NeuroImage 136 (2016) 227–257

Contents lists available at ScienceDirect

NeuroImage

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

Probabilistic delay differential equation modeling of event-related potentials

Dirk Ostwald [a,b,*], Ludger Starke [a]

[a] Computational Cognitive Neuroscience Unit, Center for Cognitive Neuroscience Berlin, Department of Education and Psychology, Freie Universität Berlin, Germany
[b] Max Planck Institute for Human Development, Center for Adaptive Rationality (ARC), Berlin, Germany

ARTICLE INFO

Article history:
Received 5 October 2015
Revised 9 April 2016
Accepted 12 April 2016
Available online 22 April 2016

ABSTRACT

"Dynamic causal models" (DCMs) are a promising approach in the analysis of functional neuroimaging data due to their biophysical interpretability and their consolidation of functional-segregative and functional-integrative propositions. In this theoretical note we are concerned with the DCM framework for electroencephalographically recorded event-related potentials (ERP-DCM). Intuitively, ERP-DCM combines deterministic dynamical neural mass models with dipole-based EEG forward models to describe the event-related scalp potential time-series over the entire electrode space. Since its inception, ERP-DCM has been successfully employed to capture the neural underpinnings of a wide range of neurocognitive phenomena. However, in spite of its empirical popularity, the technical literature on ERP-DCM remains somewhat patchy. A number of previous communications have detailed certain aspects of the approach, but no unified and coherent documentation exists. With this technical note, we aim to close this gap and to increase the technical accessibility of ERP-DCM. Specifically, this note makes the following novel contributions: firstly, we provide a unified and coherent review of the mathematical machinery of the latent and forward models constituting ERP-DCM by formulating the approach as a probabilistic latent delay differential equation model. Secondly, we emphasize the probabilistic nature of the model and its variational Bayesian inversion scheme by explicitly deriving the variational free energy function in terms of both the likelihood expectation and variance parameters. Thirdly, we detail and validate the estimation of the model with a special focus on the explicit form of the variational free energy function and introduce a conventional nonlinear optimization scheme for its maximization. Finally, we identify and discuss a number of computational issues which may be addressed in the future development of the approach.

Introduction

"Dynamic causal models" (DCMs) are a promising approach in the analysis of functional neuroimaging data due to their biophysical interpretability and their consolidation of functional-segregative and functional-integrative propositions (Friston, 2011; Friston et al., 2003). Technically, the DCM framework refers to the variational Bayesian inversion of differential equation models describing the temporal evolution of latent neural states, which are augmented with neuroimaging modality-specific forward models to probabilistically predict observed M/EEG or fMRI responses (Daunizeau et al., 2011; Friston and Dolan, 2010; Kiebel et al., 2008). While the formulation in terms of latent differential equations, forward models, and inversion by means of variational Bayes is common to all currently employed DCMs, a number of DCM-variants for different neuroimaging modalities and based on different mathematical frameworks has been proposed over the last decade: for the analysis of fMRI data there now exist, for example, the original bilinear non-linear system approximation, a nonlinear extension thereof, two-state models which represent both excitatory and inhibitory neural populations, and stochastic versions that can be applied to resting-state fMRI data (Stephan and Roebroeck, 2012). DCMs for the analysis of M/EEG and invasive electrophysiological data have become similarly diverse: there now exist convolution-based neural mass models for time and frequency domain responses, conductance-based neural mass models based on mean-field approximations, neural field models describing spatially extended cortical activity, and canonical microcircuit models that aim to describe asymmetries in cortical forward and backward message passing (Moran et al., 2013). In this theoretical study we are concerned with the mathematical details of a very specific DCM framework for event-related potentials (ERPs) that is based on neural mass models and was originally proposed by David et al. (2006). For simplicity, we will refer to this framework as ERP-DCM ("event-related potential dynamic causal modeling") in the following.

Intuitively, ERP-DCM combines deterministic dynamical neural mass models with dipole-based EEG forward models (Ilmoniemi, 1993) to describe the event-related scalp potential time-series over

* Corresponding author at: Computational Cognitive Neuroscience Unit, Center for Cognitive Neuroscience Berlin, Department of Education and Psychology, Freie Universität Berlin, Habelschwerdter Allee 45, Raum JK 26/221b, 14195 Berlin, Germany.

http://dx.doi.org/10.1016/j.neuroimage.2016.04.025
1053-8119/

the entire electrode space. Specifically, a linear combination of source-specific latent neural activity states is used to model one-dimensional dipole time-courses, which in turn are projected to electrode space using a standard lead-field approach (Grech et al., 2008; Hallez et al., 2007). Based on experimentally observed data, free parameters of the latent neural dynamics as well as the forward model (e.g., neural between-source connectivity values and dipole moments) are numerically estimated. In a Bayesian context, this estimation involves the determination of posterior parameter distributions based on suitable chosen prior distributions, which may also be conceived as regularization constraints, and the overall model plausibility or "evidence" for describing a given data set (Lindley, 1987, 2000).

Since its inception, ERP-DCM has been successfully employed to capture the neural underpinnings of a wide range of neurocognitive phenomena. In an initial series of studies, ERP-DCM was used to characterize the functional architecture of fronto-temporal networks giving rise to auditory mismatch negativity (MMN) responses (Garrido et al., 2007a, 2007b, 2008, 2009). Key insights in this line of research were that both forward (bottom-up) and backward (top-down) connectivity modulations are required for the expression of MMN effects, where the former primarily modulate early and the latter primarily late ERP components. The application of ERP-DCM to MMN effects has recently been expanded to study the changes in effective connectivity caused by age (Cooray et al., 2014, 2015), disease (Fogelson et al., 2014), and pharmacological intervention (Schmidt et al., 2013). Further empirical applications of the ERP-DCM approach have so far involved visual perception (Brown and Friston, 2012; Sharaev and Mnatsakanian, 2014), somatosensory awareness (Auksztulewicz and Blankenburg, 2013; Auksztulewicz et al., 2012) and audio-spatial processing (Dietz et al., 2014), to name just a few.

Despite the empirical popularity of the approach, the technical literature on ERP-DCM is somewhat patchy. In the seminal paper (David et al., 2006) the latent neural state model and its biological motivation are described in depth. The formulation in David et al. (2006) is based on two earlier communications, which introduced the initial development of the classical Jansen–Rit model (Grimbert and Faugeras, 2006; Jansen and Rit, 1995) to a neural state model for multiple cortical sources (David and Friston, 2003; David et al., 2005). In David et al. (2006), however, the delay differential equation form of the multiple source model is emphasized rather indirectly by an appendix, that discusses a (non-standard) approach to the integration of delay differential systems. Further, in contrast to standard implementations of the ERP-DCM approach, the EEG forward model is not explicitly parameterized, and the variational Bayesian inversion is covered only superficially. Kiebel et al. (2006) complements David et al. (2006) by introducing a parameterization of the dipole lead-fields, which assumedly forms the standard implementation of ERP-DCM in most empirical studies, but again, the variational inversion is largely only referenced and many implementational details are omitted. The most in-depth treatment of the variational Bayesian inversion scheme for model estimation in the ERP-DCM context is probably provided by Friston et al. (2007). In this communication, the authors introduced a general fixed-form variational Bayesian approach for the inversion of a variety of probabilistic models in the SPM implementation (Litvak et al., 2011). The approach rests on the analytical and numerical optimization of a variational free energy objective function. An explicit form of this function is given; the derivation of it is, however, absent from the paper. Partial derivations of this central function for model estimation in ERP-DCM are found in an unpublished internal report available from the SPM website (Stephan et al., 2005), which, however, focuses on DCMs for fMRI, and in the appendices of Penny (2012)) and Cooray et al. (in press). In addition to introducing the model inversion objective function, Friston et al. (2007) also introduces a numerical scheme for its optimization. This scheme is inspired by work on the online inferential processes that may operate in systems like the brain (Friston et al., 2008a, 2008b, 2010) and is formulated as parameter motion in continuous time. In comparison to standard Newton–Raphson methods (Nocedal and Wright, 2006), its analytical properties are thus studied to a lesser degree.

Based on the empirical success of the ERP-DCM approach on the one hand, and the patchy state of the technical ERP-DCM literature on the other hand, we reasoned that a theoretical note covering both model formulation (latent neural state model and forward model) and model estimation (variational Bayes) is warranted. Furthermore, some design choices of the ERP-DCM framework are best understood from the historical perspective of the development of the SPM toolbox (Ashburner, 2012), while they do not necessarily appear most coherent from an outsider's perspective. In our formalization, we thus took the liberty to modify the ERP-DCM approach in certain regards and hence refer to it as "probabilistic delay differential equation modeling of event-related potentials". We report smaller modifications in the technical sections below, but here would like to emphasize three principal adaptations. Firstly, we clarify the probabilistic manner in which parameters of the neural state model, EEG lead-field model, and observation error variance are handled in schemes such as "Variational-Laplace" (Friston et al., 2007) by deriving the variational free energy function in terms of both the likelihood expectation and variance parameters. Secondly, for the numerical optimization of the variational free energy function using Newton's method, we revert to a conventional backtracking step-size selection, rather than the "temporal regularization approach" introduced in Friston et al. (2007). Both modifications were motivated by the intent to relate ERP-DCM more directly to the standard literature on Bayesian estimation and numerical optimization, and thus render it more accessible for the technically-minded novice. Thirdly, we largely focus on modeling a single type of event-related potential, i.e. we omit the expression of stimulus- or condition-specific effects, and only cover their inclusion in passing. As a consequence of these modifications, the current theoretical note is presumably best understood as a simplified technical review, full derivation, and variation of the ERP-DCM approach. For less technical introductions, we refer the interested reader to (Daunizeau et al., 2011; Kiebel et al., 2008; Moran et al., 2013; Stephan et al., 2010).

In sum, the current technical note makes the following novel contributions: firstly, we provide a unified and detailed review of the mathematical machinery of the latent and forward models constituting ERP-DCM. Secondly, we emphasize the probabilistic nature of the model and its variational Bayesian inversion scheme by explicitly deriving the variational free energy function. Thirdly, we introduce a conventional and theoretically validated nonlinear optimization approach for its optimization. Finally, we identify a number of computational issues which may be addressed in the future development of the ERP-DCM framework. In order to achieve these aims, we dissolved the ERP-DCM implementation from its SPM context and provide all software underlying the reported simulations as Supplementary Material.

The outline of this technical note is as follows. In Section 2, we introduce the overall model formulation, and then detail the latent neural dynamics model and the dipole-based forward model in Sections 2.3 and 2.4, respectively. Because the latent neural dynamics model is formulated as a system of delay differential equations, we devote considerable attention to its temporal integration in Section 2.3. We conclude model formulation by embedding the forward model-augmented latent neural dynamics model in a joint distribution over data and parameters, i.e. a probabilistic model. In Section 3 we then consider the variational Bayesian estimation of this model. Specifically, we derive the variational free energy objective function, the optimization of which allows for the derivation of approximations to the model's parameter posterior distribution and log evidence (Section 3.1), and discuss a standard semi-analytical nonlinear optimization approach for its maximization (Section 3.2). In Section 4, we firstly apply the inversion framework in the context of two toy examples in order to obtain some intuition about its inner workings and validity (Section 4.1). In Section 4.2 we

then showcase the estimation of the ERP-DCM model framework of Section 2 under different model assumptions. Finally, in Section 4.3 we consider aspects of experimental applications, such as the estimation of condition-specific effects, model comparison, and an example application to real ERP data. We close with a discussion of the technical limitations and possible future developments of the current framework.

Model formulation

Notation

We aimed for standard mathematical notation throughout. Our notation of the relevant probability distributions is summarized in Supplementary Material S1. We use the *vec* operator to denote the column-wise vectorization of matrices, $|\cdot|$ to denote a matrix determinant, $tr(\cdot)$ to denote the trace of a matrix, $diag(\cdot)$ to denote diagonal matrices, I_n to denote the $n \times n$ identity matrix, and $\otimes$ to denote the Kronecker matrix product. For derivatives of vector functions with respect to time, we use the dot notation, i.e., $\dot{x}$ for $\frac{d}{dt}x : \mathbb{R} \to \mathbb{R}^n$. For a scalar-valued function $f: \mathbb{R}^m \to \mathbb{R}$, $\theta \mapsto f(\theta)$ we denote its gradient with respect to the input arguments $\tilde{\theta} \subseteq \theta$ evaluated at $\theta \in \mathbb{R}^m$ by $\nabla_{\tilde{\theta}} f(\theta)$ and its $m \times m$ Hessian matrix with respect to $\tilde{\theta}$ evaluated at $\theta \in \mathbb{R}^m$ by $H^f_{\tilde{\theta}}(\theta)$. Finally, for a vector-valued function $f: \mathbb{R}^m \to \mathbb{R}^n$, $\theta \mapsto f(\theta)$, we use $J^f(\theta)$ to denote its $n \times m$ Jacobian matrix evaluated at $\theta \in \mathbb{R}^m$. For the numerical evaluation of derivatives, we use the utilities of the SPM distribution (http://www.fil.ion.ucl.ac.uk/spm/), specifically the routine *spm_diff.m*, which is based on a finite-forward difference scheme and has recently been reviewed in Sengupta et al. (2014).

Model overview

We start our formulation of the ERP-DCM model by considering the electrode × peri-stimulus time data matrix that ERP-DCM models. Let n_e denote the number of scalp electrodes used for EEG data acquisition and n_t the number of discrete ERP peri-stimulus time bins. We denote this data matrix by:

$$Y := (y_1, \ldots, y_{n_t}) \in \mathbb{R}^{n_e \times n_t}, \tag{1}$$

where $y_t \in \mathbb{R}^{n_e}$ denotes the potential vector over electrodes for discrete time points $t = 1,\ldots,n_t$. For model inversion, we consider the data matrix Y in vectorized form:

$$y := vec(Y) \in \mathbb{R}^n, \tag{2}$$

where we defined the total number of data points by $n := n_e n_t$. We conceive $y \in \mathbb{R}^n$ as a realization of an n-dimensional "generalized linear model," i.e. a random vector distributed according to a multivariate normal distribution:

$$y = h(\theta) + \varepsilon, p(\varepsilon) = N(\varepsilon; 0, \sigma^2 I_n) \Leftrightarrow p(y) = N(y; h(\theta), \sigma^2 I_n). \tag{3}$$

In (3)

$$h : \Theta \to \mathbb{R}^n, \theta \mapsto h(\theta) := vec(g(\theta_g) f(\theta_f)) \tag{4}$$

denotes a nonlinear function of the model parameters $\theta := \{\theta_f, \theta_g\}$ generating the expectation parameter of the normal distribution in (3) and $\sigma^2 > 0$, such that $\sigma^2 I_n$ corresponds to a positive-definite covariance matrix modeling independent and identically distributed observation errors. $f(\theta_f)$ describes the latent neural dynamics model to be detailed in Section 2.3 and $g(\theta_g)$ describes the EEG forward model to be detailed in Section 2.4. In Section 2.5, we will embed (4) in a probabilistic model, i.e. conceive the parameters θ and σ^2 as random variables, rendering (4) the "likelihood" of a Bayesian scenario.

We note that our formulation of the generalized linear model in (3) deviates from the formulation of ERP-DCM in the SPM context in two regards. Firstly, in SPM error covariances are commonly formulated as linear combinations of inverse covariance, i.e. precision, basis functions (Friston et al., 2002a, 2002b). However, for ERP-DCM only one error precision parameter is optimized by assuming a single basis function. This basis function originally conformed to the $n \times n$ identity matrix equivalent to (3) in the context of relatively low temporal data resolutions (David et al., 2006), and has since been adapted to reflect an autoregressive noise process. Secondly, we omit modeling of an additive nuisance variable term in (4). In SPM this

Table 1
State variable neurobiological connotations. See also Fig. 1 for a full discussion.

State variable	Neurobiological connotation
$x_1^{(i)}$	Membrane potential of the spiny stellate cell population
$x_2^{(i)}$	Positive membrane potential of the pyramidal cell population
$x_3^{(i)}$	Negative membrane potential of the pyramidal cell population
$x_4^{(i)}$	Depolarizing current of the spiny stellate cell population
$x_5^{(i)}$	Depolarizing current of the pyramidal cell population
$x_6^{(i)}$	Hyperpolarizing current of the pyramidal cell population
$x_7^{(i)}$	Membrane potential of the inhibitory interneurons
$x_8^{(i)}$	Depolarizing current of the inhibitory interneurons
$x_9^{(i)}$	Net membrane voltage of the pyramidal neurons

term reflects the linear combination of a discrete cosine set and is usually employed to describe low-frequency data drifts. In our formulation, we assume that the electrode data has been appropriately mean-corrected/high-pass filtered during EEG data pre-processing, rendering additional nuisance variables redundant.

The neural dynamics model

The ERP-DCM neural dynamics model in delay differential equation form

Cortical neural dynamics in ERP-DCM are modeled as a system of coupled delay differential equations, where subsets of state variables represent "cortical sources" that give rise to dipole activations in an EEG forward model. Each cortical source comprises nine one-dimensional state variables that model the aggregate biophysical properties (transmembrane currents and transmembrane voltages) of three cell-type populations: spiny stellate cells, representing the input population of the cortical source, inhibitory interneurons, and pyramidal cells, representing the output population of the cortical source. Let n_s denote the number of assumed cortical sources for a given data set and let $\nu := 9n_s$ denote the total number of state variables. The values of the state variables over a time interval I can be cast as the function

$$x : I \subset \mathbb{R} \to \mathbb{R}^{\nu}, t \mapsto x(t) := \left(x_1^{(1)}(t), \ldots, x_9^{(1)}(t), \ldots, x_1^{(n_s)}(t), \ldots, x_9^{(n_s)}(t)\right)^T. \tag{5}$$

Typically, the time-interval $I \subset \mathbb{R}$ of interest corresponds to a peri-stimulus time window in the order of a couple of hundred milliseconds. For the ith source, the nine state variables $x_1^{(i)},\ldots, x_9^{(i)}$ represent the neurobiological concepts listed in Table 1. In most general form, the dynamics of the ERP-DCM neural state (1) can be expressed by a system of delay differential equations:

$$\dot{x}(t) = \phi_{\theta_\phi}\left(u(t), (x(t-d_{kl}))_{1 \le k,l \le \nu}\right), \tag{6}$$

where

$$\phi_{\theta_\phi} : \mathbb{R} \times \mathbb{R}^{\nu^2} \to \mathbb{R}^{\nu} \tag{7}$$

denotes the system's "evolution function," parameterized by the parameter set θ_ϕ. Note that here and in the following we use k and l to index state variables irrespective of their cortical source context, and i and j to index state variables with respect to their cortical source context. In (6) u denotes the time-dependent input function of the system. This function is given by

$$u : I \to \mathbb{R}, t \mapsto u(t) := 32 \exp\left(-\frac{(t-d)^2}{2s^2}\right) \tag{8}$$

and describes extrinsic input entering the system. In ERP-DCM this input, usually assumed to be of thalamic origin, is parameterized in terms of a delay variable d and a scale variable s given by

$$d := t_o + 128\rho_1 \quad \text{and} \quad s := t_w \exp(\rho_2), \tag{9}$$

respectively. In (9), t_o and t_w are fixed, user-specified scalar input-onset and input-duration time parameters, while ρ_1 and ρ_2 correspond to free parameters of the neural dynamics model that can be estimated during model inversion. Finally, d_{kl} $(1 \le k, l \le \nu)$ denote between-state variable delay parameters, some of which correspond to free parameters of the neural dynamics model as discussed below. We use $D := (d_{kl})_{1 \le k,l \le \nu} \in \mathbb{R}^{\nu \times \nu}$ to denote the complete set of delay parameters. Below, we further classify the entries of D into within- and between-source delay parameters, which will be denoted by δ_0 and δ_{ij}, respectively. Also note that the system of delay differential equations (6) is formulated in continuous time. In order to enable such a model to serve as a model of digitized, discrete EEG data of the form (1), some appropriate form of temporal integration and discretization is required. This will be discussed in detail below. For the moment, we follow the conventional portrayal of the ERP-DCM neural dynamics model in continuous time and further unpack Eq. (6). To this end, we first note that the ERP-DCM system evolution function ϕ_{θ_ϕ} partitions source-wise. This means that state variables within a source affect each other with an "intrinsic" delay δ_0, and that pyramidal cell net membrane voltages $x_9^{(j)}(i = 1,\ldots,n_s)$ enter a given partition $\phi_{\theta_\phi}^{(i)}$ with "extrinsic" between-source delays δ_{ij} $(1 \le i, j \le n_s, i \ne j)$. A more detailed account of (6) is thus given by

$$\begin{pmatrix} \dot{x}^{(1)}(t) \\ \dot{x}^{(2)}(t) \\ \vdots \\ \dot{x}^{(n_s)}(t) \end{pmatrix} = \begin{pmatrix} \phi_{\theta_\phi}^{(1)}\left(u(t), x^{(1)}(t-\delta_0), \left(x_9^{(j)}(t-\delta_{1j})\right)_{1 \le j \le n_s, j \ne i}\right) \\ \phi_{\theta_\phi}^{(2)}\left(u(t), x^{(2)}(t-\delta_0), \left(x_9^{(j)}(t-\delta_{2j})\right)_{1 \le j \le n_s, j \ne i}\right) \\ \vdots \\ \phi_{\theta_\phi}^{(n_s)}\left(u(t), x^{(n_s)}(t-\delta_0), \left(x_9^{(j)}(t-\delta_{n_s j})\right)_{1 \le j \le n_s, j \ne i}\right) \end{pmatrix}, \tag{10}$$

where the $\phi_{\theta_\phi}^{(i)}$ denote source-specific evolution functions

$$\phi_{\theta_\phi}^{(i)} : \mathbb{R} \times \mathbb{R}^9 \times \mathbb{R}^{n_s} \to \mathbb{R}^9 \quad (i = 1, \ldots, n_s). \tag{11}$$

Eq. (10) thus defines the dynamical behavior of the nine state variables of the ith source by

$$\dot{x}^{(i)}(t) := \phi_{\theta_\phi}^{(i)}\left(u(t), x^{(i)}(t-\delta_0), \left(x_9^{(j)}(t-\delta_{ij})\right)_{1 \le j \le n_s, j \ne i}\right). \tag{12}$$

In words, (12) states that the rate of change of the ith source neural state vector at time t is a function of the input state at time t, its own state at time $t - \delta_0$, and the 9th state variables $x_9^{(j)}$ of all sources $j = 1,\ldots,n_s$ at source-source specific delayed time-points $t - \delta_{ij}$. We next consider the specific functional form of the functions $\phi_{\theta_\phi}^{(i)}(i = 1,\ldots, n_s)$ in (12). These have the form

$$\phi_{\theta_\phi}^{(i)}\left(u^{(i)}(t), x^{(i)}(t-\delta_0), \left(x_9^{(j)}(t-\delta_{ij})\right)_{1 \le j \le n_s, j \ne i}\right) := \begin{pmatrix} x_4^{(i)}(t-\delta_0) \\ x_5^{(i)}(t-\delta_0) \\ x_6^{(i)}(t-\delta_0) \\ \frac{\tilde{H}_e^{(i)}}{\tilde{\tau}_e^{(i)}}\left(\sum_{j=1,j\ne i}^{n_s} a_{ij}^F S\left(x_9^{(j)}(t-\delta_{ij})\right) + \sum_{j=1,j\ne i}^{n_s} a_{ij}^L S\left(x_9^{(j)}(t-\delta_{ij})\right) + \tilde{\gamma}_1 S\left(x_9^{(i)}(t-\delta_{ij})\right) + 2c_i u(t)\right) - \frac{2x_4^{(i)}(t-\delta_0)}{\tilde{\tau}_e^{(i)}} - \frac{x_1^{(i)}(t-\delta_0)}{{\tilde{\tau}_e^{(i)}}^2} \\ \frac{\tilde{H}_e^{(i)}}{\tilde{\tau}_e^{(i)}}\left(\sum_{j=1,j\ne i}^{n_s} a_{ij}^B S\left(x_9^{(j)}(t-\delta_{ij})\right) + \sum_{j=1,j\ne i}^{n_s} a_{ij}^L S\left(x_9^{(j)}(t-\delta_{ij})\right) + \tilde{\gamma}_2 S\left(x_1^{(i)}(t-\delta_0)\right)\right) - \frac{2x_5^{(i)}(t-\delta_0)}{\tilde{\tau}_e^{(i)}} - \frac{x_2^{(i)}(t-\delta_0)}{{\tilde{\tau}_e^{(i)}}^2} \\ \frac{\tilde{H}_i}{\tilde{\tau}_i}\left(\tilde{\gamma}_4 S\left(x_7^{(i)}(t-\delta_0)\right) - 2\frac{x_6^{(i)}(t-\delta_0)}{\tilde{\tau}_i} - \frac{x_3^{(i)}(t-\delta_0)}{\tilde{\tau}_i^2}\right) \\ x_8^{(i)} \\ \frac{\tilde{H}_e^{(i)}}{\tilde{\tau}_e^{(i)}}\left(\sum_{j=1,j\ne i}^{n_s} a_{ij}^B S\left(x_9^{(j)}(t-\delta_{ij})\right) + \sum_{j=1,j\ne i}^{n_s} a_{ij}^L S\left(x_9^{(j)}(t-\delta_{ij})\right) + \tilde{\gamma}_3 S\left(x_9^{(i)}(t-\delta_0)\right)\right) - 2\frac{x_8^{(i)}(t-\delta_0)}{\tilde{\tau}_e^{(i)}} - \frac{x_7^{(i)}(t-\delta_0)}{{\tilde{\tau}_e^{(i)}}^2} \\ x_5^{(i)}(t-\delta_0) - x_6^{(i)}(t-\delta_0) \end{pmatrix}. \tag{13}$$

In (13), S denotes a function that models the conversion of neural (population) cell membrane potentials to neural (population) firing rates. This "activation function" is given by the sigmoid function

$$S : \mathbb{R} \to \mathbb{R}, \xi \mapsto S(\xi) := \frac{1}{1 + \exp(-r_1(\xi - r_2))} - \frac{1}{1 + \exp(r_1 r_2)} \tag{14}$$

and is parameterized in terms of two parameters s_1 and s_2, which enter (14) in the form

$$r_1 := \frac{2}{3}\exp(s_1) \text{ and } r_2 := \frac{1}{3}\exp(s_2). \tag{15}$$

The parameters $s_1, s_2 \in \mathbb{R}$ correspond to free parameters of the neural dynamics model that can be estimated during model inversion and are identical over all applications of the function S. In addition to S, the explicit formulation of the function $\phi_{\theta_\phi}^{(i)}$ in (13) has introduced a number of additional parameters. These parameters comprise

- three parameter sets that describe the "forward," "backward," and "lateral" connectivity of the cortical source neural dynamic model $A^F := (a_{ij}^F)_{1 \le i,j \le n_s} \in \mathbb{R}^{n_s \times n_s}$, $A^B := (a_{ij}^B)_{1 \le i,j,n_s} \in \mathbb{R}^{n_s \times n_s}$, and $A^L := (a_{ij}^L)_{1 \le i,j,n_s} \in \mathbb{R}^{n_s \times n_s}$,
- an "extrinsic input" connectivity vector $C := (c_i)_{1 \le i \le n_s} \in \mathbb{R}^{n_s}$,
- source-specific "excitatory receptor densities" $\tilde{H}^{ex} := (\tilde{H}_i^{ex})_{1 \le i \le n_s} \in \mathbb{R}_+^{n_s}$,
- source-specific "inhibitory receptor densities" $\tilde{H}^{in} := (\tilde{H}_i^{in})_{1 \le i \le n_s} \in \mathbb{R}_+^{n_s}$,

Table 2
Parameter set of the neural dynamics model and conversion formulas for positively constrained parameters.

Interpretation	Parameter set θ_ϕ	Conversion formulas
Input function parameters	$\rho := (\rho_i)_{i=1,2} \in \mathbb{R}^2$	–
Activation function parameters	$s := (s_i)_{i=1,2} \in \mathbb{R}^2$	–
Forward connectivity	$A^F := (a_{ij}^F)_{1 \le i,j \le n_s} \in \mathbb{R}^{n_s \times n_s}$	–
Backward connectivity	$A^B := (a_{ij}^B)_{1 \le i,j \le n_s} \in \mathbb{R}^{n_s \times n_s}$	–
Lateral connectivity	$A^L := (a_{ij}^L)_{1 \le i,j \le n_s} \in \mathbb{R}^{n_s \times n_s}$	–
Input connectivity	$C := (c_i)_{1 \le i \le n_s} \in \mathbb{R}^{n_s}$	–
Excitatory receptor density	$H_{ex} := (H_{ex}^i)_{1 \le i \le n_s} \in \mathbb{R}^{n_s}$	$\tilde{H}_{ex}^i = 4\,, \exp(H_{ex}^i)$
Inhibitory receptor density	$H_{in} := (H_{in}^i)_{1 \le i \le n_s} \in \mathbb{R}^{n_s}$	$\tilde{H}_{in}^i = 32\,, \exp(H_{in}^i)$
Excitatory time constants	$\tau_{ex} := (\tau_{ex}^i)_{1 \le i \le n_s} \in \mathbb{R}^{n_s}$	$\tilde{\tau}_{ex}^i = 0.008\,, \exp(\tau_e^i)$
Inhibitory time constants	$\tau_{in} := (\tau_{in}^i)_{1 \le i \le n_s} \in \mathbb{R}^{n_s}$	$\tilde{\tau}_{in}^i = 0.016\,, \exp(\tau_i^i)$
Intrinsic coupling	$\gamma := (\gamma_i)_{1 \le i \le 4} \in \mathbb{R}^4$	$\tilde{\gamma}_1 = 128\,, \exp(\gamma_1)$ $\tilde{\gamma}_2 = \frac{512}{5}\,, \exp(\gamma_2)$ $\tilde{\gamma}_3 = 32\,, \exp(\gamma_3)$ $\tilde{\gamma}_3 = 32\,, \exp(\gamma_3)$

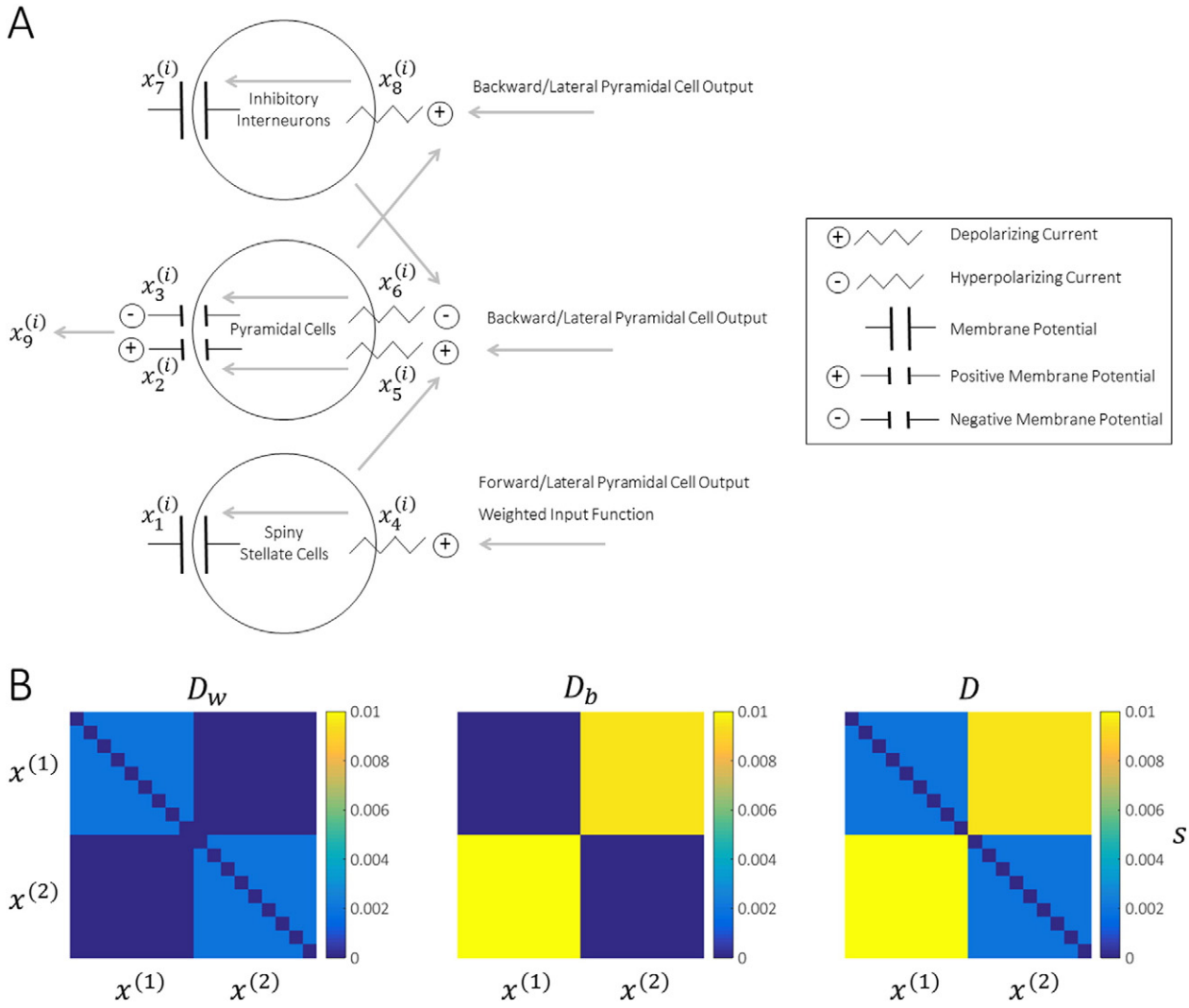

Fig. 1. State variable neurobiological connotations and delay parameters. (A) The formulation of the source-specific neural dynamics in Eq. (13) reflects the following neurobiological connotations for the neural state variables $x_1^{(i)}, x_2^{(i)}, \ldots, x_9^{(i)}$ of the ith source: (1) The rate of change of the membrane voltage of the spiny stellate cell population $\dot{x}_1^{(i)}$ is given by the depolarizing current of the spiny stellate cell population $x_4^{(i)}$. (2) The rate of change of the positive membrane potential of the pyramidal cell population $\dot{x}_2^{(i)}$ is given by the depolarizing current of the pyramidal cell population $x_5^{(i)}$. (3) The rate of change of the negative membrane potential of the pyramidal cell population $\dot{x}_3^{(i)}$ is given by the hyperpolarizing current of the pyramidal cell population $x_6^{(i)}$. (4) The rate of change of the depolarizing (excitatory) current of the spiny stellate cell population $\dot{x}_4^{(i)}$ is (a) positively modulated by the weighted sum of the delayed pyramidal cell activity in all other sources connected to the ith source by means of forward and lateral connections, the spiking output of the pyramidal cell population of the ith source, and the weighted activity in the input, and (b) negatively modulated by its current state and the membrane potential of spiny stellate cell population. (5) The rate of change of the depolarizing (excitatory) current of the pyramidal cell population $\dot{x}_5^{(i)}$ is (a) positively modulated by the spiking output of the spiny stellate cell population of the ith source and the weighted sum of the delayed pyramidal cell population activity in all other sources connected to the ith source by means of backward and lateral connections, and (b) modulated negatively by its current state and the positive membrane potential of the pyramidal cell population. (6) The rate of change of the hyperpolarizing (inhibitory) current of the pyramidal cell population $\dot{x}_6^{(i)}$ is (a) modulated positively by the weighted spiking output of the inhibitory interneuron population of the ith source and (b) modulated negatively by the hyperpolarizing current and the negative membrane potential of the pyramidal cell population. (7) The rate of change of the membrane potential of the inhibitory interneurons $\dot{x}_7^{(i)}$ is given by the depolarizing current of the inhibitory interneurons $x_8^{(i)}$. (8) The rate of change of the depolarizing current of the inhibitory interneurons $\dot{x}_8^{(i)}$ is (a) modulated positively by the spiking output of the ith-source pyramidal cell population, the weighted sum of the delayed pyramidal cell activity in all other sources connected to the ith source by means of backward and lateral connections, and (b) modulated negatively by the membrane potential and depolarizing current of the inhibitory interneurons of the ith source. (9) Finally, the rate of change of the pyramidal cell population net membrane potential $x_9^{(i)}$, which forms the basis for both the spiking output of the ith source and the forward mapping to electrode potentials, is modulated positively by the depolarizing current of the ith source pyramidal cell population $x_5^{(i)}$ and negatively by the hyperpolarizing current of the ith source pyramidal cell population $x_6^{(i)}$. (B) Delay parameters of the neural state system. The delays d_{kl} $(1 \le k,l \le \nu)$ between the $\nu := 9n_s$ state variables form a delay parameter matrix $D := (d_{kl})_{1 \le k,l \le \nu} \in \mathbb{R}^{\nu \times \nu}$ (right subpanel). This matrix is given by the sum of a within-source delay parameter matrix $D_w \in \mathbb{R}^{\nu \times \nu}$ (left subpanel) and a between-source parameter matrix $D_b \in \mathbb{R}^{\nu \times \nu}$ (middle subpanel). These two matrices in turn are constructed on the basis of a fixed within-source delay parameters δ_0 and free delay parameters δ_{ij} $(1 \le i,j \le n_s)$ as described Eqs. (32) and (33).

- source-specific "excitatory time-constants" $\tilde{\tau}^{ex} := (\tilde{\tau}_i^{ex})_{1 \le i \le n_s} \in \mathbb{R}_+^{n_s}$,
- source-specific "inhibitory time-constants" $\tilde{\tau}^{in} := (\tilde{\tau}_i^{in})_{1 \le i \le n_s} \in \mathbb{R}_+^{n_s}$, and
- four "intrinsic coupling" parameters $\gamma := (\tilde{\gamma}_1, \tilde{\gamma}_2, \tilde{\gamma}_3, \tilde{\gamma}_4)^T \in \mathbb{R}_+^4$.

We note that in contrast to the ERP-DCM implementation in SPM, we do not constrain the connectivity parameters A^F, A^B, A^L and C to be positive. This choice was made to prevent non-zero system responses to the input function for "approximately zero" connectivity parameter settings such as

$c_i = \exp(-64)$ during model inversion. All remaining parameters, denoted with a tilde, are constrained to be positive. To allow for the quantification of uncertainty about values of the parameters by means of normal distributions, these parameters derive from the weighted exponentiation of a corresponding set of parameters H^{ex}, H^{in}, τ^{ex}, τ^{in}, and γ in SPM, a convention we retain. The complete parameter set of ϕ is thus given by

$$\theta_\phi := \left\{\rho, s, A^F, A^B, A^L, C, H^{ex}, H^{in}, \tau^{ex}, \tau^{in}, \gamma\right\} \tag{16}$$

and summarized in Table 2. The specific conversion formulas from the unconstrained (tilde-free) parameters to the positive-constrained parameters above are reported in the last column of Table 2. Generally speaking, when assuming log normal distributions for non-negative (scale) parameters these parameters (which we denote by a tilde) are referred to as "log-scale parameters". The neurobiological connotations that the formulation of the neural dynamics model by means of Eq. (13) entails are discussed in Fig. 1.

We have introduced the neural dynamics model in continuous-time delay differential equation. As previously indicated, in order for such a model to serve as the basis for describing discrete-time data, it has to be integrated and appropriately discretized. In ERP-DCM, this temporal discretization and integration comprises two steps: firstly, the conversion of the delay differential equation system into a system of ordinary differential equations (David et al., 2006, Appendix A.1), and secondly the numerical integration of the resulting system of ordinary differential equations (David et al., 2006, Appendix A.2). We will discuss each step in turn.

Conversion of delay differential equations to ordinary differential equations in ERP-DCM

To add to the transparency of the following, we formulate the conversion for the special case of a single delay differential equation in Supplementary Material S2. For the conversion of the system of delay differential equations (6) into a system of ordinary differential equations, we first relabel the system's state variables and consider each variable in isolation (i.e. irrespective of its source-specific context) as x_k ($k = 1,2,\ldots,\nu := 9n_s$). We thus consider a general delay differential equation system of ν state variables

$$\dot{x}_k(t_k) = \varphi_k(x_1(t_1-d_{k1}), x_2(t_2-d_{k2}), \ldots, x_\nu(t_\nu-d_{k\nu})) \; (k = 1, 2, \ldots, \nu). \tag{17}$$

and omit the distinction between within- and between-source delay parameters for the moment. Note that in the form of Eq. (17) each state variable x_k: $\mathbb{R} \to \mathbb{R}$, $t_k \mapsto x_k(t_k)$ is considered as a function of an individual time argument t_k. For our current purposes, let $t := (t_k)_{1 \le k \le \nu} \in \mathbb{R}^\nu$ denote the time vector and $d_k := (d_{kl})_{1 \le l \le \nu} \in \mathbb{R}^\nu$ denote the vector of delay parameters of relevance for the kth state variable. We can then write the above more compactly as

$$\dot{x}_k(t_k) = \varphi_k(x(t-d_k)) \quad (k = 1, 2, \ldots, \nu). \tag{18}$$

In (18) we implicitly defined the system function as x:$\mathbb{R}^\nu \to \mathbb{R}^\nu$,$t \mapsto x(t)$. ERP-DCM converts the system of delay differential Eq. (18) into a system of ordinary differential equations by means of Taylor approximations of the functions on the right-hand side, i.e. of the composed functions:

$$(\varphi_k \circ x) : \mathbb{R}^\nu \to \mathbb{R}, t \mapsto (\varphi_k \circ x)(t) := \varphi_k(x(t)) \quad (k = 1, 2, \ldots, \nu). \tag{19}$$

Specifically, to first order

$$\varphi_k(x(t-d_k)) \approx \varphi_k(x(t)) - \frac{d}{dt}\varphi_k(x(t))d_k. \tag{20}$$

With the chain rule of differentiation, the derivative in the above takes the following explicit form

$$\frac{d}{dt}\varphi_k(x(t)) = \frac{d}{dx}\varphi_k(x(t))\frac{d}{dt}x(t) = \left(\frac{\partial}{\partial x_1}\varphi_k(x(t)) \;\cdots\; \frac{\partial}{\partial x_\nu}\varphi_k(x(t))\right)\begin{pmatrix} \frac{\partial}{\partial t_1}x_1(t_1) & \frac{\partial}{\partial t_2}x_1(t_1) & \cdots & \frac{\partial}{\partial t_\nu}x_1(t_1) \\ \frac{\partial}{\partial t_1}x_2(t_2) & \frac{\partial}{\partial t_2}x_2(t_2) & \cdots & \frac{\partial}{\partial t_\nu}x_2(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial t_1}x_\nu(t_\nu) & \frac{\partial}{\partial t_2}x_\nu(t_\nu) & \cdots & \frac{\partial}{\partial t_\nu}x_\nu(t_\nu) \end{pmatrix} \tag{21}$$

Because in (21) $\frac{\partial}{\partial t_k}x_l(t_l) = 0$ for $k \neq l$ and with the notation $\dot{x}_l(t_l) = \frac{\partial}{\partial t_l}x_l(t_l)$, the approximation in (20) takes the form:

$$\varphi_k(x(t-d_k)) \approx \varphi_k(x(t)) - \sum_{l=1}^{\nu}\frac{\partial}{\partial x_l}\varphi_k(x(t))\dot{x}_l(t_l)d_{kl}. \tag{22}$$

We next consider the concatenation of the thus approximated functions φ_k in the form

$$\varphi : \mathbb{R}^\nu \to \mathbb{R}^\nu, x \mapsto \varphi(x) := \begin{pmatrix} \varphi_1(x_1, \ldots, x_\nu) \\ \vdots \\ \varphi_\nu(x_1, \ldots, x_\nu) \end{pmatrix} \tag{23}$$

with Jacobian evaluated at x given by

$$J_x^{\varphi}(x) := \left(J_{kl}^{\varphi}\right)(x) = \begin{pmatrix} \frac{\partial}{\partial x_1}\varphi_1(x) & \cdots & \frac{\partial}{\partial x_\nu}\varphi_1(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_1}\varphi_\nu(x) & \cdots & \frac{\partial}{\partial x_\nu}\varphi_\nu(x) \end{pmatrix} \in \mathbb{R}^{\nu\times\nu}. \tag{24}$$

We can then identify the terms $\frac{\partial}{\partial x_l}\varphi_k(x(t))$ in the sum on the right-hand side of (22) with the respective entries in the Jacobian matrix of φ, yielding

$$\varphi_k(x(t-d_k)) \approx \varphi_k(x(t)) - \sum_{l=1}^{\nu}\left(J_{kl}^{\varphi}\right)(x(t))\dot{x}_l(t_l)d_{kl} \ (k = 1,\ldots,\nu). \tag{25}$$

We next treat the approximation in (25) as equality and concatenate over $k = 1,\ldots,\nu$ into vector format, yielding

$$\begin{pmatrix} \dot{x}_1(t_1) \\ \vdots \\ \dot{x}_k(t_k) \\ \vdots \\ \dot{x}_\nu(t_\nu) \end{pmatrix} = \begin{pmatrix} \varphi_1(x(t)) \\ \vdots \\ \varphi_k(x(t)) \\ \vdots \\ \varphi_\nu(x(t)) \end{pmatrix} - \begin{pmatrix} d_{11}\left(J_{11}^{\varphi}\right)(x(t)) & \cdots & d_{1\nu}\left(J_{1\nu}^{\varphi}\right)(x(t)) \\ \vdots & \ddots & \vdots \\ d_{\nu 1}\left(J_{\nu 1}^{\varphi}\right)(x(t)) & \cdots & d_{\nu\nu}\left(J_{\nu\nu}^{\varphi}\right)(x(t)) \end{pmatrix} \begin{pmatrix} \dot{x}_1(t_1) \\ \vdots \\ \dot{x}_k(t_k) \\ \vdots \\ \dot{x}_\nu(t_\nu) \end{pmatrix}. \tag{26}$$

Eq. (26) can be written more compactly with the delay parameter matrix $D = (d_{kl})_{1\le k,l\le\nu} \in \mathbb{R}^{\nu\times\nu}$ and using the element-wise (Hadamard) matrix product "$\odot$" as:

$$\dot{x}(t) = \varphi(x(t)) - \left(D \odot J^{\varphi}(x(t))\right)\dot{x}(t) \tag{27}$$

Rearranging then yields

$$\dot{x}(t) = \left(I_\nu + \left(D \odot J^{\varphi}(x(t))\right)\right)^{-1}\varphi(x(t)). \tag{28}$$

Finally, by defining:

$$Q := I_\nu + \left(D \odot J^{\varphi}(x(t))\right) \tag{29}$$

we have rewritten the system of delay differential equations (18) as a system of ordinary differential equations

$$\dot{x}(t) = Q^{-1}\varphi(x(t)), \tag{30}$$

which depends on the delay parameters D and the Jacobian of φ evaluated at $x(t)$. Finally, we clarify the relation between the entries of the delay parameter matrix D and the within- and between-source delay parameters δ_0 and δ_{ij} ($1 \le i, j \le n_s, i \ne j$). The system delay parameter matrix D is given by

$$D := D_w + D_b, \tag{31}$$

where $D_w \in \mathbb{R}^{\nu\times\nu}$ and $D_b \in \mathbb{R}^{\nu\times\nu}$ are within-source and between-source delay parameter matrices. These matrices, in turn, are given by

$$D_w := I_{n_s} \otimes (\delta_0(1_9 - I_9)) \tag{32}$$

and

$$D_b := (\delta_w(\Delta - I_{n_s}) \otimes 1_9, \tag{33}$$

In (32), δ_0 is a fixed within-source delay parameter of the ERP-DCM neural dynamics set to $\delta_0 = 0.002\ s$ and $1_9 \in \mathbb{R}^{9\times 9}$ is a matrix of all ones. Intuitively, δ_0 describes between-layer delays (Moran et al., 2013). For $n_s = 2$, the matrix D_w is visualized in the left subpanel of Fig. 1B. In (33) $\Delta := (\exp(\delta_{ij})) \in \mathbb{R}^{n_s\times n_s}$ is a matrix of exponentiated between-source delay parameters δ_{ij} ($1 \le i,j \le n_s, i \ne j$), where for $i = j$, $\delta_{ij} := 0$, such that $\Delta_{ij} = 1$ for $i = j$, and $1_9 \in \mathbb{R}^{9\times 9}$ is a matrix of all ones. Here, δ_w reflects the prior expected between-source delay and is typically set to $\delta_w = 0.016$ s (Moran et al., 2013). For $n_s = 2$, $\delta_{12} = -0.52$, and $\delta_{21} = -0.06$ the matrix D_b is visualized in the middle subpanel of Fig. 1B. The resulting delay parameter matrix is shown in the right subpanel of Fig. 1B. Notably, the non-diagonal entries of $\Delta \in \mathbb{R}^{n_s\times n_s}$ form parameters of the ERP-DCM forward model (see below).

Temporal integration of the delayed neural dynamics model

We next consider the numerical discretization and integration of systems of the form

$$\dot{x}(t) = \psi(x(t)) := Q^{-1}\varphi(x(t)) \tag{34}$$

in ERP-DCM, where we absorb the inverse of D into the function

$$\psi : \mathbb{R}^{\nu} \to \mathbb{R}^{\nu}, x \mapsto \psi(x) \tag{35}$$

and continue to omit the dependence of φ on $u(t)$ and θ_ϕ. To temporally integrate ODE systems of the form (34), ERP-DCM capitalizes on a custom Taylor-expansion-based integration scheme, which can be motivated as follows (David et al., 2006, Appendix A.2). Given the state value at time $t \in \mathbb{R}$, $x(t)$, the state value at time $t + \Delta t$, where $\Delta t > 0$ indicates the user-specified step-size of the integration scheme (e.g., corresponding to the sampling time-bin width of the recorded EEG data), can be written as a Taylor expansion about $x(t)$ as follows:

$$x(t+\Delta t) = x(t) + \frac{d}{dt}x(t)\Delta t + \frac{d^2}{dt^2}x(t)\frac{1}{2!}\Delta t^2 + \ldots \tag{36}$$

Based on:

$$\frac{d}{dt}x(t) := \dot{x}(t) := \psi(x(t)) \tag{37}$$

the right-hand side of Eq. (36) can equivalently be written as

$$x(t+\Delta t) = x(t) + \psi(x(t))\Delta t + \frac{d}{dt}\psi(x(t))\frac{1}{2!}(\Delta t)^2 + \ldots \tag{38}$$

Using the chain rule of differentiation, we obtain

$$\frac{d}{dt}\psi(x(t)) = \frac{d}{dx}\psi(x(t))\frac{d}{dt}x(t) = J^{\psi}(x(t))\,\psi(x(t)). \tag{39}$$

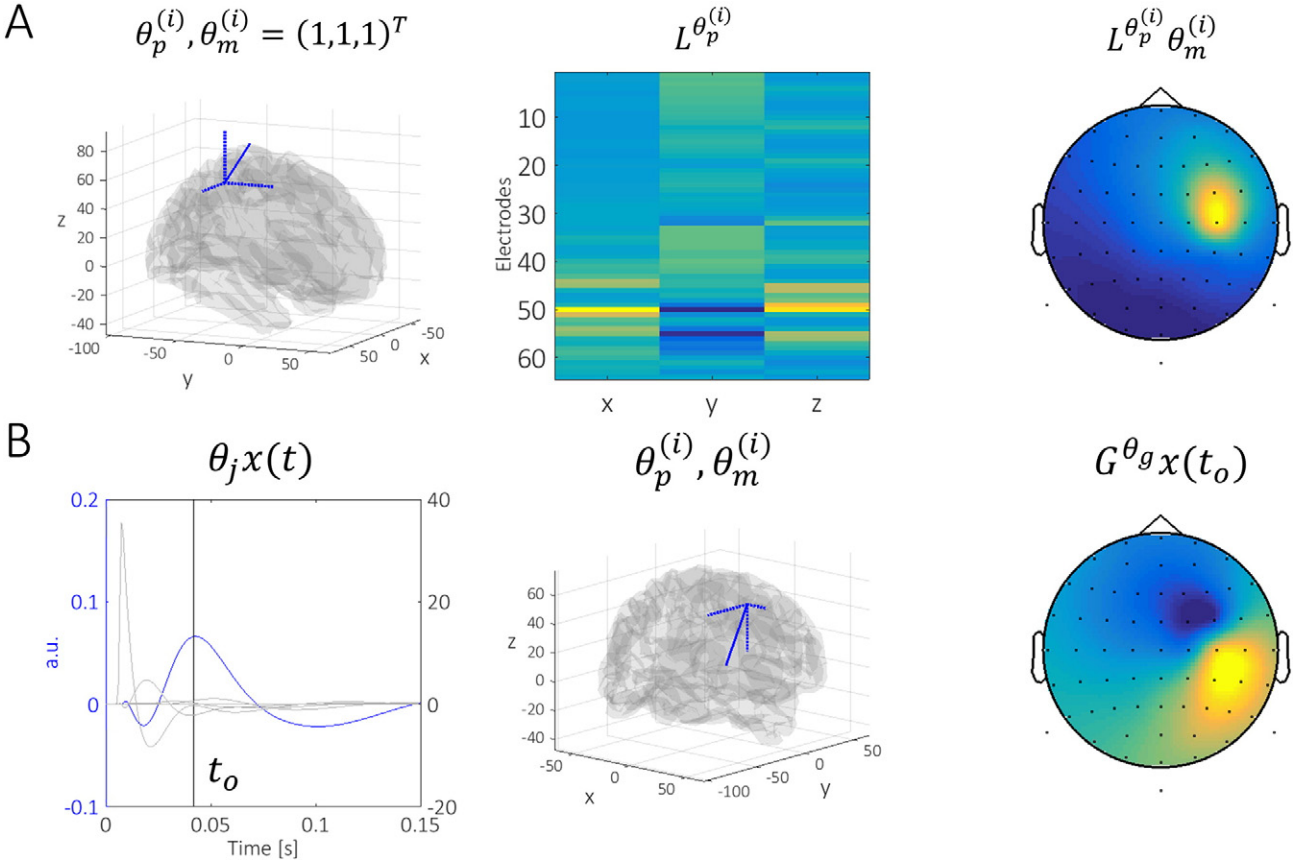

Fig. 2. EEG forward modeling. (A) The ERP-DCM forward model capitalizes on the evaluation of dipole position-dependent lead-field matrices $L^{\theta_p^{(i)}}$ for canonical dipole moments $\theta_m^{(i)(1)} = (1,0,0)^T, \theta_m^{(i)(2)} = (0,1,0)^T$, and $\theta_m^{(i)(3)} = (0,0,1)^T$. The left panel depicts these canonical moments for a dipole located at MNI coordinates $(42, -31, 58)$. Note that the dipole moment components are magnified for visualization purposes. The middle panel depicts the corresponding lead-field matrix as evaluated based on conventional solutions of the EEG forward problem implemented in the FieldTrip toolbox. The right panel depicts the projection of the canonical dipole moment of the left panel into electrode space. (B) To model ERPs over time, the latent neural state activity vector $x(t)$ is projected onto a univariate time-course by means of a neural state projection row vector $\theta_j \in \mathbb{R}^{1\times 9}$, which forms part of the EEG forward model parameter set. This univariate projection modulates the dipole moment length over time. In the left panel, the gray lines depict the dynamics of a single source's nine state variables and the blue line depicts their projection onto a univariate time-course. The middle panel depicts a dipole moment $\theta_m^{(i)}$ and location $\theta_p^{(i)}$ that does not correspond to the canonical moment. The projection of this moment based on the univariate state system projection at time $t = t_o$ is shown in the right panel. Note that the electrode space project is markedly different from the electrode space projection of the canonical moment.

Table 3
Parameter set of the EEG dipole forward model.

Interpretation	Parameter set θ_g
Dipole positions	$\theta_p := (\theta_p^{(1)},\ldots,\theta_p^{(n_s)}) \in \mathbb{R}^{3\times n_s}$
Dipole moments	$\theta_m := (\theta_m^{(1)},\ldots,\theta_m^{(n_s)}) \in \mathbb{R}^{3\times n_s}$
Neural state projection	$\theta_j \in \mathbb{R}^{1\times 9}$

Hence:

$$\begin{aligned} x(t+\Delta t) &= x(t) + \psi(x(t))\Delta t + \frac{1}{2!}J^{\psi}(x(t))\,\psi(x(t))(\Delta t)^2 + \ldots \\ &= x(t) + \left(\Delta t + \frac{1}{2!}J^{\psi}(x(t))\Delta t^2 + \ldots\right)\psi(x(t)). \end{aligned} \tag{40}$$

Finally, we can rewrite the expression

$$U := \Delta t + \frac{1}{2!}J^{\psi}(x(t))(\Delta t)^2 + \ldots \tag{41}$$

in matrix exponential form, because with the series definition of the exponential function, we have

$$\begin{aligned} U &= \Delta t + \frac{1}{2!}J^{\psi}(x(t))(\Delta t)^2 + \ldots \\ &= \left(I_\nu + \Delta t J^{\psi}(x(t)) + \frac{1}{2!}(\Delta t)^2\left(J^{\psi}(x(t))\right)^2 + \ldots - I_\nu\right)\left(J^{\psi}(x(t))\right)^{-1} \\ &= \left(\exp\left(\Delta t \cdot J^{\psi}(x(t))\right) - I_\nu\right)\left(J^{\psi}(x(t))\right)^{-1}. \end{aligned} \tag{42}$$

In the context of (stochastic) nonlinear dynamical systems, the result of Eq. (42) is known as "local linearization" (Ozaki, 1992). By choosing an integration step-size Δt and an initial condition $x(t_0)$ a recursive evaluation scheme for the neural dynamics system is thus given by

$$\text{for } j = 1, 2, \ldots, n_t$$

$$x(t_j) := x(t_{j-1}) + \exp\left(\Delta t \cdot J^{\psi}(x(t)) - I_\nu\right)\left(J^{\psi}(x(t))\right)^{-1}\psi(x(t_{j-1})) \tag{43}$$

which yields a discrete series of neural state activation vectors $x_j, j = 1,\ldots,n_t$. Because the integer index j reflects a temporal property, we will replace it by $t = 1,\ldots,n_t$ in the following, discarding the notion of continuous time.

The integrated form of the neural dynamics model

To summarize, in this section, we introduced the ERP-DCM neural dynamics model in the form of the delay differential equation system

$$\dot{x}(t) = \phi_{\theta_\phi}\left(u(t), (x(t-d_{kl}))_{1\le k,l\le \nu}\right) \tag{44}$$

in its continuous time formulation and have discussed its temporal integration and discretization. From the perspective of estimating the model parameters, the most important feature of this integration scheme is that it can be interpreted as a conversion of the neural dynamics model as a system of DDEs into a formulation of the neural dynamics model as a function that maps a parameter set $\theta_f := \{\theta_\phi, \Delta\}$, comprising the parameters of ϕ and the delay parameters Δ onto a neural state activation matrix

$$X := (x_1, x_2, \ldots, x_{n_t}) \in \mathbb{R}^{9n_s\times n_t}, \tag{45}$$

where $n_t \in \mathbb{N}$ denotes the number of peri-stimulus time samples which are chosen for the temporal integration. We write this function as

$$f : \Theta^f \to \mathbb{R}^{9n_s\times n_t}, \theta_f \mapsto f(\theta_f) := X \tag{46}$$

and denote the cardinality of the parameter $\theta_f \in \Theta_f$ by $p_f \in \mathbb{N}$. In other words, by the definition in (46), the symbol $f(\theta_f)$ introduced in Eq. (4) refers to the $9n_s \times n_t$ matrix of neural states activation time-courses over all sources. In the next section, we consider how this matrix is mapped onto electrode space time-courses by means of an EEG forward model.

The EEG forward model

Above, we have summarized the neural dynamics model as a function f that maps a parameter vector $\theta_f \in \Theta^f \subset \mathbb{R}^{p_f}$ onto a matrix

$$X = (x_1, x_2, \ldots, x_{n_t}), \text{where } x_t \in \mathbb{R}^{9n_s} \text{ for } t = 1, 2, \ldots, n_t \tag{47}$$

of discrete-time neural state activations. In this section we consider the mapping of a single neural state vector x_t onto a single electrode potential vector $y_t \in \mathbb{R}^{n_e}$, which we refer to as the "EEG forward model" of ERP-DCM. This model corresponds to a linear mapping of the neural

state vector x_t onto an electrode potential vector y_t by means of a "gain-matrix" $G \in \mathbb{R}^{n_e \times 9n_s}$

$$y_t = Gx_t \text{ for } t = 1,\ldots,n_t \tag{48}$$

Because this projection is identical for all time points t, we will omit the subscript t in the following discussion. In ERP-DCM (Kiebel et al., 2006), this gain matrix is parameter dependent, which we express by using a superscript θ_g on G, where $\theta_g \in \Theta^g \subset \mathbb{R}^{n_g}$ will denote the parameter vector and parameter set of the EEG forward model, respectively. These two notational modifications render the equation above:

$$y = G^{\theta_g}x, \text{where } y \in \mathbb{R}^{n_e}, x \in \mathbb{R}^{9n_s}, \text{and } G^{\theta_g} \in \mathbb{R}^{n_e \times 9n_s}. \tag{49}$$

In the following, we will discuss how the matrix G^{θ_g} is generated based on parameter values θ_g and introduce the subcomponents of this parameter vector. We start by considering the dipole positions and moments for the ith source ($i = 1,\ldots,n_s$) in three-dimensional (MNI) space, which we denote by

$$\theta_p^{(i)} := \begin{pmatrix} \theta_{p_1}^{(i)} \\ \theta_{p_2}^{(i)} \\ \theta_{p_3}^{(i)} \end{pmatrix} \in \mathbb{R}^3 \text{ and } \theta_m^{(i)} := \begin{pmatrix} \theta_{m_1}^{(i)} \\ \theta_{m_2}^{(i)} \\ \theta_{m_3}^{(i)} \end{pmatrix} \in \mathbb{R}^3. \tag{50}$$

The vectors $\theta_p^{(i)}$ and $\theta_m^{(i)}$ correspond to free parameters of the EEG forward model and can be estimated during model inversion (in SPM, $\theta_p^{(i)}$ is usually endowed with tight priors, such that effectively, only the dipole moments $\theta_m^{(i)}$ are free parameters). In their concatenated form over sources, i.e. as

$$\theta_p := \left(\theta_p^{(1)},\ldots,\theta_p^{(n_s)}\right) \in \mathbb{R}^{3\times n_s} \text{ and } \theta_m := \left(\theta_m^{(1)},\ldots,\theta_m^{(n_s)}\right) \in \mathbb{R}^{3\times n_s}, \tag{51}$$

they form the first two subcomponents of the forward model parameter θ_g. Note that the vectors $\theta_p^{(i)}$ and $\theta_m^{(i)}$ are source-specific, but not time-dependent: during a given modeled peri-stimulus time-period, these parameters remain constant. For a specified position vector $\theta_p^{(i)}$ the ERP-DCM approach evaluates a "canonical lead field matrix," i.e. the column-wise concatenated electrode potential distributions for "canonical dipole moments" $\theta_m^{(i)(1)} := (1,0,0)^T$, $\theta_m^{(i)(2)} := (0,1,0)^T$ and $\theta_m^{(i)(3)} := (0,0,1)^T$. We denote this canonical lead field matrix by $L^{\theta_p^{(i)}} \in \mathbb{R}^{n_e \times 3}$. Note that $L^{\theta_p^{(i)}}$ is dependent on the position of the dipole, but, as it encodes the electrode potentials for the canonical moments, independent of the moment $\theta_m^{(i)}$. The canonical lead field matrices $L^{\theta_p^{(1)}}$, …, $L^{\theta_p^{(ns)}}$ are evaluated based on conventional solutions of the EEG forward problem implemented in the FieldTrip toolbox for EEG data analysis (Oostenveld et al., 2011). In our implementation of the model, this fact requires the availability of the FieldTrip toolbox in the Matlab search path. Based on the canonical lead-field matrices $L^{\theta_p^{(i)}}$ and the moments $\theta_m^{(i)}$ of each source $j = 1,\ldots,n_s$, ERP-DCM next evaluates source-specific "dipole lead field vectors" $l^{\theta_p^{(i)},\theta_m^{(i)}} \in \mathbb{R}^{n_e}$ by means of the products

$$l^{\theta_p^{(i)},\theta_m^{(i)}} := L^{\theta_p^{(i)}}\theta_m^{(i)} \text{ for } i = 1,\ldots,n_s. \tag{52}$$

These dipole lead-field vectors may be interpreted as the electrode potentials which are evoked for the given dipole position and dipole moment settings under the assumption that each neural state variable contributes to the electrode potentials with a weighting factor of one. To render certain state variables, such as the "pyramidal cell membrane potential," important contributors to the electrode potential distribution and others negligible, the dipole lead-field vectors are firstly expanded by means of a "neural state projection row vector" to allow for a parameterized weighting of the neural state activations in their contribution to the electrode potential distribution:

$$L^{(i)} := \theta_j \otimes l^{\theta_p^{(i)},\theta_m^{(i)}} \text{ for } i = 1,\ldots,n_s \tag{53}$$

Here, $\theta_j \in \mathbb{R}^{1\times 9}$ is the third and final subcomponent of the forward model parameter θ_g, and $L^{(i)} \in \mathbb{R}^{n_e \times 9}$ corresponds to a lead-field matrix for all nine state variables of the ith source. Finally, the source-specific lead-field matrices are concatenated over sources forming the system gain matrix:

$$G^{\theta_g} := \begin{pmatrix} L^{(1)} & L^{(2)} & \cdots & L^{(n_s)} \end{pmatrix} \in \mathbb{R}^{n_e \times 9n_s} \tag{54}$$

In summary, specification of the EEG forward model parameter:

$$\theta_g := \{\theta_p, \theta_m, \theta_j\} \tag{55}$$

allows for the projection of the neural state activation vector onto an electrode potential distribution for each time-point $t = 1,\ldots,T$. We visualize the main components of the EEG forward model in Fig. 2. Note again that temporal variation in the electrode potentials is brought about only by temporal variation of the neural state activations and not by temporal variation of the EEG forward model parameter θ_g. We denote the cardinality of θ_g by $p_g \in \mathbb{N}$ and summarize the parameters of the EEG forward model in Table 3.

In summary, we may formulate the EEG forward model of ERP-DCM as a function g, which maps the parameter value θ_g onto a gain matrix G^{θ_g}, i.e., we can write

$$g : \Theta_g \subset \mathbb{R}^{p_g} \to \mathbb{R}^{n_e \times 9n_s}, \theta_g \mapsto g(\theta_g) := G^{\theta_g}. \tag{56}$$

In other words, by the definition in (56), the symbol $g(\theta_g)$ introduced in Eq. (4) refers to the $n_e \times 9n_s$ gain matrix projecting discrete-time neural state activation time-courses onto discrete-time electrode potential time courses.

Table 4
Step-size selection algorithm.

Initialization $t^{(i)} := 1$
While $-F\left(x + t^{(i)}p\right) > -F(x) + ct^{(i)}\left(\nabla F(x)^T \bar{p}_N\right)$
$t^{(i)} := \rho t^{(i)}$
End.
Initialization

Probabilistic embedding of the delay differential equation model for ERPs

In Eq. (3) we introduced the deterministic aspect of the delay differential equation model for ERPs as the function

$$h : \Theta \to \mathbb{R}^n, \theta \mapsto h(\theta) := vec(g(\theta_g)f(\theta_f)). \tag{57}$$

Thus far, we have detailed its component functions f in Section 2.3 and g in Section 2.4. Together with the assumption of Gaussian distributed observation noise, we have thus introduced the likelihood formulation of ERP-DCM as

$$p(y|\theta,\sigma^2) = N(y; h(\theta), \sigma^2 I_n) = N(y; vec(g(\theta_g)f(\theta_f)), \sigma^2 I_n), \tag{58}$$

where

$$\theta := \begin{pmatrix} \theta_f \\ \theta_g \end{pmatrix} \in \mathbb{R}^p, \tag{59}$$

and $p := p_f + p_g$. Recall that

$$g(\theta_g) = G^{\theta_g} \in \mathbb{R}^{n_e \times 9n_s} \text{ and } (\theta_f) = X \in \mathbb{R}^{9n_s \times n_t}, \tag{60}$$

such that

$$g(\theta_g)f(\theta_f) \in \mathbb{R}^{n_e \times n_t} \text{ and } vec(g(\theta_g)f(\theta_f)) \in \mathbb{R}^n \text{ with } n = n_e n_t. \tag{61}$$

To complete the specification of the ERP-DCM probabilistic model, we next specify a marginal (or "prior") distribution over θ and σ^2, such that

$$p(y,\theta,\sigma^2) = p(y|\theta,\sigma^2)p(\theta,\sigma^2). \tag{62}$$

We assume that the marginal distribution over unobserved variables factorizes over each set of unobserved variables, i.e.

$$p(\theta,\sigma^2) = p(\theta_f)p(\theta_g)p(\sigma^2). \tag{63}$$

Table 5
Fixed-form variational Bayes–Newton algorithm.

$S_\theta^{(0)} := \Sigma_\theta, m_\theta^{(0)} := \mu_\theta, m_{\sigma^2}^{(0)} := \mu_{\sigma^2}, s_{\sigma^2}^{(0)} := \varsigma_{\sigma^2}, F^{(0)} := F(m_\theta^{(0)}, S_\theta^{(0)}, m_{\sigma^2}^{(0)}, s_{\sigma^2}^{(0)})$
For $i = 0,1,2,\ldots$ until convergence
$S_\theta^{(i+1)} := (\Sigma_\theta^{-1} + , \exp(-m_{\sigma^2}^{(i)} + \frac{1}{2}s_{\sigma^2}^{(i)}) J^h(m_\theta^{(i)})^T J^h(m_\theta^{(i)})^T)^{-1}$
$m_\theta^{(0)} := m_\theta^{(i)}$
For $j = 0,1,2,\ldots$ until convergence
Evaluate $\nabla F_{m_\theta}(m_\theta^{(j)}), \tilde{H}_{m_\theta}(m_\theta^{(j)})$ and $t_{m_\theta}{}^{(j)}$
$m_\theta^{(j+1)} := m_\theta^{(j)} - t_{m_\theta}^{(j)}(\tilde{H}_{m_\theta}(m_\theta^{(j)}))^{-1}\nabla_{m_\theta} F(m_{m_\theta^{(j)}})$
end
$m_\theta^{(i+1)} := m_\theta{}^{(j)}$
$\begin{pmatrix} m_{\sigma^2}^{(0)} \\ s_{\sigma^2}^{(0)} \end{pmatrix} := \begin{pmatrix} m_{\sigma^2}^{(i)} \\ s_{\sigma^2}^{(i)} \end{pmatrix}$
For $k = 0,1,2,\ldots$ until convergence
Evaluate $\nabla F_{m_{\sigma^2}, s_{\sigma^2}}(m_{\sigma^2}^{(k)}, s_{\sigma^2}^{(k)}), \tilde{H}_{m_{\sigma^2}, s_{\sigma^2}}(m_{\sigma^2}^{(k)}, s_{\sigma^2}^{(k)})$ and $t_{m_{\sigma^2}, s_{\sigma^2}}{}^{(k)}$
$\begin{pmatrix} m_{\sigma^2}^{(k+1)} \\ s_{\sigma^2}^{(k+1)} \end{pmatrix} := \begin{pmatrix} m_{\sigma^2}^{(k)} \\ s_{\sigma^2}^{(k)} \end{pmatrix} - t_{m_{\sigma^2}, s_{\sigma^2}}^{(k)}(\tilde{H}_{m_{\sigma^2}, s_{\sigma^2}}(m_{\sigma^2}^{(k)}, s_{\sigma^2}^{(k)}))^{-1}\nabla_{m_{\sigma^2}, s_{\sigma^2}} F((m_{\sigma^2}^{(k)}, s_{\sigma^2}^{(k)}))$
end
$\begin{pmatrix} m_{\sigma^2}^{(i+1)} \\ s_{\sigma^2}^{(i+1)} \end{pmatrix} := \begin{pmatrix} m_{\sigma^2}^{(k)} \\ s_{\sigma^2}^{(k)} \end{pmatrix}$
$F^{(i+1)} := F(m_\theta^{(i+1)}, S_\theta^{(i+1)}, m_{\sigma^2}^{(i+1)}, s_{\sigma^2}^{(i+1)})$
end

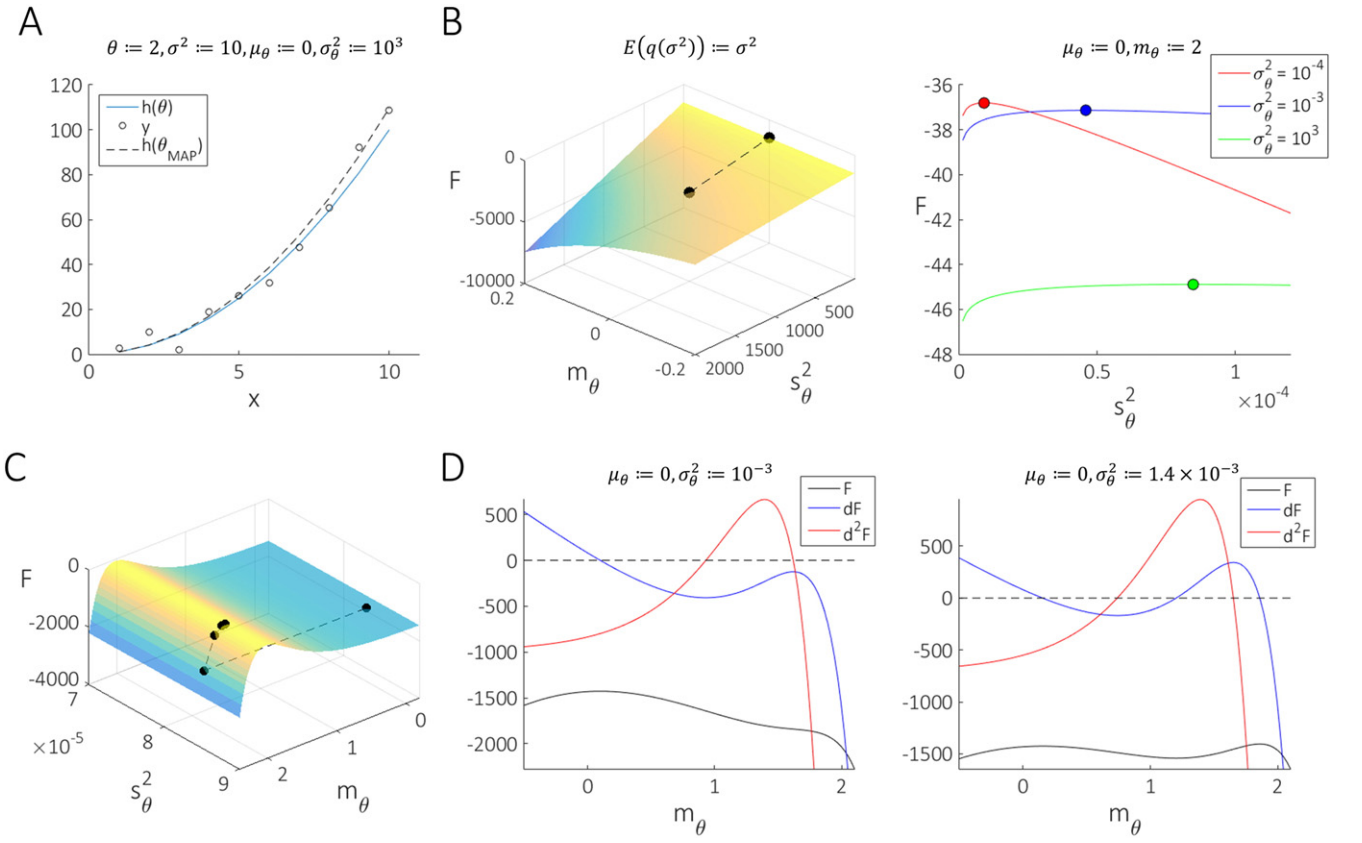

Fig. 3. Evaluation of the variational distribution $q(\theta)$ of the first toy problem (cf. Eqs. (82) and (83)). (A) Data expectation $h(\theta)$ for $\theta := 2$, sampled data y, and maximum-a-posterior model evaluation $h(\theta_{MAP})$ for prior settings of for $\mu_\theta := 0$ and $\sigma_\theta^2 := 10^3$. (B) Variational variance parameter update of the fixed-form variational Bayes–Newton algorithm (left panel) and effect of the prior variance on the variational variance setting (right panel). (C) Globalized Newton approach for free energy maximization with respect to the variational expectation parameter. (D) Variational free energy (F), and its first (dF) and second (d^2F) derivatives as a function of the prior variance setting. For the left panel a tight prior with $\sigma_\theta^2 := 10^{-4}$ was used, while for the right panel, a wider prior with $\sigma_\theta^2 := 1.4 \times 10^{-4}$ was used.

The probability distributions for θ_f and θ_g are set to mutually independent normal distributions and the probability distribution for the strictly positive error variance parameter σ^2 is set to a log-normal distribution. We thus define

$$p(\theta) := N(\theta; \mu_\theta, \Sigma_\theta), \quad \mu_\theta := \begin{pmatrix} \mu_{\theta_f} \\ \mu_{\theta_g} \end{pmatrix}, \Sigma_\theta := \begin{pmatrix} \Sigma_{\theta_f} & 0 \\ 0 & \Sigma_{\theta_g} \end{pmatrix}, \tag{64}$$

where the expectation parameters μ_{θ_f} and μ_{θ_g} are vectors in the space of the corresponding unobserved random variables, the covariance matrix parameters Σ_{θ_f} and Σ_{θ_g} are positive-definite matrices in the squared space of the corresponding unobserved variable, and the 0's indicate all zero matrices of the appropriate size. Finally, we define

$$p(\sigma^2) := LN(\sigma^2; \mu_{\sigma^2}, \varsigma_{\sigma^2}), \tag{65}$$

where $\mu_{\sigma^2}, \varsigma_{\sigma^2} \in \mathbb{R}$. Note that we use Greek letters for all parameters of the marginal distribution $p(\theta, \sigma^2)$. With (63)–(65), we have specified all probabilistic aspects of our ERP-DCM formulation and together with the structural components discussed in Sections 2.2–2.4 formulated a probabilistic delay differential equation model for ERPs.

Model inversion

Variational Bayes

The two aims of the Bayesian inversion of the probabilistic model (62) are (1.) the evaluation of the conditional distribution over unobserved random variables $p(\theta,\sigma^2|y)$, i.e. the posterior distribution quantifying the uncertainty about the values of θ and σ^2 after observing the data y, and (2.) the evaluation of the log marginal likelihood (log model evidence)

$$\ln p(y) = \ln \iint p(y, \theta, \sigma^2) d\theta d\sigma^2, \tag{66}$$

which serves as a basic measure for Bayesian model comparison (Kass and Raftery, 1995). To achieve these aims, ERP-DCM uses a variational Bayesian approach (Feynman, 1998; Neal and Hinton, 1998; Ostwald et al., 2014). We review this approach in the following, and, where necessary, adapt it for our formulation of the probabilistic model as described in Section 2.5.

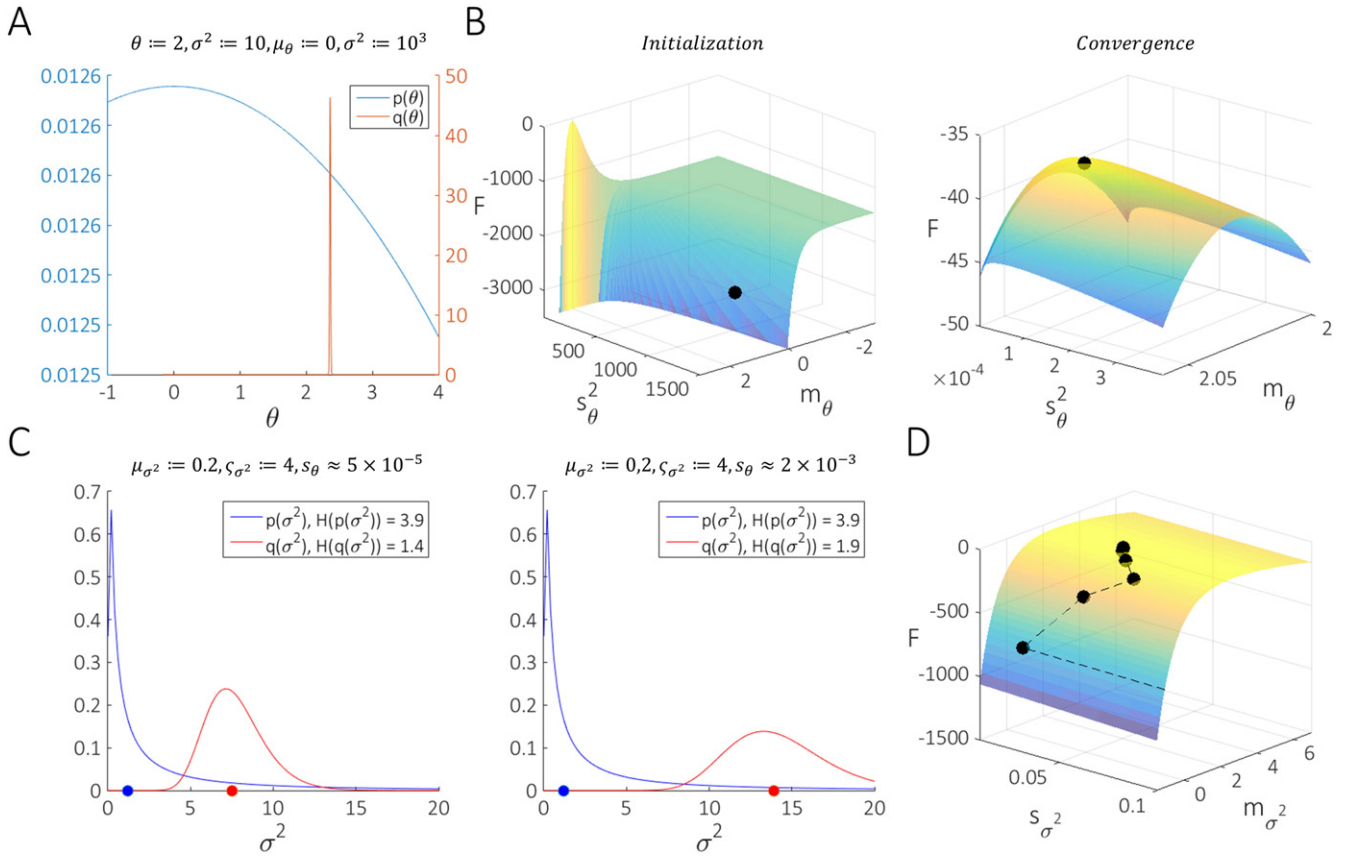

Fig. 4. Evaluation of the variational distributions $q(\theta)$ and $q(\sigma^2)$ of the first toy problem (cf. Eqs. (82) and (83)). (A) Prior and approximate posterior distributions for θ. (B) Variational parameter settings of $q(\theta)$ at initialization (left panel) and at convergence (right panel) of the globalized Newton ascent. (C) Prior and variational distribution for σ^2 for variational parameters reflecting precise knowledge of θ (left panel), and prior and variational distribution for σ^2 for variational parameters reflecting imprecise knowledge of θ (right panel). (D) Newton ascent in the parameter space of $q(\sigma^2)$.

Variational Bayes rests on the reformulation of the integration problem (66) as an optimization problem in terms of a variational free energy functional by introducing a variational distribution over the unobserved random variables. Denoting this variational distribution by $q(\theta, \sigma^2)$, the log marginal likelihood $\ln p(y)$ can be rewritten as the sum of a variational free energy functional $\mathcal{F}$ and a KL-divergence term as

$$\ln p(y) = \mathcal{F}(q(\theta,\sigma^2)) + \mathcal{KL}(q(\theta,\sigma^2)||p(\theta,\sigma^2|y)). \tag{67}$$

Because the log marginal likelihood on the left-hand side of (67) is constant, and the KL-divergence term on the right-hand side of (67) is a non-negative quantity, the iterative maximization of the variational free energy functional with respect to its argument $q(\theta, \sigma^2)$ can render the variational free energy an increasingly better approximation to the log model evidence and simultaneously decrease the KL-divergence between the variational distribution $q(\theta, \sigma^2)$ and the true conditional distribution $p(\theta,\sigma^2|y)$. In other words, upon convergence of an algorithm maximizing $\mathcal{F}(q(\theta,\sigma^2))$, one can hope to have:

$$\mathcal{F}(q(\theta,\sigma^2)) \approx \ln p(y) \tag{68}$$

and

$$q(\theta,\sigma^2) \approx p(\theta,\sigma^2|y). \tag{69}$$

For Eq. (67) to hold, the variational free energy functional is defined as:

$$\mathcal{F}(q(\theta,\sigma^2)) = \iint q(\theta,\sigma^2) \ln\left(\frac{p(y,\theta,\sigma^2)}{q(\theta,\sigma^2)}\right) d\theta d\sigma^2. \tag{70}$$

In the current application, the maximization of this functional with respect to its input argument is achieved using an iterative numeric coordinate-wise ascent in the parameters of the variational probability distribution $q(\theta, \sigma^2)$. To this end, the distribution over θ and σ^2 is assumed to factorize according to:

$$q(\theta,\sigma^2) := q(\theta)q(\sigma^2). \tag{71}$$

The variational distribution $q(\theta)$ is assumed to be of normal form and to factorize over the parameter subsets θ_f and θ_g, i.e.:

$$q(\theta) = N(\theta; m_\theta, S_\theta), m_\theta := \begin{pmatrix} m_{\theta_f} \\ m_{\theta_g} \end{pmatrix} \in \mathbb{R}^p \text{ and } S_\theta : = \begin{pmatrix} S_{\theta_f} & 0 \\ 0 & S_{\theta_g} \end{pmatrix} \in \mathbb{R}^{p\times p} \text{ p.d.}, \tag{72}$$

where the zeros indicate matrices of all zero entries of the appropriate sizes. The variational distribution for the error variance parameter σ^2 is set to a log-normal distribution:

$$q(\sigma^2) = LN(\sigma^2; m_{\sigma^2}, s_{\sigma^2}). \tag{73}$$

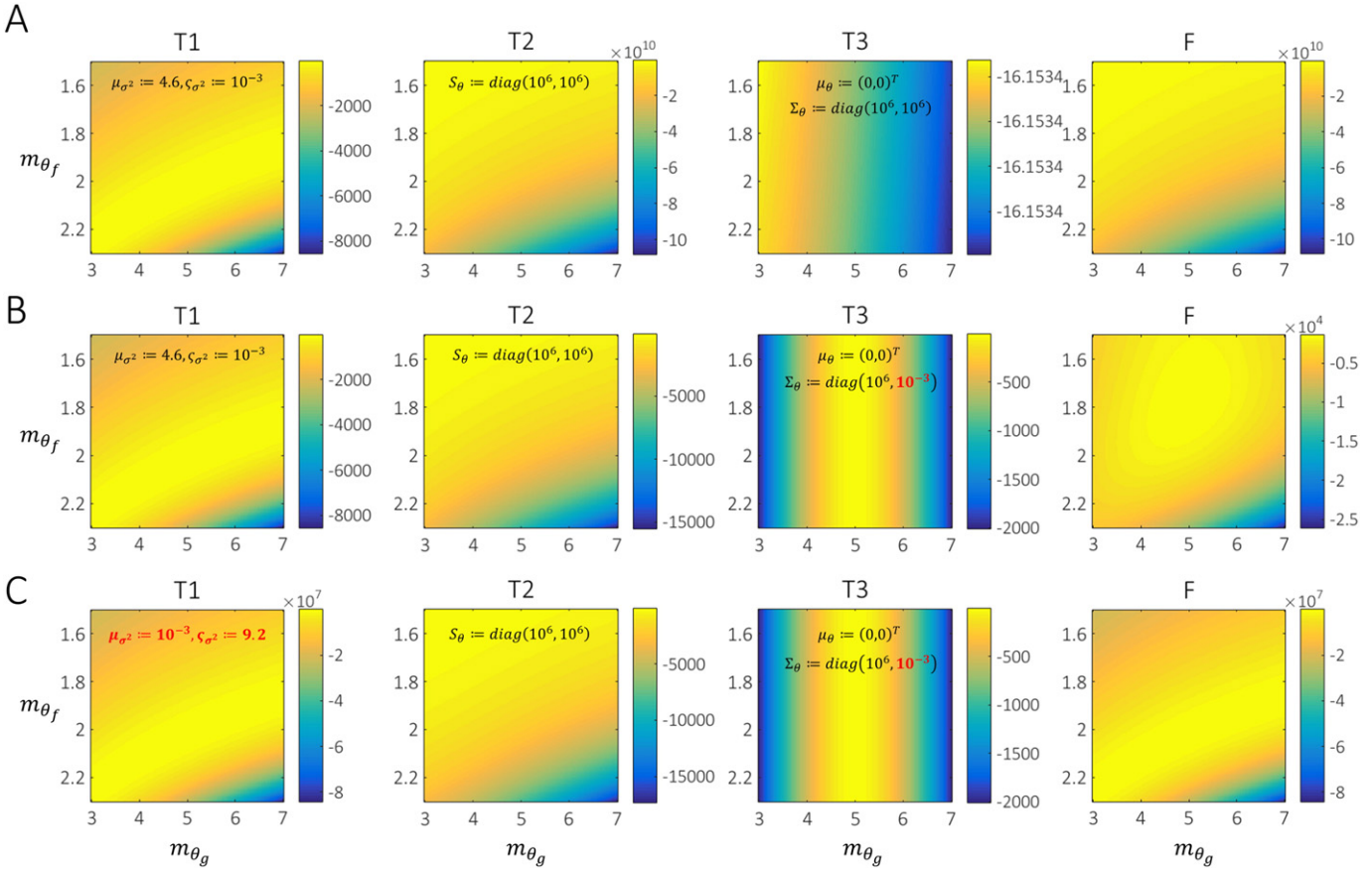

Fig. 5. Variational free energy terms for the second toy problem (cf. Eqs. (84) and (85)). All simulations shown are based on true, but unknown, parameters of $\theta := (2,5)^T$ and $\sigma^2 := 100$. (A) Variational free energy terms for isotropic, imprecise priors on θ. Note that the variational free energy F does not exhibit a unique maximum. The constant parameters for each term of the variational free energy are indicated on the respective panel, the variational free energy was evaluated based on these prior parameters. (B) Variational free energy terms for non-isotropic priors. In this case, the variational free energy F exhibits a maximum. The constant parameters for each term of the variational free energy are indicated on the respective panel, the variational free energy was evaluated based on these prior parameters. (C) Interaction effects of the prior parameter settings for θ and σ^2 on the variational free energy terms. In contrast to (B), the entropy of the prior distribution $p(\sigma^2)$ was increased, abolishing the variational free energy maximum The constant parameters for each term of the variational free energy are indicated on the respective panel, the variational free energy was evaluated based on these prior parameters.

Specifying the variational distributions directly as in (71)–(73), and not having their form determined analytically by application of the "fundamental lemma of variational Bayesian inference" (Ostwald et al., 2014) is commonly referred to as "fixed-form variational Bayes" (Honkela et al., 2010; Jordan et al., 1999). Note that, in contrast to the parameters of the marginal distributions, we use roman letters to denote the parameters of the variational (i.e. approximate posterior) distributions.

The first step towards the optimization of the variational free energy functional under the fixed-form assumptions (71)–(73) is to rewrite the functional (70) as a multivariate function, such that we are led to a standard nonlinear optimization problem of the form:

$$\max_{m_\theta, S_\theta, m_{\sigma^2}, s_{\sigma^2}} F(m_\theta, S_\theta, m_{\sigma^2}, s_{\sigma^2}), \tag{74}$$

where

$$F : \mathbb{R}^p \times \mathbb{R}^{p\times p} \times \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}, (m_\theta, S_\theta, m_{\sigma^2}, s_{\sigma^2}) \mapsto F(m_\theta, S_\theta, m_{\sigma^2}, s_{\sigma^2}) \tag{75}$$

is a real-valued function of real-valued input arguments. From (63)–(65) and (71)–(73), we have for the current application:

$$\begin{aligned} &F(m_\theta, S_\theta, m_{\sigma^2}, s_{\sigma^2}) \\ &= \int\int q(\theta)q(\sigma^2) \ln\left(\frac{p(y,\theta,\sigma^2)}{q(\theta)q(\sigma^2)}\right) d\theta d\sigma^2 = \int\int N(\theta; m_\theta, S_\theta) LN(\sigma^2; m_{\sigma^2}, s_{\sigma^2}) \\ &\ln\left(\frac{N(y; h(\theta), \sigma^2 I_n) N(\theta; \mu_\theta, \Sigma_\theta) LN(\sigma^2; \mu_{\sigma^2}, \varsigma_{\sigma^2})}{N(\theta; m_\theta, S_\theta) LN(\sigma^2; m_{\sigma^2}, s_{\sigma^2})}\right) d\theta d\sigma^2 \end{aligned} \tag{76}$$

As shown in Appendix A, the integral above can be approximated using a first-order Taylor approximation of the function h around the variational parameter m_θ (Chappell et al., 2009, Friston et al., 2007). This yields the variational free energy objective function:

$$F(m_\theta, S_\theta, m_{\sigma^2}, s_{\sigma^2}) := \tag{77}$$

$$-\frac{n}{2}\ln 2\pi - \frac{n}{2} m_{\sigma^2} - \frac{1}{2}\exp\left(-m_{\sigma^2} + \frac{1}{2}s_{\sigma^2}\right) \times \left((y - h(m_\theta))^T (y - h(m_\theta))\right) \tag{T1}$$

$$-\frac{1}{2}\exp\left(-m_{\sigma^2} + \frac{1}{2}s_{\sigma^2}\right) tr\left(J^h(m_\theta)^T J^h(m_\theta) S_\theta\right) \tag{T2}$$

$$-\frac{p}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_\theta| - \frac{1}{2}\left(tr\left(\Sigma_\theta^{-1} S_\theta\right) + (m_\theta - \mu_\theta)^T \Sigma_\theta^{-1} (m_\theta - \mu_\theta)\right) \tag{T3}$$

$$-\frac{1}{2}\ln 2\pi\varsigma_{\sigma^2} - m_{\sigma^2} - \frac{1}{2\varsigma_{\sigma^2}}\left(s_{\sigma^2} + (m_{\sigma^2} - \mu_{\sigma^2})^2\right) \tag{T4}$$

$$+\frac{1}{2}\ln|S_\theta| + \frac{p}{2}\ln(2\pi e) \tag{T5}$$

$$+\frac{1}{2} + \frac{1}{2}\ln(2\pi s_{\sigma^2}) + m_{\sigma^2} \tag{T6}$$

In (77), (T1)–(T4) are terms deriving from the energy function, i.e. the expectation of the joint distribution under the variational distribution.

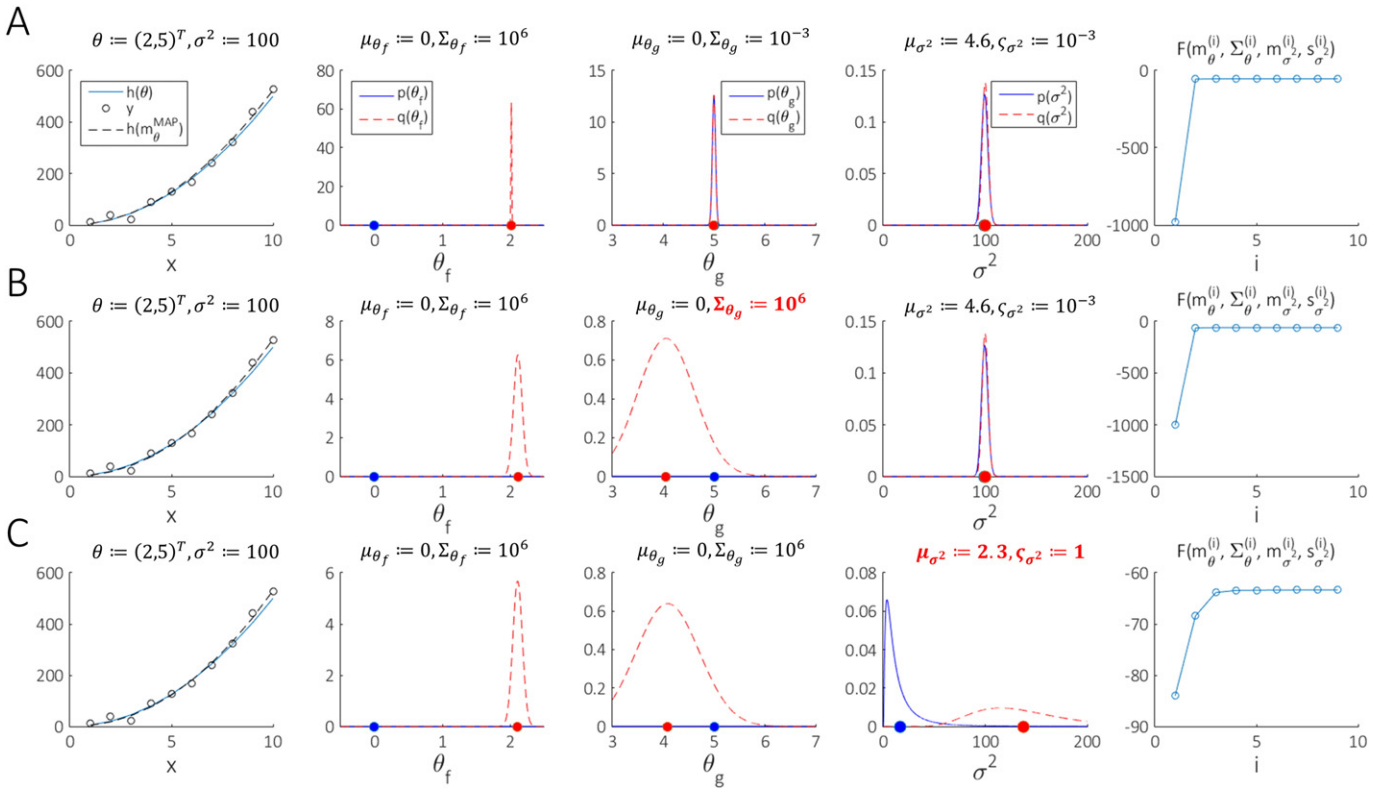

Fig. 6. Estimation of the second toy problem (cf. Eqs. (84) and (85)) For all simulations, the true, but unknown, parameter values are held constant according to their values in the leftmost panels, while subsets of the prior parameters are varied over rows of the figure. (A) Estimation for a highly precise prior on θ_g. Marginal prior parameter settings are indicated on the respective panels. (B) Estimation for a less precise prior on θ_g. Marginal prior parameter settings are indicated on the respective panels. (C) Estimation for a imprecise prior on σ^2. Marginal prior parameter settings are indicated on the respective panels.

(T1) and (T2) primarily reflect the mismatch between data and data prediction, while (T3) and (T4) reflect uncertainty-weighted contributions of the prior distributions of θ and σ^2. Finally, terms (T5) and (T6) enforce maximal entropies of the resulting variational approximation to the posterior distribution.

To summarize, in this section we have converted the problem of deriving the parameters of the conditional distribution $p(\theta,\sigma^2|y)$ and the value of $p(y)$ under the probabilistic model into a nonlinear optimization problem of the function F in terms of the parameters of the variational distributions. The solutions that this nonlinear optimization

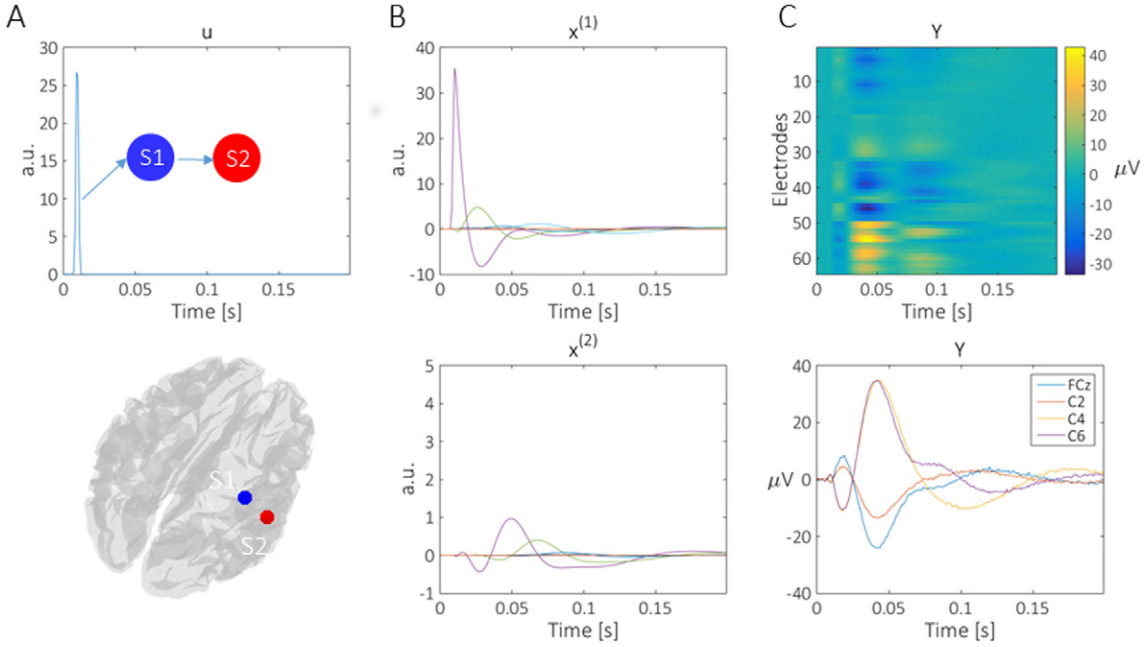

Fig. 7. Delay differential equation ERP model realization. (A) Upper panel: conceptual two-source forward architecture and input function in peri-stimulus time. Lower panel: anatomical location of the two sources overlaid on the SPM cortical mesh. (B) Latent neural dynamics of S1 (upper panel) and S2 (lower panel). (C) Full data set in matrix form (upper panel) and selected electrode time-courses (lower panel).

generates are approximations of the posterior distribution and marginal data probability under the ERP-DCM probabilistic model. In the following section, we discuss our approach to the maximization of the variational free energy function (77) with respect to m_θ, S_θ, m_{σ^2}, and s_{σ^2}.

Maximization of the variational free energy function

To maximize the variational free energy function, we use an iterative combination of analytical and numerical optimization methods. Specifically, as detailed in Appendix B, maximization with respect to the variational covariance parameter S_θ can be achieved by evaluation of the necessary condition for an extremum yielding the parameter update equation:

$$S_\theta^{(i+1)} := \left(\Sigma_\theta^{-1} + \exp\left(-m_{\sigma^2} + \frac{1}{2}s_{\sigma^2}\right) J^h(m_\theta) J^h(m_\theta)^T \right)^{-1} \tag{78}$$

on the ith iteration of the algorithm (please note that we use i to denote algorithm iterations in this section, and not sources as in Section 2). To maximize the variational free energy function with respect to m_θ and $(m_{\sigma^2}, s_{\sigma^2})^T$, we employ two nested globalized Newton descents with Hessian modification on the negative variational free energy $-F$. The parameter update equations on each iteration of the algorithm take the form

$$m_\theta^{(j+1)} := m_\theta^{(j)} - t_{m_\theta}^{(j)} \left(\tilde{H}_{m_\theta}\left(m_\theta^{(j)}\right) \right)^{-1} \nabla_{m_\theta} F\left(m_{m_\theta^{(j)}}\right) \tag{79}$$

and

$$\begin{pmatrix} m_{\sigma^2}^{(k+1)} \\ s_{\sigma^2}^{(k+1)} \end{pmatrix} := \begin{pmatrix} m_{\sigma^2}^{(k)} \\ s_{\sigma^2}^{(k)} \end{pmatrix} - t_{m_{\sigma^2}, s_{\sigma^2}}^{(k)} \left(\tilde{H}_{m_{\sigma^2}, s_{\sigma^2}} \left(\begin{pmatrix} m_{\sigma^2}^{(k)} \\ s_{\sigma^2}^{(k)} \end{pmatrix} \right) \right)^{-1} \nabla_{m_{\sigma^2}, s_{\sigma^2}} F\left(\begin{pmatrix} m_{\sigma^2}^{(k)} \\ s_{\sigma^2}^{(k)} \end{pmatrix} \right). \tag{80}$$

In (79) and (80) $\tilde{H}_{m_\theta}(m_\theta^{(j)})$ and $\tilde{H}_{m_{\sigma^2}, s_{\sigma^2}}((m_{\sigma^2}^{(k)}, s_{\sigma^2}^{(k)})^T)$ denote the negative variational free energy function's modified Hessian matrices with respect to m_θ and m_{σ^2}, s_{σ^2} and $\nabla_{m_\theta} F(m_{m_\theta^{(j)}})$ and $\nabla_{m_{\sigma^2}, s_{\sigma^2}} F((m_{\sigma^2}^{(k)}, s_{\sigma^2}^{(k)})^T)$ denote the negative variational free energy gradients with respect to m_θ and m_{σ^2},s_{σ^2} evaluated at $m_{m_\theta^{(j)}}$ and $m_{\sigma^2}^{(k)}$, $s_{\sigma^2}^{(k)}$, respectively. Finally, $t_{m_\theta}^{(j)}$ and $t_{m_{\sigma^2}, s_{\sigma^2}}^{(k)}$ denote parameter- and an iteration-specific step-sizes. In the following discussion of the motivation for (79) and (80) we abbreviate these quantities by the generic symbols $x \in \mathbb{R}^d$ for the negative free energy's argument, $\nabla F(x)$ for the negative variational free energy's gradient, $\tilde{H}^F(x)$ for the negative free energy's Hessian, and $t^{(i)}$ for the iteration specific step-size.

The parameter update rules (79) and (80) are motivated by the following considerations. Firstly, while standard Newton descents with search direction $p_N := -H^F(x)\nabla F(x)$ show good convergence properties in the vicinity of a local extremum, further away from an extremum, the Hessian matrix may become non-positive definite and the Newton descent may turn into an ascent. A number of methods have been proposed to render the Hessian positive-definite also in this case, while retaining the beneficial properties of including the local curvature during minimization (Boyd and Vandenberghe, 2004; Nocedal and Wright, 2006). In our implementation, we employ a simple modification based on the eigenspectrum of the Hessian (Nocedal and Wright, 2006, pp. 49–51). Specifically, if $(H^F(x)\nabla F(x))^T \nabla F(x) < 0$, i.e., $H^F(x)\nabla F(x)$ is

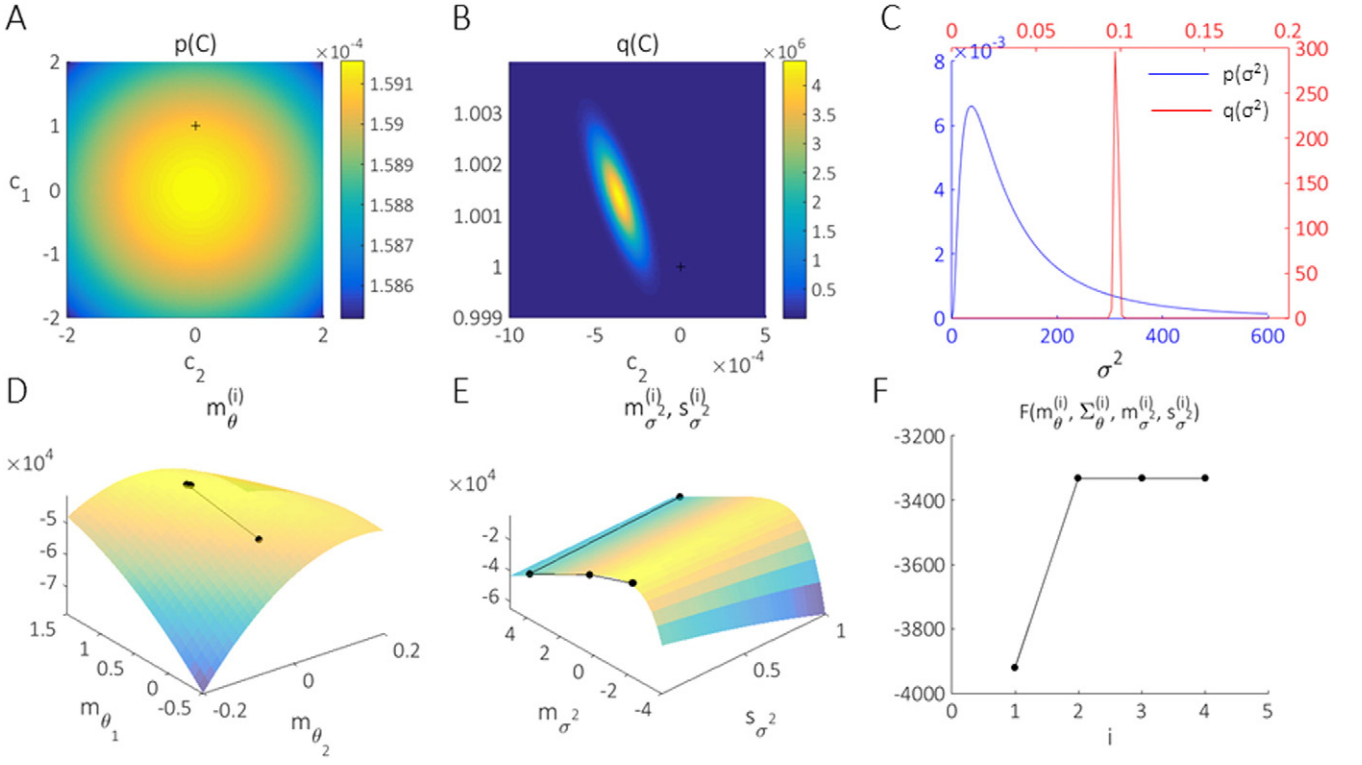

Fig. 8. Estimation of the extrinsic input connectivity in a two-source ERP-DCM model. (A) Prior distribution of the input parameter value C and its true, but unknown, value (black cross). (B) Approximated posterior distribution or distribution of the input parameter value C and its true, but unknown, value (white dot). Note the difference in scale between (A) and (B). (C) Prior and approximated posterior distribution for the variance parameter σ^2. (D) Variational free energy ascent in the expectation parameters of $q(C)$ on the first algorithm iteration. (E) Variational free energy ascent in the parameters of $q(\sigma^2)$ on the first algorithm iteration. (F) Variational free energy evolution over algorithm iterations.

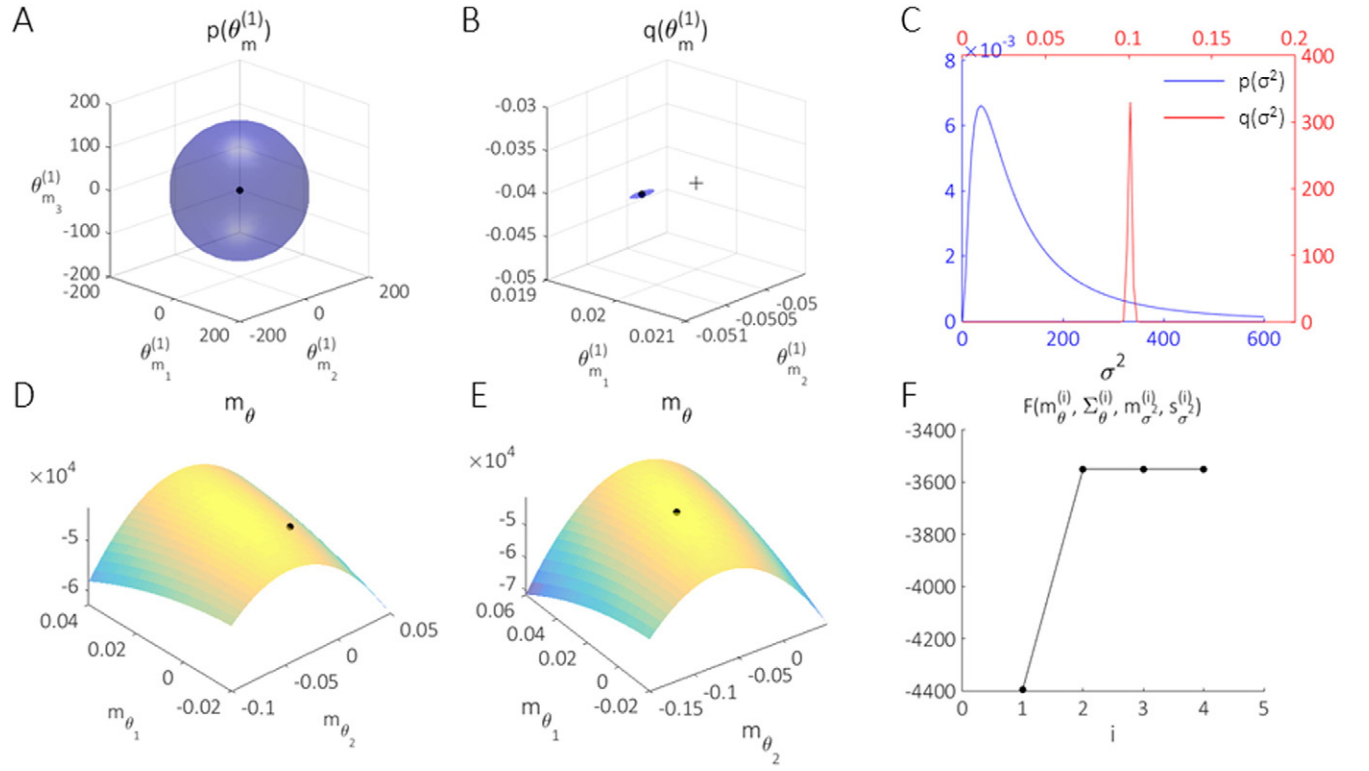

Fig. 9. Estimation of dipole moments in a two-source ERP-DCM model. (A) Prior distribution of the dipole moment $\theta_m^{(1)}$. The black dot indicates the expectation parameter and the sphere the 95% isocontour of the three-dimensional normal distribution. (B) Approximate posterior distribution of the dipole moment $\theta_m^{(1)}$. The black dot indicates the expectation parameter and the ellipse the 95% isocontour of the three-dimensional normal distribution, while the black cross indicates the true parameter value. Note the difference in scale between (A) and (B). (C) Prior and approximated posterior distribution for the variance parameter σ^2. (D) Two-dimensional variational expectation parameter projection of $q(\theta)$ on the free energy surface at initialization. (E) Variational expectation parameter of $q(\theta)$ on the free energy surface of the first algorithm iteration. (F) Variational free energy evolution.

not a descent direction, we render the Hessian positive-definite by setting:

$$\tilde{H}^F(x) := H^F(x) + \max(0, (\delta - \min_j \lambda_j))I_d, \tag{81}$$

where $\delta > 0$ denotes a small positive constant, and λ_j, $j = 1,2,\ldots,d$ denotes the set of eigenvalues of $H^F(x)$. Intuitively, (81) ensures that all negative eigenvalues of $\tilde{H}^F$ are positive by at least δ, rendering the matrix positive-definite.

Secondly, while $\tilde{p}_N := -\tilde{H}^F(x)\nabla F$ determines the search direction of a Newton descent, the convergence properties of the resulting algorithm are strongly affected by the scalar step-size $t^{(i)}$. In our implementation of the negative variational free energy descent, we employ a standard backtracking approach for the selection of $t^{(i)}$ (Nocedal and Wright, 2006). Specifically, for $\rho \in]0,1[$ and $c > 0$, a search direction $\tilde{p}_N$ and a negative variational free energy gradient of ∇F, we ensure a sufficient decrease in the Wolfe sense by means of the step-size selection algorithm listed in Table 4. We note that an intuitively similar device is used in the "Variational-Laplace" scheme by Friston et al. (2007), which invokes a backward step or decreased regularization whenever the variational free energy changes in the wrong direction. To circumvent the need for constrained optimization methods, we add the additional requirement that $s_{\sigma^2} > 0$ in the evaluation of the step-size for $(m_{\sigma^2}{}^{(i+1)}, s_{\sigma^2}{}^{(i+1)})^T$. In summary, we employ the fixed-form variational Bayesian–Newton algorithm shown in Table 5, where all variational parameters are initialized to their prior distribution correspondents.

Results

In the following, we report the results of numerical simulations implementing the probabilistic delay differential equation model for ERPs and its fixed-form variational estimation. We firstly consider the estimation scheme in the context of two toy examples, in order to validate the algorithm of Table 5 and obtain some insight into its inner workings (Section 4.1). We then apply this algorithm in the context of an ERP model comprising two cortical sources (Section 4.2). For all simulations, we took an "objective Bayesian" approach (Kass and Wasserman, 1996), in the sense that we used priors with high uncertainty and hoped to recover posterior estimate expectations close to the true, but unknown, (simulated) parameters, that are largely determined by the (simulated) data. Please note that this does neither correspond to the attempt to establish "uninformative priors" for this model class in a well-defined (e.g. Jeffreys or reference prior) sense (Bernardo and Smith, 1994), nor is it meant to imply that "objective Bayesian" approaches are superior to "subjective Bayesian" approaches. More importantly, please note that these simulations do not mean to imply that the recovery of simulated, true, but unknown, parameters by means of the current framework is guaranteed. In fact, the identifiability of model parameters for the ERP model class considered here is not established in general and constitutes an important problem for future research (see subsequent results and the Discussion section).

Two toy examples

The first toy problem we consider for the validation of the fixed-form variational Bayesian algorithm of Table 5 can be conceived as a special case of the ERP-DCM model described in Section 2 (cf. Eqs. (3), (4), (57) and (58)). It takes the form:

$$y = h(\theta) + \varepsilon, p(\varepsilon) = N(\varepsilon; 0, \sigma^2 I_{10}), \tag{82}$$

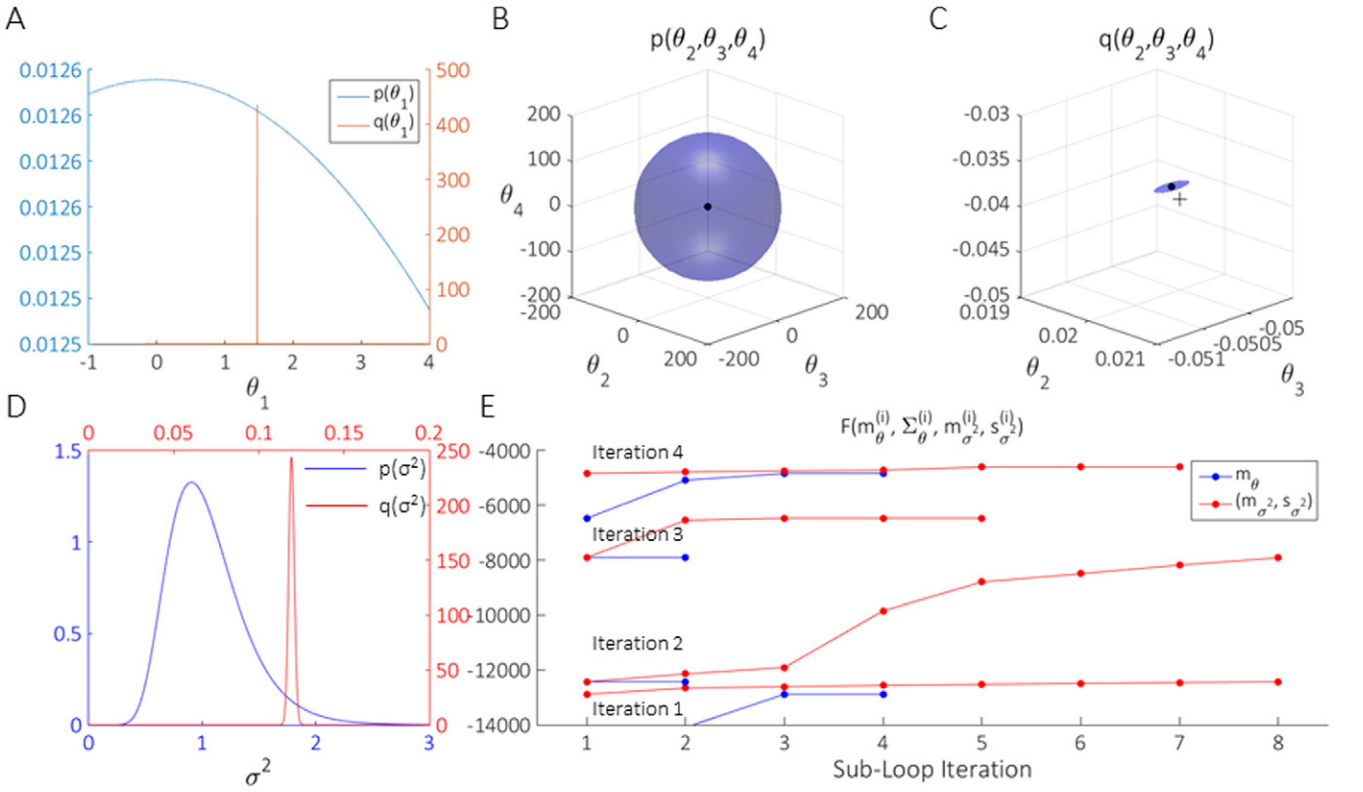

Fig. 10. Estimation of forward connectivity and dipole moments in a two-source ERP-DCM model. (A) Prior and approximated posterior distribution for θ_1 corresponding to the forward connectivity parameter a_{21}. (B) Prior distribution for $(\theta_2, \theta_3, \theta_4)$ corresponding to the dipole moment $\theta_m^{(1)}$. The black dot indicates the expectation parameter and the sphere the 95% isocontour of the three-dimensional normal distribution. (C) Approximate posterior distribution for $(\theta_2, \theta_3, \theta_4)$. The black dot indicates the expectation parameter and the ellipse the 95% isocontour of the three-dimensional normal distribution, while the black cross indicates the true parameter value. (D) Prior and approximated posterior distribution for the variance parameter σ^2. (E) Variational free energy ascent over Newton algorithm sub-loops.

where

$$h : \mathbb{R} \to \mathbb{R}^{10}, \theta \mapsto h(\theta) := \left(1^\theta, 2^\theta, \ldots, 10^\theta\right)^T \tag{83}$$

Intuitively, (82) and (83) thus corresponds to the problem of estimating the exponent parameter θ of the latent nonlinear function "x^θ" and the noise parameter $\sigma^2 > 0$ of the additive error component in a Bayesian manner. In terms of ERP-DCM, the problem may be viewed as the special case that $h = f$, i.e., g corresponds to the constant unit function.

Fig. 3 visualizes some aspects of the determination of the parameters of the univariate variational distribution $q(\theta) = N(\theta; m_\theta, s_\theta^2)$ for this problem. For this simulation, the true, but unknown, value of the exponent parameter was set to $\theta := 2$ and the variance parameter $\sigma^2 > 0$ was assumed to be known. Fig. 3A depicts the ensuing expectation $h(\theta) = (1^\theta,\ldots,10^\theta)^T$ (blue line), a set of data points sampled from $N(y; h(\theta), \sigma^2 I_{10})$ with $\sigma^2 := 10$ (blue dots), and the maximum-a-posteriori (MAP) fit based on imprecise prior settings $\mu_\theta := 0$ and $\sigma_\theta^2 := 10^3$. Due to the low prior precision, the MAP solution is dominated by the data, such that the fit is quite accurate (see below for different scenarios). The left panel of Fig. 3B depicts the effect of the initial variational variance parameter update, where the expectation of the log-normal distribution was replaced by the known variance parameter. While the update clearly maximizes the variational free energy with respect to s_θ^2, it appears as if the ensuing variational variance is located on the very edge of its parameter space. Closer inspection, however, reveals that this is not the case and unveils the influence of the prior variance σ_θ^2 on the resulting variational variance s_θ^2: as shown in the right panel of Fig. 3B for three different settings of the prior variance σ_θ^2 and a variational expectation parameter of $m_\theta = 2$, the variational free energy indeed has its maximum with respect to s_θ^2 close to the edge of parameter space, but not at its edge. Moreover, the maximum location is a function of the prior variance, as a large prior variance results in a larger maximizing value of s_θ^2. In other words, higher prior uncertainty about the value of θ leads to higher uncertainty in the variational approximation to the posterior distribution of θ, as one would expect. Fig. 3C visualizes five iterations of the globalized Newton approach for maximization of the variational free energy function with respect to m_θ and the updated variational variance parameter shown in Fig. 3B, left panel. The variational free energy has a clear maximum around the location of the true, but unknown, parameter value, and this maximum is reached by the variational expectation parameter after a few iterations. Finally, the panels of Fig. 3D visualize the effect of the prior on the variational free energy landscape and its implications for the determination of the variational expectation parameter. The left panel of Fig. 3D depicts the variational free energy as a function of m_θ for prior settings $\mu_\theta := 0$ and $\sigma_\theta^2 := 10^{-3}$ and the data sample shown in Fig. 3A (black curve). This tight prior has the consequence that the free energy has a maximum close to the location of the prior expectation, which is also evident from the zero-crossing of its first derivative (blue curve) and the negativity of its second derivative (red curve) at this location. In this case, the globalized Newton algorithm initialized at μ_θ would converge to the solution $m_\theta \approx 0$, or, in other words, due to its high precision, weigh the prior much stronger than the data. The right panel of Fig. 3D shows that, like any numerical optimization procedure, the globalized Newton algorithm for free energy maximization is not necessarily immune to convergence to local maxima. Here, for a less tight prior of $\sigma_\theta^2 := 1.4 \times 10^{-3}$ the variational free energy has two maxima over the m_θ parameter space, a local maximum close to the prior parameter $\mu_\theta = 0$ and a global maximum closer to the value of the true, but unknown,

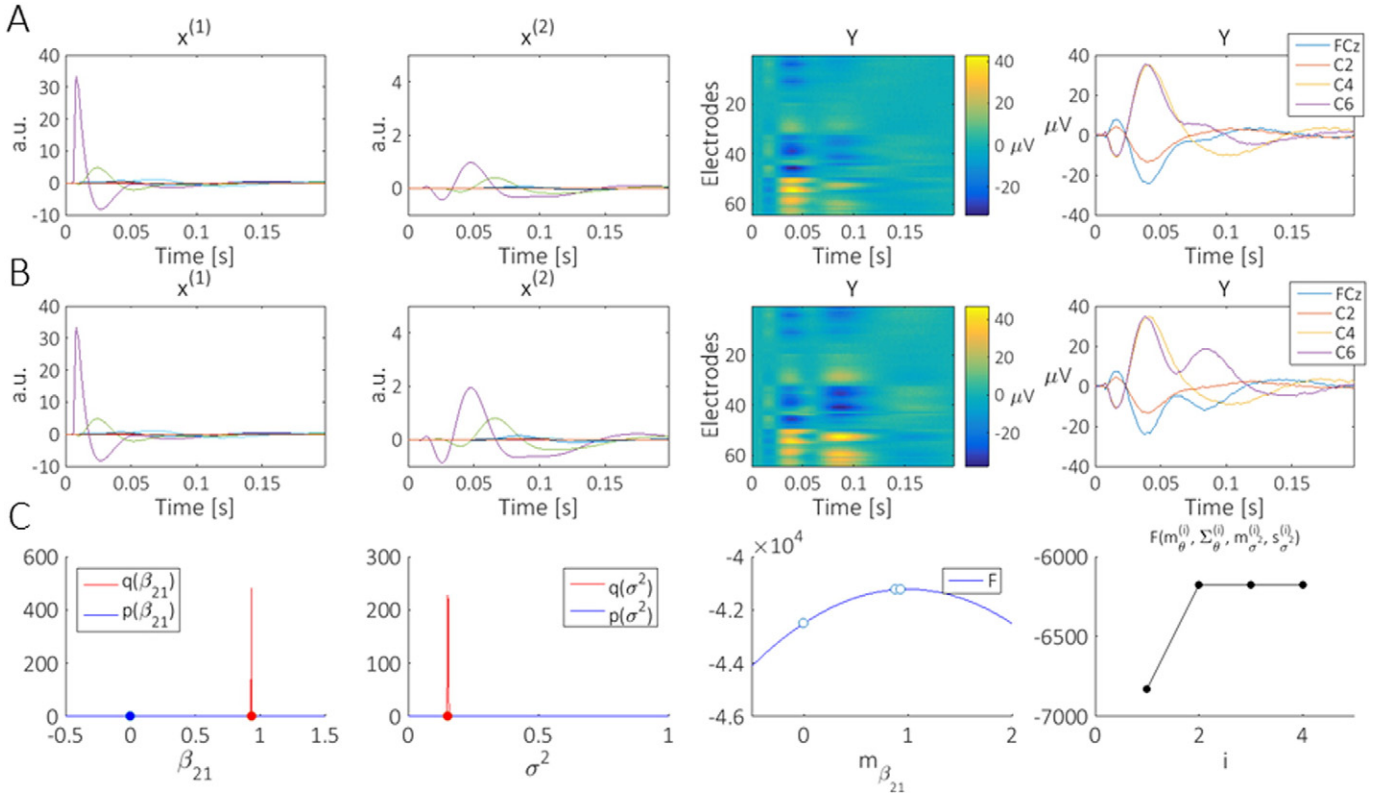

Fig. 11. Estimation of condition-specific forward connectivity in a two-source ERP-DCM model. (A) and (B) depict the latent neural state and electrode space dynamics of the two-source ERP-DCM model discussed in Section 4.2 with condition-specific parameterization of the forward connectivity matrix A^F. (A) depicts the realized dynamics for a forward parameter $a_{211} = 1$ of A_1^F, corresponding to condition $c = 1$ and (B) for a forward parameter $a_{212} = 2$ of A_2^F, corresponding to condition $c = 2$. These condition-specific connectivity parameters are parameterized by $\beta_{21} = 1$. (C) Fixed-form variational Bayes-Newton estimation of β_{21} and σ^2. The first two subpanels depict the prior and approximated posterior distributions for β_{21} and σ^2, the third panel the variational free energy ascend in the variational parameter $m_{\beta_{21}}$ over the first subiterations of the algorithm, and the right-most panel the maximization of the variational free energy over four iterations.

parameter value. In this case, initialization of the algorithm at the location of the prior expectation parameter may well result in the convergence m_θ to the local maximum in the vicinity of μ_θ and not the global maximum, which represents the intended weighting between prior and data.

Fig. 4 visualizes the joint estimation of the parameters of the variational distribution $q(\theta)$ and the parameters of the variational distribution $q(\sigma^2)$ for the toy model (82) and (83) with $\theta := 2$ and $\sigma^2 := 10$. Fig. 4A depicts the prior and converged variational distributions over θ for $\mu_\theta := 0$ and $\sigma_\theta^2 := 10^3$ (blue and green curves, respectively) under the assumption of a known variance parameter σ^2. For this wide prior distribution, the approximated posterior distribution is centered in the vicinity of the true value of θ and has high precision. The left and right panels of Fig. 4B depict the variational parameter settings of $q(\theta)$ in free energy space at initialization and upon convergence of the globalized Newton ascent, respectively. Note that the maximum of the variational free energy is located close to the boundary of the s_θ^2 parameter space. The left and right panels of Fig. 4C visualize the determination of the variational parameters m_{σ^2} and s_{σ^2}. These figures visualize the log-normal prior and variational distribution for $\mu_{\sigma^2} := 0.2$ and $\varsigma_{\sigma^2} := 4$ (blue and red curves, respectively) and the dots depict the respective distribution's median. For the left panel, the variational expectation parameters m_θ and s_θ^2 were fixed to the final results obtained under the assumption of known variance, while for the right panel, the variational variance parameter was set to $\sigma_\theta^2 := 2 \times 10^{-3}$, reflecting a larger uncertainty about θ. The assumed uncertainty about θ, which differs between the two panels of Fig. 4C has implications for the inference about σ^2: while in both cases, the entropy of the resulting variational distribution $H(q(\sigma^2))$ is lower than that of the prior $H(p(\sigma^2))$, a larger uncertainty about the value of θ implies a higher uncertainty about σ^2 and a tendency for its overestimation by mode and median. In other words, the uncertainties about θ and σ^2 interact in a meaningful manner. Finally, Fig. 4D depicts the globalized Newton ascent in the variational parameters $(m_{\sigma^2}, s_{\sigma^2})$ resulting in the final variational distribution of the right panel of Fig. 4C.

Having established our free energy maximization scheme for the estimation of both a univariate parameter θ and the noise parameter σ^2 in example (82) and (83) we next explore the performance of the approach in a second, more complex, toy example given by

$$y = h(\theta) + \varepsilon, p(\varepsilon) = N(\varepsilon; 0, \sigma^2 I_{10}), \tag{84}$$

where $\theta := (\theta_f, \theta_g) \in \mathbb{R}^2$ and

$$h : \mathbb{R} \to \mathbb{R}^{10}, \theta \mapsto h(\theta) := \theta_g \left(1^{\theta_f}, 2^{\theta_f}, \ldots, 10^{\theta_f}\right)^T. \tag{85}$$

Notably, (85) resembles the ERP-DCM model function h (cf. Eqs. (3), (4), (57) and (58)) in so far as the value $h(\theta)$ is a nonlinear function of θ_f which is multiplied by a function (here the identity) of θ_g. In the following, we first consider some features of the ensuing variational free energy as a function of the variational parameters m_{θ_f} and m_{θ_g}, before demonstrating the joint estimation of θ and σ^2 for a number of prior settings. We base our investigation on a data sample obtained from (84) and (85) with the true, but unknown, parameter settings $\theta := (2,5)^T$ and $\sigma^2 := 100$.

The first panel of Fig. 5A depicts the term T1 of the variational free energy surface as a function of the variational expectation parameter $m_\theta := (m_{\theta_f}, m_{\theta_g})$, where m_{σ^2} and s_{σ^2} are set to the prior values of $\mu_{\sigma^2} := 4.6$ and $\varsigma_{\sigma^2} := 10^{-3}$. This plot illustrates a fundamental issue with models of the type (85), namely that, based purely on the accuracy term T1 their parameters are not uniquely identifiable. Specifically, T1

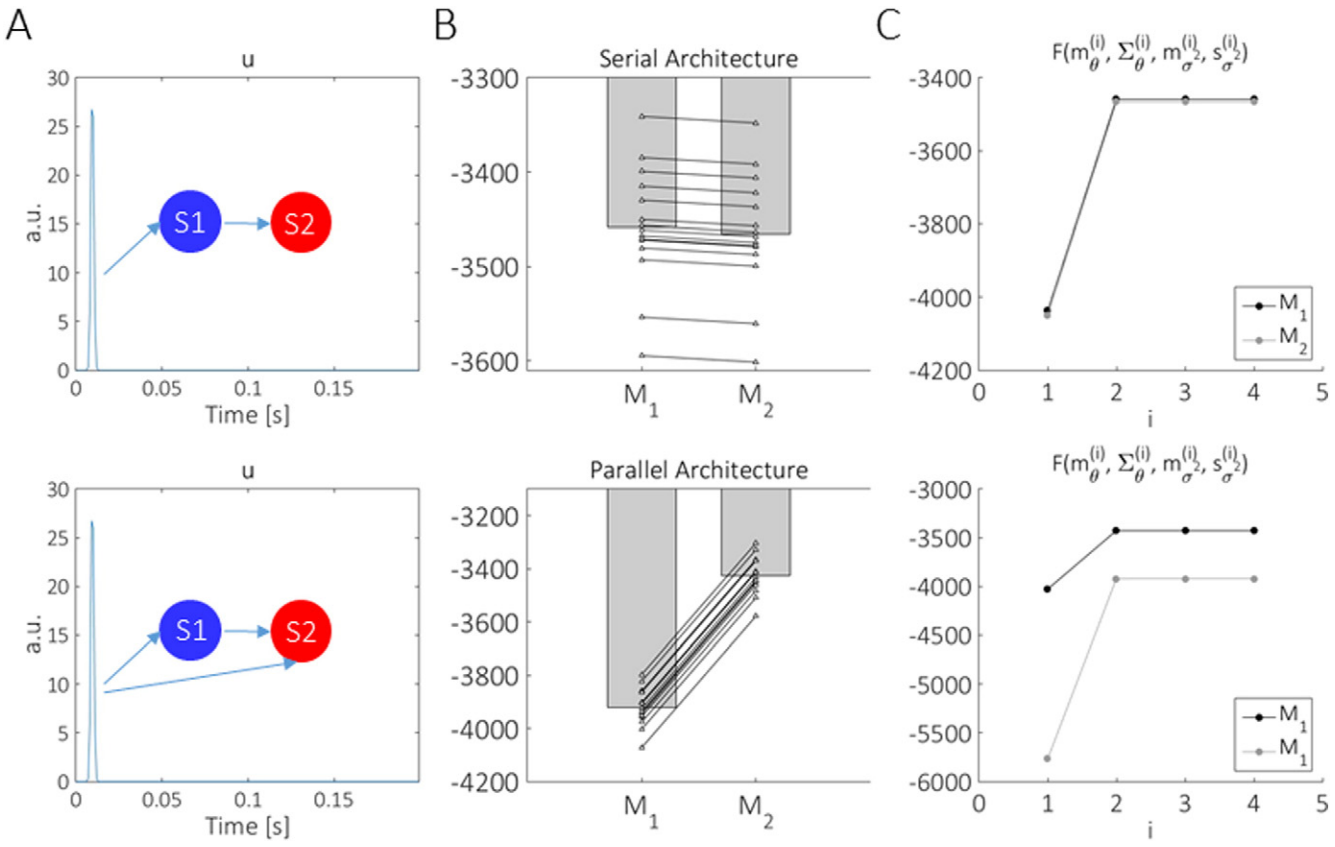

Fig. 12. Model comparison between serial and parallel architectures in a two-source ERP-DCM model. (A) True, but unknown, model architectures. To generate data, two connectivity architectures were simulated, which are parameterized by the input connectivity vectors $C^{(1)} := (1,0)^T$ and $C^{(2)} := (1,1)^T$ and can be interpreted as "serial processing" (upper panel) and "parallel processing" (lower panel) models. (B) 16 data realizations of the models depicted in (A) were obtained, and each estimated using a model M_1, which due to its prior distribution on the input connectivity is biased to a serial architecture, and a model M_2, which due to its prior distribution on the input connectivity is not biased to either architecture. The upper panel shows the final variational free energies for both estimation models for the serial processing data model (triangles) and their mean (gray bars). Notably, model M_1 which is biased towards a serial architecture takes on higher variational free energy values for all simulations than model M_2. The lower panel shows the final variational free energies for both estimation models for the true, but unknown, parallel processing architecture. Here, model M_2, which is not biased towards a serial architecture takes on the higher final variational free energy values. (C) Variational free energy evolution over evolutions of the fixed-form variational Bayes algorithm averaged over model estimations. Importantly, the differences in log marginal likelihood approximations of panel (B) are not due to non-convergence of the algorithm, but correspond to the free energy values obtained upon convergence.

exhibits a crest which assigns an identical likelihood to parameter combinations along a diagonal through parameter space. The second panel of Fig. 5A visualizes the term T2, which results from the Taylor approximation to the nonlinear function h. Notably, the contribution of this term to the overall variational free energy is a function of the variational variance parameter S_θ. For this visualization, S_θ was evaluated based on (78) for a prior variance parameter $\Sigma_\theta := diag(10^6,10^6)$, i.e., an imprecise prior for both components of m_θ. The third panel of Fig. 5A visualizes the contribution of the prior settings $\mu_\theta := (0,0)^T$ and $\Sigma_\theta := diag(10^6,10^6)$, i.e., an imprecise isotropic prior centered at the origin. Finally, the right-most panel of Fig. 5A depicts the full variational free energy surface for the aforementioned prior parameter settings. Note that the terms T4, T5 and T6 of the variational free energy function are not direct functions of the variational expectation parameter m_θ and thus constant over the space of m_{θ_f} and m_{θ_g}. As evident from the figure, the variational free energy surface is dominated by T2. Judging from its topology, a globalized Newton ascent on this surface is unlikely to identify a meaningful convergence point. As shown in Fig. 5B, the model parameters can be rendered identifiable, however, by introducing prior constraints on subsets of the parameter space. Specifically, the panels of Fig. 5B depict the identical terms as Fig. 5A, but with two modifications to the prior settings for $\theta := (\theta_f,\theta_g)^T$. Specifically, an informative prior resembling the true, but unknown, value of θ_g was chosen by setting $\mu_\theta := (0,5)^T$ and $\Sigma_\theta := diag(10^6,10^{-3})$, while the prior settings for σ^2 were held constant at $\mu_{\sigma^2} := 4.6$ and $\varsigma_{\sigma^2} := 10^{-3}$. Compared to Fig. 5A, T1 is unaffected, while the contribution of T2 to the variational free energy is diminished. T3 reflects the tight prior setting for θ_g and the imprecise prior setting for θ_f. Taken together, these changes render the variational free energy surface such that it exhibits a local maximum in the area of the true, but unknown, values. For this surface, a globalized Newton ascent is likely to identify a convergence point around $m_\theta = (2,5)^T$. Finally, Fig. 5C demonstrates that also the prior settings of μ_{σ^2} and ς_{σ^2} affect the variational free energy surface, and thus influence the convergence properties of the Newton ascent in m_θ at least on the first iteration of the overall algorithm. For the panels of Fig. 5C the expectation of the prior distribution σ^2 and the prior parameters μ_θ and Σ_θ were held constant with respect to 5B, while the entropy of the prior distribution σ^2 was increased by setting $\mu_{\sigma^2} := 10^{-3}$ and $\varsigma_{\sigma^2} := 9.2$. The higher uncertainty about σ^2 reflected by this choice of prior has the effect that the expectation of $1/\sigma^2$, entering T1, changes dramatically, such that T1 dominates the variational free energy surface, rendering the parameter m_θ non-identifiable. Note that the expectation of $1/\sigma^2$, which also enters T2 is balanced out by the modified updated value of S_θ, and that, as seen below, larger uncertainty about the noise parameter σ^2 is not necessarily detrimental for the convergence of the overall algorithm – but, as demonstrated, can affect the theoretical convergence properties of the globalized Newton descent in m_θ space at the first iteration.

We next consider some exemplary qualitative properties of the posterior and marginal likelihood approximations of the fixed-form variational Bayes approach for the estimation of the toy model (84) and (85) under different prior choices (Fig. 6). The first panel of each row of Fig. 6 depicts the data expectation $h(\theta)$, the data sample y and the

ensuing maximum-a-posteriori data expectation $h(\theta_{MAP})$. The second to fourth panels depicts the prior (blue curve) and approximated posterior (dashed red curve) distribution over θ_f, θ_g and σ^2, where the approximated posterior distributions correspond to the variational distributions at the last of 10 iterations of the fixed-form Newton algorithm. The fifth panels depicts the variational free energy, which indicates that in all cases, the algorithm has converged. For Fig. 6A – C, the prior expectation of θ was set to $\mu_\theta := (0,5)^T$, and for Fig. 5A and B the prior shape and scale parameters for σ^2 were set to $\mu_{\sigma^2} := 4.6$ and $\varsigma_{\sigma^2} := 10^{-3}$. Fig. 5A derives from using an imprecise prior for θ_f and a precise prior for θ_g given by the prior covariance parameter $\Sigma_\theta := diag(10^6,10^{-3})$. As evident from the second panel of Fig. 6A, this choice of prior results in a highly precise approximated posterior distribution over θ_f around the location of the true, but unknown, value of θ. As shown in Fig. 6B, widening the prior on θ_g by setting Σ_θ: $= diag(10^6,10^6)$, results in less precise approximate posterior distributions about θ_f and θ_g, but a marginally better data fit (visible primarily around $x = 5$ in the first panel). The resulting variational free energy, taking into account the increased posterior uncertainty, however, is marginally smaller than in the case of the tight prior of Fig. 5A (-59.8 vs. -54.1). Finally, Fig. 6C visualizes an effect of changing the prior distribution on σ^2. In comparison to Fig. 6A and B, a less precise prior with a 10-fold lower expectation resulting from $\mu_{\sigma^2} := 2.3$ and $\varsigma_{\sigma^2} := 1$ was chosen. The primary consequence of this choice of prior is that the expectation of the approximate posterior distribution on σ^2 somewhat overestimates its true, but unknown, value, while the uncertainty about the value of σ^2 increases.

In summary, from an objective Bayesian viewpoint, the fixed-form variational Bayesian algorithm developed in Section 3 performs adequately in simple toy models. Depending on the particularities of a given model, regularizing highly precise priors are sometimes necessary to render the remaining parameters identifiable and prior settings over structural and variance parameters may interact.

Estimation of probabilistic delay differential equation models for ERPs

In this section we showcase our estimation framework in the context of the delay differential equation model for ERPs discussed in Section 2. As a testbed, we employ a simple two-source model that generates peri-stimulus electrode space data which shares some intuitive similarity with somatosensory evoked potentials elicited by electrical stimulation to the left median nerve (Fig. 7) (Schomer and da Silva, 2011, Chapter 48). The full parameter value set of this simulation is documented in Supplementary Material S3. The simulation is based on a forward architecture of two sources ("S1" and "S2") located approximately in right primary and secondary somatosensory cortex (Fig. 7A). The system receives an input pulse at approximately 10 ms post-stimulus onset. The latent neural dynamics of this model are shown in Fig. 7 and involve an early strong pyramidal cell population response in S1 and a weaker and delayed pyramidal cell population response in S2. The entire response lasts for approximately 200 ms. In electrode space (Fig. 7C), this latent source-activity results in a strong left-lateralized response, with peak components around 20 ms, 60 ms, and 120 ms post-stimulus onset. Note that for demonstrative purposes the noise level in this simulation was deliberately assumed to be higher than in typical low-pass filtered event-related potentials (Fig. 7C, lower panel). Furthermore, note that in the empirical literature a number of different model architectures have been proposed for modeling somatosensory evoked potentials. These include, for example, three-source single-input models that also model contra-lateral secondary somatosensory cortical sources (Auksztulewicz et al., 2012; Kiebel et al., 2006), or two-source models that assume single or multiple

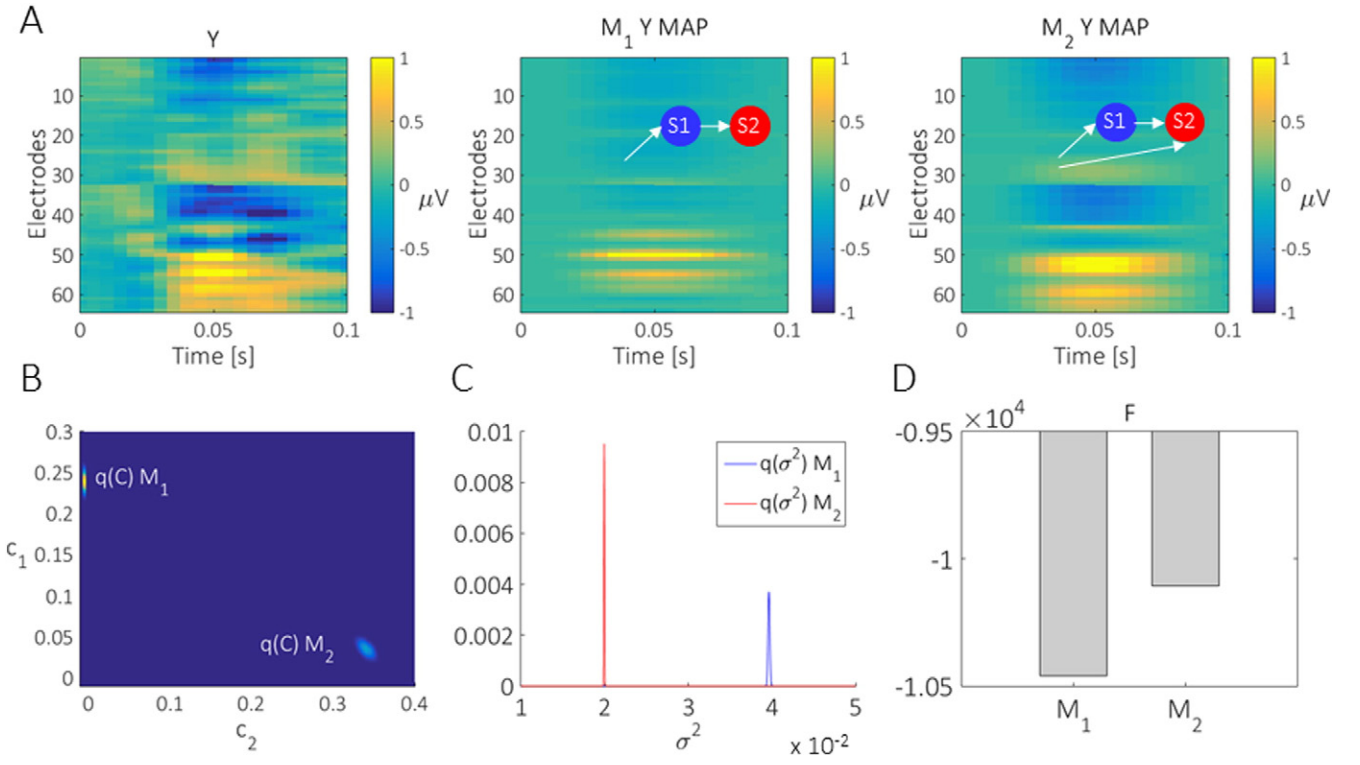

Fig. 13. Application to real data. (A) The left panel depicts the grand-mean somatosensory evoked potential data in electrode × peri-stimulus time space as reported in Ostwald et al. (2012), the middle and right panel show the maximum-a-posteriori estimate-based predictions of models M_1 and M_2, respectively. (B) Variational distributions for the input connectivity parameter upon convergence. Note that the distribution of model M_1 is strongly constrained by the tight prior distribution on input to the second source. (C) Variational distributions for the observation noise variance parameter upon convergence. In line with the visual better data fit of model M_2, the variational distribution of model M_1 is centered on a larger value than that of M_2. (D) Variational free energy approximations of the models' log marginal likelihood for the experimental data. Despite its additional flexibility, model M_2 allocates a larger marginal likelihood to the data. Based on the first 100 ms of a somatosensory evoked potential, this model comparison thus suggests a parallel input architecture of the cortical somatosensory system.

input connections (Klingner et al., 2015, see also Section 4.3). Given the undecided status of the empirical literature on the modeling of somatosensory evoked potentials, we merely view our simulation as a useful testbed for the purposes of the current theoretical study, but do not mean to imply any empirical veracity.

We first consider the Bayesian estimation of the extrinsic input connectivity vector $C := (c_1,c_2)^T$ and the variance parameter σ^2 (Fig. 8). To simulate the data shown in Fig. 7, these parameters were set to the true, but unknown, values of $C := (1,0)^T$ (modeling exclusive input to the first source) and $\sigma^2 := 0.1$. With respect to the notation in Section 3.1, we thus have $\theta := \theta_f \in \mathbb{R}^2$. For model estimation, we chose a prior with low precision by setting $\mu_\theta := (0,0)^T$ and $\Sigma_\theta := diag(10^3,10^3)$. The prior and the true, but unknown, parameter value are shown in Fig. 8A. Likewise, we chose a low precision prior for σ^2 by setting $\mu_{\sigma^2} := 4.6$ and $\varsigma_{\sigma^2} := 1$, corresponding to a log-normal distribution with expectation of approximately 164, and thus three orders of magnitude larger than the true value, and a variance of approximately 1.6×10^4 (Fig. 8C, blue curve). The approximated posterior distributions based on the variational free energy maximization algorithm discussed in Section 3.2 are shown in Fig. 8B and C. The mode of the approximated posterior distribution $q(C)$ is fairly close to the true, but unknown value with deviations on the order of 10^{-4} to 10^{-3}. Notably, some negative correlation is evident from the approximated posterior, and the resulting distribution is highly precise. Likewise, the approximated posterior distribution of the variance parameter (Fig. 8C, red curve) is fairly close to the true, but unknown value, again with high precision. Fig. 8D and E depict the iterations of the variational parameters m_θ and $(m_{\sigma^2}, s_{\sigma^2})^T$ on the first iteration of the variational Bayesian–Newton algorithm on the free energy surface. For the current estimation problem, the variational free energy has a relatively well-defined local maximum in both cases, such that the optimized variational parameter settings are attained after a few iterations of the Newton algorithm sub-loops. Finally, Fig. 8F depicts the variational free energy on the first to fourth iteration of the overall algorithm, which shows virtually no improvement after the second iteration. Note that for clarity, the variational free energy on initialization, which is more negative than the values shown by many orders of magnitudes, was omitted from the figure.

In a second simulation, we considered the estimation of the dipole moment of the first source, i.e. of $\theta := \theta_m^{(1)} = (\theta_{m_1}^{(1)},\theta_{m_2}^{(1)},\theta_{m_3}^{(1)})^T$ and the variance parameter σ^2. To this end, the true, but unknown values of the these parameters were set to $\theta_m^{(1)} := (0.02, -0.05, -0.04)^T$ and $\sigma^2 := 0.1$, and the respective prior parameters were set to the low precision priors $\mu_\theta := (0,0,0)^T$, $\Sigma_\theta := diag(10^3,10^3)$, $\mu_{\sigma^2} := 4.6$ and $\varsigma_{\sigma^2} := 1$. Fig. 9A and B depict the prior and approximate posterior distribution for $\theta_m^{(1)}$, respectively. Specifically, Fig. 9A shows the expectation parameter and the 95% isocontour of the three-dimensional normal prior distribution, while Fig. 9B depicts the expectation parameter and 95% isocontour of the posterior distribution and the true, but unknown parameter value. As in the previous case, the posterior expectation deviates only marginally from the true value and the posterior distribution is highly precise. The estimation of the variance parameter is virtually identical to the first scenario (Fig. 9C). To obtain some intuition about the validity of the estimation in the current case of a three-dimensional parameter, we examined the variational free energy ascent in the space of m_{θ_1} and m_{θ_2}. As the respective Newton algorithm converges already at the first iteration upon initialization, we depict in Fig. 9D the variational free energy surface in the variational parameters m_{θ_1} and m_{θ_2} for $m_{\theta_3} := 0$, corresponding to the prior setting of m_{θ_1} and m_{θ_2} (black dot). Upon the first iteration (Fig. 9E), the expected value of the approximate posterior has attained the maximum of the variational free energy in m_{θ_1} and m_{θ_2} space (black dot). Here, the variational free energy surface is evaluated for the converged setting of $m_{\theta_3} = -0.0403$. Finally, Fig. 9F depicts the variational free energy on the first to fourth iteration of the overall algorithm, which shows virtually no further improvement after the second iteration.

Finally, having demonstrated the estimation of parameters of the latent neural dynamics model as well as the EEG forward model individually, we consider an example for their joint estimation. Specifically, we are here interested in the estimation of the forward connectivity parameter a_{21}, the dipole moment of the first source $\theta_m^{(1)}$, and the variance parameter σ^2 (in terms of the notation of Section 3.1, we thus have $\theta := (\theta_f, \theta_g)^T = (a_{21}, \theta_m^{(1)T})^T \in \mathbb{R}^4$). The true, but unknown, values of these parameters were set to $\theta := (1{,}0.02, -0.05, -0.04)^T$ and $\sigma^2 := 0.1$. Again we chose a low precision prior for θ, depicted in Figs. 10A (blue curve) and 9B (95% isocontour of the three-dimensional normal distribution). To obtain reliable estimates for the current estimation problem under the parameter settings discussed thus far, we noted that we had to increase the prior regularization of the variance parameter. We thus selected $\mu_{\sigma^2} := 0$ and $\varsigma_{\sigma^2} := 0.1$, resulting in a log-normal distribution with expectation of approximately 1.1, i.e. an order of magnitude higher than the true, but unknown variance parameter, and variance of approximately 0.44 corresponding to a reasonable wide prior (Fig. 10D, blue curve). As in the previous simulations, the expectation of the approximated posterior distribution of θ is close to the true, but unknown values, and shows high precision (Fig. 10A, red curve and Fig. 10C, depicting the 95% isocontour of the three-dimensional normal and the true parameter value). Similarly, the approximated posterior distribution of the variance parameter σ^2 is centered at a value in the vicinity of the true, but unknown value, with (compared to the previous simulations) a slight bias towards the prior expectation, which presumably reflects the tighter prior in the current scenario. Finally, the visualization of the variational free energy ascent in the four-dimensional parameter θ is not readily achieved. To nevertheless obtain some insight into validity of the estimation, we visualized the value of the variational free energy on the Newton algorithm sub-loops for the determination of m_θ (blue curves) and $(m_{\sigma^2}, s_{\sigma^2})^T$ (red curves) in Fig. 10E over four iterations of the overall algorithm. The maximal number of iterations for these sub-loops was set to 8, and missing data points reflect convergence on sub-loops. As can be seen, the variational free energy steadily increases, and convergence is attained for both sub-loops on the fourth iteration of the overall algorithm.

In summary, from an objective Bayesian viewpoint, the fixed-form variational Bayesian algorithm developed in Section 3 performs adequately for a number of parameter estimation problems in the current probabilistic delay differential equation ERP model.

Experimental applications of probabilistic delay differential equation models for ERPs

In Section 4.2 we provided a detailed demonstration of parameter estimation and log marginal likelihood approximation for a delay differential equation ERP model. While the issues discussed are highly relevant for the neurobiological interpretation of empirical results obtained using this approach, they may not be of primary concern in experimental contexts. In this section, we thus demonstrate three additional issues that may be of higher immediate concern for experimental applications: the estimation of condition-specific effects, the comparison of different model architectures, and an application to real, not simulated, data. Because a general theoretical coverage of these topics is beyond the scope of the current note, we take a less formal approach than in the preceding sections.

Estimation of condition-specific effects

ERP studies typically rest on the comparison of ERPs acquired under different experimental conditions, e.g., stimulus types or cognitive contexts (Luck, 2014). Like the standard ERP-DCM approach, the model discussed in Section 2 can be extended to the estimation of condition-specific effects by including modulatory effects on its parameter set and concatenation of the resulting model predictions. To generalize our

model to a scenario of multiple ERP conditions, let n_c denote the number of conditions, and let $c = 1, 2, \ldots, n_c$ be the condition index. We firstly note that in the case of more than one ERP, the data takes the form

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_{n_c} \end{pmatrix} \in \mathbb{R}^{n_e n_t n_c}, \tag{86}$$

where $y_c \in \mathbb{R}^{n_e n_t}$ is the vectorized form of the electrode × peri-stimulus time data matrix acquired under experimental condition c. As previously, we define $n := n_e n_t n_c$. Correspondingly, the data model for multiple ERPs takes the form (cf. Eq. (4)):

$$h : \Theta \to \mathbb{R}^n, \theta \mapsto h(\theta) := \begin{pmatrix} vec\left(g\left(\theta_g^{(1)}\right) f\left(\theta_f^{(1)}\right)\right) \\ \vdots \\ vec\left(g\left(\theta_g^{(n_c)}\right) f\left(\theta_f^{(n_c)}\right)\right) \end{pmatrix}, \tag{87}$$

where the parameter set θ partitions into the condition-specific latent neural and forward model parameters $\theta_f^{(1)}, \theta_f^{(2)}, \ldots, \theta_f^{(n_c)}$ and $\theta_g^{(1)}, \theta_g^{(2)}, \ldots, \theta_g^{(n_c)}$, respectively. In principle, any of the elements of $\theta_f^{(c)}$ and $\theta_g^{(c)}$ could be endowed with condition-specific effects. Customary, however, ERP-DCM has parameterized condition-specific effects at the level of the models connectivity architecture, i.e. the forward, backward, and lateral connectivity matrices A^F, A^B, and A^L (cf. Eq. (16)) (David et al., 2006). In the following, we discuss a parameterized approach to endow the forward connectivity matrix A^F with condition-specific effects and demonstrate this for the two-source model of Section 4.2.

To implement the estimation of condition-specific effects we firstly assume that all parameters of the data model (87) stay constant over conditions, except for the forward connectivity matrix, which now takes on condition-specific values $A_1^F, A_2^F, \ldots, A_{n_c}^F$. The elements $a_{ijc}^F \in \mathbb{R}$ (where $1 \le i,j \le n_s$ and $c = 1,\ldots, n_c$) of these matrices can be parameterized in an affine-linear additive manner by defining a reference connectivity matrix $A_1^F := (a_{ij1}^F) \in \mathbb{R}^{n_s \times n_s}$, a "design vector" $M := (0,1,2,\ldots,n_c - 1) \in \mathbb{R}^{n_c \times 1}$, an effect matrix $B := (\beta_{ij})_{1 \le i,j \le n_s}$, and setting

$$a_{ijc} := a_{ij1} + m_c \beta_{ij} (c = 1, \ldots, n_c). \tag{88}$$

In words, the forward connectivity parameter between the jth and ith source in condition c is given by the "baseline connectivity" a_{ij1} (corresponding to the forward connectivity in the first condition) plus the product of the design vector entry m_c and a source-source connectivity specific parameter β_{ij}. For the simulation shown in Fig. 11, we consider $n_c = 2$ conditions and defined:

$$A_1^F := \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, M := \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \text{and } B := \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \tag{89}$$

such that:

$$A_2^F = \begin{pmatrix} 0 & 0 \\ 2 & 0 \end{pmatrix}, \tag{90}$$

which corresponds to a doubled forward connectivity from source 1 to 2 in the case of condition $c = 2$. The remaining parameters remained constant between conditions and were set as in Section 4.2 (see Supplementary Material S3 for details). Fig. 11A and B depict the simulation of the latent neural state dynamics and their resulting observed responses for forward connectivity matrices A_1^F and A_2^F, respectively. The leftmost subpanels depict $x^{(1)}$, the latent dynamics of the first source, which is unaffected by the condition-specific variation of the forward connectivity. The second subpanels depict $x^{(2)}$, the latent dynamics of the second source. Notably, because the forward influence of $x^{(1)}$ on $x^{(2)}$ is doubled under the A_2^F, the amplitude of the latent dynamics of the second source are doubled for the second condition. The final two subpanels visualize the result of this variation across the entire electrode space and for selected electrodes, respectively. Here, the increase in $x^{(2)}$ amplitude affects primarily the electrode responses in the time-window of 50 to 100 ms.

Panel C of Fig. 11 depicts the estimation of the corresponding condition-specific effect parameter β_{21} and the variance parameter σ^2. For the realization of Panels A and B, the true, but unknown, values of these parameters were set to ($\beta_{21} := 1$ (cf. Eq. (82)) and $\sigma^2 := 0.1$. The left-most subpanel depicts the prior distribution $p(\beta_{21})$, where $\mu_{\beta_{21}} := 0$ and $\Sigma_{\beta_{21}} := 10^3$, and the approximation to the posterior distribution $q(\beta_{21})$, which, for the current data realization, results in a MAP estimate of $m_{\beta_{21}} \approx 0.94$. The next subpanel depicts the prior distribution $p(\sigma^2)$, where $\mu_{\sigma^2} := 4.6$ and $\varsigma_{\sigma^2}{}^2 := 1$, and the approximation to the posterior distribution $q(\sigma^2)$, which, for the current data realization, results in a MAP estimate of $E(q(\sigma^2)) \approx 0.15$. The first subpanels on the right depicts the globalized Newton algorithms iterations for the optimization of $m_{\beta_{21}}$ on the first iteration of the fixed-form variational Bayes–Newton algorithm. Notably, the variational free energy F has a well-defined maximum and this is reached after three iterations. Finally, the right-most subpanel depicts the maximization of the variational free energy over iterations of the fixed-form variational Bayes–Newton algorithm, which is achieved effectively after two iterations. In summary, the ERP-DCM approach discussed herein can naturally be extended to the recovery of condition-specific effects. Given its extensive mathematical documentation, targeting other parameters than those encoding effective connectivity for the expression of condition-dependent ERP differences is readily possible. Of course, an affine-linear regression-type parameterization of condition specific effects as in Eq. (88) is just one possible scenario, and categorical ANOVA-type parameterization are readily conceivable.

Model comparison

One of the central features of the variational Bayes approach is that it affords both posterior distribution and log marginal likelihood ("log model evidence") approximation. Upon convergence of a variational Bayes algorithm, the posterior parameter distribution approximation can be used for inferences about the probability of parameters to take on values in parameter space, while the log marginal likelihood approximation (an approximation to the probability of the observed data under the model), can be used for model comparison. So far, we have been concerned with the details and properties of the approximation of the posterior distribution in terms of variational distributions and primarily considered the variational free energy as an indicator of algorithmic convergence. In this section, we explicitly demonstrate how it can be used for model comparison.

Inspired by a recent report that demonstrated evidence for a parallel input architecture of the cortical somatosensory system using MEG and the ERP-DCM SPM12 implementation (Klingner et al., 2015), we investigated if we can recover the true, but unknown, input architecture of a two-source ERP model based on the comparison of variational free energy log marginal likelihood approximations. To this end, we simulated data using the parameter set of Section 4.2 (see Supplementary Material S3 for details) in two scenarios: firstly, assuming a "serial architecture," i.e. limiting the effect of the input function on the first source, and secondly, assuming a "parallel architecture," i.e. allowing for an effect of the input function on both sources (Fig. 12A). In parameterized form, these scenarios correspond to setting the true, but unknown, input connectivity vector $C \in \mathbb{R}^{n_s}$ to $C^{(1)} := (1,0)^T$ and $C^{(1)} := (1,1)^T$, respectively. We next analyzed realizations from both scenarios using two probabilistic models that are distinguished by their marginal (prior) distributions for the input connectivity vector as reported by Klingner et al. (2015). Both models, which we denote by M_1 and M_2, have a prior expectation of $\mu_C^{M_1} := \mu_C^{M_2} := (0,0)^T$ for $C := (c_1,c_2)^T$. However, they are distinct in their prior covariances given by

$$\Sigma_C^{M_1} := \begin{pmatrix} 10^3 & 0 \\ 0 & 10^{-3} \end{pmatrix} \text{ and } \Sigma_C^{M_2} := \begin{pmatrix} 10^3 & 0 \\ 0 & 10^3 \end{pmatrix}. \tag{91}$$

Model M_1 thus has an imprecise marginal prior for the input to the first source and a precise marginal prior for the input to the second source, while model M_2 has imprecise marginal priors for inputs to both sources. Intuitively, under model M_1 a non-zero value of c_1 is thus more likely than a non-zero value of c_2, while under model M_2 no such bias exists.

The upper and lower panels of Fig. 12B depict the results of the corresponding model comparisons. For the upper panel, 16 data realizations based on the true, but unknown, parameter $C^{(1)}$ were obtained, models M_1 and M_2 fit to the data by approximating the variational distribution $q(C)$ and $q(\sigma^2)$ as described in previous sections, and the final variational free energy assessed. For all realizations, model M_1, which embeds a bias for a serial architecture, achieves a higher variational free energy than model M_2, which embeds no such bias. In other words, the true, but unknown, serial architecture underlying the data realization is correctly recovered. For the lower panel, 16 data realizations based on the true, but unknown, parameter $C^{(2)}$ were obtained. In this case, the model M_2, which, in contrast to model M_1, renders a parallel architecture not virtually impossible, assumes the higher variational free energy values throughout. Again, the true, but unknown, parallel architecture underlying the data realization is thus correctly recovered. Finally, the upper and lower panels of Fig. 12C depict the average variational free energy evolution over iterations of the fixed-form variational Bayes algorithm for both true, but unknown, scenarios and estimation models. Notably, the differences in variational free energy observed in Fig. 12B are not due to non-convergence of any of the algorithms, but correspond to the final variational free energy values upon convergence. In summary, the ERP-DCM approach discussed herein can be used for model comparison and, for the scenario considered, performs reliably.

Application to real data

Finally, we were interested in the applicability of our approach to experimental data. To this end, we reconsidered the somatosensory evoked potential (SEP) data reported in Fig. 3A of Ostwald et al. (2012). In brief, these data were obtained using a standard electrical stimulation protocol of the left-median nerve in a group of 15 participants (0.2 ms pulse duration, ≈ 5 mA amplitude, 650 ms interstimulus interval, ≈ 5000 stimuli per participant) and recorded using a 64-channel active electrode system (ActiveTwo, BioSemi) at 2048 Hz with electrodes placed according to the extended 10–20 system. Data processing using SPM8 (Litvak et al., 2011) included down-sampling to 512 Hz, band-pass filtering at 1–40 Hz, eye-movement correction, epoching from −100 ms to 600 ms, and averaging. For full methodological details, please refer to (Ostwald et al., 2012). Again inspired by the work of Klingner et al. (2015), we here focus on the 0–100 ms time-window of the resulting grand-mean SEP, which is shown in the left subpanel of Fig. 13A in electrode × peri-stimulus time space. This time-window includes an N20/P20 response and a stronger N45/P60 component, both primarily over contralateral electrodes (electrode numbers 33–64). Like (Klingner et al., 2015) and as in our simulations above, we model these data with a two-source dipole model, and focus on the estimation of the input connectivity parameters $C \in \mathbb{R}^2$. All other neural parameters were set to their SPM default values, and the parameters of the forward model corresponded to that used in Ostwald et al. (2012) (see Supplementary Material S4 for details). As in the previous section, we created two probabilistic models M_1 and M_2 with prior expectations of $\mu_C^{M_1} := \mu_C^{M_2} := (0,0)^T$ for $C := (c_1, c_2)^T$. Again, we chose an imprecise marginal prior for the input to the first source and a precise marginal prior for the input to the second source for model M_1 and imprecise marginal priors for inputs to both sources for model M_2 by setting

$$\Sigma_C^{M_1} := \begin{pmatrix} 10^6 & 0 \\ 0 & 10^{-6} \end{pmatrix} \text{ and } \Sigma_C^{M_2} := \begin{pmatrix} 10^6 & 0 \\ 0 & 10^6 \end{pmatrix}. \tag{92}$$

The stylized connectivity architectures resulting from these prior settings are visualized as inlets in the middle and right subpanels of Fig. 13A, respectively. Estimating both models based on the experimental data using the fixed-form variational Bayes–Newton algorithm results in the converged variational approximations of the posterior parameter distributions $q(C)$ and $q(\sigma^2)$ shown for both models in Fig. 13B and C, respectively. In line with the highly precise prior for the input connectivity parameter to the second source (S2) in model M_1, the variational distribution is centered on 0 for c_2 and assumes a non-zero value for c_1, i.e., input to the first source, only. Complementarily, the imprecise priors for both input connection parameters in M_2 result in variational distributions that assume non-zero values for both parameters, with a larger value for input to the second source. Based on the expectations of these variational distributions, the maximum-a-posteriori predictions of models M_1 and M_2 are shown in the middle and right-most subpanels of Fig. 13A. Visual comparison with the experimental data shown in the left-most panel shows a better predictive correspondence of model M_2. In line with this result, the converged variational distribution for model M_2 is centered on a smaller value for the observation noise parameter σ^2 than for model M_1, as shown in Fig. 13C. Finally, comparing the variational free energy approximations to the log marginal likelihood yields a higher value for model M_2, reflecting the better data fit, which outscores the higher model flexibility given by the imprecise prior distribution of c_2. Like (Klingner et al., 2015), we may thus conclude that a parallel architecture of the somatosensory system is more supported by our grand-mean SEP data than a serial architecture.

Discussion

In this note, we have reviewed the technical framework of ERP-DCM. Further, we have modified the framework on a number of occasions, most prominently with respect to the numerical optimization of the variational free energy. Using low precision priors, we have shown that the ensuing numerical implementation of the ERP-DCM estimation algorithm is able to recover, true, but unknown, simulated model parameter values in a number of model scenarios. In addition, we considered aspects in the experimental applicability of the approach. In the following we, elaborate on the technical modification we implemented in our approach to ERP-DCM, consider some limitations of the current framework, and finally point towards possible future developments that may overcome these.

Our use of a conventional globalized Newton algorithm with backtracking step-size selection for maximization of the variational free energy function was motivated by the comprehensive technical literature on this approach. For example, in the limit of convex objective functions, these and related approaches have theoretically established convergence properties (Nocedal and Wright, 2006). With our approach we thus would like to emphasize that the fixed-form variational Bayes approach condenses to a standard nonlinear optimization problem, which, given the novelty and rootedness of the approach in statistical physics and probabilistic machine learning, is not necessarily immediately clear. In other words, having formulated approximate Bayesian inference in terms of optimizing variational free energy, one is free to use any standard optimization scheme to minimize free energy and implicitly maximize Bayesian model evidence. If we further assume that the form of the posterior distribution is Gaussian, this leads to an enormous simplification of the optimization and lends itself to the standard descent schemes of the sort that we have described. Importantly, this was the basic idea behind the original introduction of the "Variational-Laplace" scheme proposed by Friston et al. (2007). Based on this insight, future developments of approximate Bayesian estimation techniques can thus follow a simple two-step procedure: firstly, the analytical derivation of the model-specific variational free energy function based on its functional form, and secondly, its numerical optimization using a suitable nonlinear optimization routine. To this end, it is worth

noting that the globalized Newton approach for the numerical optimization of the variational free energy function employed herein is obviously not without alternatives. While trust-region methods (Conn et al., 1987) offer an unexplored alternative to the step-size selection procedures employed here, larger progress may be afforded by the deployment of constrained optimization algorithms which explicitly take into account (positivity) constraints imposed on subsets of variational parameters (Nocedal and Wright, 2006), by global optimization approaches (Thoai et al., 2008), or by probabilistic numerics (Hennig et al., 2015) (see (Lomakina et al., 2015) for first developments in this direction in a related model class). Finally, this theme also points towards an important mathematical issue for the future development of the fixed-form variational Bayes method for nonlinear models: the mathematical properties of the variational free energy objective function and the influence of approximations in its derivation for both its statistical validity and its proneness to exhibiting local extrema (e.g. Wipf and Nagarajan, 2009).

The delay differential equation character of the neural dynamics model in ERP-DCM is fundamental. It may well be the case that delay differential equations represent an optimal compromise between ordinary and partial differential equations for modeling the spatiotemporal dynamics of non-invasive neuroimaging signals (Jirsa, 2009; Pinotsis et al., 2013). To this end, it would be a worthwhile endeavor to evaluate the performance of the approximate integration approach currently employed to standard approaches in the numerical treatment of delay differential equation systems (in't Hout, 1996; Shampine and Thompson, 2001), also with respect to stability considerations (Erneux, 2009). Furthermore, many aspects of the ERP-DCM framework rest on numerical differentiation. The currently employed finite (forward) difference technique (Sengupta et al., 2014) may well be improved on, and state-of-the art algorithmic-differentiation techniques offer a promising alternative (Griewank and Walther, 2008). In our experimentation with the current approach, we made the occasional observation that changing the forward step-size of the numerical differentiation routine resulted in vastly different performance of the fixed-form variational Bayes algorithm, which is of course undesirable.

While these numerical developments will likely improve the robustness of the ERP-DCM estimation framework, a more fundamental concern pertains to its parameter identifiability. By parameter identifiability we mean, intuitively, the property of a probabilistic model to generate different probability distributions of its observed random variables given different values of the parameters governing its unobserved random variables. This issue is starting to be addressed in the application of DCM to fMRI data (Arand et al., 2015; Frässle et al., 2015), but its systematic investigation is outstanding for ERP-DCM, including our current modification. This is particularly important, if maximum-a-posteriori parameter estimates are being employed to test hypotheses about condition-specific connectivity architectures (e.g. Auksztulewicz et al., 2012; Boly et al., 2011, 2012). Two principal strategies are conceivable in this regard. Firstly, we believe that the neural dynamics model currently employed in ERP-DCM could potentially be simplified without compromising its biological plausibility by treating a number of parameters as fixed constants. For example, Spiegler et al. (2010) have shown that upon reduction of the dimensionality of a single source's parameter space from 33 to 5, the ensuing system can still display waxing and waning of alpha activity, epileptic spiking-like activity, and noise-driven REM-sleep like activity. To our knowledge, similar theoretical analyses of multiple source ERP-DCM models in the context of ERPs have not been performed yet and may help to identify critical and unique parameter values for the expression of biologically observed effects. Secondly and more pragmatically, empirical applications of the ERP-DCM framework could be supplemented by parameter-recovery simulations, or, more directly, by visualizations of the variational free energy surface, affirming the uniqueness of parameter estimates that are used for subsequent neurocognitive inferences. Finally, we predict that questions of parameter identifiability will also play a central role in ongoing efforts to establish construct validity for DCM based on invasive electrophysiological recordings (e.g. David et al., 2008; Moran et al., 2008, 2011a, 2011b). While parameter identifiability as considered here is a purely technical question (i.e. a parameter values may be uniquely identifiable without bearing any construct validity), ideally the parameters of biologically-plausible neuroimaging data would be both uniquely identifiability and validated based on independent evidence.

In summary, a decade after its inception, the ERP-DCM framework enjoys widespread empirical popularity and forms an important cornerstone in the coalescence of computational neuroscience and non-invasive functional neuroimaging. Many more ramifications of this framework can be and are being explored, including its generalization to stochastic dynamics (e.g. in the spirit of Daunizeau et al. (2009, 2012)) or its estimation using non-variational Bayesian methods, such as MCMC (Chumbley et al., 2007; Sengupta et al., 2015). With the current technical note we hope to ease the technical access to this promising methodology and contribute to its appreciation as a standard probabilistic delay differential equation scheme for the modeling of event-related electroencephalographic potentials.

Software note

The Matlab code implementing the approach and simulations of this technical note is included as Supplementary Material S5 and available for download from the corresponding author's webpage.

Appendix A

For the derivation of (77), we require the following property of expectations of multivariate random variables $x \in \mathbb{R}^d$ under normal distributions, which we state without proof.

A.1. Normal expectation theorem

For $x, m, \mu \in \mathbb{R}^d$, $\Sigma \in \mathbb{R}^{d\times d}$ *p.d.* and $A \in \mathbb{R}^{d\times d}$

$$\left\langle (x-m)^T A(x-m) \right\rangle_{N(x;\mu,\Sigma)} = (\mu-m)^T A(\mu-m) + tr(A\Sigma). \tag{A1}$$

Proofs of (A1) can be found for example in Petersen et al. (2006) and in the references therein.

We are concerned with the variational free energy functional (76) for the following joint distribution:

$$p(y,\theta,\sigma^2) = p(y|\theta,\sigma^2)p(\theta,\sigma^2) = p(y|\theta,\sigma^2)p(\theta)p(\sigma^2), \tag{A2}$$

where

$$p(y|\theta,\sigma^2) := N(y; h(\theta), \sigma^2 I_n), p(\theta) := N(\theta;\mu_\theta,\Sigma_\theta) \text{ and } p(\sigma^2) := LN(\sigma^2;\mu_{\sigma^2},\varsigma_{\sigma^2}) \tag{A3}$$

and the variational distributions

$$q(\theta) := N(\theta; m_\theta, S_\theta) \text{ and } q(\sigma^2) := LN(\sigma^2; m_{\sigma^2}, s_{\sigma^2}). \tag{A4}$$

Using the properties of the logarithm and the linearity of the integral, we first note that

$$\begin{aligned}\mathcal{F}(q(\theta)q(\sigma^2)) &= \int q(\theta)q(\sigma^2)\ln\left(\frac{p(y,\theta,\sigma^2)}{q(\theta)q(\sigma^2)}\right)d\theta d\sigma^2 \\ &= \int q(\theta)q(\sigma^2)(\ln p(y,\theta,\sigma^2) - \ln q(\theta) - \ln q(\sigma^2))d\theta d\sigma^2 \\ &= \int q(\theta)q(\sigma^2)\ln p(y,\theta,\sigma^2)d\theta d\sigma^2 - \int q(\theta)\ln q(\theta)d\theta - \int q(\sigma^2)\ln q(\sigma^2)d\sigma^2,\end{aligned} \tag{A5}$$

where the last equality follows with $\int q(\sigma^2)d\sigma^2 = 1$ and $\int q(\theta)d\theta = 1$. Of the remaining three integral terms, the latter two correspond to the differential entropy of a multivariate normal distribution and a log-normal distribution, both of which are well-known to correspond to nonlinear functions of their variational parameters:

$$-\int q(\theta)\ln q(\theta)d\theta = \mathcal{H}(N(\theta; m_\theta, S_\theta)) = \frac{1}{2}\ln|S_\theta| + \frac{m}{2}\ln(2\pi e) \tag{A6}$$

and

$$-\int q(\sigma^2)\ln q(\sigma^2)d\sigma^2 = \mathcal{H}(LN(\sigma^2; m_{\sigma^2}, s_{\sigma^2})) = \frac{1}{2} + \frac{1}{2}\ln(2\pi s_{\sigma^2}) + m_{\sigma^2}. \tag{A7}$$

There thus remains the evaluation of the first integral term, which corresponds to the expectation of the log joint probability density of the observed and unobserved random variables under the variational distribution of the unobserved random variables, which we denote using the bracket notation for expectations $\langle f(x)\rangle_{p(x)} = \int p(x)f(x)dx$ as

$$\int q(\theta)q(\sigma^2)\ln p(y,\theta,\sigma^2)d\theta d\sigma^2 = \langle \ln p(y,\theta,\sigma^2)\rangle_{q(\theta)q(\sigma^2)}. \tag{A8}$$

Substitution of the functional form of $p(y, \theta, \sigma^2)$ then results in:

$$\begin{aligned}\langle \ln p(y,\theta,\sigma^2)\rangle_{q(\theta)q(\sigma^2)} &= \langle \ln(N(y; h(\theta), \sigma^2 I_n)N(\theta; \mu_\theta, \Sigma_\theta)LN(\sigma^2; \mu_{\sigma^2}, \varsigma_{\sigma^2}))\rangle_{N(\theta;m_\theta,S_\theta)LN(\sigma^2;m_{\sigma^2},s_{\sigma^2})} \\ &= \langle \ln(N(y; h(\theta), \sigma^2 I_n))\rangle_{N(\theta;m_\theta,S_\theta)LN(\sigma^2;m_{\sigma^2},s_{\sigma^2})} \\ &\quad + \langle \ln(N(\theta; \mu_\theta, \Sigma_\theta))\rangle_{N(\theta;m_\theta,S_{x0})} \\ &\quad + \langle \ln(LN(\sigma^2; \mu_{\sigma^2}, \varsigma_{\sigma^2}))\rangle_{LN(\sigma^2;m_{\sigma^2},s_{\sigma^2})} \\ &= \left\langle \ln\left((2\pi)^{-\frac{n}{2}}|\sigma^2 I_n|^{-\frac{1}{2}}\exp\left(-\frac{1}{2\sigma^2}(y-h(\theta))^T(y-h(\theta))\right)\right)\right\rangle_{N(\theta;m_\theta,S_\theta)LN(\sigma^2;m_{\sigma^2},s_{\sigma^2})} \\ &\quad + \left\langle \ln\left((2\pi)^{-\frac{p}{2}}|\Sigma_\theta|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}(\theta-\mu_\theta)^T\Sigma_\theta^{-1}(\theta-\mu_\theta)\right)\right)\right\rangle_{N(\theta;m_\theta,S_\theta)} \\ &\quad + \left\langle \ln\left((2\pi\varsigma_{\sigma^2})^{-\frac{1}{2}}(\sigma^2)^{-1}\exp\left(-\frac{1}{2\varsigma_{\sigma^2}}(\ln\sigma^2-\mu_{\sigma^2})^2\right)\right)\right\rangle_{LN(\sigma^2;m_{\sigma^2},s_{\sigma^2})} \\ &= \left\langle -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}(y-h(\theta))^T(y-h(\theta))\right\rangle_{N(\theta;m_\theta,S_\theta)LN(\sigma^2;m_{\sigma^2},s_{\sigma^2})} \\ &\quad + \left\langle -\frac{p}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_\theta| - \frac{1}{2}(\theta-\mu_\theta)^T\Sigma_\theta^{-1}(\theta-\mu_\theta)\right\rangle_{N(\theta;m_\theta,S_\theta)} \\ &\quad + \left\langle -\frac{1}{2}\ln 2\pi\varsigma_{\sigma^2} - \ln\sigma^2 - \frac{1}{2\varsigma_{\sigma^2}}(\ln\sigma^2-\mu_{\sigma^2})^2\right\rangle_{LN(\sigma^2;m_{\sigma^2},s_{\sigma^2})} \\ &= -\frac{n}{2}\ln 2\pi - \frac{n}{2}\langle \ln\sigma^2\rangle_{LN(\sigma^2;m_{\sigma^2},s_{\sigma^2})} - \frac{1}{2}\left\langle(\sigma^2)^{-1}\right\rangle_{LN(\sigma^2;m_{\sigma^2},s_{\sigma^2})}\left\langle(y-h(\theta))^T(y-h(\theta))\right\rangle_{N(\theta;m_\theta,S_\theta)} \\ &\quad - \frac{p}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_\theta| - \frac{1}{2}\left\langle(\theta-\mu_\theta)^T\Sigma_\theta^{-1}(\theta-\mu_\theta)\right\rangle_{N(\theta;m_\theta,S_\theta)} \\ &\quad - \frac{1}{2}\ln 2\pi\varsigma_{\sigma^2} - \langle \ln\sigma^2\rangle_{LN(\sigma^2;m_{\sigma^2},s_{\sigma^2})} - \frac{1}{2\varsigma_{\sigma^2}}\left\langle(\ln\sigma^2-\mu_{\sigma^2})^2\right\rangle_{LN(\sigma^2;m_{\sigma^2},s_{\sigma^2})}.\end{aligned} \tag{A9}$$

There thus remain five different integral terms. We consider each of these in turn.

- $$\langle \ln\sigma^2\rangle_{LN(\sigma^2;m_{\sigma^2},s_{\sigma^2})} \tag{I1}$$

In (I1) the random variable σ^2 is log-normally distributed with parameters m_{σ^2} and s_{σ^2}, which implies that the logarithm of σ^2 is normally distributed with parameters m_{σ^2} and s_{σ^2}. It thus follows that the expectation of $\ln \sigma^2$ under $LN(\sigma^2;m_{\sigma^2},s_{\sigma^2})$ is given by:

$$\left\langle \ln\sigma^2 \right\rangle_{LN\left(\sigma^2;m_{\sigma^2},s_{\sigma^2}\right)} = m_{\sigma^2}. \tag{A10}$$

- $$\left\langle \left(\sigma^2\right)^{-1} \right\rangle_{LN\left(\sigma^2;m_{\sigma^2},s_{\sigma^2}\right)} \tag{I2}$$

The integral (I2) corresponds to the first inverse moment of σ^2 under $LN(\sigma^2;m_{\sigma^2},s_{\sigma^2})$, such that with Supplementary Material S1:

$$\left\langle \left(\sigma^2\right)^{-1} \right\rangle_{LN\left(\sigma^2;m_{\sigma^2},s_{\sigma^2}\right)} = \exp\left((-1)m_{\sigma^2} + \frac{1}{2}(-1)^2 s_{\sigma^2}\right) = \exp\left(-m_{\sigma^2} + \frac{1}{2}s_{\sigma^2}\right). \tag{A11}$$

- $$\left\langle (y-h(\theta))^T (y-h(\theta)) \right\rangle_{N(\theta;m_\theta,S_\theta)} \tag{I3}$$

We have:

$$\left\langle (y-h(\theta))^T (y-h(\theta)) \right\rangle_{N(\theta;m_\theta,S_\theta)} = \left\langle y^T y - y^T h(\theta) - h(\theta)^T y + h(\theta)^T h(\theta) \right\rangle_{N(\theta;m_\theta,S_\theta)} = y^T y - 2y^T \left\langle h(\theta) \right\rangle_{N(\theta;m_\theta,S_\theta)} + \left\langle h(\theta)^T h(\theta) \right\rangle_{N(\theta;m_\theta,S_\theta)} \tag{A12}$$

and are hence led to the evaluation of the expectation of a normal random variable θ under the nonlinear transformation h. The apparent idea of the ERP-DCM literature is to approximate the function h by a multivariate first-order Taylor expansion in order to evaluate the remaining expectations (see (Chappell et al., 2009) for an explicit discussion of this approach). Denoting the Jacobian matrix of h evaluated at the variational expectation parameter m_θ by $J^h(m_\theta)$ we thus write (Magnus, 1999):

$$h(\theta) \approx h(m_\theta) + J^h(m_\theta)(\theta - m_\theta). \tag{A13}$$

By replacing $h(\theta)$ in the first expectation of the right-hand side of (A12) with the approximation (A13), we then obtain:

$$\begin{aligned} \left\langle h(\theta) \right\rangle_{N(\theta;m_\theta,S_\theta)} &\approx \left\langle h(m_\theta) + J^h(m_\theta)(\theta - m_\theta) \right\rangle_{N(\theta;m_\theta,S_\theta)} \\ &= h(m_\theta) + J^h(m_\theta)\left(\left\langle \theta - m_\theta \right\rangle_{N(\theta;m_\theta,S_\theta)}\right) \\ &= h(m_\theta) + J^h(m_\theta)\left(\left\langle \theta \right\rangle_{N(\theta;m_\theta,S_\theta)} - m_\theta\right) \\ &= h(m_\theta) + J^h(m_\theta)(m_\theta - m_\theta) \\ &= h(m_\theta). \end{aligned} \tag{A14}$$

Further, replacing $h(\theta)$ in the second expectation of the right-hand side of (A12) with the approximation (A13), we obtain

$$\begin{aligned} \left\langle h(\theta)^T h(\theta) \right\rangle_{N(\theta;m_\theta,S_\theta)} &\approx \left\langle \left(h(m_\theta) + J^h(m_\theta)(\theta - m_\theta)\right)^T \left(h(m_\theta) + J^h(m_\theta)(\theta - m_\theta)\right) \right\rangle_{N(\theta;m_\theta,S_\theta)} \\ &= \left\langle h(m_\theta)^T h(m_\theta) + 2h(m_\theta)^T J^h(m_\theta)(\theta - m_\theta) + \left(J^h(m_\theta)(\theta - m_\theta)\right)^T \left(J^h(m_\theta)(\theta - m_\theta)\right) \right\rangle_{N(\theta;m_\theta,S_\theta)} \\ &= h(m_\theta)^T h(m_\theta) + 2h(m_\theta)^T J^h(m_\theta)\left\langle (\theta - m_\theta) \right\rangle_{N(\theta;m_\theta,S_\theta)} + \left\langle \left(J^h(m_\theta)(\theta - m_\theta)\right)^T \left(J^h(m_\theta)(\theta - m_\theta)\right) \right\rangle_{N(\theta;m_\theta,S_\theta)}. \end{aligned} \tag{A15}$$

Considering the first remaining expectations yields:

$$\left\langle (\theta - m_\theta) \right\rangle_{N(\theta;m_\theta,S_\theta)} = \left\langle \theta \right\rangle_{N(\theta;m_\theta,S_\theta)} - m_\theta = m_\theta - m_\theta = 0. \tag{A16}$$

To evaluate the second remaining expectation, we first rewrite it as:

$$\left\langle \left(J^h(m_\theta)(\theta - m_\theta)\right)^T \left(J^h(m_\theta)(\theta - m_\theta)\right) \right\rangle_{N(\theta;m_\theta,S_\theta)} = \left\langle (\theta - m_\theta)^T J^h(m_\theta)^T J^h(m_\theta)(\theta - m_\theta) \right\rangle_{N(\theta;m_\theta,S_\theta)} \tag{A17}$$

and note that $(\theta - m_\theta)^T \in \mathbb{R}^{1\times p}$ $J^h(\theta)^T \in \mathbb{R}^{p\times n}$ $J^h(m_\theta) \in \mathbb{R}^{n\times p}$ and $(\theta - m_\theta) \in \mathbb{R}^{p\times 1}$. Application of the Normal expectation theorem (A1) then yields:

$$\left\langle (\theta - m_\theta)^T J^h(m_\theta)^T J^h(m_\theta)(\theta - m_\theta) \right\rangle_{N(\theta;m_\theta,S_\theta)} = (m_\theta - m_\theta)^T J^h(m_\theta)^T J^h(m_\theta)(m_\theta - m_\theta) + tr\left(J^h(m_\theta)^T J^h(m_\theta) S_\theta\right) = tr\left(J^h(m_\theta)^T J^h(m_\theta) S_\theta\right). \tag{A18}$$

We thus have

$$\left\langle h(\theta)^T h(\theta)\right\rangle_{N(\theta;m_\theta,S_\theta)} \approx h(m_\theta)^T h(m_\theta) + tr\left(J^h(m_\theta)^T J^h(m_\theta) S_\theta\right). \tag{A19}$$

In summary, we obtain the following approximation for the integral (I3):

$$\begin{aligned}
&\left\langle (y-h(\theta))^T (y-h(\theta))\right\rangle_{N(\theta;m_\theta,S_\theta)} \\
&\quad = y^T y - 2y^T \langle h(\theta)\rangle_{N(\theta;m_\theta,S_\theta)} + \left\langle h(\theta)^T h(\theta)\right\rangle_{N(\theta;m_\theta,S_\theta)} \\
&\quad \approx y^T y - 2y^T h(m_\theta) + h(m_\theta)^T h(m_\theta) + tr\left(J^h(m_\theta)^T J^h(m_\theta) S_\theta\right) \\
&\quad = (y-h(m_\theta))^T (y-h(m_\theta)) + tr\left(J^h(m_\theta)^T J^h(m_\theta) S_\theta\right).
\end{aligned} \tag{A20}$$

- $\left\langle (\theta-\mu_\theta)^T \Sigma_\theta^{-1} (\theta-\mu_\theta)\right\rangle_{N(\theta;m_\theta,S_\theta)}$ (I4)

Using the Normal expectation theorem (A1), we have:

$$\begin{aligned}
\left\langle (\theta-\mu_\theta)^T \Sigma_\theta^{-1} (\theta-\mu_\theta)\right\rangle_{N(\theta;m_\theta,S_\theta)} &= \left\langle \theta^T\Sigma_\theta^{-1}\theta - \theta^T\Sigma_\theta^{-1}\mu_\theta - \mu_\theta^T\Sigma_\theta^{-1}\theta + \mu_\theta^T\Sigma_\theta^{-1}\mu_\theta\right\rangle_{N(\theta;m_\theta,S_\theta)} \\
&= \left\langle \theta^T\Sigma_\theta^{-1}\theta\right\rangle_{N(\theta;m_\theta,S_\theta)} - 2\mu_\theta^T\Sigma_\theta^{-1}\langle\theta\rangle_{N(\theta;m_\theta,S_\theta)} + \mu_\theta^T\Sigma_\theta^{-1}\mu_\theta \\
&= tr\left(\Sigma_\theta^{-1}S_\theta\right) + m_\theta^T\Sigma_\theta^{-1}m_\theta - 2\mu_\theta^T\Sigma_\theta^{-1}m_\theta + \mu_\theta^T\Sigma_\theta^{-1}\mu_\theta \\
&= tr\left(\Sigma_\theta^{-1}S_\theta\right) + (m_\theta-\mu_\theta)^T\Sigma_\theta^{-1}(m_\theta-\mu_\theta).
\end{aligned} \tag{A21}$$

- $\left\langle \left(\ln\sigma^2-\mu_{\sigma^2}\right)^2\right\rangle_{LN(\sigma^2;m_{\sigma^2},s_{\sigma^2})}$ (I5)

We have

$$\begin{aligned}
\left\langle \left(\ln\sigma^2-\mu_{\sigma^2}\right)^2\right\rangle_{LN(\sigma^2;m_{\sigma^2},s_{\sigma^2})} &= \left\langle \left(\ln\sigma^2\right)^2 - 2\mu_{\sigma^2}\ln\sigma^2 + \mu_{\sigma^2}^2\right\rangle_{LN(\sigma^2;m_{\sigma^2},s_{\sigma^2})} \\
&= \left\langle \left(\ln\sigma^2\right)^2\right\rangle_{LN(\sigma^2;m_{\sigma^2},s_{\sigma^2})} - 2\mu_{\sigma^2}\left\langle \ln\sigma^2\right\rangle_{LN(\sigma^2;m_{\sigma^2},s_{\sigma^2})} + \mu_{\sigma^2}^2 \\
&= \left\langle \left(\ln\sigma^2\right)^2\right\rangle_{LN(\sigma^2;m_{\sigma^2},s_{\sigma^2})} - 2\mu_{\sigma^2}m_\sigma^2 + \mu_{\sigma^2}^2
\end{aligned} \tag{A22}$$

The term $\langle(\ln\sigma^2)^2\rangle_{LN(\sigma^2;m_\sigma{}^2,s_\sigma{}^2)}$ corresponds to the expectation of the square of the normally distributed random variable $\ln\sigma^2$ under $N(\ln\sigma^2;m_{\sigma^2},s_{\sigma^2})$ and is thus given by:

$$\left\langle \left(\ln\sigma^2\right)^2\right\rangle_{LN(\sigma^2;m_{\sigma^2},s_{\sigma^2})} = m_{\sigma^2}^2 + s_{\sigma^2}. \tag{A23}$$

We thus obtain

$$\begin{aligned}
\left\langle \left(\ln\sigma^2-m_{\sigma^2}\right)^2\right\rangle_{LN(\sigma^2;m_{\sigma^2},s_{\sigma^2})} &= m_{\sigma^2}^2 + s_{\sigma^2} - 2\mu_{\sigma^2}m_\sigma^2 + \mu_{\sigma^2}^2 \\
&= s_{\sigma^2} + (m_{\sigma^2}-\mu_{\sigma^2})^2.
\end{aligned} \tag{A24}$$

Finally, concatenating the results we have obtained the following approximation of the variational free energy functional as a real-valued function of the variational parameters (ref. Eq. (77) of the main text)

$$\begin{aligned}
F(m_\theta, S_\theta, m_{\sigma^2}, s_{\sigma^2}) = &-\frac{n}{2}\ln 2\pi - \frac{n}{2}m_{\sigma^2} - \frac{1}{2}\exp\left(-m_{\sigma^2}+\frac{1}{2}s_{\sigma^2}\right)\left((y-h(m_\theta))^T(y-h(m_\theta)) + tr\left(J^h(m_\theta)^T J^h(m_\theta)^T S_\theta\right)\right) \\
&-\frac{p}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_\theta| - \frac{1}{2}\left(tr\left(\Sigma_\theta^{-1}S_\theta\right) + (m_\theta-\mu_\theta)^T\Sigma_\theta^{-1}(m_\theta-\mu_\theta)\right) - \frac{1}{2}\ln 2\pi\varsigma_{\sigma^2} - m_{\sigma^2} - \frac{1}{2\varsigma_{\sigma^2}}\left(s_{\sigma^2} + (m_{\sigma^2}-\mu_{\sigma^2})^2\right) \\
&+\frac{1}{2}\ln|S_\theta| + \frac{p}{2}\ln(2\pi e) + \frac{1}{2} + \frac{1}{2}\ln(2\pi s_{\sigma^2}) + m_{\sigma^2}.
\end{aligned} \tag{A25}$$

Appendix B

Evaluation of the first partial derivative of F with respect to S_θ yields

$$\begin{aligned}\frac{\partial}{\partial S_\theta}F(m_\theta,S_\theta,m_{\sigma^2},s_{\sigma^2}) &= -\frac{1}{2}\exp\left(-m_{\sigma^2}+\frac{1}{2}s_{\sigma^2}\right)\frac{\partial}{\partial S_\theta}tr\left(J^h(m_\theta)^TJ^h(m_\theta)S_\theta\right)\\ &\quad -\frac{1}{2}\frac{\partial}{\partial S_\theta}tr\left(\Sigma_\theta^{-1}S_\theta\right)+\frac{1}{2}\frac{\partial}{\partial S_\theta}\ln|S_\theta|\\ &= -\frac{1}{2}\exp\left(-m_{\sigma^2}+\frac{1}{2}s_{\sigma^2}\right)J^h(m_\theta)^TJ^h(m_\theta)-\frac{1}{2}\Sigma_\theta^{-1}\\ &\quad +\frac{1}{2}S_\theta^{-1}\end{aligned} \tag{B1}$$

Setting this derivative to S_θ to zero and solving for $S_\theta^{(i+1)}$ then yields the update rule for the variational covariance parameter as follows:

$$\begin{aligned}\frac{\partial}{\partial S_\theta}F\left(m_\theta,S_\theta^{(i+1)},m_{\sigma^2},s_{\sigma^2}\right) &= 0\Leftrightarrow -\frac{1}{2}\exp\left(-m_{\sigma^2}+\frac{1}{2}s_{\sigma^2}\right)J^h(m_\theta)^TJ^h(m_\theta)\\ &\quad -\frac{1}{2}\Sigma_\theta^{-1}+\frac{1}{2}\left(S_\theta^{(i+1)}\right)^{-1}=0\Leftrightarrow\left(S_\theta^{(i+1)}\right)^{-1}\\ &= \Sigma_\theta^{-1}+\exp\left(-m_{\sigma^2}+\frac{1}{2}s_{\sigma^2}\right)J^h(m_\theta)^TJ^h(m_\theta)\Leftrightarrow S_\theta^{(i+1)}\\ &= \left(\Sigma_\theta^{-1}+\exp\left(-m_{\sigma^2}+\frac{1}{2}s_{\sigma^2}\right)J^h(m_\theta)\,J^h(m_\theta)^T\right)^{-1}\end{aligned} \tag{B2}$$

Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2016.04.025.

References

Arand, C., Scheller, E., Seeber, B., Timmer, J., Klöppel, S., Schelter, B., 2015. Assessing parameter identifiability for dynamic causal modeling of fMRI data. Front. Neurosci. 9, 43. http://dx.doi.org/10.3389/fnins.2015.00043.

Ashburner, J., 2012. SPM: a history. NeuroImage 62 (2), 791–800. http://dx.doi.org/10.1016/j.neuroimage.2011.10.025.

Auksztulewicz, R., Blankenburg, F., 2013. Subjective rating of weak tactile stimuli is parametrically encoded in event-related potentials. J. Neurosci. 33 (29), 11878–11887. http://dx.doi.org/10.1523/JNEUROSCI.4243-12.2013.

Auksztulewicz, R., Spitzer, B., Blankenburg, F., 2012. Recurrent neural processing and somatosensory awareness. J. Neurosci. 32 (3), 799–805. http://dx.doi.org/10.1523/JNEUROSCI.3974-11.2012.

Bernardo, J.M., Smith, A.F.M., 1994. Bayesian Theory. John Wiley & Sons Canada, Limited.

Boly, M., Garrido, M.I., Gosseries, O., Bruno, M.-A., Boveroux, P., Schnakers, C., ... Friston, K., 2011. Preserved feedforward but impaired top-down processes in the vegetative state. Science (New York, N.Y.) 332 (6031), 858–862. http://dx.doi.org/10.1126/science.1202043.

Boly, M., Moran, R., Murphy, M., Boveroux, P., Bruno, M.-A., Noirhomme, Q., ... Friston, K., 2012. Connectivity changes underlying spectral EEG changes during propofol-induced loss of consciousness. J. Neurosci. 32 (20), 7082–7090. http://dx.doi.org/10.1523/JNEUROSCI.3769-11.2012.

Boyd, S., Vandenberghe, L., 2004. Convex Optimization. Cambridge University Press, Cambridge, UK; New York.

Brown, H.R., Friston, K.J., 2012. Dynamic causal modelling of precision and synaptic gain in visual perception — an EEG study. NeuroImage 63 (1), 223–231. http://dx.doi.org/10.1016/j.neuroimage.2012.06.044.

Chappell, M.A., Groves, A.R., Whitcher, B., Woolrich, M.W., 2009. Variational Bayesian inference for a nonlinear forward model. IEEE Trans. Signal Process. 57 (1), 223–236. http://dx.doi.org/10.1109/TSP.2008.2005752.

Chumbley, J.R., Friston, K.J., Fearn, T., Kiebel, S.J., 2007. A Metropolis–Hastings algorithm for dynamic causal models. NeuroImage 38 (3), 478–487. http://dx.doi.org/10.1016/j.neuroimage.2007.07.028.

Conn, A.R., Gould, N.I.M., Toint, P.L., 1987. Trust Region Methods. Society for Industrial and Applied Mathematics, Philadelphia, PA.

Cooray, G., Garrido, M.I., Hyllienmark, L., Brismar, T., 2014. A mechanistic model of mismatch negativity in the ageing brain. Clin. Neurophysiol. 125 (9), 1774–1782. http://dx.doi.org/10.1016/j.clinph.2014.01.015.

Cooray, G.K., Garrido, M.I., Brismar, T., Hyllienmark, L., 2015. The maturation of mismatch negativity networks in normal adolescence. Clin. Neurophysiol. http://dx.doi.org/10.1016/j.clinph.2015.06.026.

Cooray, G.K., Sengupta, B., Douglas, P., Friston, K., 2016. Dynamic causal modelling of electrographic seizure activity using Bayesian belief updating. NeuroImage 125, 1142–1154. http://dx.doi.org/10.1016/j.neuroimage.2015.07.063.

Daunizeau, J., Friston, K.J., Kiebel, S.J., 2009. Variational Bayesian identification and prediction of stochastic nonlinear dynamic causal models. Phys. D Nonlinear Phenom. 238 (21), 2089–2118. http://dx.doi.org/10.1016/j.physd.2009.08.002.

Daunizeau, J., David, O., Stephan, K.E., 2011. Dynamic causal modelling: a critical review of the biophysical and statistical foundations. NeuroImage 58 (2), 312–322. http://dx.doi.org/10.1016/j.neuroimage.2009.11.062.

Daunizeau, J., Stephan, K.E., Friston, K.J., 2012. Stochastic dynamic causal modelling of fMRI data: should we care about neural noise? NeuroImage 62 (1), 464–481. http://dx.doi.org/10.1016/j.neuroimage.2012.04.061.

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. http://dx.doi.org/10.1016/j.neuroimage.2004.12.030.

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. http://dx.doi.org/10.1016/j.neuroimage.2005.10.045.

David, O., Guillemain, I., Saillet, S., Reyt, S., Deransart, C., Segebarth, C., Depaulis, A., 2008. Identifying neural drivers with functional MRI: an electrophysiological validation. PLoS Biol. 6 (12), 2683–2697. http://dx.doi.org/10.1371/journal.pbio.0060315.

Dietz, M.J., Friston, K.J., Mattingley, J.B., Roepstorff, A., Garrido, M.I., 2014. Effective connectivity reveals right-hemisphere dominance in audiospatial perception: implications for models of spatial neglect. J. Neurosci. 34 (14), 5003–5011. http://dx.doi.org/10.1523/JNEUROSCI.3765-13.2014.

Erneux, T., 2009. Applied Delay Differential Equations (2009. Aufl.). Springer, New York.

Feynman, 1998. Statistical Mechanics: a Set of Lectures (RevisEd.). Westview Press, Boulder, Colo.

Fogelson, N., Litvak, V., Peled, A., Fernandez-del-Olmo, M., Friston, K., 2014. The functional anatomy of schizophrenia: a dynamic causal modeling study of predictive coding. Schizophr. Res. 158 (1–3), 204–212. http://dx.doi.org/10.1016/j.schres.2014.06.011.

Frässle, S., Stephan, K.E., Friston, K.J., Steup, M., Krach, S., Paulus, F.M., Jansen, A., 2015. Test–retest reliability of dynamic causal modeling for fMRI. NeuroImage 117, 56–66. http://dx.doi.org/10.1016/j.neuroimage.2015.05.040.

Friston, K.J., 2011. Functional and effective connectivity: a review. Brain Connect. 1 (1), 13–36. http://dx.doi.org/10.1089/brain.2011.0008.

Friston, K.J., Dolan, R.J., 2010. Computational and dynamic models in neuroimaging. NeuroImage 52 (3), 752–765. http://dx.doi.org/10.1016/j.neuroimage.2009.12.068.

Friston, K.J., Glaser, D.E., Henson, R.N.A., Kiebel, S., Phillips, C., Ashburner, J., 2002a. Classical and Bayesian inference in neuroimaging: applications. NeuroImage 16 (2), 484–512. http://dx.doi.org/10.1006/nimg.2002.1091.

Friston, K.J., Penny, W., Phillips, C., Kiebel, S., Hinton, G., Ashburner, J., 2002b. Classical and Bayesian inference in neuroimaging: theory. NeuroImage 16 (2), 465–483. http://dx.doi.org/10.1006/nimg.2002.1090.

Friston, K.J., Harrison, L., Penny, W., 2003. Dynamic causal modelling. NeuroImage 19 (4), 1273–1302.

Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J., Penny, W., 2007. Variational free energy and the Laplace approximation. NeuroImage 34 (1), 220–234. http://dx.doi.org/10.1016/j.neuroimage.2006.08.035.
Friston, K.J., Trujillo-Barreto, N., Daunizeau, J., 2008. DEM: a variational treatment of dynamic systems. NeuroImage 41 (3), 849–885. http://dx.doi.org/10.1016/j.neuroimage.2008.02.054.
Friston, K., Stephan, K., Li, B., Daunizeau, J., 2010. Generalised filtering. Math. Probl. Eng. 2010 http://dx.doi.org/10.1155/2010/621670.
Garrido, M.I., Kilner, J.M., Kiebel, S.J., Friston, K.J., 2007a. Evoked brain responses are generated by feedback loops. Proc. Natl. Acad. Sci. U. S. A. 104 (52), 20961–20966. http://dx.doi.org/10.1073/pnas.0706274105.
Garrido, M.I., Kilner, J.M., Kiebel, S.J., Stephan, K.E., Friston, K.J., 2007b. Dynamic causal modelling of evoked potentials: a reproducibility study. NeuroImage 36 (3), 571–580. http://dx.doi.org/10.1016/j.neuroimage.2007.03.014.
Garrido, M.I., Friston, K.J., Kiebel, S.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 (2), 936–944. http://dx.doi.org/10.1016/j.neuroimage.2008.05.018.
Garrido, M.I., Kilner, J.M., Stephan, K.E., Friston, K.J., 2009. The mismatch negativity: a review of underlying mechanisms. Clin. Neurophysiol. 120 (3), 453–463. http://dx.doi.org/10.1016/j.clinph.2008.11.029.
Grech, R., Cassar, T., Muscat, J., Camilleri, K.P., Fabri, S.G., Zervakis, M., ... Vanrumste, B., 2008. Review on solving the inverse problem in EEG source analysis. J. Neuroeng. Rehabil. 5, 25. http://dx.doi.org/10.1186/1743-0003-5-25.
Griewank, A., Walther, A., 2008. Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation, Second Edition Second Edition by Griewank, Andreas, Walther, Andrea (2008) Paperback. Manchester University Press.
Grimbert, F., Faugeras, O., 2006. Bifurcation analysis of Jansen's neural mass model. Neural Comput. 18 (12), 3052–3068. http://dx.doi.org/10.1162/neco.2006.18.12.3052.
Hallez, H., Vanrumste, B., Grech, R., Muscat, J., De Clercq, W., Vergult, A., ... Lemahieu, I., 2007. Review on solving the forward problem in EEG source analysis. J. Neuroen. Rehabil. 4, 46. http://dx.doi.org/10.1186/1743-0003-4-46.
Hennig, P., Osborne, M.A., Girolami, M., 2015. Probabilistic numerics and uncertainty in computations arXiv:1506.01326 [cs, math, stat]. Abgerufen von http://arxiv.org/abs/1506.01326.
Honkela, A., Raiko, T., Kuusela, M., Tornio, M., & Karhunen, J. Approximate Riemannian conjugate gradient learning for fixed-form variational Bayes. J. Mach. Learn. Res., 2010 (o. J.).
Ilmoniemi, R.J., 1993. Models of source currents in the brain. Brain Topogr. 5 (4), 331–336.
in't Hout, K.J., 1996. On the stability of adaptations of Runge–Kutta methods to systems of delay differential equations. Appl. Numer. Math. 22 (1–3), 237–250. http://dx.doi.org/10.1016/S0168-9274(96)00035-9.
Jansen, B.H., Rit, V.G., 1995. Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. Biol. Cybern. 73 (4), 357–366.
Jirsa, V.K., 2009. Neural field dynamics with local and global connectivity and time delay. Philos. Transact. A Math. Phys. Eng. Sci. 367 (1891), 1131–1143. http://dx.doi.org/10.1098/rsta.2008.0260.
Jordan, M.I., Ghahramani, Z., Jaakkola, T.S., Saul, L.K., 1999. An introduction to variational methods for graphical models. Mach. Learn. 37 (2), 183–233. http://dx.doi.org/10.1023/A:1007665907178.
Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90 (430), 773–795. http://dx.doi.org/10.1080/01621459.1995.10476572.
Kass, R.E., Wasserman, L., 1996. The selection of prior distributions by formal rules. J. Am. Stat. Assoc. 91 (435), 1343–1370. http://dx.doi.org/10.1080/01621459.1996.10477003.
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. http://dx.doi.org/10.1016/j.neuroimage.2005.12.055.
Kiebel, S.J., Garrido, M.I., Moran, R.J., Friston, K.J., 2008. Dynamic causal modelling for EEG and MEG. Cogn. Neurodyn. 2 (2), 121–136. http://dx.doi.org/10.1007/s11571-008-9038-0.
Klingner, C.M., Brodoehl, S., Huonker, R., Götz, T., Baumann, L., Witte, O.W., 2015. Parallel processing of somatosensory information: evidence from dynamic causal modeling of MEG data. NeuroImage 118, 193–198. http://dx.doi.org/10.1016/j.neuroimage.2015.06.028.
Lindley, D.V., 1987. Bayesian Statistics, a Review. Society for Industrial and Applied Mathematics, Philadelphia.
Lindley, D.V., 2000. The philosophy of statistics. J. R. Stat. Soc. D 49 (3), 293–337. http://dx.doi.org/10.1111/1467-9884.00238.
Litvak, V., Mattout, J., Kiebel, S., Phillips, C., Henson, R., Kilner, J., ... Friston, K., 2011. EEG and MEG data analysis in SPM8. Comput. Intell. Neurosc. 2011, 852961. http://dx.doi.org/10.1155/2011/852961.
Lomakina, E.I., Paliwal, S., Diaconescu, A.O., Brodersen, K.H., Aponte, E.A., Buhmann, J.M., Stephan, K.E., 2015. Inversion of hierarchical Bayesian models using Gaussian processes. NeuroImage 118, 133–145. http://dx.doi.org/10.1016/j.neuroimage.2015.05.084.
Luck, S.J., 2014. An Introduction to the Event-Related Potential Technique. second ed. The Mit Press, Cambridge, Massachusetts.
Magnus, J.R., 1999. Matrix Differential Calc with Apps Rev (2. Auflage). Wiley, New York.
Moran, R.J., Stephan, K.E., Kiebel, S.J., Rombach, N., O'Connor, W.T., Murphy, K.J., ... Friston, K.J., 2008. Bayesian estimation of synaptic physiology from the spectral responses of neural masses. NeuroImage 42 (1), 272–284. http://dx.doi.org/10.1016/j.neuroimage.2008.01.025.
Moran, R.J., Jung, F., Kumagai, T., Endepols, H., Graf, R., Dolan, R.J., ... Tittgemeyer, M., 2011a. Dynamic causal models and physiological inference: a validation study using isoflurane anaesthesia in rodents. PLoS One 6 (8), e22790. http://dx.doi.org/10.1371/journal.pone.0022790.
Moran, R.J., Symmonds, M., Stephan, K.E., Friston, K.J., Dolan, R.J., 2011b. An in vivo assay of synaptic function mediating human cognition. Curr. Biol. 21 (15), 1320–1325. http://dx.doi.org/10.1016/j.cub.2011.06.053.
Moran, R., Pinotsis, D.A., Friston, K., 2013. Neural masses and fields in dynamic causal modeling. Front. Comput. Neurosci. 7, 57. http://dx.doi.org/10.3389/fncom.2013.00057.
Neal, R., Hinton, G.E., 1998. A view of the EM algorithm that justifies incremental, sparse, and other variants. Learning in Graphical Models (S. 355–368). Kluwer Academic Publishers.
Nocedal, J., Wright, S., 2006. Numerical Optimization. second ed. Springer, New York.
Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.-M., 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci. 2011, 156869. http://dx.doi.org/10.1155/2011/156869.
Ostwald, D., Spitzer, B., Guggenmos, M., Schmidt, T.T., Kiebel, S.J., Blankenburg, F., 2012. Evidence for neural encoding of Bayesian surprise in human somatosensation. NeuroImage 62 (1), 177–188. http://dx.doi.org/10.1016/j.neuroimage.2012.04.050.
Ostwald, D., Kirilina, E., Starke, L., Blankenburg, F., 2014. A tutorial on variational Bayes for latent linear stochastic time-series models. J. Math. Psychol. 60, 1–19. http://dx.doi.org/10.1016/j.jmp.2014.04.003.
Ozaki, T., 1992. A bridge between nonlinear time series models and nonlinear stochastic dynamical systems: a local linearization approach. Stat. Sin. 2 (1), 113–135.
Penny, W.D., 2012. Comparing dynamic causal models using AIC, BIC and free energy. NeuroImage 59 (1), 319–330. http://dx.doi.org/10.1016/j.neuroimage.2011.07.039.
Petersen, K.B., Pedersen, M.S., Larsen, J., Strimmer, K., Christiansen, L., Hansen, K., ... The, W., 2006. The Matrix Cookbook.
Pinotsis, D.A., Schwarzkopf, D.S., Litvak, V., Rees, G., Barnes, G., Friston, K.J., 2013. Dynamic causal modelling of lateral interactions in the visual cortex. NeuroImage 66, 563–576. http://dx.doi.org/10.1016/j.neuroimage.2012.10.078.
Schmidt, A., Diaconescu, A.O., Kometer, M., Friston, K.J., Stephan, K.E., Vollenweider, F.X., 2013. Modeling ketamine effects on synaptic plasticity during the mismatch negativity. Cereb. Cortex 23 (10), 2394–2406. http://dx.doi.org/10.1093/cercor/bhs238.
Schomer, D.L., da Silva, F.L., 2011. Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields. 6th Revised ed. Lippincott Williams&Wilki.
Sengupta, B., Friston, K.J., Penny, W.D., 2014. Efficient gradient computation for dynamical models. NeuroImage 98, 521–527. http://dx.doi.org/10.1016/j.neuroimage.2014.04.040.
Sengupta, B., Friston, K.J., Penny, W.D., 2016. Gradient-based MCMC samplers for dynamic causal modelling. NeuroImage 125, 1107–1118.
Shampine, L.F., Thompson, S., 2001. Solving DDEs in Matlab. Appl. Numer. Math. 37 (4), 441–458. http://dx.doi.org/10.1016/S0168-9274(00)00055-6.
Sharaev, M.G., Mnatsakanian, E.V., 2014. Dynamic causal modeling of brain electrical responses elicited by simple stimuli in visual oddball paradigm. Zh. Vyssh. Nerv. Deiat. Im. I. P. Pavlova 64 (6), 627–638.
Spiegler, A., Kiebel, S.J., Atay, F.M., Knösche, T.R., 2010. Bifurcation analysis of neural mass models: impact of extrinsic inputs and dendritic time constants. NeuroImage 52 (3), 1041–1058. http://dx.doi.org/10.1016/j.neuroimage.2009.12.081.
Stephan, K.E., Roebroeck, A., 2012. A short history of causal modeling of fMRI data. NeuroImage 62 (2), 856–863. http://dx.doi.org/10.1016/j.neuroimage.2012.01.034.
Stephan, K., Friston, K., Penny, W., 2005. Computing the objective function in DCM available from http://www.fil.ion.ucl.ac.uk/spm/doc/papers/stephan_DCM_ObjFcn_tr05.pdf (o. J.).
Stephan, K.E., Penny, W.D., Moran, R.J., den Ouden, H.E.M., Daunizeau, J., Friston, K.J., 2010. Ten simple rules for dynamic causal modeling. NeuroImage 49 (4), 3099–3109. http://dx.doi.org/10.1016/j.neuroimage.2009.11.015.
Thoai, N.V., Horst, R., Pardalos, P.M., 2008. Introduction to Global Optimization (1995. Aufl.). Springer.
Wipf, D., Nagarajan, S., 2009. A unified Bayesian framework for MEG/EEG source imaging. NeuroImage 44 (3), 947–966. http://dx.doi.org/10.1016/j.neuroimage.2008.02.059.