explained by differential adaptation of two pairs of bilateral temporal sources. In a recent paper (Garrido et al. in press), we have compared models derived from both hypotheses: (i) the model-adjustment hypothesis and (ii) the adaptation hypothesis. Here, we will constrain ourselves to demonstrate inference based on DCMs derived from the model-

adjustment hypothesis only, which involves a fronto-temporal network.

### *Experimental design*

We studied a group of 13 healthy volunteers aged 24–35 (5 female). Each subject gave signed informed consent before the study, which proceeded under local ethical committee guidelines. Subjects sat on a comfortable chair in front of a desk in a dimly illuminated room. Electroencephalographic activity was measured during an auditory ‘oddball’ paradigm, in which subjects heard of “standard” (1,000 Hz) and “deviant” tones (2,000 Hz), occurring 80% (480 trials) and 20% (120 trials) of the time, respectively, in a pseudo-random sequence. The stimuli were presented binaurally via headphones for 15 min every 2 s. The duration of each tone was 70 ms with 5 ms rise and fall times. The subjects were instructed not to move, to keep their eyes closed and to count the deviant tones.

EEG was recorded with a Biosemi system with 128 scalp electrodes. Data were recorded at a sampling rate of 512 Hz. Vertical and horizontal eye movements were monitored using EOG (electro-oculograms) electrodes. The data were epoched offline, with a peri-stimulus window of –100 to 400 ms, down-sampled to 200 Hz, band-pass filtered between 0.5 and 40 Hz and re-referenced to the average of the right and left ear lobes. Trials in which the absolute amplitude of the signal exceeded 100  $\mu$ V were excluded. Two subjects were eliminated from further analysis due to excessive trials containing artefacts. In the remaining subjects, an average 18% of trials were excluded.

### *Specification of dynamic causal model*

In this section, we specify three plausible models defined under a given architecture and dynamics. The network architecture was motivated by recent electrophysiological and neuroimaging studies looking at the sources underlying the MMN (Doeller et al. 2003; Opitz et al. 2002). We assumed five sources, modelled as ECDs, over left and right primary auditory cortices (A1), left and right superior temporal gyrus (STG) and right inferior frontal gyrus (IFG), see Fig. 2. Our mechanistic model attempts to explain the generation of each individual response, i.e., responses to standards and deviants. Therefore, left and right primary auditory cortex (A1) were chosen as cortical input stations for processing the auditory information. Opitz et al. (2002) identified sources for the differential response, with fMRI and EEG measures, in both left and right STG, and right IFG. Here we employ the coordinates reported by Opitz et al. (2002) (for left and right STG and right IFG) and Rademacher et al. (2001) (for left and right A1) as prior source location means, with a prior variance of 32 mm. We

converted these coordinates, given in the literature in Talairach space, to MNI space using the algorithm described in <http://imaging.mrc-cbu.cam.ac.uk/imaging/MniTalairach>. The moment parameters had prior mean of 0 and a variance of 8 in each direction. We have used these parameters as priors to estimate, for each individual subject, the posterior locations and moments of the ECDs (Table 2). Using these sources and prior knowledge about the functional anatomy we constructed the following DCM: An extrinsic input entered bilaterally to A1, which were connected to their ipsilateral STG. Right STG was connected with the right IFG. Inter-hemispheric (lateral) connections were placed between left and right STG. All connections were reciprocal (i.e., connected with forward and backward connections or with bilateral connections). Given this connectivity graph, specified in terms of its nodes and connections, we tested three models. These models differed in the connections which could show putative learning-related changes, i.e., differences between listening to standard or deviant tones. Models F, B and FB allowed changes in forward, backward and both forward and backward connections, respectively (see Fig. 2). All three models were compared against a baseline or null model. The null model had the same architecture described above but precluded any coupling changes between standard and deviant trials.

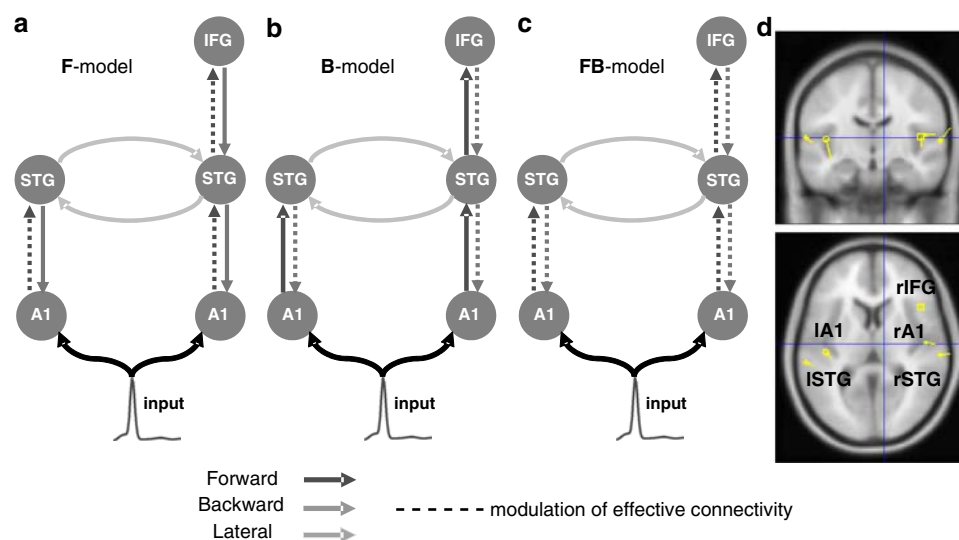
## Results

The difference between the ERPs evoked by the standard and deviant tones revealed a standard MMN. This negativity was present from 90 to 190 ms and had a broad spatial pattern,

**Table 2** Prior coordinates for the locations of the ECDs in Montreal Neurology Institute (MNI) space [mm]

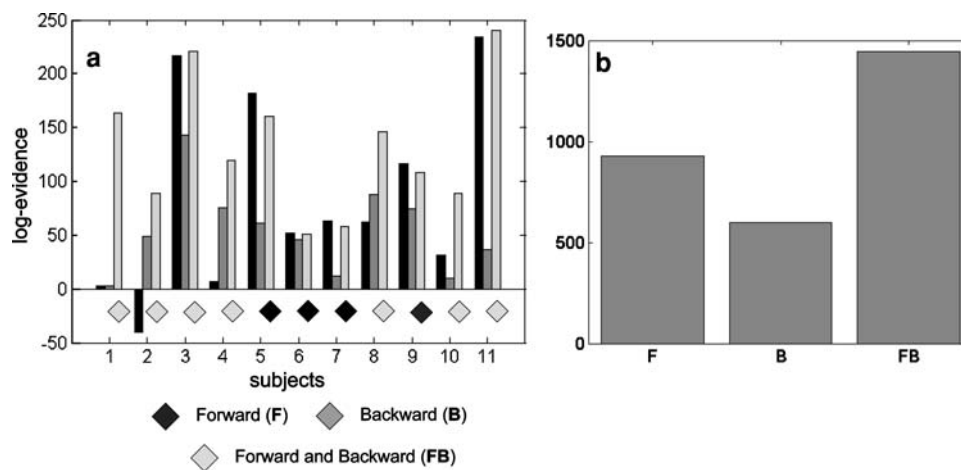
Left primary auditory cortex (lA1)	−42, −22, 7
Right primary auditory cortex (rA1)	46, −14, 8
Left superior temporal gyrus (lSTG)	−61, −32, 8
Right superior temporal gyrus (rSTG)	59, −25, 8
Right inferior frontal gyrus (rIFG)	46, 20, 8

encompassing electrodes previously associated with auditory and frontal areas. Four different DCMs, forward only (F-model), backward only (B-model), forward and backward (FB-model) and the null model were inverted for each subject. Figure 3 illustrates the model comparison based on the increase in log-evidence over the null model, for all subjects. Figure 3a shows the log-evidence for the three models, relative to the null model, for each subject, revealing that the three models were significantly better than the null in all subjects. The diamond attributed to each subject identifies the best model on the basis of the highest log-evidence. The FB-model was significantly better in seven out of 11 subjects. The F-model was better in four subjects but only significantly so in three (for one of these subjects [subject 6], model comparison revealed only weak evidence in favour of the F-model over the FB-model, though still very strong evidence over the B-model). In all but one subject, the F and FB-models were better than the B-model. Figure 3b shows the log-evidences for the three models at the group level. The log-evidence for the group is the sum of the log-evidences from all subjects, because of the independent measures over subjects. Both F and FB are clearly more likely than B and,



**Fig. 2** Model specification. The sources comprising the network are connected with forward (dark grey), backward (grey) or lateral (light grey) connections as shown. A1: primary auditory cortex, STG: superior temporal gyrus, IFG: inferior temporal gyrus. Three different models were tested within the same architecture (a–c), allowing for

learning-related changes in forward F, backward B and forward and backward FB connections, respectively. The broken lines indicate the connections we allowed to change. (d) Sources of activity, modelled as dipoles (estimated posterior moments and locations), are superimposed in an MRI of a standard brain in MNI space



**Fig. 3** Bayesian model selection among DCMs for the three models, F, B and FB, expressed relative to a DCM in which no connections were allowed to change (null model). The graphs show the free energy approximation to the log-evidence. **(a)** Log-evidence for

models F, B and FB for each subject (relative to the null model). The diamond attributed to each subject identifies the best model on the basis of the subject's highest log-evidence. **(b)** Log-evidence at the group level, i.e., pooled over subjects, for the three models

over subjects, there is very strong evidence in favour of model FB over model F. Figure 4a shows, for the best model FB, the predicted responses at each node of the network for each trial type (i.e., standard or deviant) for a single subject (subject 9). For each connection in the network, the plot shows the coupling gains and the conditional probability that the gains are different from one. For example, a coupling change of 2.04 from IA1 to ISTG means that the effective connectivity increased 104% for rare events relative to frequent events. The response, in measurement space, of the three principal spatial modes is shown on the right (Fig. 4b). This figure shows a remarkable agreement between predicted (solid) and observed (dotted) responses. Figure 5 summarises the conditional densities of the coupling parameters for the F-model (Fig. 5a) and FB-model (Fig. 5b). For each connection in the network, the plot shows the coupling gains and the conditional probability that the gains are different from one, pooled over subjects. For the F-model the effective connectivity has increased in all connections with a conditional probability of almost 100%. For the FB-model the effective connectivity has changed in all forward and backward connections with a probability of almost 100%. Equivalently, and in accord with theoretical predictions, all extrinsic connections (i.e., influences) were modulated for rare events as compared to frequent events.

#### Steady-state responses

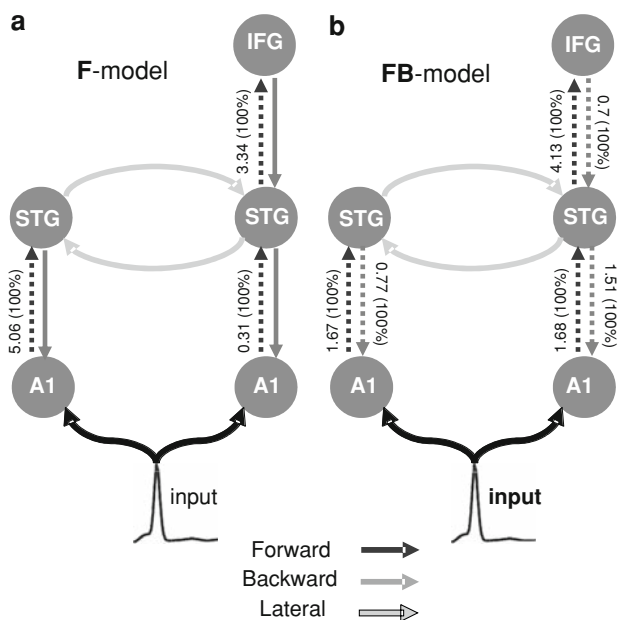
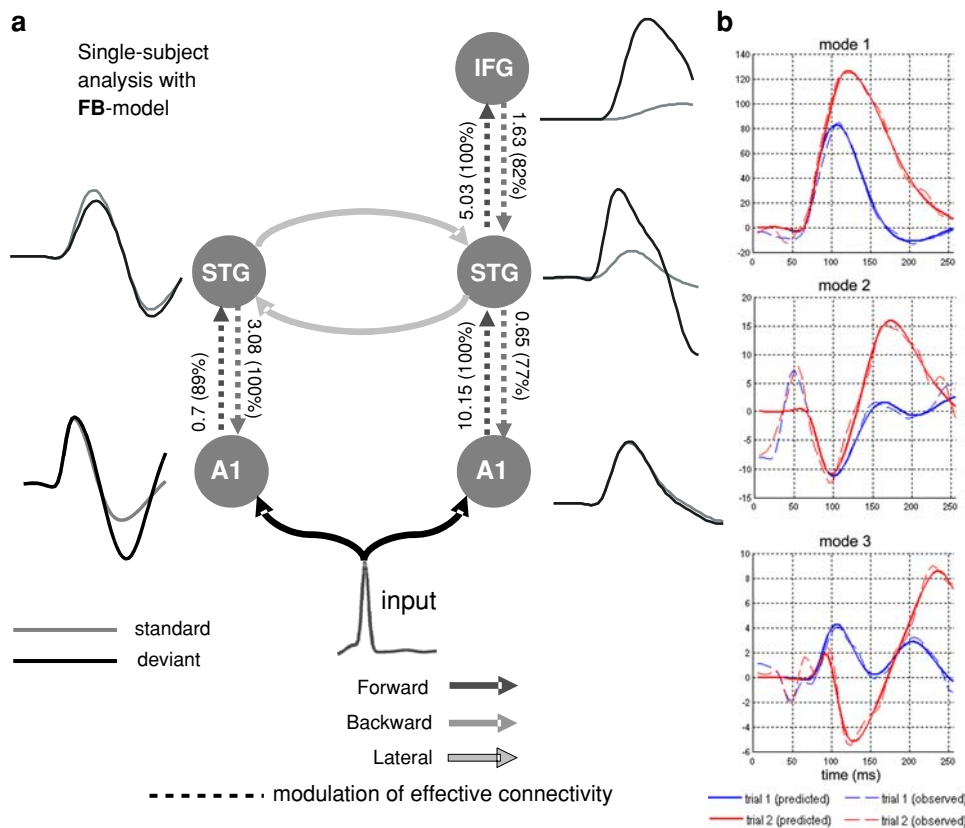
Here, we describe briefly an experiment using local field potential (LFP) data. The same technique can be applied to M/EEG data, after source reconstruction. LFP recordings were taken from embedded electrodes in the prefrontal cortex of normal rats and isolation reared counterparts. The latter animals are a well-established model of

schizophrenia-like sensorimotor deficits as measured by pre-pulse inhibition of startle (Geyer et al. 1993). Moreover, these animals were recently reported to show profound reduction in prefrontal glutamate levels as measured by microdialysis (Table 3) (Mean GABA levels were also reduced but variability among the isolated group meant that these differences were not significant). This sort of reduction in extracellular neurotransmitter levels usually leads to an up-regulation of neurotransmitter uptake processes and a sensitization of post-synaptic receptor mechanisms. In the current context, this suggests that we should see an increase in the amplitude ( $H_e$ ) of synaptic kernels and an increase in the coupling parameters ( $\gamma_1, \gamma_2, \gamma_3$ ) in the isolated group, relative to the social control group.

Empirical LFP data were acquired with a Data Science International radio-telemetric system; collecting LFPs over a 24 h period from the prefrontal cortex of six social and six isolation reared animals. These animals were moving freely in their home cage and not exposed to external stimuli. The data analyzed here was an average spectral response over a 10-min period. Pre-processing involved a Fast Fourier Transform of the data, using the same frequencies as above.

The inversion was performed separately using each rat's spectral response. See Fig. 6 for an exemplar fit. To speed the inversion, the number of parameters was reduced by setting prior variances on parameters  $H_i, \tau_i, \gamma_4, \gamma_5$  to zero. The model could then account for differences in spectral response, between the two groups, using the excitatory parameters  $H_e, \tau_e, \gamma_1, \gamma_2, \gamma_3$  and  $\rho_2$ . Population differences between their MAP estimates were significant in the case of  $H_e$  and  $\rho_2$  ( $p < 0.05$ ). Group parameter means and their respective  $p$ -values are illustrated in Fig. 7.

**Fig. 4** DCM results for a single subject [subject 9] (FB model). (a) Reconstructed responses for each source and changes in coupling during oddball processing relative to standards. The numbers next to each connection are the gain modulation in connection strength and the posterior probability that the modulation is different from one. The mismatch response is expressed in nearly every source. (b) Predicted (solid) and observed (broken) responses in measurement space, which result from a projection of the scalp data onto their first three spatial modes



**Fig. 5** Coupling gains and their posterior probability estimated over subjects for each connection in the network for models F (a) and FB (b). There are widespread learning-related changes in all connections, expressed as modulations of coupling for deviants relative to standards

The picture from LFP modelling corresponds to the microdialysis predictions on two levels, first the MAP estimates suggest a sensitization of post-synaptic

**Table 3** Microdialysis measures of extracellular glutamate neurotransmitter levels from two groups (social and isolated) of Wistar rats

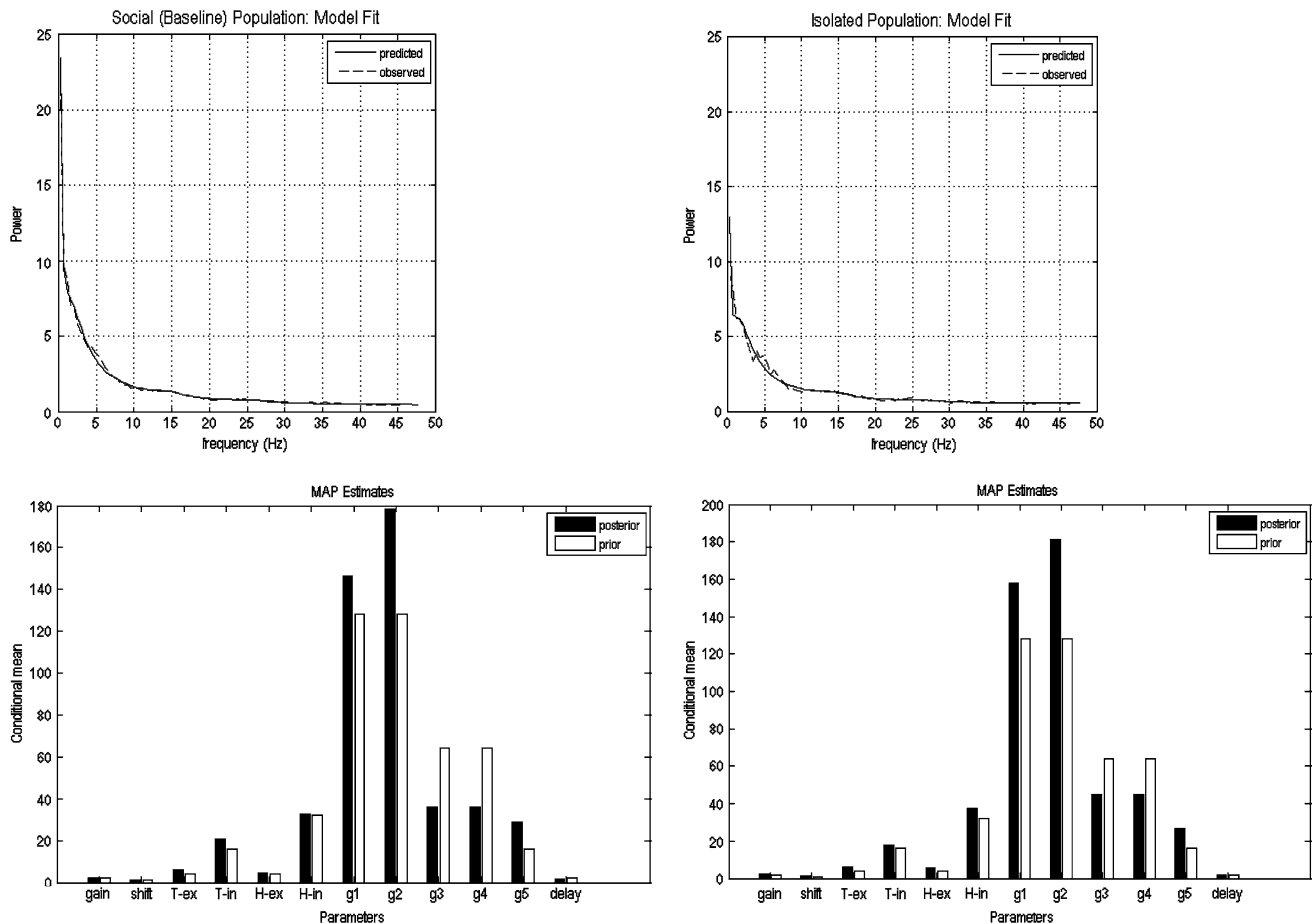
	Glutamate
Social	4.2 ± 1.4 μM (100%)
Isolated	1.5 ± 0.8 μM (36%)

Measurements were taken from the medial prefrontal cortex

responses; with increases in  $H_e$  (and excitatory intrinsic connections) in the isolated animals, secondly an overall decrease in firing rate for that group with the increase in  $\rho_2$  point to a low excitatory field. This parameter is a proxy for neuronal adaptation and highlights a greater adaptation in the low-glutamate “schizophrenic” rat group. This is consistent with reduced levels of extracellular neurotransmitter. For a more detailed discussion of these results see Moran et al. (in press).

**Discussion**

Dynamic Causal Modelling for M/EEG entails the inversion of informed spatiotemporal models of observed responses. The idea is to model condition-specific responses over channels and peri-stimulus time with the same model, where the differences among conditions are explained by changes in only a few key parameters. The face and predictive



**Fig. 6** Social (left) and isolated (right) parameter estimates from the steady-state LFP data analysis. The top panels illustrate the predicted and actual (dashed line) spectra. The bottom panels show the prior (in

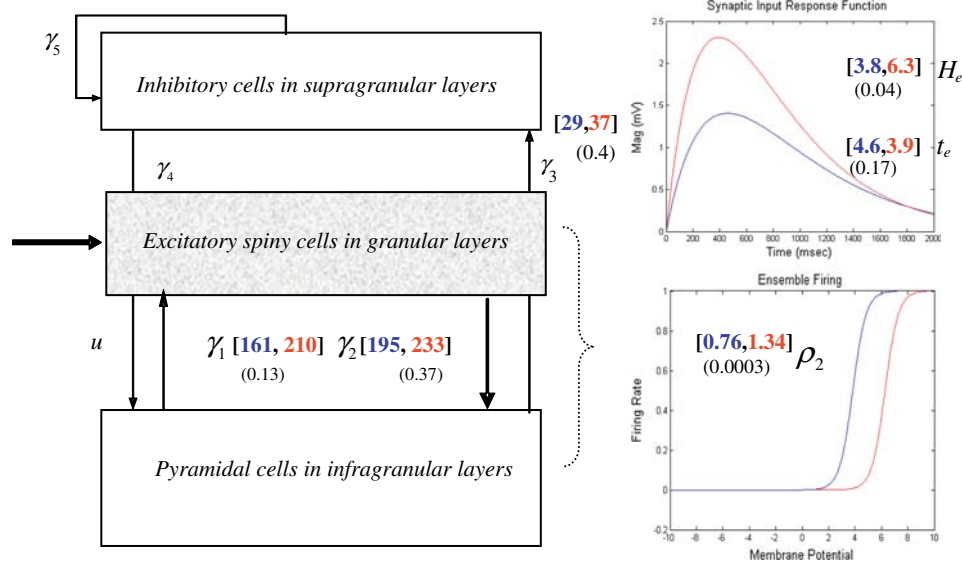
white) and posterior (in black) mean for each parameter. Parameters here are  $\rho_1, \rho_2, \tau_e, \tau_i, H_e, H_i, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, d$ , see also Fig. 7

validity of DCM have been established, which makes it a potentially useful tool for group studies (David et al. 2006; Garrido et al. 2007; Kiebel et al. 2006). In principle, the same approach can be applied to the analysis of single trials, where one would use a parametric modulation of parameters to model the effects of trial-to-trial changes in an experimental variable (e.g., reaction time or forgotten vs. remembered). Furthermore, we have described how DCM can be extended to cover source-reconstructed M/EEG or LFP steady-state responses under simple assumptions about the statistical distribution of the input.

One can also view DCM for evoked responses as a source reconstruction device using biophysically informed temporal constraints. This is because DCM has two components; a neural-mass model of the interactions among a small number of dipole sources and a classical electromagnetic forward model that links these sources to extracranial measurements. Inverting the DCM implicitly optimises the location and orientation of the sources. This is in contrast to traditional ECD fitting approaches, where dipoles are fitted sequentially to the data; using user-

selected periods and/or channels of the data. Classical approaches have to proceed in this way, because there is usually too much spatial and temporal dependency among the sources to identify the parameters precisely. With our approach, we place temporal constraints on the model that are consistent with the way that signals are generated biophysically. As we have shown, these allow simultaneous fitting of multiple dipoles to the data.

We used the ECD model because it is analytic, fast to compute and a quasi-standard when reconstructing evoked responses. However, the ECD model is just one candidate for spatial forward models. Given the lead-field, one can use any spatial model in the observation equation (Eq. 6). A further example would be some linear distributed approach (Baillet et al. 2001; Phillips et al. 2005; Daunizeau et al. 2006), where a ‘patch’ of dipoles, confined to the cortical surface, would act as the spatial expression of one area. With DCM, one could also use different forward models for different areas (hybrid models). For example, one could employ the ECD model for early responses while using a distributed forward model for higher areas.



**Fig. 7** Results of steady-state LFP data analysis. The left panel shows the connection parameters of the different cell groups within the modelled source. Parameters were inferred with inhibitory connectivity (and impulse response) prior parameter variances set to zero. The mean estimates of the connectivity's  $\gamma_1$ ,  $\gamma_2$ ,  $\gamma_3$  are shown with the associated  $p$ -values. The right panels display the expected excitatory impulse response functions and sigmoid firing functions for

both groups. These are constructed using the maximum a posteriori (MAP) estimates of the excitatory synaptic kernel amplitude and time-constant ( $H_e$ ,  $\tau_e$ ) in the former and the MAP estimate of  $\rho_2$  in the latter. The control group estimates are shown in blue and the isolated animals in red, with  $p$ -values in parentheses. Note that for steady-state models, we have added an inhibitory–inhibitory connection ( $\gamma_5$ ), which is not used for the evoked response models

More generally, we anticipate that Bayesian model comparison will become a ubiquitous tool in M/EEG. This is because further development of M/EEG models and their fusion with other imaging modalities requires more complex models embodying useful constraints. The appropriateness of such models for any given data cannot necessarily be intuited, but can be assessed formally using Bayesian model comparison. The key is to compute the model evidence  $p(y|m)$  (Eq. 12), for using a variational approach (see above) or as described in Sato et al. (2004), or by employing sampling approaches like the Monte Carlo Markov Chain (MCMC) techniques as in Auranen et al. (2007) and Jun et al. (2005). In principle, one can compare models based on different concepts or, indeed, inversion schemes, for a given data set  $y$ . For example, one can easily compare different types of source reconstruction (ECD versus source imaging) with DCM. This cannot be done with classical, non-Bayesian approaches, for which model comparisons are only feasible under certain constraints ('nested models'); precluding comparisons among qualitatively different models. Although other approximations to the model evidence exist, e.g. the Akaike Information Criterion, they are not generally useful with informative priors (Beal 2003).

Currently, the DCM framework is deterministic, i.e., it allows for observation noise in the sensors but does not consider noise at the level of the neuronal dynamics. This is part of ongoing work; several groups are developing

variational techniques that invert stochastic DCMs based on stochastic differential equations with both nonlinear evolution and observation functions (c.f., Eq. 1).

### Software note

All procedures described in this note have been implemented as Matlab (MathWorks) code. The source code is freely available in the DCM and neural model toolboxes of the Statistical Parametric Mapping package (SPM5) under <http://www.fil.ion.ucl.ac.uk/spm/>.

**Acknowledgments** This work was supported by the Wellcome Trust. MIG holds a FCT doctoral scholarship, Ministry of Science, Portugal. RJM was supported by the Irish Research Council for Science Engineering and Technology.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

### References

- Auranen T, Nummenmaa A, Hamalainen MS, Jaaskelainen IP, Lampinen J, Vehtari A, Sams M (2007) Bayesian inverse analysis of neuromagnetic data using cortically constrained multiple dipoles. *Hum Brain Mapp* 28:979–994

- Baillet S, Garnero L (1997) A Bayesian approach to introducing anatomic-functional priors in the EEG/MEG inverse problem. *IEEE Trans Biomed Eng* 44:374–385
- Baillet S, Mosher JC, Leahy RM (2001) Electromagnetic brain mapping. *IEEE Signal Process Mag* 18:14–30
- Beal MJ (2003) Variational algorithms for approximate Bayesian inference. University College London
- Breakspear M, Roberts JA, Terry JR, Rodrigues S, Mahant N, Robinson PA (2006) A unifying explanation of primary generalized seizures through nonlinear brain modeling and bifurcation analysis. *Cereb Cortex* 16:1296–1313
- Daunizeau J, Friston KJ (2007) A mesostate-space model for EEG and MEG. *Neuroimage* 38:67–81
- Daunizeau J, Mattout J, Clonda D, Goulard B, Benali H, Lina JM (2006) Bayesian spatio-temporal approach for EEG source reconstruction: conciliating ECD and distributed models. *IEEE Trans Biomed Eng* 53:503–516
- Daunizeau J, Grova C, Marrelec G, Mattout J, Jbabdi S, Pelegrini-Issac M, Lina JM, Benali H (2007) Symmetrical event-related EEG/fMRI information fusion in a variational Bayesian framework. *Neuroimage* 36:69–87
- David O, Friston KJ (2003) A neural mass model for MEG/EEG: coupling and neuronal dynamics. *Neuroimage* 20:1743–1755
- David O, Harrison L, Friston KJ (2005) Modelling event-related responses in the brain. *Neuroimage* 25:756–770
- David O, Kiebel SJ, Harrison LM, Mattout J, Kilner JM, Friston KJ (2006) Dynamic causal modeling of evoked responses in EEG and MEG. *Neuroimage* 30:1255–1272
- Doeller CF, Opitz B, Mecklinger A, Krick C, Reith W, Schroger E (2003) Prefrontal cortex involvement in preattentive auditory deviance detection: neuroimaging and electrophysiological evidence. *Neuroimage* 20:1270–1282
- Escera C, Yago E, Corral MJ, Corbera S, Nunez MI (2003) Attention capture by auditory significant stimuli: semantic analysis follows attention switching. *Eur J Neurosci* 18:2408–2412
- Fahrmeir L, Tutz G (1994) Multivariate statistical modelling based on generalized linear models. Springer-Verlag, New York
- Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex* 1:1–47
- Friston KJ, Harrison L, Penny W (2003) Dynamic causal modelling. *Neuroimage* 19:1273–1302
- Friston K, Henson R, Phillips C, Mattout J (2006) Bayesian estimation of evoked and induced responses. *Hum Brain Mapp* 27:722–735
- Garrido MI, Kilner JM, Kiebel SJ, Stephan KE, Friston KJ (2007) Dynamic causal modelling of evoked potentials: a reproducibility study. *Neuroimage* 36:571–580
- Geyer MA, Wilkinson LS, Humby T, Robbins TW (1993) Isolation rearing of rats produces a deficit in prepulse inhibition of acoustic startle similar to that in schizophrenia. *Biol Psychiatry* 34:361–372
- Giard MH, Perrin F, Pernier J, Bouchet P (1990) Brain generators implicated in the processing of auditory stimulus deviance: a topographic event-related potential study. *Psychophysiology* 27:627–640
- Grimbert F, Faugeras O (2006) Bifurcation analysis of Jansen's neural mass model. *Neural Comput* 18:3052–3068
- Gross J, Kujala J, Hamalainen M, Timmermann L, Schnitzler A, Salmelin R (2001) Dynamic imaging of coherent sources: studying neural interactions in the human brain. *Proc Natl Acad Sci USA* 98:694–699
- Jaaskelainen IP, Ahveninen J, Bonmassar G, Dale AM, Ilmoniemi RJ, Levanen S, Lin FH, May P, Melcher J, Stufflebeam S, Tiitinen H, Belliveau JW (2004) Human posterior auditory cortex gates novel sounds to consciousness. *Proc Natl Acad Sci USA* 101:6809–6814
- Jansen BH, Rit VG (1995) Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biol Cybern* 73:357–366
- Jemel B, Achenbach C, Muller BW, Ropcke B, Oades RD (2002) Mismatch negativity results from bilateral asymmetric dipole sources in the frontal and temporal lobes. *Brain Topogr* 15:13–27
- Jun SC, George JS, Pare-Blagoev J, Plis SM, Ranken DM, Schmidt DM, Wood CC (2005) Spatiotemporal Bayesian inference dipole analysis for MEG neuroimaging data. *Neuroimage* 28:84–98
- Kiebel SJ, David O, Friston KJ (2006) Dynamic causal modelling of evoked responses in EEG/MEG with lead field parameterization. *Neuroimage* 30:1273–1284
- Kiebel SJ, Friston KJ (2004) Statistical parametric mapping for event-related potentials (II): a hierarchical temporal model. *Neuroimage* 22:503–520
- Kiebel SJ, Garrido MI, Friston KJ (2007) Dynamic causal modelling of evoked responses: the role of intrinsic connections. *Neuroimage* 36:332–345
- Liljestrom M, Kujala J, Jensen O, Salmelin R (2005) Neuromagnetic localization of rhythmic activity in the human brain: a comparison of three methods. *Neuroimage* 25:734–745
- Maess B, Jacobsen T, Schroger E, Friederici AD (2007) Localizing pre-attentive auditory memory-based comparison: magnetic mismatch negativity to pitch change. *Neuroimage* 37:561–571
- Makeig S, Westerfield M, Jung TP, Enghoff S, Townsend J, Courchesne E, Sejnowski TJ (2002) Dynamic brain sources of visual evoked responses. *Science* 295:690–694
- Mattout J, Phillips C, Penny WD, Rugg MD, Friston KJ (2006) MEG source localization under multiple constraints: an extended Bayesian framework. *Neuroimage* 30:753–767
- Moran RJ, Stephan KE, Kiebel SJ, Rombach N, O'Connor WT, Murphy KJ, Reilly RB, Friston KJ (in press) Bayesian estimation of synaptic physiology from the spectral responses of neural masses. *Neuroimage*
- Moran RJ, Kiebel SJ, Stephan KE, Reilly RB, Daunizeau J, Friston KJ (2007) A neural mass model of spectral responses in electrophysiology. *Neuroimage* 37:706–720
- Mosher JC, Leahy RM, Lewis PS (1999) EEG and MEG: forward solutions for inverse methods. *IEEE Trans Biomed Eng* 46:245–259
- Naatanen R, Winkler I (1999) The concept of auditory stimulus representation in cognitive neuroscience. *Psychol Bull* 125:826–859
- Nummenmaa A, Auranen T, Hamalainen MS, Jaaskelainen IP, Lampinen J, Sams M, Vehtari A (2007) Hierarchical Bayesian estimates of distributed MEG sources: theoretical aspects and comparison of variational and MCMC methods. *Neuroimage* 35:669–685
- Opitz B, Rinne T, Mecklinger A, von Cramon DY, Schroger E (2002) Differential contribution of frontal and temporal cortices to auditory change detection: fMRI and ERP results. *Neuroimage* 15:167–174
- Penny WD, Stephan KE, Mechelli A, Friston KJ (2004) Comparing dynamic causal models. *Neuroimage* 22:1157–1172
- Penny WD, Kilner J, Blankenburg F (2007) Robust Bayesian general linear models. *Neuroimage* 36:661–671
- Phillips C, Mattout J, Rugg MD, Maquet P, Friston KJ (2005) An empirical Bayesian solution to the source reconstruction problem in EEG. *Neuroimage* 24:997–1011
- Picton TW, Bentin S, Berg P, Donchin E, Hillyard SA, Johnson R Jr, Miller GA, Ritter W, Ruchkin DS, Rugg MD, Taylor MJ (2000) Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology* 37:127–152
- Poline JB, Strother SC, Dehaene-Lambertz G, Egan GF, Lancaster JL (2006) Motivation and synthesis of the FIAC experiment:

- reproducibility of fMRI results across expert analyses. *Hum Brain Mapp* 27:351–359
- Rademacher J, Morosan P, Schormann T, Schleicher A, Werner C, Freund HJ, Zilles K (2001) Probabilistic mapping and volume measurement of human primary auditory cortex. *Neuroimage* 13:669–683
- Rinne T, Alho K, Ilmoniemi RJ, Virtanen J, Naatanen R (2000) Separate time behaviors of the temporal and frontal mismatch negativity sources. *Neuroimage* 12:14–19
- Rodrigues S, Terry JR, Breakspear M (2006) On the genesis of spike-wave oscillations in a mean-field model of human thalamic and corticothalamic dynamics. *Phys Lett A* 355:352–357
- Rugg MD, Curran T (2007) Event-related potentials and recognition memory. *Trends Cogn Sci* 11:251–257
- Sams M, Paavilainen P, Alho K, Naatanen R (1985) Auditory frequency discrimination and event-related potentials. *Electroencephalogr Clin Neurophysiol* 62:437–448
- Sato MA, Yoshioka T, Kajihara S, Toyama K, Goda N, Doya K, Kawato M (2004) Hierarchical Bayesian estimation for MEG inverse problem. *Neuroimage* 23:806–826
- Sotero RC, Trujillo-Barreto NJ, Iturria-Medina Y, Carbonell F, Jimenez JC (2007) Realistically coupled neural mass models can generate EEG rhythms. *Neural Comput* 19:478–512
- Sussman E, Winkler I (2001) Dynamic sensory updating in the auditory system. *Brain Res Cogn Brain Res* 12:431–439
- Tiitinen H, May P, Reinikainen K, Naatanen R (1994) Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature* 372:90–92
- van Wassenhove V, Grant KW, Poeppel D (2005) Visual speech speeds up the neural processing of auditory speech. *Proc Natl Acad Sci USA* 102:1181–1186
- Winkler I, Karmos G, Naatanen R (1996) Adaptive modeling of the unattended acoustic environment reflected in the mismatch negativity event-related potential. *Brain Res* 742:239–252
- Zhang Z (1995) A fast method to compute surface potentials generated by dipoles within multilayer anisotropic spheres. *Phys Med Biol* 40:335–349
- Zumer JM, Attias HT, Sekihara K, Nagarajan SS (2007) A probabilistic algorithm integrating source localization and noise suppression for MEG and EEG data. *Neuroimage* 37:102–115