Journal of Neuroscience Methods 269 (2016) 6–20

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

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

A hierarchical model for integrating unsupervised generative embedding and empirical Bayes

Sudhir Raman[a,*], Lorenz Deserno[b,c,d], Florian Schlagenhauf[b,c], Klaas Enno Stephan[a,e,f]

[a] Translational Neuromodeling Unit (TNU), Institute for Biomedical Engineering, University of Zurich and ETH Zurich, Switzerland
[b] Department of Psychiatry and Psychotherapy, Charité Universitätsmedizin Berlin, Berlin, Germany
[c] Max Planck Fellow Group "Cognitive and Affective Control of Behavioral Adaptation", Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany
[d] Department of Neurology, Otto-von-Guericke University, Magdeburg, Germany
[e] Wellcome Trust Centre for Neuroimaging, University College London, UK
[f] Max Planck Institute for Metabolism Research, Cologne, Germany

HIGHLIGHTS

- A novel unified hierarchical framework for DCM is presented.
- Simultaneous parameter inference, unsupervised learning and empirical Bayes.
- MCMC sampling for inference.
- Improved model evidence over non-hierarchical DCM.

ARTICLE INFO

Article history:
Received 9 November 2015
Received in revised form 23 April 2016
Accepted 24 April 2016
Available online 30 April 2016

Keywords:
Clustering
Dynamic causal modelling
DCM
Markov chain Monte Carlo sampling
MCMC
Mixture model
Psychiatric spectrum diseases
Schizophrenia

ABSTRACT

Background: Generative models of neuroimaging data, such as dynamic causal models (DCMs), are commonly used for inferring effective connectivity from individual subject data. Recently introduced "generative embedding" approaches have used DCM-based connectivity parameters for supervised classification of individual patients or to find unknown subgroups in heterogeneous groups using unsupervised clustering methods.

New method: We present a novel framework which combines DCMs with finite mixture models into a single hierarchical model. This approach unifies the inference of connectivity parameters in individual subjects with inference on population structure, i.e. the existence of subgroups defined by model parameters, and allows for empirical Bayesian estimates of a subject's connectivity based on subgroup-specific prior distributions. We introduce a Markov chain Monte Carlo sampling method for inversion of this hierarchical generative model.

Results: This paper formally introduces the idea behind our novel concept and demonstrates the face validity of the model in application to both simulated data as well as an empirical fMRI dataset from healthy controls and patients with schizophrenia.

Comparison with existing method(s): The analysis of our empirical fMRI data demonstrates that our approach results in superior model evidence than the conventional non-hierarchical inversion of DCMs.

Conclusions: In this paper, we have presented a novel unified framework to jointly infer the effective connectivity parameters in DCMs for multiple subjects and, at the same time, discover connectivity-defined cluster structure of the whole population, using a mixture model approach.

1. Introduction

Over the past two decades, neuroimaging has had tremendous success in elucidating principles and mechanisms of cognition. While these insights were mainly derived from group analyses of

* Corresponding author at: Translational Neuromodeling Unit (TNU), Institute of Biomedical Engineering, University of Zurich & Swiss Federal Institute of Technology (ETH Zurich), Wilfriedstr. 6, 8032 Zurich, Switzerland. Fax: +41 44 634 9131.
E-mail address: ssudhir@ethz.ch (S. Raman).

http://dx.doi.org/10.1016/j.jneumeth.2016.04.022
0165-0270/© 2016 Elsevier B.V. All rights reserved.

healthy volunteers, it has proven far more difficult to establish clinically useful neuroimaging procedures that operate at the level of single subjects (for review, see Klöppel et al., 2012; Orrù et al., 2012; Wolfers et al., 2015). Despite the development of powerful supervised (classification) and unsupervised (clustering) methods for neuroimaging data analysis (e.g. Schrouff et al., 2013), few practical applications have managed to enter clinical practice. These successful examples can mainly be found in neurology where, for example, EEG-based classification has enabled brain–computer-interfaces for patients with locked-in syndrome (Sellers et al., 2014) or differential diagnosis of comatose states (Cruse et al., 2011).

By contrast, we presently lack convincing neuroimaging-based clinical tools for diagnosis and prediction in major psychiatric diseases. One major reason for this is that although numerous classification analyses of neuroimaging data, in particular fMRI, exist, these have largely been cross-sectional studies which attempted to differentiate conventionally defined psychiatric patients from healthy volunteers. This, however, ignores the fact that our current disease classifications in psychiatry rest on syndromatic concepts which define diseases as clusters of symptoms over certain periods. These disease definitions likely group together heterogeneous patient groups characterised by a diversity of pathophysiological mechanisms (Casey et al., 2013; Krystal and State, 2014). This probable heterogeneity in terms of disease mechanisms is the reason why many major diseases are increasingly conceptualised as “spectrum” diseases and explains why existing disease definitions in psychiatry lack predictive validity, i.e. assignment to a diagnostic category predicts neither clinical trajectory nor treatment response (Kapur et al., 2012).

As a consequence, a longstanding debate has concerned the question how psychiatric diseases should be redefined. Previous proposals have referred to genetics (Smoller, 2013) or classical cognitive and neuroimaging methods (Cuthbert and Insel, 2013). An alternative “translational neuromodelling” strategy promotes the use of neurocomputational models for discovery of mechanistically more homogenous patient groups (Stephan and Mathys, 2014; Stephan et al., 2015). Here, model-based estimates of patient-specific disease mechanisms, obtained from individual neuroimaging and/or behavioural data, are used as a basis for splitting a heterogeneous spectrum disease into subgroups or defining general dimensions of a disease across the spectrum.

A practical strategy for implementing this general idea is “generative embedding” (Brodersen et al., 2011, 2014). This approach rests on using generative models of neuroimaging data and behaviour to estimate parameters encoding subject-specific mechanisms underlying the individual measurements. The ensuing parameter estimates serve to define a feature space as a basis for subsequent supervised (classification) or unsupervised (clustering) learning. This approach has two major strengths. First, the generative model serves as an informed dimensionality reduction device, providing a compact summary of how data are generated and removing uninformative noise. Second, the resulting classes or clusters have mechanistic interpretations because the underlying feature space is directly connected to the model. The advantages of generative embedding have been demonstrated in two fMRI studies where generative embedding (based on dynamic causal models, DCMs, of fMRI data) not only provided a more mechanistic interpretation of classification/clustering results compared to conventional approaches based on local activation measures or functional connectivity, but also demonstrated significantly superior performance for classifying/clustering patients with stroke (Brodersen et al., 2011) and schizophrenia (Brodersen et al., 2014), respectively.

So far, generative embedding analyses of neuroimaging data have employed a two-step procedure where a given DCM was initially fitted to data from each individual separately and the resulting subject-specific parameter vectors were subsequently fed into a classification (support vector machine, Brodersen et al., 2011) or clustering (Gaussian mixture model, Brodersen et al., 2014) procedure. In this article, we extend the scope of generative embedding for unsupervised clustering and present a novel hierarchical generative model which can be applied to data from all individuals at once. This approach allows for simultaneous inference on individual parameter estimates and group structure (number and composition of clusters). An important motivation for this extension is that in our unified hierarchical model, these two aspects are allowed to interact. That is, the hierarchical structure of our model allows for an empirical Bayesian type of inference, where subgroup-specific priors are estimated from the group data and inform parameter estimates of individual subjects; conversely, definition of subgroups (clustering) is informed by parameter estimates across subjects. This mutual dependency between finding subgroups in the sample and regularisation by subgroup-dependent priors is a novel concept with potential for future clinical applications.

The goal of this paper is to introduce the general idea behind this novel concept and to demonstrate its practical feasibility. By contrast, it does not aim for demonstration of its general superiority compared to established methods, nor does the model formulation presented in this paper already address all facets of the general problem of combining hierarchical inference and subgroup detection (see Section 4).

This paper is structured as follows. In Section 2, we first provide a summary of DCM for fMRI, followed by an outline of how single subject inference (parameter estimation) can be achieved by Markov chain Monte Carlo (MCMC). This serves to introduce the notation which is then used for defining a novel multi-subject hierarchical model for joint parameter estimation and clustering. We conclude this section by outlining our approach to inference based on an MCMC algorithm. In Section 3, we present two application examples. First, we demonstrate the face validity of our models by applying it to simulated fMRI data. In a second application to empirical fMRI data from patients with schizophrenia and healthy controls, we find that our method has comparable performance in separating patients and controls, in an unsupervised way, as previous analyses. Importantly, model comparison demonstrates that the full hierarchical model (combining empirical Bayes and clustering) outperforms a model using empirical Bayes alone; furthermore, the latter is superior to the conventional DCM approach which considers each subject in isolation. In Section 4, we evaluate the pros and cons of our method compared to “classical” generative embedding, consider current limitations of our method, and outline further developments for the future.

2. Methods

2.1. Dynamic causal modelling

Dynamic causal models (DCM) are generative models of neuroimaging data like fMRI (Friston et al., 2003) or EEG (David et al., 2006). They usually serve to infer on the effective connectivity between neuronal populations and are applied to single-subject data. Their structure consists of two hierarchical layers, a model of hidden neuronal states and an observation or forward model. At the neuronal level, differential equations describe a dynamical system of interacting brain regions; the forward model describes how the ensuing neuronal states give rise to observed measurements, e.g.

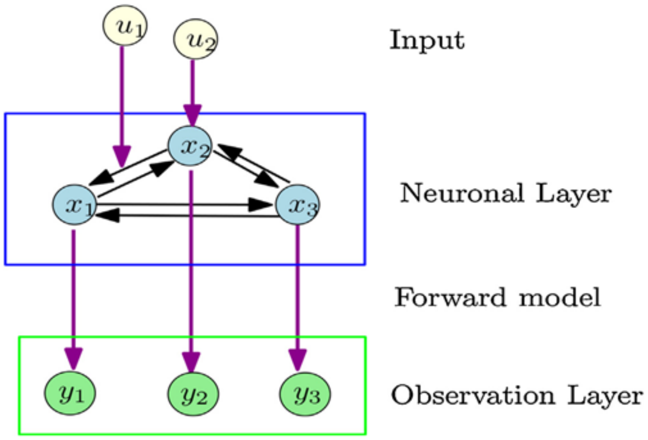


Fig. 1. An abstract graphical representation of a single-subject DCM where x denotes neuronal states, y refers to region-specific measurements, and u represent the inputs to the system.

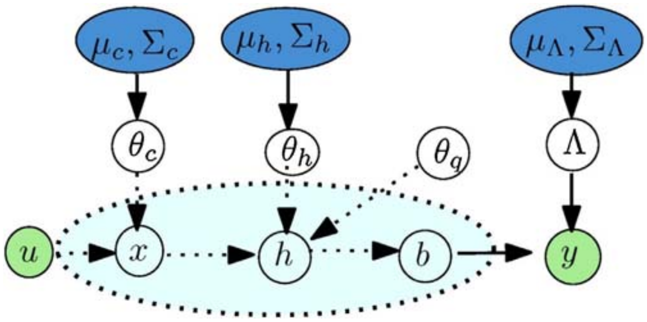


Fig. 2. A probabilistic graphical model for the generation of BOLD signals from a DCM. The dotted lines indicate a deterministic relationship whereas the solid arrows indicate a probabilistic dependency. The intermediate dotted region represents the function which is derived by integrating the dynamical system described in the text.

electrical signals measured at EEG sensors or blood oxygen level dependent (BOLD) signals measured by fMRI. This forward model takes into account the biophysical foundations of the respective imaging modality as well as measurement noise.

In contrast to statistical characterisations of measured data, DCM thus describes (unobservable) neuronal processes underlying the observed measurements. Although the mathematical equations represent a highly simplified model of neuronal dynamics, they provide a mechanistic perspective on how measured data were generated, with estimates of neurobiologically interpretable quantities such as the effective connection strength between distinct neuronal populations. An example of a single-subject DCM with three regions or nodes is given in Fig. 1. The neuronal layer describes the dynamics between the three regions, and the forward model captures the generation of observations given the neuronal states of each region.

The models described in this article will focus entirely on fMRI data where the forward model describes the generation of BOLD signals (Friston et al., 2000; Stephan et al., 2007). The full generative model for an fMRI-DCM can be summarised as follows:

$$\begin{aligned} \theta_{\mathbf{c}} &\sim \mathrm{Normal}(\mu_{\mathbf{c}}, \Sigma_c) \\ \theta_{\mathbf{h}} &\sim \mathrm{logNormal}(\mu_{\mathbf{h}}, \Sigma_{\mathrm{h}}) \\ \frac{\mathrm{dx}}{\mathrm{dt}} &= f_1\left(\mathrm{x}, \theta_{\mathbf{c}}, \mathbf{u}\right) \\ \frac{\mathrm{dh}}{\mathrm{dt}} &= f_2\left(\mathrm{h}, \mathbf{x}, \theta_{\mathbf{q}}, \theta_{\mathbf{h}}\right) \\ \mathbf{b} &= g\left(\mathrm{h}, \theta_{\mathbf{h}}, \theta_{\mathbf{q}}\right) \\ \lambda &\sim \mathrm{logNormal}(\mu_{\Lambda}, \Sigma_{\Lambda}) \\ \mathbf{y} &\sim \mathrm{Normal}\left(\mathrm{b}, \Lambda^{-1}\right). \end{aligned} \tag{1}$$

Here, θ_c are parameters of the neuronal layer (connectivity and input strengths), θ_h are the forward hemodynamic model parameters, θ_q are biophysical constants of the forward model, $\mathbf{u}$ are external (e.g. sensory) inputs over time, $(\mu_{\Lambda}, \Sigma_{\Lambda})$ are the parameters of the Normal distribution, $\mathbf{y}$ represents measured data, and Λ is the precision matrix of observation noise (with hyperparameters λ, described below). Where it is necessary to restrict parameter values to positive values, the prior normal distributions are over the log of the parameters rather than the parameters themselves. The deterministic dynamical system, represented by $f_1(.)$, models the hidden dynamics of the neuronal layer, where $\mathbf{x}$ denotes the neuronal state variables (e.g. one per region or neuronal population), and $\mathbf{h}$ denotes the state variables of the hemodynamic forward model.

The neuronal state equation $f_1(.)$ can be derived from a low-order Taylor series approximation to any nonlinear dynamical system (where the expansion point is at rest, i.e. $x = 0$, $u = 0$). For example, using an approximation to second order yields (Stephan et al., 2008):

$$\frac{\mathrm{dx}}{\mathrm{d}t} = A\mathbf{x} + \sum_{m=1}^{M} u_m B_m \mathbf{x} + C\mathbf{u} + \sum_{r=1}^{R} x_r D_r \mathbf{x}, \tag{2}$$

where M is the number of inputs, R is the number of regions, A describes connections between regions, B describes modulatory effects of inputs on connections between regions, C describes the driving effect of inputs on regions, and D describes the modulatory effect of regional activity on connections. Linear DCMs use A and C matrices only, bilinear DCMs contain a non-zero B matrix, and non-linear DCMs contain a non-zero D matrix.

The hemodynamic model has separate parameters for each region in the model. This is important as it allows for taking into account regional variations in the hemodynamic response function (cf. David et al., 2008). The version presently used in DCM rests on the so-called “Balloon model” initially proposed by Buxton et al. (1998) and later extended by Friston et al. (2000) and Stephan et al. (2007). Details of the hemodynamic equations used in the present paper can be found in Stephan et al. (2007); a brief summary is provided in the Appendix A. The function g(·) is a nonlinear static output function operating on states (volume and deoxyhemoglobine content) provided by the nonlinear differential equations of hemodynamics represented by $f_2(.)$. Since it is not analytically feasible to solve these differential equations, suitable numerical integration schemes (like the matrix exponential under bilinear approximations or Euler integration) are used in practice (cf. Friston, 2002; Daunizeau et al., 2014).

The graphical model for a single subject DCM is given in Fig. 2. The corresponding joint distribution of parameters, hyperparameters and data from R regions is as follows:

$$\begin{aligned} p\left(\mathrm{y}, \theta_{\mathbf{c}}, \theta_{\mathbf{h}}, \theta_{\mathbf{q}}, \Lambda\right) \propto\ & \mathrm{Normal}(\mathrm{y}\,|\,g\left(\theta_{\mathbf{c}}, \theta_{\mathbf{h}}, \theta_{\mathbf{q}}\right), \Lambda^{-1}) \\ & \mathrm{Normal}(\theta_{\mathbf{c}}|\mu_{\mathbf{c}}, \Sigma_c)\mathrm{logNormal}(\theta_{\mathbf{h}}|\mu_{\mathbf{h}}, \Sigma_{\mathrm{h}}) \\ & \prod_{r=1}^{R} \mathrm{logNormal}(\lambda_r|\mu_{\Lambda}, \sigma_{\Lambda}^2). \end{aligned} \tag{3}$$

Table 1
Priors over connection and hemodynamic parameters for a single-subject DCM. The hemodynamic parameters κ, τ, ε are described in detail in the Appendix A.

Parameters	Mean	Variance	log-mean	log-var
A (self-connections)	−0.5	1/(8R)		
A (other connections)	1/(64R)	8/R		
B	0	1		
C	0	1		
D	0	1		
κ			log(0.64)	0.0025
τ			log(2)	0.0025
ε			0	0.0025
λ_r			0	1

2.2. Priors over model parameters

There are three sets of parameters in the model, the neuronal parameters $\square_c = (A,B,C,D)$, the hemodynamic parameters $\square_h = (\kappa,\tau,\varepsilon)$, and the hyperparameters λ_r of the observation noise precision matrix □ (see Eq. (4) below). For most parameters, the priors are normal distributions. Only in cases where positivity is to be enforced, the parameters follow a log-Normal distribution (see Table 1). For the connection matrices, there are informed priors on self-connections and shrinkage priors on all other connections; please see Table 1 for details. To facilitate comparison with previous work, the priors for noise and hemodynamic parameters were matched to SPM8 release 5236 (http://www.fil.ion.ucl.ac.uk/spm/).

The observation noise precision matrix Λ is represented as a linear combination of predefined matrices (inverse covariance are precision components Q_r) whose contributions are scaled by hyperparameters λ_r. The Q_r's can be defined to account for regional differences in signal variance and to capture temporal autocorrelation (for details, see Friston et al., 2002, 2003). In the current model, we assume that the time series have been whitened (as, for example, can be done automatically in SPM when extracting timeseries for DCM) and only deal with the region-specific variances in BOLD signal. In other words, for a single-subject DCM, we consider a precision (inverse covariance) matrix Λ which has diagonal structure, with region-specific precisions along the diagonal

$$\Lambda = \Sigma_{r=1}^{R} \square_r Q_{r'} \tag{4}$$

where R is the number of regions. Here, each Q_r is simply a diagonal matrix where diagonal elements belonging to region r have the value 1 (and zero elsewhere). The resulting precision matrix Λ is then a diagonal with T repeated values λ_r for each region r, where T is the number of scans. To provide an example, for the specific case of two regions, this matrix is:

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 & \ldots & 0 \\ 0 & \square_1 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & \square_2 & 0 \\ 0 & 0 & 0 & \ldots & \square_2 \end{bmatrix}. \tag{5}$$

2.3. Parameter inference

Parameter estimation in DCMs rests on Bayesian inference (Gelman et al., 1995). Model inversion yields the posterior distribution over the parameters, given the prior distribution and the observations. Using Bayesian inference, we can not only obtain point estimates of parameters, but their full posterior distribution, including estimates of uncertainty and parameter interdependency (posterior covariance). Since the nonlinearities in the state equations above prevent an analytic derivation for the posterior distribution of DCMs, a standard method for inverting DCMs is variational Bayes (VB) under a Laplace approximation (Friston et al., 2003, 2007). While VB is very fast, it is susceptible to local extrema and can be affected by suboptimal assumptions made about the posterior distribution (see the discussion of its use for DCM in Daunizeau et al., 2011 and Lomakina et al., 2015). Furthermore, it can be difficult to derive the required update equations for inference, particularly for complex models as in this paper.

Here, we consider the use of Markov Chain Monte Carlo (MCMC) sampling, based on a Metropolis–Hastings scheme, which involves generating approximate samples from the posterior distribution and then using these samples to estimate the properties of the posterior distribution (for previous applications to DCM, see Chumbley et al., 2007 and Sengupta et al., 2015). For the model specified above, the goal is to obtain the posterior distribution over the (hyper) parameters $\square_c$, $\square_h$ and Λ, where the connection weights $\square_c$ are of primary interest. The following equations summarise MCMC sampling in our context:

$$\begin{aligned} p(\theta_{\mathbf{c}}|\bullet) &\propto p\,\square\, y|g(\bullet), \Lambda^{-1}\,\square\, p(\square_{\mathbf{c}}|\square_{\mathbf{c}}, \Sigma_{\mathrm{c}}) \\ p(\theta_{\mathbf{h}}|\bullet) &\propto p\,\square\, y|g(\bullet), \Lambda^{-1}\,\square\, p(\square_{\mathbf{h}}|\square_{\mathbf{h}}, \Sigma_{\mathrm{h}}) \\ p(\Lambda|\bullet) &\propto p\,\square\, y|g(\bullet), \Lambda^{-1}\,\square\, p(\Lambda|\square_{\square}, \Sigma_{\Lambda}), \end{aligned} \tag{6}$$

where "•" denotes all the remaining variables.

The advantage of using MCMC is that we can avoid distributional assumptions and complex derivations; furthermore, it is asymptotically exact, i.e. as the sample size goes to infinity it converges to the true posterior distribution. On the other hand, MCMC sampling can be computationally expensive and may require a long time to converge.

2.4. A hierarchical model with embedded clustering

Above, we have described the existing framework for single-subject connectivity inference in DCMs. It is possible to extend this framework for multiple subjects using the notion of generative embedding for model-based inference. Generative embedding involves a two-stage process: (1) inference on model parameters and (2) using these parameters to construct a feature space for (un)supervised learning problems like classification or clustering (Brodersen et al., 2011). In essence, this involves embedding the observed data into a parameter space and performing learning in this new feature space. This method comes under the umbrella of emerging hybrid discriminative-generative approaches, as discussed in Doyle et al. (2013).

In this work, we extend the notion of generative embedding and construct a multi-subject model which unifies inference on subject-specific parameters and detection of subgroups in the population. This is done under an empirical Bayesian inversion scheme in which the latent layer of the prior distributions over subject parameters is also estimated. This combines the two-stage process of generative embedding into one unified hierarchical generative model.

Practically, this is achieved by combining the existing framework of DCMs with the framework of finite Gaussian mixture models. This allows for simultaneous inference on the connectivity parameters for each subject, given their individual fMRI measurements, and defines clusters or subgroups of subjects based on their patterns of connectivity parameter estimates. From a mixture-model perspective, the 'observations' are the subject-specific vectors of connectivity parameter estimates. From the perspective of the DCM framework, this requires a re-definition of the connection priors for each subject as a mixture of normal distributions rather than the previously defined normal distribution.

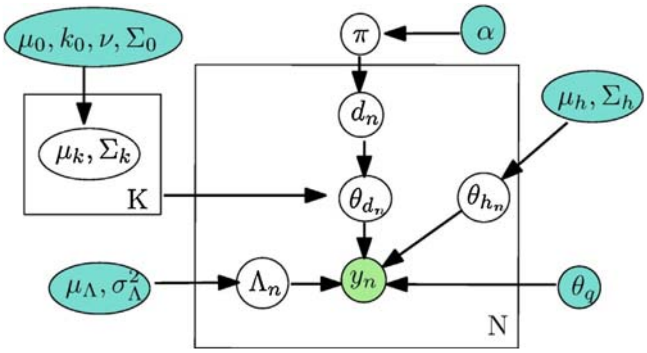


Fig. 3. A unified probabilistic model for joint clustering (with respect to K mixing components) and parameter estimation in N subjects. The cyan shaded regions are fixed hyperparameters or biophysical constants, the green shaded region represents observed data. Square regions with variables indicate multiple copies of the variable.

These parameters further form the basis of a generative model for a DCM through which the truly observed values, like BOLD signals in the case of fMRI data, are generated.

Assuming that the numbers of clusters K are set in advance, the full generative model is specified as follows (see Fig. 3 for a representation as graphical model):

$$
\begin{aligned}
&\Box && \sim \text{Dirichlet}(\alpha,\ \mathbf{M})(\text{mixing probabilities})\\
&\Box\, d_n \Box_{n=1}^{N} && \sim \text{Categorical}(\Box)(\text{cluster assignment index})\\
&\Box\, \mu_{\mathbf{k}}, \Sigma_k \Box_{k=1}^{K} && \sim \text{NormalInvWishart}(\Box_0, \mathrm{k}_0, \nu, \Sigma_0)\ (\text{moments of Gaussians})\\
&\Box\, \theta_{\mathbf{d_n}} \Box_{n=1}^{N} && \sim \text{Normal}(\mu_{\mathbf{d_n}}, \Sigma_{d_n})\ (\text{connection parameters})\\
&\Box\, \theta_{\mathbf{h_n}} \Box_{n=1}^{N} && \sim \text{logNormal}(\mu_{\mathrm{h}}, \Sigma_{\mathrm{h}})(\text{hemodynamic model parameters})\\
&\Box\, \lambda_r \Box_{r=1}^{R} && \sim \text{logNormal}\ \Box\, \mu_\Lambda, \sigma^2_\Lambda \Box\ (\text{noise parameters})\\
&\mathbf{y_n} && \sim \text{Normal}\ \Box\, g\ \Box\, \Box_{\mathrm{d}\,\mathbf{n}}, \Box_{\mathbf{h_n}}, \Box_{\mathbf{q}}\ \Box, \Lambda_n^{-1}\ \Box\ (\text{BOLD signals})
\end{aligned}
\tag{7}
$$

where R and N represent the number of regions and subjects, respectively. $\Box_{\mathbf{d_n}}$ and $\Box_{\mathbf{h_n}}$ are the neuronal and hemodynamic parameters, respectively, for subject n. ($\Box_k$, Σ_k) are the parameters for the k-th cluster whose prior is defined as the conjugate Normal-InverseWishart distribution. $\Box_q$ are hemodynamic constants as defined in the Appendix A. d_n is the n-th subject's cluster assignment which is based on the categorical distribution $\Box$ which describes the mixing proportions of the clusters. The prior over the mixing proportions is a Dirichlet distribution where $\mathbf{M}$ is a categorical distribution and α is a scalar quantity which represents the level of confidence in the prior mean $\mathbf{M}$ of the Dirichlet distribution.

Notably, in the present work, the clustering of subjects is only informed by the neuronal parameter estimates. This is not a fixed property of the method, however, and it would be perfectly possible to inform clustering by hemodynamic parameters as well. This might be of utility for studying diseases where contributions of vascular impairments are known to differ across subforms, for example, dementia. Generally, whether the inclusion of hemodynamic parameters improves the overall model, can be tested by model selection based on the (log) model evidence (see below).

Fig. 3 shows the graphical model for these conditional distributions. The noise and hemodynamic parameters are estimated independently for each subject. The final equation for $\mathbf{y}_n$ describes the generative model of a single-subject DCM (see previous section). The priors are the same as listed in Table 1; the key difference concerns the prior distributions over connection parameters A, B and C which are now defined over clusters of parameters rather than over subjects' parameters directly.

Such a unified hierarchical model offers the opportunity of using regularities across subjects and cluster structure of the group in order to select more informed priors for estimating DCM parameters in each individual subject. That is, in our generative model, each cluster represents a prior distribution for the parameters of the subjects assigned to that cluster. Hence, we can view this as an implicit learning of hyperparameters of the prior distribution of connectivity parameters in single-subject DCMs by pooling the data across all subjects while respecting the cluster structure of the population. This amounts to an interaction between 'empirical Bayesian' inference and unsupervised generative embedding which is enabled by the hierarchical structure of our model. Another advantage of our hierarchical model is with respect to the use of point estimates (like posterior expectations) as compared to entire posterior distributions for clustering purposes. In the conventional two-stage process, point estimates like posterior expectations have been used so far (Brodersen et al., 2011, 2014). By contrast, our unified model uses all the available distributional information to specify clusters.

A special case of this unified model is a single cluster model where all subjects have the same latent prior distribution. This drops the ambition to detect structure in the population but still represents a significant step beyond current non-hierarchical schemes for DCM inversion. Specifically, the introduction of a latent layer essentially means that we adopt an empirical Bayesian perspective, and makes it possible to infer the prior distribution and subject posterior estimates under the same scheme. Below, we use an empirical dataset to demonstrate that such a model enhancement can lead to a significant increase in model evidence.

2.5. MCMC for inference

We now describe the inference for the finite mixture framework using MCMC sampling. The posterior distribution of our finite mixture model is not analytically tractable. Approximations to the posterior distribution could be obtained using either MCMC sampling or variational Bayes. In this work, we chose MCMC sampling; variational Bayes will be considered in future work. We constructed a blocked Gibbs sampler mixed with Metropolis–Hastings steps whenever exact posterior conditional distributions are not easily derived as in the sampling of the subject specific parameters. The joint model can be described as follows:

$$
\begin{aligned}
&p(\{\mathrm{y}_{\mathbf{n}}\}, \Box_{\mathbf{d}}, \Box_{\mathbf{h}}, \Box, \mathbf{d}, \Box\, \mu_{\mathbf{k}}, \Sigma_k \Box, \Lambda|\bullet) \propto\\
&\Box_{n=1}^{N} \text{Normal}(\mathbf{y_n}|g\ \Box\, \theta_{\mathbf{d_n}}, \Box_{\mathbf{h_n}}, \Box_{\mathbf{q}}\ \Box, \Lambda_n^{-1})\\
&\cdot \Box_{n=1}^{N} \text{Normal}(\Box_{\mathbf{d_n}}|\Box_{\mathbf{d_n}}, \Sigma_{d_n})\\
&\cdot \Box_{n=1}^{N} \text{logNormal}(\Box_{\mathbf{h_n}}|\Box_{\mathbf{h}}, \Sigma_h)\\
&\cdot \Box_{n=1}^{N} \Box_{r=1}^{R} \text{logNormal}(\Box_{n,r}|\Box_\Lambda, \Box^2_\Lambda)\\
&\cdot \Box_{n=1}^{N} \Box_{d_n} \Box_{k=1}^{K} \text{NormalInvWishart}(\mu_{\mathbf{k}}, \Sigma_k|\bullet)\\
&\cdot \text{Dir}(\Box|\Box),
\end{aligned}
\tag{8}
$$

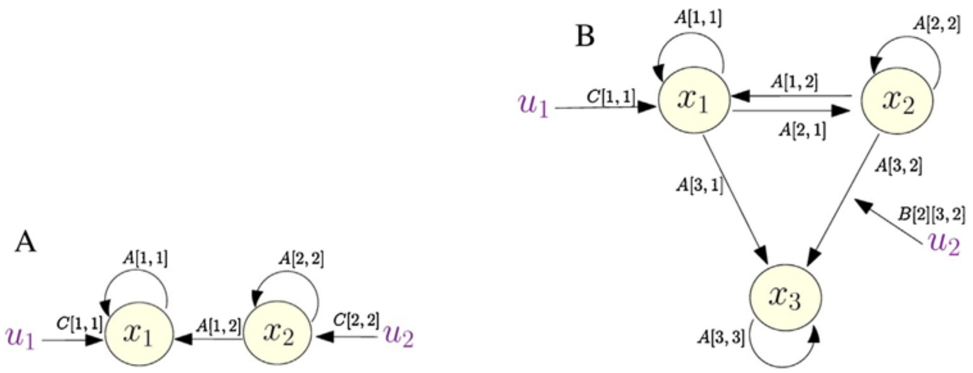


Fig. 4. The figure shows the two DCMs used for simulations. (A) A two-region linear DCM with two inputs. (B) A three-region bilinear DCM with two inputs.

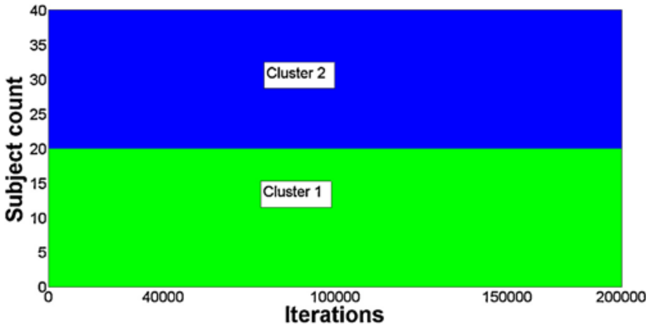


Fig. 5. Clustering results for the simulation based on the two-region DCM (Fig. 4A). Colours indicate the two clusters. The trace of the cluster sizes over the MCMC iterations indicate rapid convergence to two clusters and is stable across the entire chain. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The posterior sampling distributions for the Gaussian distributions describing each cluster given subject-specific connectivity parameter estimates can be expanded as follows:

$$p(\Sigma_k|\bullet)\sim \quad \mathrm{InvWishart}(\Sigma_k|\Sigma_0 + C_k + ..$$

$$.. + \frac{\square_0 n_k}{\square_0 + n_k}\left(\bar{\theta}_{\mathbf{k}} - \square_0\right)(\bar{\theta}_{\mathbf{k}} - \square_0)^T, (\nu + n_k)\Big) \tag{9}$$

$$p(\mu_{\mathbf{k}}|\Sigma,\bullet)\sim \quad \mathrm{Normal}\left(\mu_{\mathbf{k}}|\frac{\kappa_0\square_0 + n'_k\bar{\square}_k}{\kappa_0 + n'_k}, \frac{1}{\kappa_0 + n'_k}\Sigma_k\right),$$

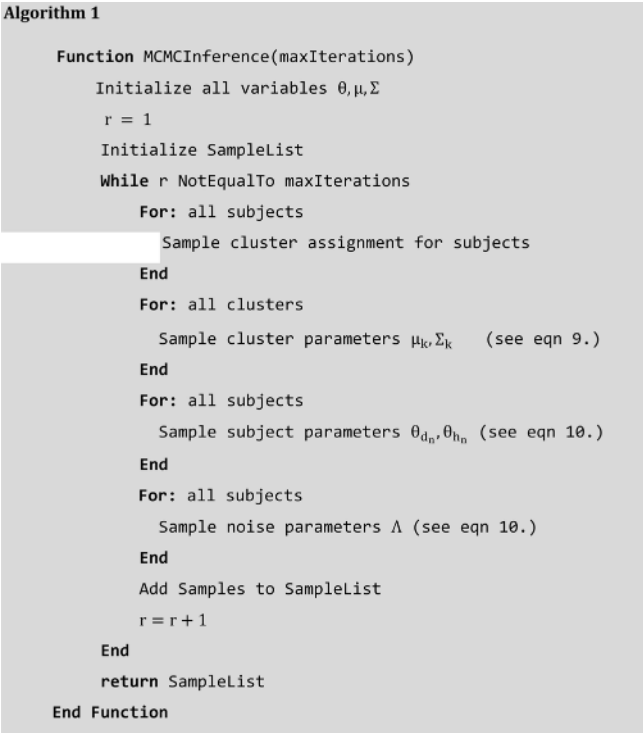
where C_k is the covariance matrix of the connection parameters of subjects assigned to the k-th cluster and $\bar{\theta}_{\mathbf{k}}$ is the mean of the $\square_{\mathbf{d_n}}$'s of subjects assigned to cluster d_n, i.e. $\bar{\theta}_{\mathbf{k}} = 1/n_k\Sigma_{d_{n=k}}\square_{\mathbf{d_n}}$, where n_k is the number of subjects assigned to cluster k. The posterior sampling distributions for the remaining variables (DCM parameters and hyperparameters) are:

$$p(\theta_{\mathbf{d_n}}, \square_{\mathbf{h_n}}|\bullet) \propto p(y_{\mathbf{n}}|\bullet)\mathrm{Normal}(\square_{\mathbf{d_n}}|\square_{\mathbf{k}}, \Sigma_k)\mathrm{logNormal}(\square_{\mathbf{h_n}}|\square_{\mathbf{h}}, \Sigma_{\mathrm{h}})$$

$$p(\log(\lambda_{\mathbf{n}})|\bullet) \propto \quad p(y_{\mathbf{n}}|\bullet) \square_{r=1}^{R} \mathrm{Normal}(\log(\lambda_{n,r})|\square_{\Lambda}, \square_{\Lambda}^2). \tag{10}$$

Finally, the sampling of subject-wise cluster assignments is done by computing the probabilities of assigning a subject to each cluster given all the other parameter values.

Based on the above list of posteriors, it is straightforward to construct an MCMC algorithm that samples from the posterior distributions. The algorithm is described in Algorithm 1.



Algorithm 1

```
Function MCMCInference(maxIterations)
    Initialize all variables $\theta,\mu,\Sigma$
    r = 1
    Initialize SampleList
    While r NotEqualTo maxIterations
        For: all subjects
            Sample cluster assignment for subjects
        End
        For: all clusters
            Sample cluster parameters $\mu_k,\Sigma_k$    (see eqn 9.)
        End
        For: all subjects
            Sample subject parameters $\theta_{d_n},\theta_{h_n}$ (see eqn 10.)
        End
        For: all subjects
            Sample noise parameters $\Lambda$ (see eqn 10.)
        End
        Add Samples to SampleList
        r = r + 1
    End
    return SampleList
End Function
```

The MCMC algorithm described above is implemented in Matlab. In the current version of the code, sampling is initialised based on prior means and by learning the single subject estimates without taking clustering into consideration; empirically, this proved to ensure faster convergence. Furthermore in the present code, the numerical integration scheme used to solve the differential equations is Euler's method (step size of roughly 0.1 s) which is implemented in *C* for computational efficiency. Our model will be made available as part of the open source toolbox TAPAS (http://www.translationalneuromodeling.org/tapas).

3. Results

3.1. Simulations

We tested the ability of our model inversion scheme to provide veridical estimates of cluster assignments and individual parameter estimates in simulations. To this end we generated synthetic

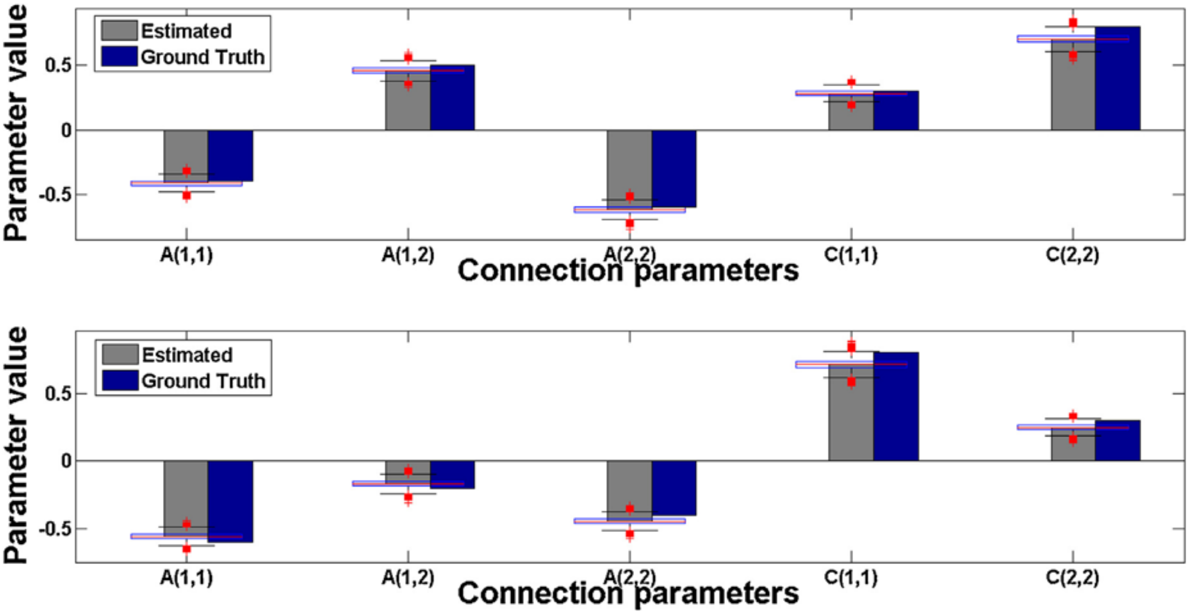


Fig. 6. Parameter recovery for the simulation based on the two-region DCM (Fig. 4A). The figure plots the estimated means of the cluster parameters for all clusters (in this case two), along with a box plot to illustrate the variance in the estimate. It can be seen that all the parameter estimates are close to the values used for the simulation.

multi-subject BOLD data, with added observation noise, using two different DCMs. To ensure that data were simulated from a stable system, we chose A matrices in the simulations according to the stability criterion for dynamical systems in continuous time, i.e. ensuring that the largest eigenvalue of the coupling matrix was negative.

First, we simulated data using a two-region linear DCM with two inputs, as shown in Fig. 4A. The BOLD signal data were simulated for 40 subjects with signal-to-noise ratio (SNR) = 1, time to repetition (TR) = 2, number of scans per subject = 256, number of clusters = 2 (20 subjects from each cluster), cluster variance = 0.05, cluster means—$\{A_1 = [-0.4\ 0.5;0\ -0.6], C_1 = [0.3\ 0;0\ 0.8]\}$, $\{A_2 = [-0.6\ -0.2;0\ -0.4], C_2 = [0.8\ 0;\ 0\ 0.3]\}$. The inputs were boxcar functions with a time-step of 0.125 s. Using the fixed cluster parameters, the parameters of each subject were sampled from the respective cluster distributions. Notably, in our simulations, SNR was defined as the ratio of signal standard deviation to noise standard deviation (cf. definition 4 in Welvaert and Rosseel 2013). The log of the square of this ratio corresponds to a decibel value (note the standard definition of decibel as the logarithmic unit of the ratio of powers or intensities). In our case, the chosen SNR of 1 corresponds to 0 db. Considering that DCM analyses use regional time series that are de-noised by taking the first principal component (eigenvariate) over tens to hundreds of voxels, our scenario represents a relatively challenging case.

The MCMC algorithm was applied to the synthetic data from all virtual subjects and executed for 200,000 iterations, where the burn-in was taken to be 100,000 iterations. The results are summarised by Figs. 5–7 showing the recovery of the clusters (Fig. 5), cluster parameters (Fig. 6), and subject assignments (Fig. 7).

Our current implementation requires that the number of clusters K is predefined. The question how many clusters are plausible, given the data, is basically a question of model selection. That is, inverting the model under different values of K and comparing the respective model evidence enables one to determine the most likely number of clusters. In the context of the present example, we simulated this scenario, running the inference for $K \in \{1, 2, 3, 4\}$ and computed the model evidence for each value of K using thermodynamic integration (see Aponte et al., 2016). The results are shown in Fig. 8 and indicate that in this ground truth scenario the model correctly detects that the data were generated under the existence of two clusters ($K = 2$).

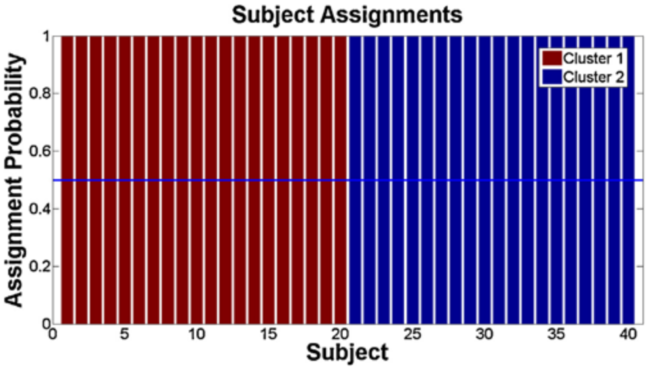


Fig. 7. Cluster assignments for the simulation based on the two-region DCM (Fig. 4A). The plot shows inferred cluster assignments for each subject. It matches the ground truth perfectly (data were generated with the first half of subjects assigned to cluster 1 and the second half to cluster 2).

In a second simulation study, we repeated the same procedure for the three-region bilinear DCM shown in Fig. 4B. Data for 40 subjects was simulated using two clusters (20 subjects each) of three region bilinear DCMs. This is a more challenging case than the linear DCM since there are more parameters to be estimated with a relatively small number of subjects and low SNR. The data was generated using the same procedure and parameters as for the two-region linear DCM, except that the cluster means in this case were $\{A_1 = [-0.5\ -0.01\ 0;0.2\ -0.6\ 0;0.01\ 0.40\ -0.4], C_1 = [0.3\ 0;0\ 0;0\ 0], B2_{(3,2)} = 0.6\}$ and $\{A_2 = [-0.6\ -0.4\ 0;0.01\ -0.4\ 0;0.3\ -0.01\ -0.6], C_2 = [0.6\ 0;0\ 0;0\ 0], B2_{(3,2)} = 0.3\}$ and the cluster variance = 0.01. The results are shown in Figs. 9–11. We can see that the procedure does detect the presence of two major clusters. Moreover, subject assignments are recovered with high accuracy along with some of the cluster parameters. However, in this more challenging simulation, not all parameters were recovered well (Fig. 10); the possible reasons for this are addressed in Section 4.

Additionally, as in the case of the simulated linear DCM, we inverted the model under different values of $K \in \{1, 2, 3, 4\}$ and

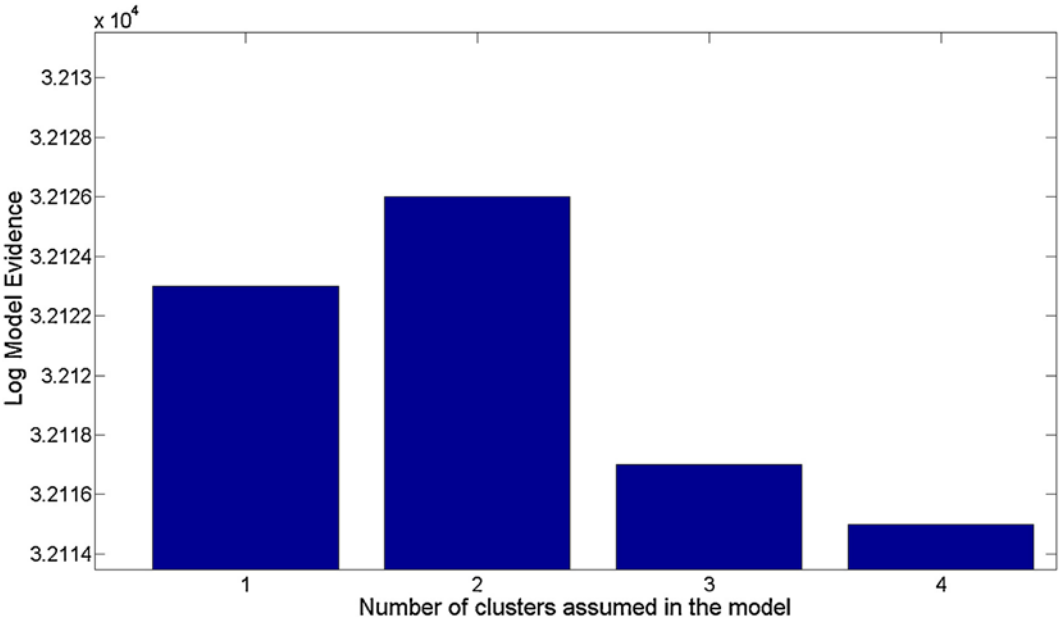


Fig. 8. Comparison of models (linear DCM) with different assumptions about the underlying number of clusters (K = 1–4). We see that the model with K = 2, which corresponds to the ground truth, is correctly identified as the most plausible model.

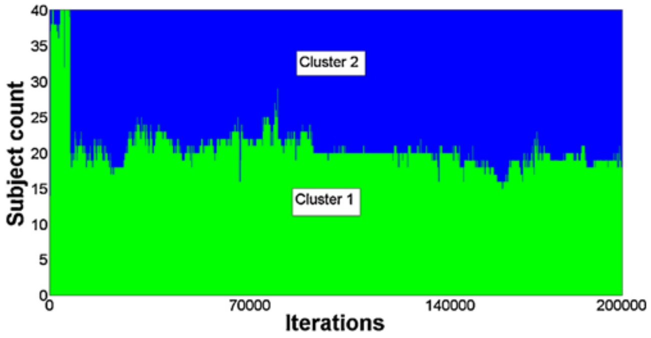


Fig. 9. Clustering results for the simulation based on the three-region DCM (Fig. 4B). The colours indicate the clusters. We observe that with increasing MCMC iterations the solution stabilizes to 2 dominating clusters. The convergence was assessed based on the Geweke's convergence diagnostic (Geweke, 1992).

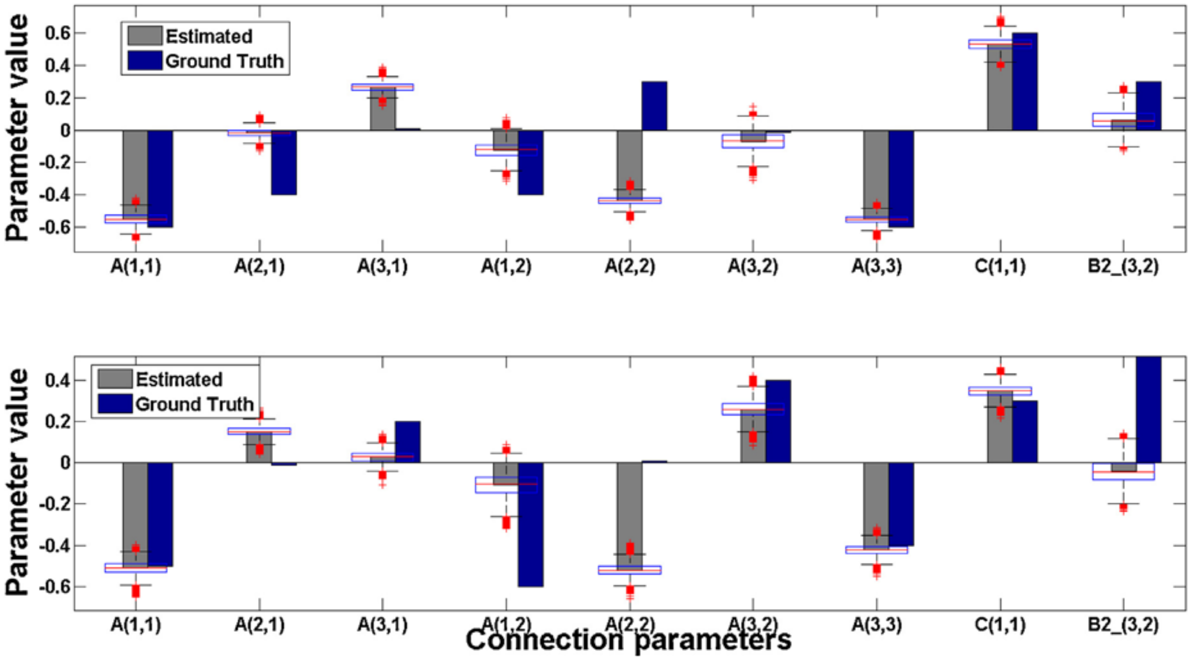


Fig. 10. Parameter estimates for the simulation based on the three-region DCM (Fig. 4B). The figure shows the estimated mean cluster parameters for all clusters (in this case two) along with the box plot to visualise the variance in the estimates. Although the clusters have been identified well, not all parameters are close to the ground truth.

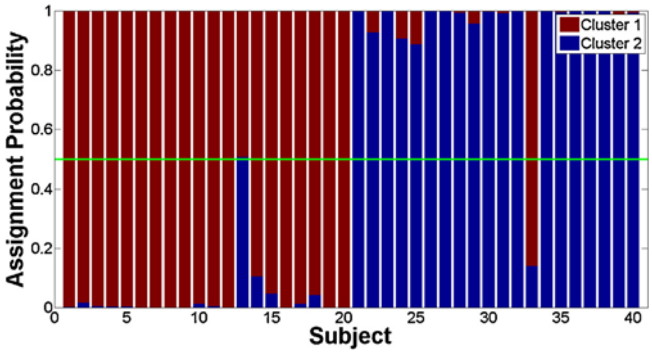


Fig. 11. Cluster assignment results for the simulation based on the three-region DCM (Fig. 4B). The inferred assignments of subjects to clusters show high accuracy with respect to the ground truth. The data were generated with subjects 1–20 belonging to cluster 1 and subjects 21–40 belong to cluster 2.

compared the respective model evidences to determine the most likely number of clusters. The results are shown in Fig. 12 and indicate that in this ground truth scenario the model correctly detects that the data were generated under the existence of two clusters ($K = 2$).

3.2. Empirical fMRI data: working memory task in patients with schizophrenia and healthy controls

Following the above simulations, we applied our unified hierarchical model for embedded clustering to an empirical fMRI dataset consisting of 83 subjects engaged in a working memory task (Deserno et al., 2012). This sample comprised 41 patients diagnosed with schizophrenia according to DSM-IV (10 female; mean age 34.1 years; SD 10.4) and 42 healthy controls (19 female; mean age 35.4; SD 12.2). Full details of this dataset can be found in Deserno et al. (2012). Here we used this dataset to test whether our model could identify the two groups (patients and healthy individuals) in an unsupervised way, under the assumption of a particular DCM (i.e. fixed network structure). Additionally, we also tested different models for parameter estimation to verify the gains achieved by using our novel framework. For these analyses, the priors for noise and hemodynamic parameters were matched to the SPM version used by the original analyses by Deserno et al. (2012), i.e. SPM8 release 4010. The model that is used for these analyses is the same model as in Brodersen et al. (2014) and is shown in Fig. 13.

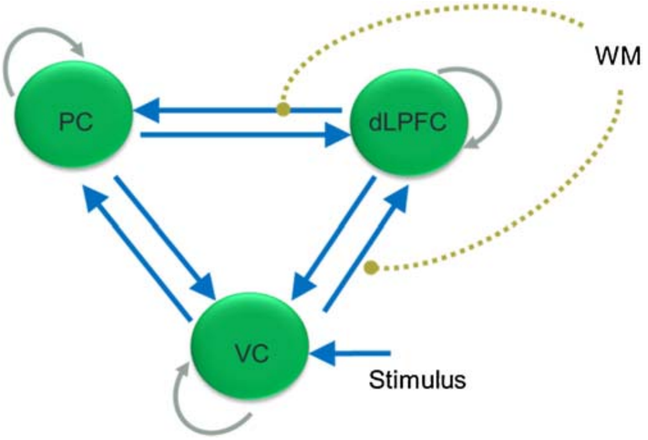


Fig. 13. This plot shows the DCM network model used for the empirical analyses. This is the winning model chosen in Brodersen et al. (2014).

3.3. Unsupervised discrimination of healthy subjects vs. patients

We tested how well our model would discriminate, in an unsupervised way, the two groups from the fMRI data, i.e. patients and healthy individuals. In this context, it should be mentioned that, in distinction to previous analyses of this dataset, our current methodology does not yet allow for a straightforward correction for potential confound variables. For example, since age and sex are major determinants of working memory processes (e.g. Pauls et al., 2013; Spencer-Smith et al., 2013), a previous cluster analysis of this dataset using the generative embedding approach employed multiple regression to adjust individual connectivity estimates for age, sex and handedness before applying clustering (Brodersen et al., 2014).

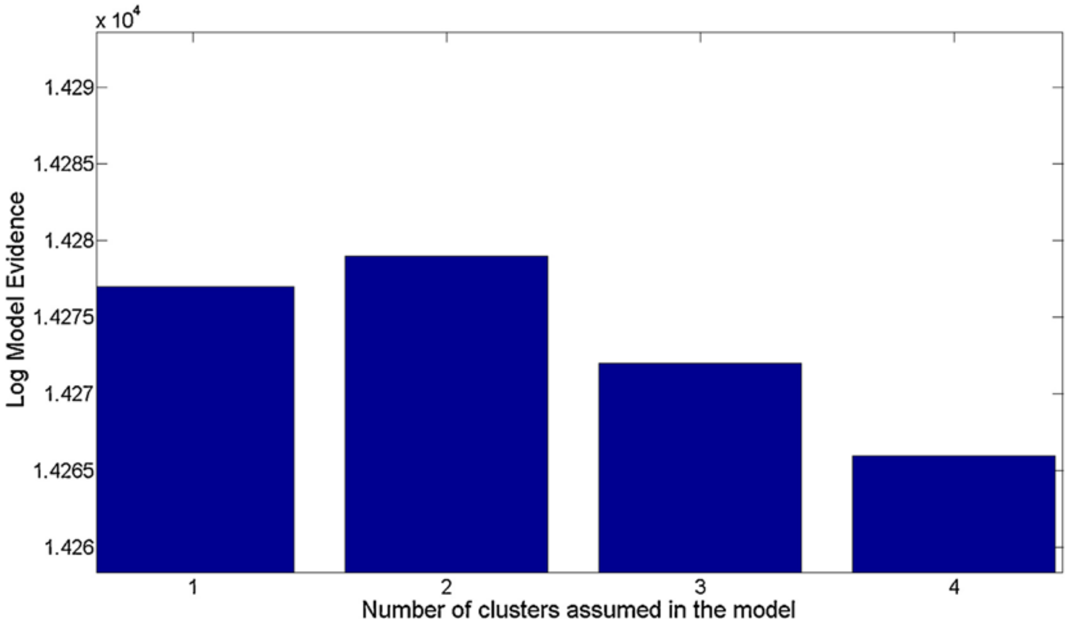


Fig. 12. Comparison of models (bilinear DCM) with different assumptions about the underlying number of clusters ($K = 1–4$). We see that the model with $K = 2$, which corresponds to the ground truth, is correctly identified as the most plausible model.

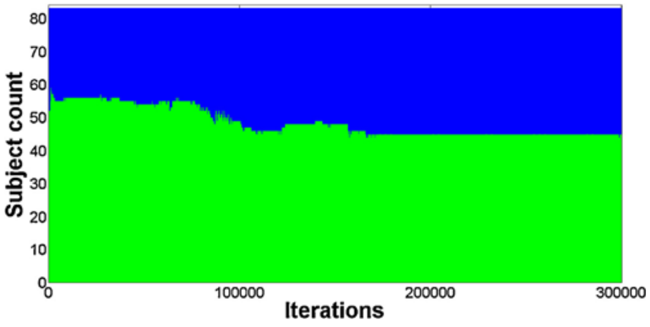


Fig. 14. This plot shows the cluster evolution for the empirical fMRI dataset. One can see two stable cluster sizes emerge.

In the present analysis, application of our finite mixture model (200,000 MCMC iterations) yielded two stable clusters. The evolution of cluster size, cluster assignment, and parameter estimates in both clusters are shown in Figs. 14–16.

By labelling the clusters based on the diagnostic status of the majority of subjects contained by them (see Fig. 16), we can test whether the clustering allows for accurate distinctions. Specifically, the degree to which the inferred labels agree with the true diagnostic labels can be measured using the "balanced purity" (Brodersen et al., 2014). This criterion improves upon the more conventional "purity" measure, which indicates how well the cluster composition matches an external class label. While simple, purity is strongly affected by imbalance in the data, e.g. when subgroups differ considerably in size. Balanced purity shields against such distortions and provides an unbiased estimate for cluster validation. In our analysis, we obtained a balanced purity of 65%. Using the distribution over randomly assigned labels as null distribution, this clustering-based discrimination of groups is highly significant ($p = 0.0032$).

3.4. Model comparison

One might wonder whether the relatively complex methodology we have presented here conveys any advantages over conventional approaches to DCM. This question can be addressed by model comparison: the fact that our framework rests on a generative model enables us to compute the evidence for different models and thus formally decide whether a hierarchical (empirical Bayesian) formulation conveys an advantage over the conventional non-hierarchical approach, as well as to investigate whether simultaneously taking into account group substructure (clustering) is advantageous. We addressed this question both in the context of our simulations and the above empirical dataset, comparing three models of the data. The first model used the conventional non-hierarchical approach in which parameter estimates are obtained independently for each subject, with fixed priors over the parameters. The second model was of the hierarchical form introduced in this paper, but assuming that all subjects came from the same group (i.e. their connectivity parameters were sampled from the same prior). This represents a special case of our framework where the cluster size is set to one. Finally, the third model exploited the full functionality of our approach, adopting a hierarchical (empirical Bayesian) perspective while allowing for simultaneous clustering into two subgroups; in other words, the procedure determined subgroup-specific priors in order to compute subject-specific parameter estimates.

Table 2
Log model evidences for the three different modelling approaches with different assumptions about the prior distributions and group structure using the simulated examples for linear and bilinear DCMs described earlier and the empirical dataset for the working memory task. The hierarchical mixture model performs best, followed by the hierarchical single-cluster model. Both these hierarchical extensions in turn perform far better than the non-hierarchical model. Please note that a log evidence difference of three is considered a strong indication of one model being superior to another (Kass and Raftery, 1995).

Log model evidence	Non-hierarchical	Hierarchical (single cluster)	Hierarchical (mixture model)
Linear DCM (simulated)	32,015.0	32,123.0	32,126.0
Bilinear DCM (simulated)	14,058.0	14,277.0	14,279.0
Working memory(empirical)	−4282.7	−4080.9	−4064.6

In order to obtain an approximation to the log evidence as a basis for model comparison, we employed thermodynamic integration (TI) (Gelman and Meng, 1998). Our implementation of TI is an extension to the single chain MCMC approach, using multiple chains at different temperatures in order to obtain a robust estimate of the model evidence. Details of this TI implementation are described elsewhere (Aponte et al., 2016; Raman et al., in preparation).

For both linear and bilinear DCM simulations as well as for the empirical working memory dataset, the ensuing model comparison, based on log evidence estimates obtained by TI, indicates that a hierarchical model is clearly superior to conventional inversion of DCMs; additionally, the mixture model formulation (with two clusters) is found to be a more adequate explanation of the data than a single cluster formulation. These results are described in Table 2.

For the empirical dataset, the results are summarised by Table 2; here the benefit obtained from the interaction between an empiri-

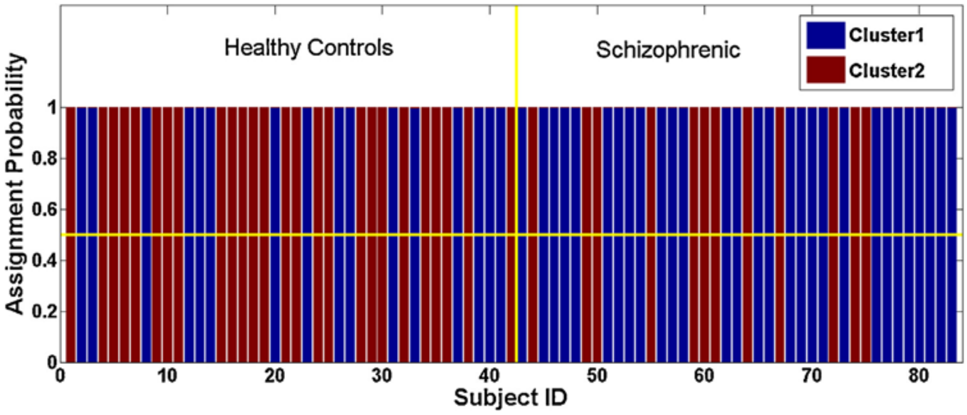


Fig. 15. This plot shows the assignment of subjects to the two clusters.

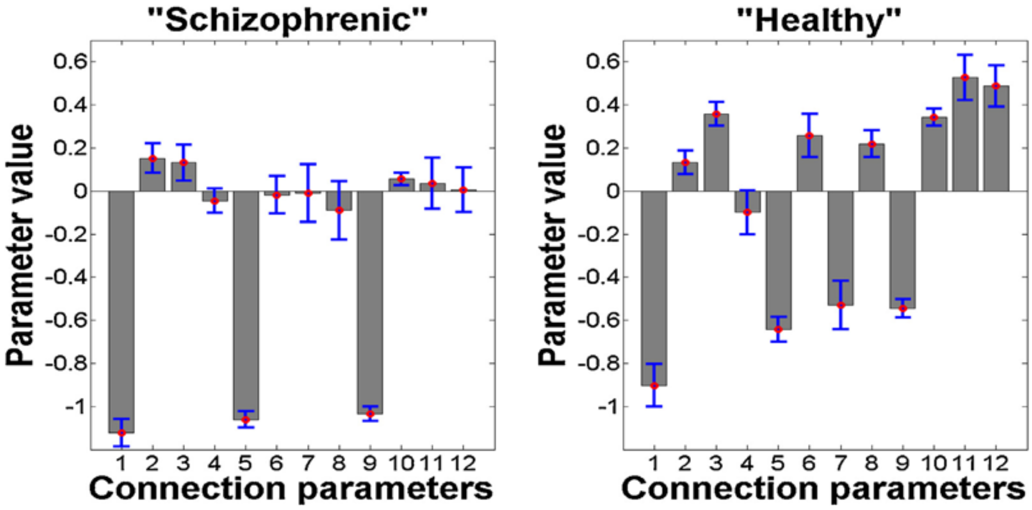


Fig. 16. This figure plots the average connectivity parameter estimates for the two clusters (using the labels "healthy" and "patients" based on the true diagnostic label of the majority of subjects in the cluster).

Table 3
Root mean square error (RMSE) averaged over all subjects for the two simulated DCMs (linear and bilinear) and for the three different modelling approaches with different assumptions about the prior distributions and group structure. There is a clear improvement of parameter value recovery (RMSE) when moving from conventional DCM to a hierarchical model, with a further, less pronounced improvement when adding the mixture model formulation.

RMSE (average)	Non-hierarchical	Hierarchical (single cluster)	Hierarchical (mixture model)
Linear DCM (simulated)	2.3530	0.1165	0.0894
Bilinear DCM (simulated)	2.7939	0.3793	0.3625

cal Bayesian perspective and clustering is even more beneficial: the log evidence differences between the best model (the hierarchical mixture model assuming two groups) and its two competitors are approximately 16 and 218 and thus considerably higher than the conventional requirement of a log evidence difference of 3 for distinguishing models (Kass and Raftery, 1995). In other words, for this dataset, it is advantageous to learn the prior distribution over parameters using the data from all subjects but, at the same time, determine the priors in a group-specific way.

Furthermore, we investigated how accurately subject wise parameter values were recovered, using the standard root mean squared error (RMSE) criterion for estimation accuracy. Specifically, we computed and report the average RMSE (between the estimated and true connection parameters) over all subjects. Of course, this can only be done for the simulated data since the true (generative) parameter values are not known for the empirical data. The results for the two simulated datasets show a clear improvement when using the hierarchical model, with an additional, albeit less dramatic benefit under the mixture model (Table 3).

4. Discussion

In this paper, we have presented a novel framework to jointly infer the effective connectivity parameters in DCMs for multiple subjects and, at the same time, discover connectivity-defined cluster structure of the whole population, using a mixture model approach. While mixture models have found successful applications in a variety of scenarios (de la Torre et al., 2006; Hurn et al., 2003; Lenk and DeSarbo, 2000; Qi et al., 2007), they have only relatively recently been introduced to neuroimaging data analysis (e.g. Thirion et al., 2007; Woolrich et al., 2005; Lashkari et al., 2010; Stingo et al., 2013). Our model is novel in that the features which enter a generative mixture model represent the full posterior parameter distributions from another (embedded) generative model, i.e. a DCM for fMRI data. This represents a hierarchical extension of previously introduced generative embedding procedures (Brodersen et al., 2011, 2014) and unifies inference on individual connectivity with inference on population subgroups.

In addition to introducing the mathematical details of our new approach, this paper used simulations and empirical data to probe the validity of our model. In this context, one needs to distinguish between different kinds of validity (face, construct, and predictive validity). Our simulations were designed to test face validity. That is, we asked: if data are generated by the generative process embodied by the model, are we able to recover the parameter values and model structure (e.g. true number of clusters)? This is not necessarily given. For example, our implementation could be flawed (coding errors); or the model might have a likelihood function which introduces profound convergence problems during optimisation or significant non-identifiability problems. It is these type of problems which can be detected by face validity tests based on simulated data. Furthermore, our empirical analyses assessed predictive validity: here, we used the empirical schizophrenia dataset to assess the correspondence of our clustering solution to external labels (diagnostic status). By contrast, construct validity concerns the question to what degree a model delivers comparable results as alternative techniques (cf. Lee et al., 2006). In this paper, we do not address construct validity with regard to non-DCM generative models, but hope to address this issue in future work.

It is worth reiterating that the focus of this initial paper is not on establishing a procedure with higher accuracy than existing schemes; instead, the main purpose of this paper is to introduce the idea behind our novel approach and demonstrate its practical feasibility. To this end, we assessed the accuracy of model inversion using simulations and applied our model to an empirical fMRI dataset consisting of two groups, healthy controls and schizophrenic patients performing a working memory task. The simulations in this paper demonstrate that our current implementation performs well, even under a relatively challenging scenario of non-trivial observation noise (compare Figs. 5–10). The application to the empirical data by Deserno et al. (2012) produced results similar to those from an earlier analysis of this dataset by Brodersen et al. (2014). Specifically, our model was able to distinguish patients and controls in an unsupervised way with highly significant accuracy (65%; $p = 0.0032$). This is similar to the 71% classification accuracy reported by Brodersen et al. (2014). Generally, given the simplicity of our cortical circuit model (comprising merely 6 connections and 3 inputs), the accuracy reported here does not fare too badly compared to previous attempts of distinguishing between schizophrenic patients and healthy controls on the basis of whole-brain connectivity estimates, resulting in accuracies in the range of 60–90% (e.g. Anticevic et al., 2015; Arbabshirani et al., 2013; Venkataraman et al., 2012; Fekete et al., 2013).

Several reasons might explain the slightly lower performance of our method, compared to Brodersen et al. (2014), in this particular application case. First, our model inversion scheme uses a less accurate method (Euler integration, instead of the matrix exponential) for integrating the differential equations of the neuronal state equations. This choice was motivated by the goal of reducing compute times as much as possible. In future implementations, GPU-based implementations will allow for more accurate integration schemes without significant increase in compute time (cf. Aponte et al., 2016). Second, and more importantly, in contrast to Brodersen et al. (2014), the implementation presented in this paper does not yet incorporate a correction for confounding variables (such as age, handedness or sex) which might have considerable impact on the definition of subgroups. This is an important extension of our method which we hope to present in future work and which is of particular relevance for identification of subgroups in heterogeneous spectrum diseases (see Wiecki et al., 2015). That is, unless inter-individual variability due to disease-irrelevant confounds can be distinguished from the variability due to disease-relevant factors, putative cluster structures in data from patient populations with known heterogeneity are difficult to interpret or trust.

Importantly, our new framework rests on a single generative model that can be reduced to two special cases: (i) a hierarchical which assumes that all subjects come from a single cluster; this corresponds to an empirical Bayesian perspective without embedded clustering; and (ii) a non-hierarchical framework; this is identical to the conventional approach where parameters from each subject are estimated independently under fixed priors and without clustering. This allows one to use model comparison to address the question which of these perspectives is most appropriate for explaining multi-subject fMRI data with DCM. Here, we have shown that for the dataset by Deserno et al. (2012) our unified framework has considerably higher evidence when compared to either of the other two cases. In other words, allowing for regularisation of subject-specific parameters by estimating subgroup-specific priors is, at least for this application, superior to the conventional DCM approach.

Compared to previous two-step generative embedding procedures (Brodersen et al., 2011, 2014) our model has two distinct advantages. First, the hierarchical structure of our model allows one to exploit across-subject information in order to define cluster-specific priors for DCM parameter estimation. This corresponds to an 'empirical Bayesian' inference scheme and may result in more appropriate priors than the currently used fixed shrinkage prior. Second, our hierarchical model uses the entire information from the posterior distributions for clustering, not only point estimates (like posterior expectations). More recently, an elegant and computationally extremely efficient Bayesian reduction strategy (Friston et al., 2016) has been proposed for empirical Bayesian estimates in DCM: this corresponds to the special case of a single cluster in our model. In comparison, our approach is computationally far more expensive. However, our approach is more general in that it does not assume that the models are nested; furthermore, it does not make any distributional assumptions regarding the posterior distribution of the connectivity parameters. This may be an important asset when dealing with highly non-linear models such as conductance-based DCMs (Moran et al., 2011). Second, our framework can be extended to an infinite mixture model, thus providing a principled solution to tackling the general and difficult problem of inferring the number of subgroups within a population of subjects. In the present formulation, determining the number of clusters corresponds to a classical model selection problem: it can be achieved by running the model under different assumed values of K and comparing the resulting log evidences (see Fig. 8). It is worth emphasising that this type of model comparison is critically important, because wrong assumptions about the number of clusters can have deleterious consequences on inference. That is, under false assumptions about the number of clusters erroneous merging or splitting of clusters will necessarily occur, with possible impact on parameter estimates of individual subjects within the respective clusters. However, our current implementation not only allows for model selection with regard to cluster number, but also estimates the uncertainty (posterior variance) of the cluster assignments and thus allows for detecting potentially problematic cluster solutions.

Having said this, the implementation of our approach presented in this paper also has several important limitations (as already touched upon above) which we will address in future work. Five further aspects of the model are particularly prominent targets for improvement. The first concerns a reduction of the computational cost of the inference scheme. Although MCMC is an attractive inference scheme due to its simplicity, it is computationally expensive. This may become prohibitive with large DCMs and large subject population. For larger studies, this can be partly rectified by parallelizing the sampling of connectivity parameters of individual subjects (cf. Aponte et al., 2016). To further improve scalability to large DCMs, as a next step, we are currently pursuing a variational Bayesian approach to inference and will examine whether this allows for substantial computational gains without compromising the accuracy of inference. Another alternative is to employ Gaussian processes for model inversion, an approach we have recently introduced to DCM (Lomakina et al., 2015). Additionally, we aim to improve upon the current single chain MCMC implementation by extending the population MCMC scheme in Aponte et al. (2016) which uses multiple chains for better convergence.

Second, an extension to population MCMC may provide a remedy for a potential problem we encountered above. This concerns the fact that our more complex simulations of a bilinear DCM indicated that not all generative parameter values were recovered correctly. There could be multiple reasons for this. One possibility is that the more complex likelihood function induces partial parameter dependencies; such dependencies can lead to mathematically entirely correct but seemingly counterintuitive deviations of posterior estimates from "ground truth" parameter values (see the discussion of a similar observation in Lomakina et al., 2015). We checked the resulting posterior covariance matrices and found only relatively moderate dependencies. An alternative explanation is that, for finite runtime, there is no guarantee that MCMC is not getting stuck in certain parts of the posterior distributions. The likelihood of this possibility can be reduced significantly by population MCMC, and we will explore the benefits of this extension in future work.

A third extension of the model concerns feature selection and multi-view clustering (Niu et al., 2012). This is also an ongoing project within our group, where we hope to find that both feature selection and multi-view clustering conveys further improvements with respect to finding unknown subgroups in subject populations.

Fourth, we will extend the current formulation from finite mixture to infinite mixture models (Rasmussen and Ghahramani, 2002) based on Dirichlet processes (Neal, 2000). This will eschew the necessity of specifying the number of clusters in advance and make model comparison (between models with different values of cluster number K) superfluous. Instead, this formulation would allow for inferring the number of clusters together with all other parameters.

Finally, as mentioned earlier and perhaps most urgently, we strive to extend the model such that potential confounds – e.g. sex, age, handedness, medication, etc. – can be removed which might otherwise affect the clustering results (cf. Stingo et al., 2013). The importance of confound removal is an important theme in current discussions of model-based clustering (e.g. Brodersen et al., 2014; Wiecki et al., 2015), and is likely to improve the results of the preliminary analyses of empirical data described in this study. Extending the current model to deal with confound variables, will also enable us to investigate the heterogeneity within the patient group itself, similar to Brodersen et al. (2014), without being affected by differences in age, sex, and medication.

It should be emphasised that the framework presented in this paper is neither restricted to DCM nor to indices of connectivity, but can, in principle, be applied to any generative model. This is important to highlight because connectivity estimates are not the only pathophysiologically relevant indices for describing disease mechanisms and defining patient subgroups. The translational neuromodelling strategy we pursue is equally interested in computational characterisations of individual patients, and how these may be linked to neurophysiological processes such as neuromodulation (Stephan and Mathys, 2014; Schlagenhauf et al., 2014). For example, generative models of behaviour can be applied to individual responses, yielding subject-specific trajectories of prediction errors and uncertainty (or its inverse, precision) which, in turn, have been found to correlate with BOLD signals in neuromodulatory nuclei like the dopaminergic midbrain or the cholinergic basal forebrain (D'Ardenne et al., 2008; Iglesias et al., 2013; Schwartenbeck et al., 2014), or in dopaminoceptive regions like the ventral striatum (O'Doherty et al., 2003; Deserno et al., 2015). Furthermore, a recent pharmacological MEG study using L-Dopa has shown that single-region conductance-based DCMs for electrophysiological responses can provide plausible estimates of dopaminergically mediated changes in glutamatergic receptor conductances (Moran et al., 2011). Integrating these generative models for behavioural or MEG/EEG data into the hierarchical model presented in this paper might allow for clustering patients according to individual profiles of neuromodulation. This is an intriguing possibility of high clinical relevance, given that the large majority of drugs used in psychiatry target neuromodulatory mechanisms (e.g. dopamine receptor antagonists in psychosis, blockade of serotonin or noradrenaline reuptake in depression, inhibition of acetylcholine breakdown in dementia). In other words, unsupervised clustering based on indices of neuromodulation might delineate patient subgroups with different treatment predictions (Stephan et al., 2015). Given the modular structure and the generic and robust inference scheme of the unified framework for model-based clustering presented in this paper, it can easily be extended to other generative models. In this regard, we plan to implement this framework for other DCMs, such as DCM for event-related responses (David et al., 2006) and conductance-based DCMs (Moran et al., 2011), as well as hierarchical Bayesian models of learning (Mathys et al., 2011).

In summary, this paper has introduced a novel hierarchical model for simultaneous inference on single-subject connection strengths and cluster structure in the population. We hope that this model will become a useful tool for detecting pathophysiological subgroups in psychiatric and neurological spectrum diseases and for assigning single patients to such subgroups. While many future improvements and extensions are conceivable and desirable, a clear strategy exists for evaluating the utility and robustness of this approach in practice, i.e. prospective patient studies in which the predictive strength of individual subgroup assignment can be tested against clinically relevant outcome criteria, such as individual treatment response (Stephan and Mathys, 2014).

Software note

The code of the hierarchical model presented in this paper will be made available as part of the open source software TAPAS (http://www.translationalneuromodeling.org/tapas).

Acknowledgements

This work is supported by core funding from the University of Zurich (KES), the Clinical Research Priority Program (CRPP) "Multiple Sclerosis" (SR, KES), by the René and Susanne Braginsky Foundation (KES) and Max Planck Institute for Human Cognitive and Brain Sciences (LD, FS).

Appendix A. Hemodynamic equations in DCM for fMRI

This section provides a brief summary of the hemodynamic forward model in DCM which translates neuronal population activity (neuronal state x) into an observable BOLD signal y. This model was inspired by the Balloon model initially proposed by Buxton et al. (1998) and then subsequently extended by Friston et al. (2000) and Stephan et al. (2007) as follows:

$$\begin{aligned}
\frac{\mathrm{d}s}{\mathrm{d}t} &= x - \kappa s - \square(f-1)\\
\frac{\mathrm{d}f}{\mathrm{d}t} &= s\\
\tau\frac{\mathrm{d}\mathrm{v}}{\mathrm{d}t} &= f - \mathrm{v}^{1/\square}\\
\tau\frac{\mathrm{d}q}{\mathrm{d}t} &= f\,\frac{1-(1-E_0)^{1/f}}{E_0} - \mathrm{v}^{1/\alpha}\frac{q}{\mathrm{v}}\\
y = \frac{\Delta S}{S_0} &= V_0\;^{\square}\,k_1(1-q) + k_2\;^{\square}\,1-\frac{q}{\mathrm{v}}\;^{\square} + k_3(1-\mathrm{v})^{\square}
\end{aligned} \tag{11}$$

where s, f, v, q represent vasodilatory signal, blood flow, and deoxyhemoglobine content. $V_0 = 4$, $E_0 = 0.4$, $\alpha = 0.32$, $\gamma = 0.32$, $\rho_0 = 40.3$, $\mathrm{TE} = 0.04$, $r_0 = 25$ are biophysical constants (see Stephan et al., 2007 for details). Furthermore, the final line of Eq. (11) represents a nonlinear static output equation which links blood volume and deoxyhemoglobine content to the observed BOLD signal y. This output equation is determined by three coefficients:

$$\begin{aligned}
k_1 &= 4.3\rho_0 E_0 \mathrm{TE}\\
k_2 &= \varepsilon r_0 E_0 \mathrm{TE}\\
k_3 &= \varepsilon - 1,
\end{aligned} \tag{12}$$

The remaining quantities in Eq. (11) represent subject- and region-specific parameters that can either be fixed to typical values or treated as free parameters (see Stephan et al., 2007 for details). In the present application, we matched the hemodynamic model to the version used by Deserno et al. (2012), allowing for the estimation of three free parameters, i.e. $\theta_h = (\kappa, \tau, \varepsilon)$, with prior distributions as indicated in Table 1. Notably, there is a separate parameter set for each region, i.e. regional differences in hemodynamics are taken into account by the model.

References

Anticevic, A., Hu, X., Xiao, Y., Hu, J., Li, F., Bi, F., Cole, M.W., Savic, A., Yang, G.J., Repovs, G., Murray, J.D., Wang, X.-J., Huang, X., Lui, S., Krystal, J.H., Gong, Q., 2015. Early-course unmedicated schizophrenia patients exhibit elevated prefrontal connectivity associated with longitudinal change. J. Neurosci. 35, 267–286, http://dx.doi.org/10.1523/JNEUROSCI.2310-14.2015.

Aponte, E.A., Raman, S., Sengupta, B., Penny, W.D., Stephan, K.E., Heinzle, J., 2016. mpdcm: a toolbox for massively parallel dynamic causal modeling. J. Neurosci. Methods 257 (January), 7–16, http://dx.doi.org/10.1016/j.jneumeth.2015.09.009, ISSN 0165-0270.

Arbabshirani, M.R., Kiehl, K.A., Pearlson, G.D., Calhoun, V.D., 2013. Classification of schizophrenia patients based on resting-state functional network connectivity. Front. Neurosci. 7, 133, http://dx.doi.org/10.3389/fnins.2013.00133.

Brodersen, K.H., Deserno, L., Schlagenhauf, F., Lin, Z., Penny, W.D., Buhmann, J.M., Stephan, K.E., 2014. Dissecting psychiatric spectrum disorders by generative embedding. NeuroImage: Clin. 4, 98–111.

Brodersen, K.H., Schofield, T.M., Leff, A.P., Ong, C.S.S., Lomakina, E.I., Buhmann, J.M., Stephan, K.E., 2011. Generative embedding for model-based classification of fMRI data. PLoS Comput. Biol. 7, e1002079+.

Buxton, R., Wong, E., Frank, L., 1998. Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. Magn. Reson. Med. 39, 855–864.

Casey, B.J., Craddock, N., Cuthbert, B.N., Hyman, S.E., Lee, F.S., Ressler, K.J., 2013. DSM-5 and RDoC: progress in psychiatry research? Nat. Rev. Neurosci. 14, 810–814, http://dx.doi.org/10.1038/nrn3621.

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

Cruse, D., Chennu, S., Chatelle, C., Bekinschtein, T.A., Fernández-Espejo, D., Pickard, J.D., Laureys, S., Owen, A.M., 2011. Bedside detection of awareness in the vegetative state: a cohort study. Lancet 378, 2088–2094, http://dx.doi.org/10.1016/S0140-6736(11)61224-5.

Cuthbert, B., Insel, T., 2013. Toward the future of psychiatric diagnosis: the seven pillars of rDoC. BMC Med. 11, 126, http://dx.doi.org/10.1186/1741-7015-11-126.

Daunizeau, J., Adam, V., Rigoux, L., 2014. VBA: a probabilistic treatment of nonlinear models for neurobiological and behavioural data. PLoS Comput. Biol. 10, e1003441, http://dx.doi.org/10.1371/journal.pcbi.1003441.

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.

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, e315, http://dx.doi.org/10.1371/journal.pbio.0060315.

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, 1255–1272, http://dx.doi.org/10.1016/j.neuroimage.2005.10.045.

Deserno, L., Huys, Q.J.M., Boehme, R., Buchert, R., Heinze, H.-J., Grace, A.A., Dolan, R.J., Heinz, A., Schlagenhauf, F., 2015. Ventral striatal dopamine reflects behavioral and neural signatures of model-based control during sequential decision making. Proc. Natl. Acad. Sci. U. S. A. 112, 1595–1600, http://dx.doi.org/10.1073/pnas.1417219112.

Deserno, L., Sterzer, P., Wüstenberg, T., Heinz, A., Schlagenhauf, F., 2012. Reduced prefrontal-parietal effective connectivity and working memory deficits in schizophrenia. J. Neurosci. 32, 12–20.

Doyle, O.M., Tsaneva-Atansaova, K., Harte, J., Tiffin, P.A., Tino, P., Diaz-Zuccarini, V., 2013. Bridging paradigms: hybrid mechanistic-discriminative predictive models. IEEE Trans. Biomed. Eng. 60, 735–742.

D'Ardenne, K., McClure, S.M., Nystrom, L.E., Cohen, J.D., 2008. BOLD responses reflecting dopaminergic signals in the human ventral tegmental area. Science 319, 1264–1267, http://dx.doi.org/10.1126/science.1150605.

Fekete, T., Wilf, M., Rubin, D., Edelman, S., Malach, R., Mujica-Parodi, L.R., 2013. Combining classification with fMRI-derived complex network measures for potential neurodiagnostics. PLoS One 8 (5), e62867, http://dx.doi.org/10.1371/journal.pone.0062867.

Friston, K., 2002. Bayesian estimation of dynamical systems: an application to fMRI. NeuroImage 16, 513–530, http://dx.doi.org/10.1006/nimg.2001.1044.

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

Friston, K., Harrison, L., Penny, W., 2003. Dynamic causal modelling. NeuroImage 19, 1273–1302, http://dx.doi.org/10.1016/S1053-8119(03)00202-7.

Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J., Penny, W., 2007. Variational free energy and the Laplace approximation. NeuroImage 34, 220–234.

Friston, K.J., Mechelli, A., Turner, R., Price, C.J., 2000. Nonlinear responses in fMRI: the balloon model Volterra kernels, and other hemodynamics. NeuroImage 12, 466–477, http://dx.doi.org/10.1006/nimg.2000.0630.

Friston, K.J., Litvak, V., Oswal, A., Razi, A., Stephan, K.E., van Wijk, B.C.M., Ziegler, G., Zeidman, P., 2016. Bayesian model reduction and empirical Bayes for group (DCM) studies. Neuroimage 128, http://dx.doi.org/10.1016/j.neuroimage.2015.11.015, 413–31 [Epub 2015 November 11].

Gelman, A., Carlin, J., Stern, H., Rubin, D., 1995. Bayesian Data Analysis. Chapman & Hall.

Gelman, A., Meng, X.-L., 1998. Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. Stat. Sci. 13, 163–185, http://dx.doi.org/10.1214/ss/1028905934.

Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In: Bayesian Statistics 4. Oxford University Press, Oxford, pp. 169–193.

Orrù, G., Pettersson-Yeo, W., Marquand, A.F., Sartori, G., Mechelli, A., 2012. Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. Neurosci. Biobehav. Rev. 36 (April (4)), 1140–1152, http://dx.doi.org/10.1016/j.neubiorev.2012.01.004.

Hurn, M., Justel, A., Robert, C.P., 2003. Estimating mixtures of regressions. J. Comput. Graph. Stat. 12, 55–79, http://dx.doi.org/10.2307/1391069.

Iglesias, S., Mathys, C., Brodersen, K.H., Kasper, L., Piccirelli, M., den Ouden, H.E., Stephan, K.E., 2013. Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. Neuron 80, 519–530, http://dx.doi.org/10.1016/j.neuron.2013.09.009.

Kapur, S., Phillips, A., Insel, T., 2012. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? Mol. Psychiatry 17, 1174–1179.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90, 773–795.

Klöppel, S., Abdulkadir, A., Jack, C.R., Koutsouleris, N., Mourao-Miranda, J., Vemuri, P., 2012. Diagnostic neuroimaging across diseases. NeuroImage 61 (2), 457–463, http://dx.doi.org/10.1016/j.neuroimage.2011.11.002.

Krystal, J.H., State, M.W., 2014. Psychiatric disorders: diagnosis to therapy. Cell 157, 201–214, http://dx.doi.org/10.1016/j.cell.2014.02.042.

Lashkari, D., Vul, E., Kanwisher, N., Golland, P., 2010. Discovering structure in the space of fMRI selectivity profiles. NeuroImage 50, 1085–1098, http://dx.doi.org/10.1016/j.neuroimage.2009.12.106.

Lee, L., Friston, K., Horwitz, B., 2006. Large-scale neural models and dynamic causal modelling. NeuroImage 30 (4), 1243–1254, http://dx.doi.org/10.1016/j.neuroimage.2005.11.007.

Lenk, P., DeSarbo, W., 2000. Bayesian inference for finite mixtures of generalized linear models with random effects. Psychometrika 65, 93–119.

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.

Mathys, C., Daunizeau, J., Friston, K.J., Stephan, K.E., 2011. A Bayesian foundation for individual learning under uncertainty. Front. Hum. Neurosci. 5, http://dx.doi.org/10.3389/fnhum.2011.00039.

Moran, R.J., Symmonds, M., Stephan, K.E., Friston, K.J., Dolan, R.J., 2011. An in vivo assay of synaptic function mediating human cognition. Curr. Biol. 21, 1320–1325.

Neal, R.M., 2000. Markov chain sampling methods for Dirichlet process mixture models. J. Comput. Graph. Stat. 9, 249–265.

Niu, D., Dy, J.G., Ghahramani, Z., 2012. A nonparametric Bayesian model for multiple clustering with overlapping feature views. JMLR Proceedings, 814–822 JMLR.org.

O'Doherty, J.P., Dayan, P., Friston, K., Critchley, H., Dolan, R.J., 2003. Temporal difference models and reward-related learning in the human brain. Neuron 38, 329–337, http://dx.doi.org/10.1016/s0896-6273(03)00169-7.

Pauls, F., Petermann, F., Lepach, A.C., 2013. Gender differences in episodic memory and visual working memory including the effects of age. Memory 21, 857–874, http://dx.doi.org/10.1080/09658211.2013.765892.

Qi, Y., Paisley, J.W., Carin, L., 2007. Music analysis using hidden Markov mixture models. IEEE Trans. Signal Process. 55, 5209–5224.

Raman, Sudhir, Aponte, Eduardo A., Heinzle, Jakob, Sengupta, Biswa. Will Penny and Klaas Enno Stephan Thermodynamic integration for dynamic causal models. (in preparation).

Rasmussen, C.E., Ghahramani, Z., 2002. Infinite mixtures of Gaussian process experts. In: Advances in Neural Information Processing Systems 14. MIT Press, pp. 881–888.

Schlagenhauf, F., Huys, Q.J., Deserno, L., Rapp, M.A., Beck, A., Heinze, H.-J., Dolan, R., Heinz, A., 2014. Striatal dysfunction during reversal learning in unmedicated schizophrenia patients. NeuroImage 89, 171–180, http://dx.doi.org/10.1016/j.neuroimage.2013.11.034.

Schrouff, J., Rosa, M.J., Rondina, J.M., Marquand, A.F., Chu, C., Ashburner, J., Phillips, C., Richiardi, J., Mourao-Miranda, J., 2013. PRoNTo: pattern recognition for neuroimaging toolbox. Neuroinformatics 11 (3), 319–337.

Schwartenbeck, P., FitzGerald, T.H.B., Mathys, C., Dolan, R., Friston, K., 2014. The dopaminergic midbrain encodes the expected certainty about desired outcomes. Cereb. Cortex, http://dx.doi.org/10.1093/cercor/bhu159.

Sellers, E.W., Ryan, D.B., Hauser, C.K., 2014. Noninvasive brain-computer interface enables communication after brainstem stroke. Sci. Transl. Med. 6, 257re7, http://dx.doi.org/10.1126/scitranslmed.3007801.

Sengupta, B., Friston, K.J., Penny, W.D., 2015. Gradient-free MCMC methods for dynamic causal modelling. NeuroImage 112, 375–381, http://dx.doi.org/10.1016/j.neuroimage.2015.03.008.

Smoller, J.W., 2013. Disorders and borders: psychiatric genetics and nosology. Am. J. Med. Genet. B: Neuropsychiatr. Genet. 162, 559–578, http://dx.doi.org/10.1002/ajmg.b.32174.

Spencer-Smith, M., Ritter, B.C., Murner-Lavanchy, I., El-Koussy, M., Steinlin, M., Everts, R., 2013. Age, sex, and performance influence the visuospatial working memory network in childhood. Dev. Neuropsychol. 38, 236–255, http://dx.doi.org/10.1080/87565641.2013.784321.

Stephan, K.E., Iglesias, S., Heinzle, J., Diaconescu, A.O., 2015. Translational perspectives for computational neuroimaging. Neuron 87 (August (4)), 716–732, http://dx.doi.org/10.1016/j.neuron.2015.07.008, ISSN 0896-6273.

Stephan, K.E., Kasper, L., Harrison, L., Daunizeau, J., den Ouden, H., Breakspear, M., Friston, K., 2008. Nonlinear dynamic causal models for fMRI. NeuroImage 42, 649–662, http://dx.doi.org/10.1016/j.neuroimage.2008.04.262.

Stephan, K.E., Mathys, C., 2014. Computational approaches to psychiatry. Curr. Opin. Neurobiol. 25, 85–92, http://dx.doi.org/10.1016/j.conb.2013.12.007.

Stephan, K.E., Weiskopf, N., Drysdale, P.M., Robinson, P.A., Friston, K.J., 2007. Comparing hemodynamic models with DCM. NeuroImage 38, 387–401.

Stingo, F.C., Guindani, M., Vannucci, M., Calhoun, V.D., 2013. An integrative Bayesian modeling approach to imaging genetics. J. Am. Stat. Assoc. 108, 876–891, http://dx.doi.org/10.1080/01621459.2013.804409, 2013-09-01T00:00:00.

Thirion, B., Tucholka, A., Keller, M., Pinel, P., Roche, A., Mangin, J.-F., Poline, J.-B., 2007. High level group analysis of fMRI data based on Dirichlet process mixture models. Inf. Process. Med. Imaging 4584, 482–494, http://dx.doi.org/10.1007/978-3-540-73273-0_40, Lecture notes in computer science.

de la Torre, F., Kanade, T., 2006. Discriminative cluster analysis. In: Machine Learning, Proceedings of the Twenty-Third International Conference (ICML 2006), Pittsburgh, Pennsylvania, USA, June 25–29, 2006, pp. 241–248. doi:10.1145/1143844.1143875.

Venkataraman, A., Whitford, T.J., Westin, C.-F., Golland, P., Kubicki, M., 2012. Whole brain resting state functional connectivity abnormalities in schizophrenia. Schizophr. Res. 139 (1–3), 7–12, http://dx.doi.org/10.1016/j.schres.2012.04.021.

Welvaert, M., Rosseel, Y., 2013. On the definition of signal-to-noise ratio and contrast-to-noise ratio for fMRI data. PLoS One 8 (11), e77089, http://dx.doi.org/10.1371/journal.pone.0077089.

Wiecki, T., Poland, J., Frank, M., 2015. Model-based cognitive neuroscience approaches to computational psychiatry: clustering and classification. Clin. Psychol. Sci. 3 (3), 378–399.

Wolfers, T., Buitelaar, J.K., Beckmann, C.F., Franke, B., Marquand, A.F., 2015. From estimating activation locality to predicting disorder: a review of pattern recognition for neuroimaging-based psychiatric diagnostics. Neurosci. Biobehav. Rev., http://dx.doi.org/10.1016/j.neubiorev.2015.08.001, ISSN 0149-7634.

Woolrich, M.W., Behrens, T.E.J., Beckmann, C.F., Smith, S.M., 2005. Mixture models with adaptive spatial regularization for segmentation with an application to fMRI data. IEEE Trans. Med. Imaging 24, 1–11, http://dx.doi.org/10.1109/TMI.2004.836545.