NeuroImage ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

NeuroImage

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

Regression DCM for fMRI

Stefan Frässle [a,*,1], Ekaterina I. Lomakina [a,b,1], Adeel Razi [c,d], Karl J. Friston [c], Joachim M. Buhmann [b], Klaas E. Stephan [a,c]

[a] Translational Neuromodeling Unit (TNU), Institute for Biomedical Engineering, University of Zurich & ETH Zurich, Wilfriedstrasse 6, 8032, Zurich, Switzerland
[b] Department of Computer Science, ETH Zurich, 8032, Zurich, Switzerland
[c] Wellcome Trust Centre for Neuroimaging, University College London, London WC1N 3BG, United Kingdom
[d] Department of Electronic Engineering, NED University of Engineering & Technology, Karachi, Pakistan

ARTICLE INFO

Article history:
Received 6 November 2016
Accepted 28 February 2017

Keywords:
Bayesian regression
Dynamic causal modeling
Variational Bayes
Generative model
Effective connectivity
Connectomics

ABSTRACT

The development of large-scale network models that infer the effective (directed) connectivity among neuronal populations from neuroimaging data represents a key challenge for computational neuroscience. Dynamic causal models (DCMs) of neuroimaging and electrophysiological data are frequently used for inferring effective connectivity but are presently restricted to small graphs (typically up to 10 regions) in order to keep model inversion computationally feasible. Here, we present a novel variant of DCM for functional magnetic resonance imaging (fMRI) data that is suited to assess effective connectivity in large (whole-brain) networks. The approach rests on translating a linear DCM into the frequency domain and reformulating it as a special case of Bayesian linear regression. This paper derives regression DCM (rDCM) in detail and presents a variational Bayesian inversion method that enables extremely fast inference and accelerates model inversion by several orders of magnitude compared to classical DCM. Using both simulated and empirical data, we demonstrate the face validity of rDCM under different settings of signal-to-noise ratio (SNR) and repetition time (TR) of fMRI data. In particular, we assess the potential utility of rDCM as a tool for whole-brain connectomics by challenging it to infer effective connection strengths in a simulated whole-brain network comprising 66 regions and 300 free parameters. Our results indicate that rDCM represents a computationally highly efficient approach with promising potential for inferring whole-brain connectivity from individual fMRI data.

Introduction

The human brain is organized as a network of local circuits that are interconnected via long-range fiber pathways, providing the structural backbone for the functional cooperation of distant specialized brain systems (Passingham et al., 2002; Sporns et al., 2005). Understanding both structural and functional integration among neuronal populations is indispensable for deciphering mechanisms of both normal cognition and brain disease (Bullmore and Sporns, 2009). Neuroimaging techniques, such as functional magnetic resonance imaging (fMRI), have contributed substantially to this endeavor. While the early neuroimaging era focused mainly on localizing cognitive processes in specific brain areas (functional specialization), the last decade has seen a fundamental shift towards the study of connectivity as the fundament for functional integration (Friston, 2002; Smith, 2012).

Three different aspects of brain connectivity are typically distinguished. First, structural connectivity – that is, anatomical connections such as long-range projections that make up white matter and link cortical and subcortical regions. Structural connectivity is typically inferred from human diffusion-weighted imaging data or from tract tracing studies in animals. Second, functional connectivity describes interactions among neuronal populations (brain regions) as statistical relations. Functional connectivity can be computed in numerous ways, including correlation, mutual information, or spectral coherence (Friston, 2011). Third, effective connectivity is based on a model of the interactions between neuronal populations and how the ensuing neuronal dynamics translate into measured signals.

While structural and functional connectivity methods have provided valuable insights into the wiring and organization of the human brain both in health and disease (for reviews, see Buckner et al. (2013), Bullmore and Sporns (2009), Fornito et al. (2015), Sporns et al. (2005)), they are essentially descriptive and do not

* Corresponding author.
E-mail address: stefanf@biomed.ee.ethz.ch (S. Frässle).
[1] Contributed equally to this work.

http://dx.doi.org/10.1016/j.neuroimage.2017.02.090

allow for mechanistic accounts of a neuronal circuit – that is, what computations are performed and how they are implemented physiologically. By contrast, models of effective connectivity that rest upon a generative model can, in principle, infer upon the latent (neuronal or computational) mechanisms that underlie measured brain activity. In other words, models of effective connectivity seek explanations of the data, not statistical characterizations. This is not only fundamentally important for basic neuroscience, but also offers tremendous opportunities for clinical applications (Stephan et al., 2015).

The last decade has seen enormous interest and activity in developing methods for inferring directed connection strengths from fMRI data, such as Granger causality (GC; Roebroeck et al., 2005) and dynamic causal modeling (DCM; Friston et al., 2003). While GC operates directly on the data and quantifies connectivity in terms of temporal dependencies, DCM rests on a generative model, allowing for inference on latent neuronal states that cause observations (Friston et al., 2013). While these methods have already made fundamental contributions to our understanding of functional integration in the human brain, existing methods are still subject to major limitations (for reviews on strengths and challenges of models of effective connectivity, see Daunizeau et al. (2011a), Friston et al. (2013), Stephan and Roebroeck (2012), Valdes-Sosa et al. (2011)). For example, there is a fundamental trade-off between the complexity of a model and parameter estimability: while biophysical network models (BNMs) capture many anatomical and physiological details (Deco et al., 2013a; Jirsa et al., 2016), their nonlinear functional forms, very large number of parameters, and pronounced parameter interdependencies usually render parameter estimation an extremely challenging computational problem (for discussion, see Stephan et al. (2015)). At present, large-scale BNMs are therefore usually used for simulating data, as opposed to inferring the strength of individual connections.

In contrast to biophysical network models, generative models like DCM rest on a forward model (from hidden neuronal circuit dynamics to measured data) that is inverted using Bayesian principles in order to compute the posterior probability distributions of the model parameters (model inversion). To render this challenge computationally feasible, DCM typically deals with small networks consisting of no more than 10 regions (but see Seghier and Friston (2013)) whose activity has been perturbed by carefully designed experimental manipulations. DCM for fMRI has also been extended to cover resting state fMRI time series by modelling endogenous fluctuations in neuronal activity. These neuronal fluctuations can be treated as hidden or latent neuronal states (leading to stochastic DCM; Daunizeau et al., 2009). Alternatively, the second order statistics of neuronal fluctuations can be treated deterministically within DCM for cross-spectral responses (Friston et al., 2014a, 2014b). Irrespective of the particular form of DCM, the restriction to a small number of nodes can be a major limitation; for example, for clinical applications concerned with whole-brain physiological phenotyping of patients in terms of directed connectivity.

In this paper, we introduce a novel variant of DCM for fMRI that has the potential to overcome this bottleneck and is suitable, in principle, to assess effective connectivity in large (whole-brain) networks. Put simply, the approach rests upon shifting the formulation of DCM from the time to the frequency domain and casting model inversion as a problem of Bayesian regression. More specifically, we reformulate the neuronal state equation of a linear DCM in the time domain as an algebraic expression in the frequency domain. This transformation rests on solving differential equations using the Fourier transformation. Using this approach from the signal processing literature (e.g., Bracewell, 1999; Oppenheim et al., 1999), we show that – under a few assumptions and modifications to the original framework – the problem of model inversion in DCM for fMRI can be cast as a special case of a Bayesian linear regression problem (Bishop, 2006). This *regression DCM* (rDCM) is computationally extremely efficient, enabling its potential use for inferring effective connectivity in whole-brain networks. Note that rDCM is conceptually not unrelated to DCM for cross-spectral responses mentioned above in the sense that both approaches use spectral data features. However, rDCM is formally distinct from cross-spectral DCM because it models the behavior of each system node (in the frequency domain) rather than the cross spectral density as a compact summary (of functional connectivity among system nodes).

In what follows, we first introduce the theoretical foundations of rDCM and highlight where the approach deviates from the standard DCM implementation. We then demonstrate the face validity and practical utility of rDCM for a small six-region network, testing the robustness of both parameter estimation and model selection under rDCM using simulations and an empirical fMRI dataset on face perception (Frässle et al., 2016b, 2016c). Having established the validity of rDCM for small networks, we then proceed to simulations that provide a proof-of-principle for the utility of rDCM for assessing effective connectivity in large networks. The simulations use a whole-brain parcellation (66 regions) and empirical connectivity matrix that was introduced by Hagmann et al. (2008) and has been used by several modeling studies since (e.g., Deco et al., 2013b; Honey et al., 2009), resulting in a model with 300 free parameters.

Methods and materials

Dynamic causal modeling

DCM is a generative modeling framework for inferring hidden neuronal states from measured neuroimaging data by quantifying the effective (directed) connectivity among neuronal populations (Friston et al., 2003). Specifically, DCM explains changes in neuronal population dynamics as a function of the network's connectivity (endogenous connectivity A) and some experimental manipulations. These experimental manipulations u_j can either directly influence neuronal activity in the network's regions (driving inputs C) or perturb the strength of the endogenous connections among regions (modulatory influences B). This can be cast in terms of the following bilinear state equation:

$$\frac{dx}{dt} = \left(A + \sum_{j=1}^{m} u_j B^{(j)}\right)x + Cu \tag{1}$$

This neuronal model is then coupled to a weakly nonlinear hemodynamic forward model that maps hidden neuronal dynamics to observed BOLD signal time series (Buxton et al., 1998; Friston et al., 2000; Havlicek et al., 2015; Stephan et al., 2007). In brief, the hemodynamic model describes how changes in neuronal states induce changes in cerebral blood flow, blood volume and deoxyhemoglobin content. The latter two variables enter an observation equation to yield a predicted BOLD response. For reviews on the biophysical and statistical foundations, see Daunizeau et al. (2011a) and Friston et al. (2013).

Inference proceeds in a fully Bayesian setting, using an efficient variational Bayesian approach under the Laplace approximation (VBL) – meaning that prior and posterior densities are assumed to have a Gaussian fixed form Friston et al. (2007). This scheme provides two estimates: (i) The sufficient statistics of the posterior distributions of model parameters (i.e., conditional mean and covariance), and (ii) the negative free energy, a lower-bound approximation to the log model evidence (i.e., the probability of the

Please cite this article as: Frässle, S., et al., (2017), http://dx.doi.org/10.1016/j.neuroimage.2017.02.090

data given the model). The negative free energy provides a principled trade-off between a model's accuracy and complexity, and serves as a measure for testing competing models (hypotheses) about network architecture by means of Bayesian model comparison (Penny et al., 2004; Stephan et al., 2009a).

Regression DCM

Neuronal state equation in frequency domain

In this section, we introduce a new formulation of DCM that essentially reformulates model inversion as a special case of Bayesian linear regression (for further details of the derivation, see Lomakina (2016)). We thus refer to this approach as *regression DCM* (rDCM). This approach rests on several modifications of the original DCM implementation that include: (i) translation from the time domain to the frequency domain, (ii) linearizing the hemodynamic forward model, (iii) assuming partial independence between connectivity parameters, and (iv) using a Gamma prior for noise precision. These changes allow us to derive an algebraic expression for the likelihood function and a variational Bayesian scheme for inference, which convey a highly significant increase in computational efficiency by several orders of magnitude. This potentially enables a number of innovative applications – most importantly, it renders rDCM a promising tool for studying effective connectivity in whole-brain networks. The massive increase in computational efficiency afforded by rDCM rests on the fact that standard (VBL) inversion schemes require one to integrate a deterministic system of neuronal dynamics to produce a predicted (hemodynamic) response. This integration can be computationally demanding, especially for long time series. The beauty of summarizing a time series (and underlying latent states) with its Fourier transform is that one eludes the problem of solving differential equations, enabling the solution of a compact, static regression model.

In this initial paper, we focus on the simplest case – a linear DCM – because bilinear models aggravate the derivation of an algebraic expression for the likelihood function (but see the Discussion for potential future extensions of rDCM). Linear DCMs are described by the following neuronal state equation

$$\frac{dx}{dt} = Ax + Cu \tag{2}$$

This differential equation can be translated to the frequency domain by means of a Fourier transformation. As the Fourier transform is a linear operator, this results in the following expression

$$\widehat{\frac{dx}{dt}} = A\hat{x} + C\hat{u} \tag{3}$$

where the Fourier transform is denoted by the hat symbol. We can now apply the differential property of the Fourier transform

$$\widehat{\frac{dx}{dt}} = i\omega\hat{x} \tag{4}$$

where $i = \pm\sqrt{-1}$ is the imaginary number and ω the Fourier coordinate. Substituting Eq. (4) into Eq. (3) leads to the representation of the neuronal state equation as an algebraic system in the frequency domain:

$$i\omega\hat{x} = A\hat{x} + C\hat{u} \tag{5}$$

The system described in Eq. (5) is still linear with respect to the model parameters, and the meaning of the parameters is preserved.

Observation model and measurement noise

Having re-expressed the neuronal state equation in the frequency domain, we now turn to the observation model that links hidden neuronal dynamics to measured BOLD signals. In classical DCM, the observation model consists of a cascade of nonlinear differential equations describing the hemodynamics and a nonlinear static BOLD signal equation (Friston et al., 2000; Stephan et al., 2007). These nonlinearities pose a problem for our approach because they prevent a straightforward translation to the frequency domain. One possibility would be to linearize these equations (as in Stephan et al. (2007)). In this initial paper, however, we adopt a simpler approach: convolution with a fixed hemodynamic response function (HRF). Multiplying Eq. (5) with the Fourier transform of the HRF and making use of the fact that a multiplication of the Fourier transforms of two functions is equivalent to the Fourier transform of the convolution of these two functions, one arrives at the following algebraic system:

$$\begin{aligned} i\omega\left(h \circledast x\right) &= A\left(h \circledast x\right) + C\hat{h}\hat{u} \\ i\omega\hat{y}_B &= A\hat{y}_B + C\hat{h}\hat{u} \end{aligned} \tag{6}$$

Here, $\circledast$ denotes the convolution and $\hat{y}_B$ is the deterministic (noise-free) prediction of the data. However, Eq. (6) is not an accurate description for measured fMRI data, for two reasons: First, while the neuronal activity x and BOLD response y_B are continuous signals, our measurements or observations are discrete. Second, measured fMRI data is inevitably affected by noise.

To account for the discrete nature of the data (and the fact that computers can only represent discrete rather than continuous data), we use a discretized version of Eq. (6). This necessitates the use of the discrete Fourier transform (DFT) and the discretization of frequency and time:

$$i\omega := i\boldsymbol{m}\Delta\omega = 2\pi i\frac{\boldsymbol{m}}{NT} \approx \frac{1}{T}\left(e^{2\pi i\frac{\boldsymbol{m}}{N}} - 1\right) \tag{7}$$

where N represents the number of data points, T the time interval between subsequent points, $\Delta\omega$ the frequency interval, and $\boldsymbol{m} = [0, 1, \ldots, N-1]$ a vector of frequency indices. In Eq. (7), we have made use of a linear approximation to the exponential function to obtain the final expression, which is also known as the difference operator of the DFT. Plugging Eq. (7) into Eq. (6) leads to the discrete representation of the (deterministic) BOLD equation in the frequency domain:

$$\left(e^{2\pi i\frac{\boldsymbol{m}}{N}} - 1\right)\frac{\hat{y}_B}{T} = A\hat{y}_B + C\hat{h}\hat{u} \tag{8}$$

where the hat symbol now denotes the discrete Fourier transform.

Having obtained an expression for discrete data, we now augment the model with observation or measurement noise. Here, similar to the setting in classical DCM, we assumed the measurement noise to be white for each region $i = [1, \ldots, R]$ (or, more precisely, the hemodynamic responses at each region to be whitened following an estimation of their temporal autocorrelations) with region-specific noise variances σ_i^2:

$$y_i = y_{B,i} + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}\left(0, \sigma_i^2 I_{N\times N}\right) \tag{9}$$

where $I_{N\times N}$ is the identity matrix. Inserting Eq. (9) into Eq. (8) gives an expression for the measured fMRI signal

$$\left(e^{2\pi i\frac{\boldsymbol{m}}{N}} - 1\right)\frac{\hat{y}}{T} = A\hat{y} + C\hat{h}\hat{u} + \upsilon \tag{10}$$

The form of Eq. (10) is reminiscent of structural equation models (McIntosh, 1998) and multivariate autoregressive models (Roebroeck et al. 2005) in the frequency domain. In Eq. (10), υ is a

noise vector of the following form:

$$\upsilon = \left(e^{2\pi i\frac{\boldsymbol{m}}{N}} - 1\right)\frac{\hat{\epsilon}}{T} - A\hat{\epsilon} \tag{11}$$

While υ also has white noise properties, its dependence on the endogenous connectivity parameters (A matrix) complicates the derivation of an analytical expression for the likelihood function. We circumvent this problem by an approximation, introducing a partial independence assumption that regards υ_i as an independent random vector with a noise precision parameter τ_i. This approximation means that potential dependencies amongst parameters affecting different regions are discarded (i.e., inter-dependencies are only considered for parameters entering the same region). Effectively, this constitutes a mean field approximation in which the (approximate) posterior factorizes among sets of connections providing inputs to each node. This assumption allows for an extremely efficient (variational) inversion of our DCM. Heuristically, because DCM models changes in activity caused by hidden states in other regions, this approximation means that Eq. (10) can estimate the strengths of connections to any given region by, effectively, minimizing the difference between observed changes in responses and those predicted by observed activity elsewhere.

Given this approximation, we can re-write Eq. (10) as a standard multiple linear regression problem:

$$Y = X\theta + \upsilon, \quad \upsilon \sim \mathcal{N}\left(\upsilon|0, \tau^{-1}I_{N\times N}\right) \tag{12}$$

Here, we have defined Y as the dependent variable, X as the design matrix (set of regressors) and θ as the parameter vector as follows:

$$\begin{aligned} p\left(Y \mid \theta, \tau, X\right) &= \prod_{i=1}^{R} \mathcal{N}\left(Y_i \mid X\theta_i, \tau_i^{-1}I_{N\times N}\right) \\ Y_i &:= \left(e^{2\pi i\frac{\boldsymbol{m}}{N}} - 1\right)\frac{\hat{y}_i}{T} \\ X &:= \left[\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_R, \hat{h}\,\hat{u}_1, \hat{h}\,\hat{u}_2, \ldots, \hat{h}\,\hat{u}_K\right] \\ \theta_i &:= \left[a_{i1}, a_{i2}, \ldots, a_{iR}, c_{i1}, c_{i2}, \ldots, c_{iK}\right] \end{aligned} \tag{13}$$

where y_i represents the measured signal in region i and u_k the kth experimental input to that region.

This derivation completes the reformulation of a linear DCM in the time domain as a general linear model (GLM) in the frequency domain. The resulting algebraic expression in Eq. (12) offers many advantages, such as extremely efficient computation and exploitation of existing statistical solutions. The transfer to the frequency domain also means that we can exploit knowledge about the frequencies that contain useful information in fMRI; these are constrained by the low-pass filter properties of neurovascular coupling and the sampling frequency (cf. Nyquist theorem), respectively. This means that sampling rate (TR) becomes an important factor, something that will be considered in our simulations below.

Specification of regression DCM as a generative model

In order to turn the GLM in Eq. (12) into a full generative model (Bayesian regression), we need to specify priors for parameters and hyperparameters. While we keep the zero-mean Gaussian shrinkage priors on connectivity parameters from classical DCM, we chose a Gamma prior on the noise precision τ (not a log-normal prior as in classical DCM). This change was motivated by the fact that Gamma priors serve as conjugate priors on precision for a Gaussian likelihood, which simplifies the derivation of an expression for the posterior:

$$\begin{aligned} p(\theta_i) &= \mathcal{N}\left(\theta_i; \mu_0^i, \Sigma_0^i\right) = (2\pi)^{-\frac{D_i}{2}}\left|\Sigma_0^i\right|^{-\frac{1}{2}} exp^{-\frac{1}{2}\left(\theta_i-\mu_0^i\right)^T\Sigma_0^{i\,-1}\left(\theta_i-\mu_0^i\right)} \\ p(\tau_i) &= Gamma\left(\tau_i; \alpha_0, \beta_0\right) = \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)}\tau_i^{\alpha_0-1}exp^{-\beta_0\tau_i} \end{aligned} \tag{14}$$

Here, μ_0 and Σ_0 are the mean and covariance of the Gaussian prior on connectivity parameters, α_0 and β_0 are the shape and rate parameters of the Gamma prior on noise precision, and Γ is the Gamma function. In this paper, we adopted the standard neuronal priors from DCM10 as implemented in the Statistical Parametric Mapping software package SPM8 (version R4290; www.fil.ion.ucl.ac.uk/spm). For the noise precision, we used α_0=2 and β_0=1 to match the Gamma distribution closely to the first two moments of the standard log-normal prior from DCM10.

Under this choice of priors, the posterior distribution over connections to each region i and for the entire model, respectively, then takes the form:

$$\begin{aligned} p\left(\theta_i, \tau_i|Y_i, X\right) &\propto p\left(Y_i|X, \theta_i, \tau_i\right)p(\theta_i)p(\tau_i) \\ p\left(\theta, \tau|Y, X\right) &\propto \prod_{i=1}^{R} p\left(Y_i|X, \theta_i, \tau_i\right)\prod_{i=1}^{R}\left(p(\theta_i)p(\tau_i)\right) \end{aligned} \tag{15}$$

As already highlighted above, the formulation of rDCM represents a special case of Bayesian linear regression (Bishop, 2006). If the noise precision were known, Eq. (15) could be solved exactly. Since rDCM does not make this assumption, an analytical solution is not possible and approximate inference procedures are needed instead. Here, we chose a variational Bayesian approach under the Laplace approximation (VBL; compare Friston et al. (2007)) to derive an iterative optimization scheme.

Variational Bayes: a brief summary

This section provides a summary of variational Bayes (VB) that hopes to enable readers with limited experience in variational Bayes to follow the derivation of the update equations for rDCM below. Comprehensive introductions to VB can be found elsewhere (e.g., Bishop, 2006). Generally speaking, VB is a framework for transforming intractable integrals into tractable optimization problems. The main idea of this approach is to approximate the true posterior $p(\theta, \tau|y, m)$ by a simpler distribution $q(\theta, \tau|y, m)$. For VBL, $q(\theta, \tau|y, m)$ is assumed to have a Gaussian form and can thus be fully described by its sufficient statistics – that is, the conditional mean and covariance (Friston et al., 2007).

Given such an approximate density, VB allows for achieving two things simultaneously: (i) model inversion, i.e., estimating the best approximation to the true posterior (under the chosen form of q), and (ii) obtaining an approximation to the log model evidence (the basis for Bayesian model comparison). This can be seen by decomposing the log model evidence as follows:

$$\begin{aligned} \ln p(y|m) &= \iint q(\theta, \tau|y, m)\ln p(y|m)\frac{q(\theta, \tau|y, m)}{q(\theta, \tau|y, m)}d\theta d\tau \\ &= \iint q(\theta, \tau|y, m)\ln\frac{p(y, \theta, \tau|m)q(\theta, \tau|y, m)}{p(\theta, \tau|y, m)q(\theta, \tau|y, m)}d\theta d\tau \\ &= \iint q(\theta, \tau|y, m)\ln\frac{p(y, \theta, \tau|m)}{q(\theta, \tau|y, m)}d\theta d\tau \\ &\quad + \iint q(\theta, \tau|y, m)\ln\frac{q(\theta, \tau|y, m)}{p(\theta, \tau|y, m)}d\theta d\tau \\ &= F + KL\left[q(\theta, \tau|y, m)\middle\|p(\theta, \tau|y, m)\right] \end{aligned} \tag{16}$$

where F is known as the negative free energy and the second term is the Kullback-Leibler (KL) divergence between the approximate posterior and the true posterior density. Because the KL divergence is always positive or zero (Mackay, 2003), the negative free energy provides a lower bound on the log model evidence. The KL term can thus be minimized (implicitly) by maximizing the negative free energy. The latter is feasible because F does not depend on the true (but unknown) posterior, but only on the approximate posterior (see Eq. (16)), and can be maximized by gradient ascent (with regard to the sufficient statistics of q).

To facilitate finding the q that maximizes F, a mean field approximation to $q(\theta, \tau|y, m)$ is typically chosen. For example, one might assume that q factorizes into marginal posterior densities of parameters and hyperparameters:

$$q(\theta, \tau|y, m) = q(\theta|y, m)q(\tau|y, m) \tag{17}$$

Under this mean field approximation, the approximate marginal posteriors that maximize F can be found by iteratively applying the following two update equations

$$q(\theta|y, m) = \exp\left[\left\langle \ln p(y, \theta, \tau|m)\right\rangle_{q(\tau|y,m)}\right]$$
$$q(\tau|y, m) = \exp\left[\left\langle \ln p(y, \theta, \tau|m)\right\rangle_{q(\theta|y,m)}\right] \tag{18}$$

where, $\langle\cdot\rangle_q$ denotes the expectation with respect to q. While deriving the right-hand side terms (the so-called "variational energies") can be complicated, once known they enable a very fast optimization scheme.

Variational Bayes for regression DCM

Having outlined the basic concepts of variational Bayes, we now present an efficient VBL approach for rDCM, which results in a set of analytical update equations for the model parameters and an expression for the negative free energy. Under the VBL assumptions, update equations for $q(\theta|Y, X)$ and $q(\tau|Y, X)$ – that is, for connectivity parameters and noise precision, respectively – can be derived as shown by Eqs. (19) and (20).

Given the mean field approximation or factorization of the approximate posterior over subsets of connections (see above), optimization can be performed for each region independently. Technically, this enables us to dissolve the problem of inverting a full adjacency matrix of endogenous connectivity strengths into a series of variational updates in which the posterior expectations of each subset (rows of the A matrix) are optimized successively. Hence, without loss of generality, we restrict the following derivation of the update equations to a single region.

Update equation of θ:

$$\begin{aligned}
\ln q(\theta|Y, X) &= \left\langle \ln p(\theta, \tau, Y|X)\right\rangle_{q(\tau)} + const \\
&= \left\langle \ln \mathcal{N}\left(Y|X\theta, \tau^{-1}I_{N\times N}\right) + \ln \mathcal{N}\left(\theta|\mu_0, \Sigma_0\right)\right\rangle_{q(\tau)} + const \\
&= -\frac{\langle\tau\rangle_{q(\tau)}}{2}\left(Y - X\theta\right)^T\left(Y - X\theta\right) - \frac{1}{2}\left(\theta - \mu_0\right)^T\Sigma_0^{-1}\left(\theta - \mu_0\right) + const \\
&= -\frac{\alpha_{\tau|y}}{2\beta_{\tau|y}}\theta^T X^T X\theta + \frac{\alpha_{\tau|y}}{\beta_{\tau|y}}\theta^T X^T Y - \frac{1}{2}\theta^T\Sigma_0^{-1}\theta + \theta^T\Sigma_0^{-1}\mu_0 + const \\
&= -\frac{1}{2}\theta^T\left(\frac{\alpha_{\tau|y}}{\beta_{\tau|y}}X^T X + \Sigma_0^{-1}\right)\theta + \theta^T\left(\frac{\alpha_{\tau|y}}{\beta_{\tau|y}}X^T Y + \Sigma_0^{-1}\mu_0\right) + const
\end{aligned} \tag{19}$$

where Σ_0^{-1} is the inverse prior covariance matrix on connectivity parameters, and $\alpha_{\tau|y}$ and $\beta_{\tau|y}$ are the posterior shape and rate parameters of the Gamma distribution on noise precision, respectively. Here, we made use of $\langle\tau\rangle_{q(\tau)} = \frac{\alpha_{\tau|y}}{\beta_{\tau|y}}$, with $\langle\cdot\rangle$ denoting the expected value, and the fact that all terms independent of θ can be absorbed by the constant term.

Update equation of τ:

$$\begin{aligned}
\ln q(\tau|Y, X) &= \left\langle \ln p(\theta, \tau, Y|X)\right\rangle_{q(\theta)} + const \\
&= \left\langle \ln \mathcal{N}\left(Y|X\theta, \tau^{-1}I_{N\times N}\right) + \ln Gamma\left(\tau|\alpha_0\beta_0\right)\right\rangle_{q(\theta)} + const \\
&= \frac{N}{2}\ln\tau - \frac{\tau}{2}\left\langle\left(Y - X\theta\right)^T\left(Y - X\theta\right)\right\rangle_{q(\theta)} + \left(\alpha_0 - 1\right)\ln\tau - \beta_0\tau + const \\
&= \frac{N}{2}\ln\tau - \frac{\tau}{2}\left\langle\theta^T X^T X\theta - 2\theta^T X^T Y + Y^T Y\right\rangle_{q(\theta)} + \left(\alpha_0 - 1\right)\ln\tau - \beta_0\tau + const \\
&= \frac{N}{2}\ln\tau - \frac{\tau}{2}\left(Y - X\mu_{\theta|y}\right)^T\left(Y - X\mu_{\theta|y}\right) - \frac{\tau}{2}trace\left(X^T X\Sigma_{\theta|y}\right) \\
&\quad + \left(\alpha_0 - 1\right)\ln\tau - \beta_0\tau + const
\end{aligned} \tag{20}$$

where N is the number of data points, and $\mu_{\theta|y}$ and $\Sigma_{\theta|y}$ are the mean and covariance of the posterior (Gaussian) density on connectivity parameters, respectively. Here, we made use of $\langle\theta\rangle_{q(\theta)} = \mu_{\theta|y}$ and the fact that all terms independent of τ can be absorbed by the constant term. Comparing Eqs. (19) and (20) to the logarithm of the multivariate normal distribution and to the logarithm of the Gamma distribution, respectively, allows one to derive a set of simple update equations for the sufficient statistics of the approximate posterior densities $q(\theta|Y, X)$ and $q(\tau|Y, X)$.

Final iterative scheme:

$$\begin{aligned}
\Sigma_{\theta|y} &= \left(\frac{\alpha_{\tau|y}}{\beta_{\tau|y}}X^T X + \Sigma_0^{-1}\right)^{-1} \\
\mu_{\theta|y} &= \Sigma_{\theta|y}\left(\frac{\alpha_{\tau|y}}{\beta_{\tau|y}}X^T Y + \Sigma_0^{-1}\mu_0\right) \\
\alpha_{\tau|y} &= \alpha_0 + \frac{N}{2} \\
\beta_{\tau|y} &= \beta_0 + \frac{1}{2}\left(Y - X\mu_{\theta|y}\right)^T\left(Y - X\mu_{\theta|y}\right) + \frac{1}{2}trace\left(X^T X\Sigma_{\theta|y}\right)
\end{aligned} \tag{21}$$

Since the update equations for $q(\theta|Y, X)$ and $q(\tau|Y, X)$ are mutually dependent on each other, we iterate their updates until convergence to obtain the optimal approximate distributions. More precisely, in the current implementation of rDCM, the iterative scheme proceeds until the change in τ falls below a specified threshold (i.e., 10^{-10}). In future implementations, we will explore the utility of using changes in variational free energy within each iteration as the criterion for convergence (for comparison, VBL typically uses a change in free energy of 1/8 or less to terminate the iterations).

Having obtained expressions for the approximate posterior densities for the connectivity and noise parameters, one can derive an expression for the negative free energy F. As described above, F serves as a lower-bound approximation to the log model evidence which represents a measure of the "goodness" of a model, taking into account both its accuracy and complexity (Friston et al., 2007; Mackay, 1992; Penny et al., 2004; Stephan et al., 2009a). F is thus routinely used to formally compare different candidate models and decide which of them provides the most plausible explanation for the observed data (Bayesian model selection, BMS). To do so, one needs to

compute the actual value of F, given the data and (a current estimate of the) approximate posterior; the following equations show how this is done in the case of rDCM.

As can be seen from Eq. (16), the negative free energy can be cast in terms of the difference of the expected energy of the system (i.e., log-joint) and the entropy of the approximate posterior:

$$\begin{aligned} F &= max_{q(\theta,\tau|y,X)}\left[-\iint q(\theta,\tau|y,X)\ln\frac{q(\theta,\tau|y,X)}{p(\theta,\tau,Y|X)}d\theta d\tau \right] \\ &= \left\langle \ln p(\theta,\tau,Y|X)\right\rangle_{q(\theta,\tau)} - \left\langle \ln q(\theta,\tau|y,X)\right\rangle_{q(\theta,\tau)} \\ &= \left\langle \ln p(Y|\theta,\tau,X)\right\rangle_{q(\theta,\tau)} + \left\langle \ln p(\theta)\right\rangle_{q(\theta,\tau)} + \left\langle \ln p(\tau)\right\rangle_{q(\theta,\tau)} \\ &\quad - \left\langle \ln q(\theta|Y,X)\right\rangle_{q(\theta,\tau)} - \left\langle \ln q(\tau|Y,X)\right\rangle_{q(\theta,\tau)} \end{aligned} \tag{22}$$

In the following, we outline the derivation of the individual components of the negative free energy (see Lomakina (2016)):

Expectation of the likelihood:

$$\begin{aligned} \left\langle \ln p\,(Y|\theta,\tau,X)\right\rangle_{q(\theta,\tau)} &= \left\langle \ln \mathcal{N}\left(Y|X\theta,\tau^{-1}I_{N\times N}\right)\right\rangle_{q(\theta,\tau)} \\ &= \left\langle -\frac{N}{2}\ln 2\pi - \frac{1}{2}\ln\left|\tau^{-1}I_{N\times N}\right| - \frac{\tau}{2}(Y-X\theta)^T(Y-X\theta)\right\rangle_{q(\theta,\tau)} \\ &= -\frac{N}{2}\ln 2\pi + \frac{N}{2}\langle\ln\tau\rangle_{q(\tau)} - \frac{\langle\tau\rangle_{q(\tau)}}{2}\left\langle (Y-X\theta)^T(Y-X\theta)\right\rangle_{q(\theta)} \\ &= -\frac{N}{2}\ln 2\pi + \frac{N}{2}\left(\Psi(\alpha_{\tau|y}) - \ln\beta_{\tau|y}\right) \\ &\quad - \frac{\alpha_{\tau|y}}{2\beta_{\tau|y}}\left(Y-X\mu_{\theta|y}\right)^T\left(Y-X\mu_{\theta|y}\right) - \frac{\alpha_{\tau|y}}{2\beta_{\tau|y}}trace\left(X^TX\Sigma_{\theta|y}\right) \end{aligned} \tag{23}$$

where Ψ denotes the digamma function.

Expectation of the prior on θ:

$$\begin{aligned} \left\langle \ln p(\theta)\right\rangle_{q(\theta,\tau)} &= \left\langle \ln \mathcal{N}(\theta|\mu_0,\Sigma_0)\right\rangle_{q(\theta,\tau)} \\ &= \left\langle -\frac{D}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_0| - \frac{1}{2}(\theta-\mu_0)^T\Sigma_0^{-1}(\theta-\mu_0)\right\rangle_{q(\theta)} \\ &= -\frac{D}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_0| - \frac{1}{2}\left\langle (\theta-\mu_0)^T\Sigma_0^{-1}(\theta-\mu_0)\right\rangle_{q(\theta)} \\ &= -\frac{D}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_0| - \frac{1}{2}\left(\mu_{\theta|y}-\mu_0\right)^T\Sigma_0^{-1}\left(\mu_{\theta|y}-\mu_0\right) \\ &\quad - \frac{1}{2}trace\left(\Sigma_0^{-1}\Sigma_{\theta|y}\right) \end{aligned} \tag{24}$$

where D is the number of connections entering the region.

Expectation of the prior on τ:

$$\begin{aligned} \left\langle \ln p(\tau)\right\rangle_{q(\theta,\tau)} &= \left\langle \ln Gamma\left(\tau|\alpha_0\beta_0\right)\right\rangle_{q(\theta,\tau)} \\ &= \left\langle \alpha_0\ln\beta_0 - \ln\Gamma(\alpha_0) + (\alpha_0-1)\ln\tau - \beta_0\tau\right\rangle_{q(\tau)} \\ &= \alpha_0\ln\beta_0 - \ln\Gamma(\alpha_0) + (\alpha_0-1)\langle\ln\tau\rangle_{q(\tau)} - \beta_0\langle\tau\rangle_{q(\tau)} \\ &= \alpha_0\ln\beta_0 - \ln\Gamma(\alpha_0) + (\alpha_0-1)\left(\Psi(\alpha_{\tau|y}) - \ln\beta_{\tau|y}\right) - \beta_0\frac{\alpha_{\tau|y}}{\beta_{\tau|y}} \end{aligned} \tag{25}$$

where Γ is the Gamma function.

Entropy of θ:

$$\begin{aligned} -\left\langle \ln q(\theta|Y,X)\right\rangle_{q(\theta,\tau)} &= -\left\langle \ln \mathcal{N}\left(\theta|\mu_{\theta|y},\Sigma_{\theta|y}\right)\right\rangle_{q(\theta,\tau)} \\ &= -\left\langle -\frac{D}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_{\theta|y}| - \frac{1}{2}\left(\theta-\mu_{\theta|y}\right)^T\Sigma_{\theta|y}^{-1}\left(\theta-\mu_{\theta|y}\right)\right\rangle_{q(\theta)} \\ &= \frac{D}{2}\ln 2\pi + \frac{1}{2}\ln|\Sigma_{\theta|y}| + \frac{1}{2}\left\langle \left(\theta-\mu_{\theta|y}\right)^T\Sigma_{\theta|y}^{-1}\left(\theta-\mu_{\theta|y}\right)\right\rangle_{q(\theta)} \\ &= \frac{D}{2}\ln 2\pi + \frac{1}{2}\ln|\Sigma_{\theta|y}| + \frac{1}{2}\left(\mu_{\theta|y}-\mu_{\theta|y}\right)^T\Sigma_{\theta|y}^{-1}\left(\mu_{\theta|y}-\mu_{\theta|y}\right) \\ &\quad + \frac{1}{2}trace\left(\Sigma_{\theta|y}^{-1}\Sigma_{\theta|y}\right) \\ &= \frac{D}{2}(1+\ln 2\pi) + \frac{1}{2}\ln|\Sigma_{\theta|y}| \end{aligned} \tag{26}$$

Entropy of τ:

$$\begin{aligned} -\left\langle \ln q(\tau|Y,X)\right\rangle_{q(\theta,\tau)} &= -\left\langle \ln Gamma\left(\tau|\alpha_{\tau|y}\beta_{\tau|y}\right)\right\rangle_{q(\theta,\tau)} \\ &= -\left\langle \alpha_{\tau|y}\ln\beta_{\tau|y} - \ln\Gamma\left(\alpha_{\tau|y}\right) + \left(\alpha_{\tau|y}-1\right)\ln\tau - \beta_{\tau|y}\tau\right\rangle_{q(\tau)} \\ &= -\alpha_{\tau|y}\ln\beta_{\tau|y} + \ln\Gamma\left(\alpha_{\tau|y}\right) - \left(\alpha_{\tau|y}-1\right)\langle\ln\tau\rangle_{\tau} + \beta_{\tau|y}\langle\tau\rangle_{q(\tau)} \\ &= \alpha_{\tau|y} - \alpha_{\tau|y}\ln\beta_{\tau|y} + \ln\Gamma\left(\alpha_{\tau|y}\right) - \left(\alpha_{\tau|y}-1\right)\left(\Psi\left(\alpha_{\tau|y}\right) - \ln\beta_{\tau|y}\right) \\ &= \alpha_{\tau|y} - \ln\beta_{\tau|y} + \ln\Gamma\left(\alpha_{\tau|y}\right) - \left(\alpha_{\tau|y}-1\right)\Psi\left(\alpha_{\tau|y}\right) \end{aligned} \tag{27}$$

Summing up the components from Eqs. (23)–(27) yields an estimate of the negative free energy for each individual region. The negative free energy for the full model can then be computed by summing over all regions of the model

$$F = \sum_{i=1}^{R} F_i \tag{28}$$

Synthetic data: six-region DCM

We assessed the face validity of rDCM in systematic simulation studies, generating synthetic data for which the ground truth (i.e., the network architecture and parameter values) was known. More precisely, we generated data from 5 synthetic linear DCMs with identical driving inputs but distinct endogenous connectivity architectures (Fig. 1). For all models, two block input regressors u_1 and u_2 served as driving inputs and were specified to elicit activity in x_1 and x_2, respectively. Each activation block lasted 14.5 s and alternated with baseline periods of the same length.

While driving inputs were kept identical across models, varying the endogenous connectivity patterns yielded models of different complexity, with the most complex model consisting of 20 parameters (model 1) and the sparsest model of 12 parameters (model 5). Specifically, for model 1, feedforward and feedback endogenous connections were set between $x_{1/2}$ and $x_{3/4}$, and between $x_{3/4}$ and $x_{5/6}$. Additionally, reciprocal connections were assumed between x_3 and x_4, as well as between x_5 and x_6. Model 2 resulted from model 1 by discarding feedback connections, model 3 and 4 from further removing reciprocal connections either at the highest or intermediate hierarchical level, respectively, and model 5 by considering only feedforward connections from $x_{1/2}$ to $x_{3/4}$, and from $x_{3/4}$ to $x_{5/6}$.

For each of these five models, 20 different sets of observations were generated. To ensure the data were realistic, we sampled the generating ("true") parameter values of each simulation from the posterior distributions of the endogenous and driving input parameters reported in Frässle et al. (2016b). For each set of models and observations, synthetic BOLD data was then simulated under different conditions where we systematically varied the signal-to-noise ratio (SNR=[1, 3, 5, 10, 100]) and repetition time

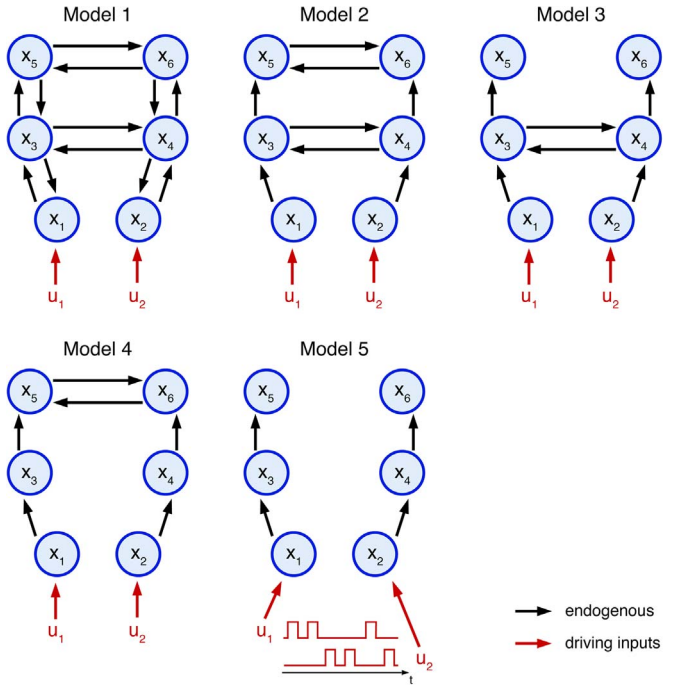

Fig. 1. Five models encoding different effective connectivity patterns of a six-region network utilized for generating synthetic data. For all models, driving inputs (C matrix) were identical – that is, the two block input regressors u_1 and u_2 were assumed to modulate neuronal activity in x_1 and x_2, respectively. On the contrary, endogenous connectivity patterns varied across the five models, ranging from a (relatively) complex model with 20 free parameters to a sparse model with 12 free parameters. For the most complex model (i.e., model 1), feedforward and feedback endogenous connections were set between $x_{1/2}$ and $x_{3/4}$, and between $x_{3/4}$ and $x_{5/6}$. Additionally, reciprocal connections were assumed between x_3 and x_4, as well as between x_5 and x_6. For the sparsest model (i.e., model 5), connections were restricted to feedforward connections from $x_{1/2}$ to $x_{3/4}$, and from $x_{3/4}$ to $x_{5/6}$. The number of free parameters for the remaining models (i.e., models 2–4) ranged between these two "extremes" of the complexity spectrum. Note that the two blocked input regressors u_1 and u_2 shown here represent only a section of the driving input regressors (exemplifying the temporal relationship between the two inputs), rather than the entire time course. The model architecture and the generating parameter values were motivated from a recent study on the effective connectivity in the core face perception network (Frässle et al., 2016b, 2016c).

(TR=[2 s, 1 s, 0.5 s, 0.25 s, 0.1 s]). Here, SNR was defined as the ratio between standard deviation of the signal and standard deviation of the noise (i.e., SNR = $\sigma_{signal}/\sigma_{noise}$), where the noise term is specified as additive white Gaussian noise with zero mean. This definition offers an intuitive measure of the ratio of the variability of signal and noise, is a standard SNR measure in DCM and well established for fMRI analyses more generally (Welvaert and Rosseel, 2013). Under this definition, SNR levels of fMRI time series used for DCM are often 3 or higher; this is because these extracted time series result from a principal component analysis (over numerous voxels in local volumes of interest) that suppresses noise. Evaluating the accuracy of parameter estimation and model selection under the different settings of SNR and TR allowed us to assess the performance of rDCM as a function of data quality and sampling rate, respectively. Note that in all simulations of this initial paper, we used a fixed (canonical) hemodynamic response function. In future work, we will extend the model to account for variations in HRF over regions (see Discussion).

Empirical data: core face perception network

For application of rDCM to empirical fMRI data, we used a previously published fMRI dataset from a simple face perception paradigm (the same dataset from which the generating parameter values in the simulations above were sampled). A comprehensive description of the experimental design and data analysis can be found elsewhere (Frässle et al., 2016b, 2016c); here, we only briefly summarize the most relevant information (Figs. 2–4).

Participants and experimental design

Twenty right-handed subjects viewed either gray-scale neutral faces (F), objects (O), or scrambled (Fourier-randomized) images (S) in the left (LVF) or right visual field (RVF), while fixating a central cross (cf. Fig. 5A). Stimuli were presented in a block design, with each block lasting 14.5 s during which 36 stimuli of the same condition were shown (150 ms, ISI=250 ms). Subsequent stimulus blocks were interleaved with a resting period of the same length where only the fixation cross was shown.

Data acquisition and analysis

For each subject, a total of 940 functional images were acquired on a 3-T MR scanner (Siemens TIM Trio, Erlangen, Germany) using a T_2^*-weighted single-shot gradient-echo echo-planar-imaging (EPI) sequence (30 slices, TR=1450 ms, TE=25 ms, matrix size 64 × 64 voxels, voxel size 3 × 3 × 4 mm^3, FoV=192 × 192 mm^2, flip angle 90°). BOLD activation patterns were analyzed using a first-level GLM (Friston et al., 1995) to identify brain regions sensitive to the processing of faces ([2*F]-[O+S]), as well as to the visual field baseline contrasts (RVF, LVF). Six regions of interest (ROIs) were selected, representing occipital face area (OFA; Puce et al., 1996), fusiform face area (FFA; Kanwisher et al., 1997), and primary visual cortex (V1), each in both hemispheres (Fig. 5A). Peak coordinates of the ROIs were identified for each subject individually (to account for inter-subject variability in the exact locations). From the individual ROIs, time series were extracted (removing signal mean and correcting for head movements), which then entered rDCM analyses.

rDCM analysis

The endogenous and driving input connectivity of the DCM was specified as follows (Fig. 5B; model A): First, intra-hemispheric endogenous forward connections were set between V1 and OFA, and between OFA and FFA. Furthermore, reciprocal inter-hemispheric endogenous connections were set among the homotopic face-sensitive regions (Catani and Thiebaut de Schotten, 2008; Park et al., 2008; Van Essen et al., 1982; Zeki, 1970). Second, inputs representing the visual field of stimulus presentation drove neuronal activity in the contralateral V1 (i.e., RVF influenced left V1, LVF influenced right V1). Third, driving inputs representing the presentation of faces (FP) elicited activity in the face-sensitive areas OFA and FFA in both hemispheres (Frässle et al., 2016c).

Synthetic data: whole-brain DCM

In a final simulation analysis, we assessed the ability of rDCM to infer effective connectivity in a large (whole-brain) network. To this end, synthetic data was generated from a linear DCM including 66 brain regions (Fig. 6A and Supplementary Table S1). The network was defined on the basis of the Hagmann parcellation (Hagmann et al., 2008), which has been utilized frequently for whole-brain connectomics (e.g., Deco et al., 2013b; Honey et al., 2009). To adequately capture the network characteristics of the human brain – for instance, with regard to small-world architecture, node degree, path length, centrality of nodes, or modularity (Bullmore and Sporns, 2009) – the endogenous connectivity architecture of our whole-brain DCM was based on the average structural connectome provided by the diffusion-weighted imaging work by Hagmann et al. (2008). Specifically, we used the matrix of average inter-regional fiber densities (Fig. 4 in

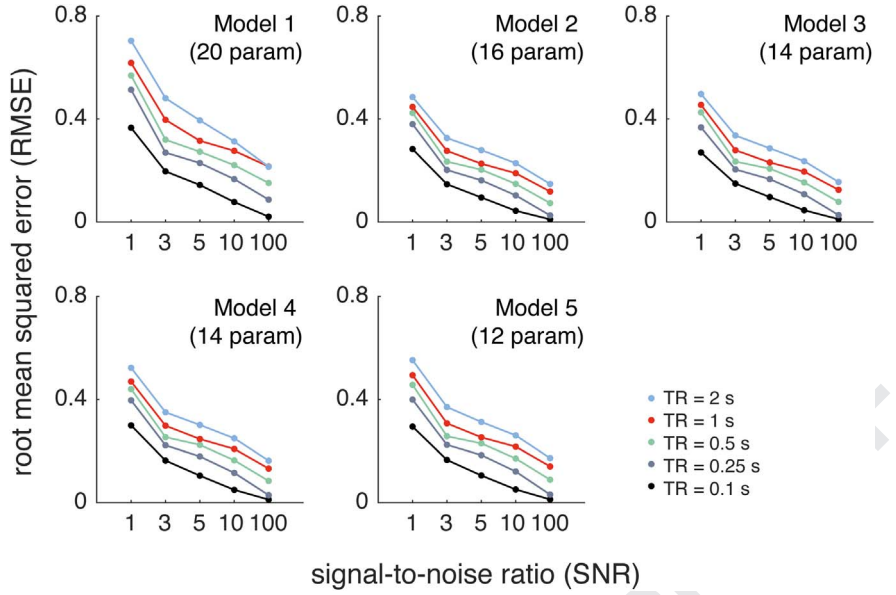

Fig. 2. Parameter recovery of rDCM in terms of the root mean squared error (RMSE). Each of the five subplots illustrates the results for one model (see Fig. 1 for a visualization of the model space). Within each subplot, the RMSE is shown for various combinations of the signal-to-noise ratio (SNR) and the repetition time (TR) of the synthetic fMRI data. The various settings of SNR (i.e., 1, 3, 5, 10, and 100) are shown along the x-axis of each subplot. The different TR settings are illustrated by the differently colored curves and were as follows: 2 s (blue), 1 s (red), 0.5 s (green), 0.25 s (grey), and 0.1 s (black). Results show a clear (and expected) dependence of the RMSE on SNR and TR, with parameter recovery becoming more accurate for better data quality (i.e., higher SNR) and higher sampling rates (i.e., shorter TR).

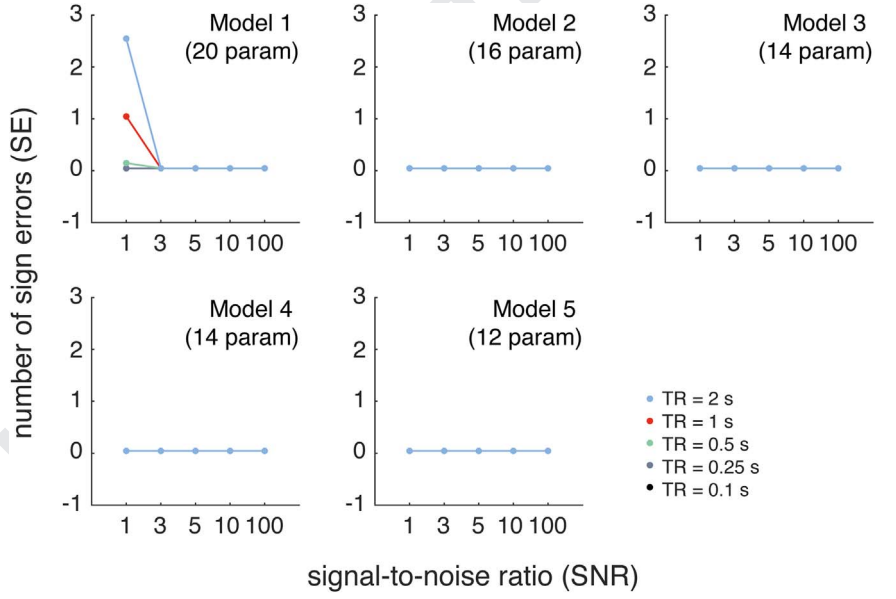

Fig. 3. Parameter recovery of rDCM in terms of the number of sign errors (SE). Each of the five subplots illustrates the results for one model (see Fig. 1 for a visualization of the model space). Within each subplot, the number of sign errors is shown for various combinations of the signal-to-noise ratio (SNR) and the repetition time (TR) of the synthetic fMRI data. The various settings of SNR (i.e., 1, 3, 5, 10, and 100) are shown along the x-axis of each subplot. The different TR settings are illustrated by the differently colored curves and were as follows: 2 s (blue), 1 s (red), 0.5 s (green), 0.25 s (grey), and 0.1 s (black). Results indicate that rDCM recovers whether a connection was excitatory or inhibitory with high precision, with sign errors only occurring for the most complex model (model 1). Notably, for model 1, sign errors were only observed for challenging noise scenarios (i.e., SNR = 1). In all other cases, rDCM accurately recovered the sign of the true generating parameter.

Hagmann et al. (2008)) and included all connections with a weight larger than 0.06. This threshold ensured that, under randomly sampling connection strengths from the prior densities, the system remained stable (i.e., all eigenvalues of the endogenous connectivity matrix were negative). As diffusion-weighted imaging does not allow for detecting the directionality of fibers, connected nodes were always coupled by reciprocal connections (i.e., two separate parameters). This resulted in 298 connections, each of which was represented by a free parameter in the endogenous connectivity (A) matrix of rDCM.

Additionally, two block input regressors, mimicking the presentation of visual stimuli in the left (LVF) and right visual field (RVF), served as driving inputs, driving neuronal activity in the right and left cuneus (primary visual cortex), respectively. In total, this resulted in 300 neuronal parameters that had to be estimated by rDCM (Fig. 6A).

For this model, 25 simulations with 20 observations ("subjects") each were created by sampling the generating parameter values

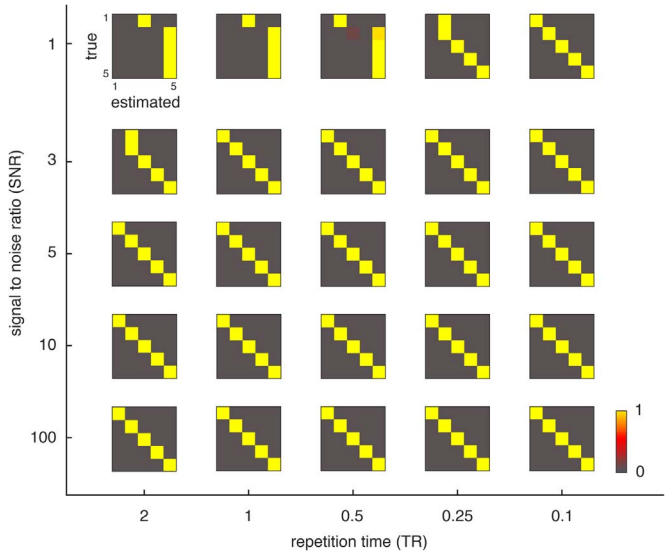

Fig. 4. Accuracy of Bayesian model comparison for rDCM. Each subplot illustrates the model comparison results for a specific combination of the signal-to-noise ratio (SNR) and the repetition time (TR) of the synthetic fMRI data. The various settings of TR (i.e., 2 s, 1 s, 0.5 s, 0.25 s, and 0.1 s) are shown along the x-axis and the different SNR settings (i.e., 1, 3, 5, 10, and 100) are shown along the y-axis. For each combination of TR and SNR (i.e., each subplot), a matrix is shown that summarizes the fixed effects Bayesian model selection results for each of the five different models (see Fig. 1 for a visualization of the model space). Specifically, each row in these matrices represents the posterior model probabilities of all DCMs that were used for model inversion (estimated) of the DCM that was used to actually generate the synthetic fMRI data (true). Hence, each row signals whether rDCM was able to recover the true data-generating model architecture among the five competing alternatives. Hence, a diagonal structure (i.e., highest posterior probability on the diagonal) indicates that rDCM was able to recover the model that actually generated the data. Note that higher posterior probabilities are color-coded in warm colors (yellowish).

from the prior density over model parameters (multivariate normal distribution, see above). The 25 simulations differed with regard to the SNR (i.e., 1, 3, 5, 10, and 100) and TR (i.e., 2 s, 1 s, 0.5 s, 0.25 s, and 0.1 s) settings that were used for generating synthetic BOLD data.

Results

Synthetic data: six-region DCM

Model parameter estimation

First, we tested, using 5 different models and under various settings of SNR and TR, whether rDCM can reliably recover the generating ("true") parameter values in a small network of 6 regions (Fig. 1). We quantified the accuracy of parameter recovery by (i) the root mean squared error (RMSE) between true and estimated parameter values, and (ii) the number of sign errors (SE), that is, the number of parameters for which the estimated sign differed from ground truth. The latter is a metric of interest because the interpretation of effective connectivity often boils down to whether directed influences are excitatory (positive) or inhibitory (negative).

As expected, we found a dependence of the RMSE on both the SNR and TR, with the overall pattern being highly consistent across the different models (Fig. 2): RMSE decreased with higher SNR and shorter TR (higher sampling rate). Notably, in the case of high SNR data (SNR = 100) and ultra-fast data acquisition (TR = 0.1 s), rDCM recovered the connection strengths of the generating parameters almost perfectly (mean RMSE $\leq$ 0.02, for all models). While these settings are not realistic for fMRI experiments, this is an important observation because it serves as a sanity check for our rDCM implementation. For more realistic settings (TR = 1 s, SNR = 3), we found the mean RMSE to range from 0.28 ± 0.01 (mean $\pm$ std) for model 2 to 0.40 ± 0.02 for model 1. With regard to sign errors, rDCM could recover whether a connection was excitatory or inhibitory with high precision (Fig. 3). More precisely, for four models (i.e., models 2–5), we have not observed any sign errors regardless of the SNR and TR. For model 1, sign errors occurred only for an SNR of 1, which represents a challenging SNR scenario (Welvaert and Rosseel, 2013).

For comparison, we also assessed the accuracy of parameter recovery using the default VBL implementation in DCM10 (as implemented in SPM8, version R4290). In brief, VBL recovered the generating parameter values with high accuracy with a mean RMSE $\leq$ 0.2 and hardly any sign error, regardless of the particular model. Sign errors only occurred for model 2 for the most challenging scenario (TR = 2 s, SNR = 1). This excellent performance of VBL is not surprising given that VBL does not need to resort to simplifying assumptions as made in the current implementation of rDCM (in particular, assumptions of partial independence between connectivity parameters). These assumptions affect the accuracy of rDCM most severely for challenging scenarios with low SNR and long TR. On the other hand, our analyses also indicated that for short TRs ($\leq$ 0.5 s), and thus cases with a large number of data points, VBL rarely converged within the limit of SPM8's default upper bound on iterations, which deteriorates parameter recovery performance. Overall, our simulations demonstrate that for realistic settings for SNR and TR, rDCM and VBL yield qualitatively comparable results in terms of parameter recovery.

Bayesian model selection

In a next step, we aimed to establish the face validity of rDCM with regard to Bayesian model selection under the different settings of SNR and TR. To this end, we tested whether the model that actually generated the data (the "true" model) was assigned the largest model evidence – that is, for each of the 25×5 synthetic datasets (with 20 observations/"subjects" for each dataset), we inverted the five different DCMs and compared the negative free energies by means of fixed effects BMS (Stephan et al., 2009a). To this end, we computed the posterior model probability of the estimated model (Fig. 4). Notably, to rule out that any of our BMS results were confounded by outliers (against which fixed effects analyses are vulnerable), we additionally compared negative free energies by means of random effects BMS (Stephan et al., 2009a; as implemented in SPM12, version: R6685) and found highly consistent results (data not shown).

As above, we observed the expected dependence of model selection performance on SNR and sampling rate. Specifically, model selection became more accurate for higher SNR and shorter TR. For challenging scenarios with low signal-to-noise ratios (i.e., SNR = 1), rDCM frequently failed to identify the correct model, except for extremely fast data acquisitions (TR = 0.1 s) where we find perfect recovery of the data-generating model architecture (Fig. 4, top row). More specifically, in the case of noisy data, rDCM showed a tendency to selecting the simplest of all models in our model space (Model 5) for TR $\geq$ 0.5 s. This "Bayesian illusion" – where a simpler model, nested in a more complex data-generating model, has higher evidence – is not an infrequent finding when dealing with nested models that are only distinguished by parameters with weak effects or strongly correlated parameters, whose effects become difficult to detect in the presence of noise.

Having said this, given a reasonably high signal-to-noise ratio in the synthetic fMRI data (i.e., SNR $\geq$ 3), rDCM recovered the true model in the vast majority of cases with hardly any model selection error (Fig. 4, rows 2–5). Even for relatively slow data acquisitions (TR = 2 s), there was only one case for which a "wrong"

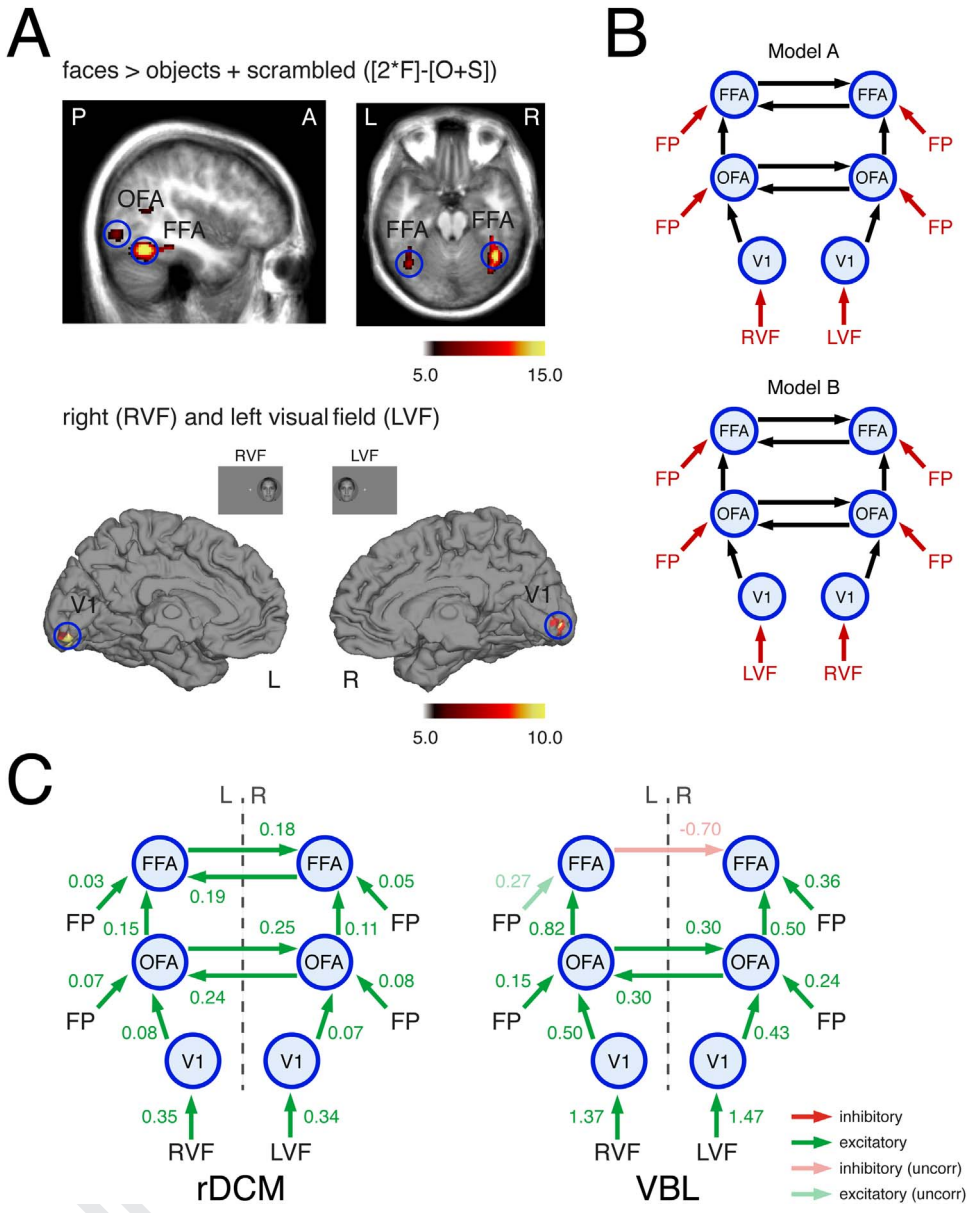

Fig. 5. Effective connectivity in the core face perception network as assessed with rDCM for an empirical fMRI dataset. **(A)** BOLD activation pattern shows regions that were more activated during the perception of faces as compared to objects and scrambled images, as determined by the linear face-sensitive contrast: [2*F]-[O+S]. Reproduced, with permission, from Frässle et al. (2016b) (*top*), as well as regions that were activated when stimuli were presented in the right (*bottom, left*) or left visual field (*bottom, right*). Results are thresholded at a voxel-level threshold of $p < 0.05$ (FWE-corrected). **(B)** Two alternative models for explaining effective connectivity in the core face perception network. Both models assumed the same endogenous connectivity (A matrix) – that is, intra-hemispheric feedforward connections from V1 to OFA and from OFA to FFA in both hemispheres, as well reciprocal inter-hemispheric connections among the face-sensitive homotopic regions. Additionally, both models assumed driving inputs (C matrix) to the four face-sensitive regions (i.e., OFA and FFA, each in both hemispheres) by the processing of faces (FP). Critically, model A (*top*) and model B (*bottom*) differed in their driving inputs to left and right V1. While model A was biologically plausible by assuming that stimuli in the left (LVF) and right visual field (RVF) modulated activity in the contralateral V1, model B assumed these driving inputs to be swapped. **(C)** Group level parameter estimates for the endogenous and driving input connectivity of model A as estimated using rDCM (*left*) and VBL (*right*). Results are remarkably consistent across the two methods. The strength of each connection is displayed in terms of the mean coupling parameter (in [Hz]). Significant ($p < 0.05$, Bonferroni-corrected) connections are shown in full color; connections significant at an uncorrected threshold ($p < 0.05$) are shown in faded colors. L=left hemisphere; R=right hemisphere; A=anterior; P=posterior; LVF=left visual field; RVF=right visual field.

model had higher model evidence compared to the generating ("true") model. More specifically, in the case of model 1 being the "true" model, rDCM falsely assigned highest evidence to model 2, suggesting that the presence of feedback connections can sometimes be difficult to detect – a finding which has been highlighted previously (Daunizeau et al., 2011b).

Again, we compared rDCM to VBL by assessing model selection performance for the default implementation of DCM10. As expected, VBL recovered the true model perfectly with no model selection error. Consistent with our observations on parameter recovery reported above, differences between rDCM and VBL were thus only observed for challenging scenarios (SNR=1), where the limitations of the current version of rDCM are likely to become most severe. On the contrary, both rDCM and VBL provide accurate model selection results for realistic SNR and TR settings.

Computational burden

To illustrate the computational efficiency of rDCM, we compared run-times for rDCM with the time required to perform the respective model inversion using the default VBL implementation in DCM10 (as implemented in SPM8, version R4290). Specifically, we evaluated the run-times for all different settings of TR, under a

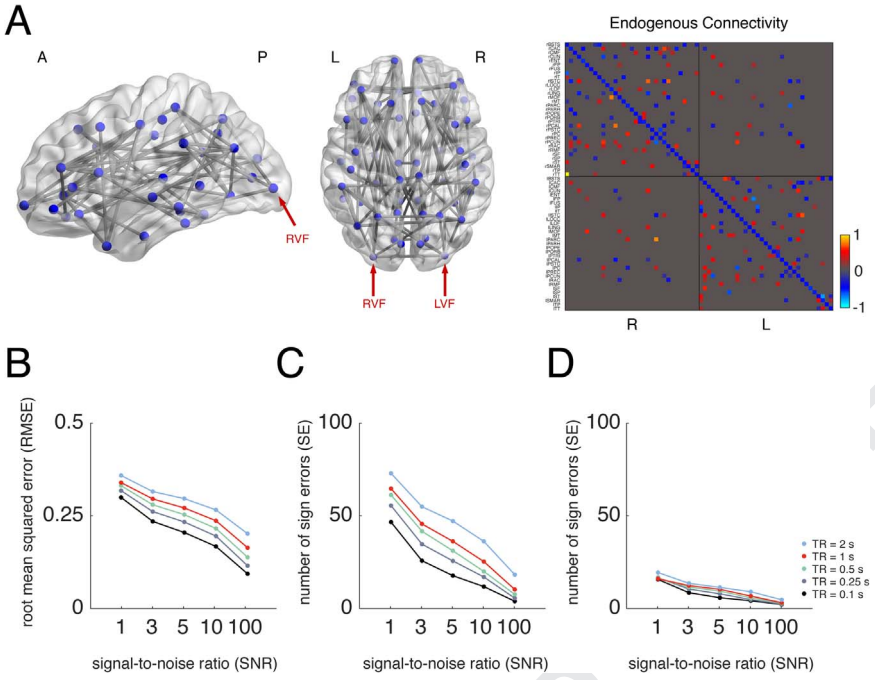

Fig. 6. Parameter recovery of rDCM in terms of the root mean squared error (RMSE) and the number of sign errors (SE) for the large (whole-brain) network. **(A)** Endogenous connectivity architecture (A matrix) among the 66 brain regions from the Hagmann parcellation. Endogenous connectivity was restricted to the most pronounced edges of the human structural connectome by only selecting those connections for which an average inter-regional fiber densities larger than 0.06 has been reported in Hagmann et al. (2008). Additionally, two block input regressors u_1 and u_2, mimicking the effect of visual stimulation in the right and the left visual field, were assumed to modulate neuronal activity in left and right cuneus (primary visual cortex), respectively. The brain network was visualized with the BrainNet Viewer (Xia et al., 2013), which is available as open-source software for download (http://www.nitrc.org/projects/bnv/) (*left*). An actual "observation" of the endogenous connectivity, generated by sampling connection strengths from the prior density on the endogenous parameters (*right*). A complete list of the anatomical labels of the 66 parcels can be found in the Supplementary Table S1. **(B)** The RMSE and **(C)** the number of sign errors are shown for various combinations of the signal-to-noise ratio (SNR) and the repetition time (TR) of the synthetic fMRI data. The various settings of SNR (i.e., 1, 3, 5, 10, and 100) are shown along the x-axis of each subplot. The different TR settings are illustrated by the differently colored curves and were as follows: 2 s (blue), 1 s (red), 0.5 s (green), 0.25 s (grey), and 0.1 s (black). **(D)** Number of sign errors (SE) for the different SNR and TR settings when restricting the analysis to parameter estimates that showed a non-negligible effect size (i.e., the 95% Bayesian credible interval of the posterior not containing zero). For these parameters, the number of SE was considerably reduced, suggesting that the sign of an endogenous influences (i.e., inhibitory vs. excitatory) could be adequately recovered for parameters of large effect size. L=left hemisphere; R=right hemisphere; A=anterior; P=posterior; LVF=left visual field; RVF=right visual field.

fixed SNR of 3 because we expected SNR to exert a smaller impact on run-times as compared to TR, which essentially determines the number of data points. Our run-time results should be interpreted in a comparative, not absolute, manner, given their dependency on computer hardware.

Generally, model inversion under rDCM was considerably faster than VBL across all TR values (Table 1). For instance, for TR=1 s, rDCM was on average four orders of magnitude faster than VBL. Additionally, our analyses suggested that the computational efficiency and feasibility of VBL-based DCM for large numbers of data points (as would also be the case for DCMs with many regions) rapidly diminishes. This behavior was not only reflected by the long run-times reported in Table 1, which ranged from approximately 1,000–40,000 s per model inversion, but also (as mentioned above) by the fact that for short TRs (and hence many data points), the VBL algorithm rarely converged within the limit of SPM8's default upper bound on iterations (see the results for TR $\leq$ 0.5 s in Table 1).

Table 1
Computational burden of rDCM and VBL quantified in terms of the approximate run-times (in s) required for estimating a single linear DCM (i.e., model 1–5, here labelled as m1–m5, respectively) under various settings of the repetition time (TR) of the synthetic fMRI data. Note that run-times are reported exemplarily for a realistic signal-to-noise ratio (SNR) of 3. Run-times are given as the mean ± standard deviation of the 20 simulations, as well as the range.

		Computational burden (s)									
		TR=2 s		TR=1 s		TR=0.5 s		TR=0.25 s		TR=0.1 s	
		Mean ± std	Range	Mean ± std	Range	Mean ± std	Range	Mean ± std	Range	Mean ± std	Range
rDCM	m1	0.20 ± 0.12	0.14–0.72	0.20 ± 0.12	0.14–0.69	0.23 ± 0.13	0.17–0.79	0.24 ± 0.12	0.19–0.74	0.37 ± 0.41	0.26–2.10
	m2	0.18 ± 0.11	0.13–0.63	0.19 ± 0.10	0.14–0.62	0.21 ± 0.12	0.15–0.72	0.26 ± 0.12	0.20–0.78	0.37 ± 0.37	0.26–1.94
	m3	0.16 ± 0.11	0.11–0.62	0.16 ± 0.009	0.12–0.53	0.18 ± 0.11	0.13–0.62	0.20 ± 0.10	0.15–0.62	0.33 ± 0.32	0.23–1.70
	m4	0.18 ± 0.11	0.12–0.62	0.19 ± 0.10	0.14–0.61	0.21 ± 0.12	0.15–0.74	0.27 ± 0.14	0.20–0.85	0.37 ± 0.38	0.26–1.98
	m5	0.19 ± 0.11	0.12–0.67	0.20 ± 0.11	0.15–0.68	0.23 ± 0.14	0.17–0.82	0.27 ± 0.13	0.20–0.81	0.38 ± 0.40	0.26–2.08
VBL	m1	1157 ± 477	830–2498	3589 ± 1153	2432–5448	7932 ± 1169	6303–9986	17704 ± 1124	16484–19143	43633 ± 3904	40449–51119
	m2	1970 ± 563	893–2452	3566 ± 597	2380–4048	7811 ± 939	6859–10349	14804 ± 760	14229–16357	37322 ± 2367	35248–41464
	m3	1592 ± 436	1075–2294	3497 ± 567	2507–4209	7228 ± 444	6715–7935	13741 ± 801	13198–16924	35738 ± 3265	32106–43923
	m4	1177 ± 277	987–2283	2649 ± 469	2214–4295	6654 ± 693	5946–8289	14301 ± 1265	13387–18147	33791 ± 1838	32455–37599
	m5	819 ± 55	733–920	2166 ± 161	1930–2574	5725 ± 560	5217–7184	13691 ± 1165	12422–15907	32014 ± 1934	30354–35322

By comparison, rDCM handles large amounts of data more gracefully due to the algebraic form of the likelihood function. In rDCM, model inversion took less than half a second, with no significant increase in computation time when increasing the number of data points. This highlights the computational efficiency of rDCM and points towards its suitability for studying effective connectivity in very large networks, a theme we return to below.

It should be noted that in our comparative evaluation of rDCM and VBL, there is an additional computational saving under rDCM's fixed form assumptions for the HRF. In other words, conventional DCM optimizes the parameters of the hemodynamic response function separately for each region. This means that the number of parameters – that determines the number of solutions or integrations required to estimate free energy gradients – increases linearly for hemodynamic parameters and quadratically for neuronal (connectivity) parameters. This means that the VBL analyses could be made more efficient by adopting the rDCM (fixed form) assumptions for the HRF; however, the computational savings would not be very marked because the key (quadratic) determinant of computation time depends upon the number of connections.

Empirical data: core face perception network

Model parameter estimation

We applied rDCM to an empirical fMRI dataset of a simple face perception task, which had been used previously to investigate intra- and inter-hemispheric integration in the core face perception network (Frässle et al., 2016a, 2016b, 2016c). Individual connectivity parameters were estimated using model A (Fig. 5B, top), which then entered summary statistics at the group level (one-sample *t*-tests, Bonferroni-corrected for multiple comparisons). We found all parameter estimates to be excitatory except for the inhibitory self-connections (Fig. 5C, left). Specifically, we found excitatory driving influences of visual stimuli (LVF and RVF) on activity in right and left V1, respectively. Additionally, and in line with the well-established role of bilateral OFA and FFA in face processing, we found excitatory face-specific driving influences on activity in all four regions of the core face perception network. With regard to the endogenous connectivity, we found excitatory connections both within each hemisphere and between homotopic face-sensitive regions in both hemispheres, suggesting inter-hemispheric integration within the core face perception network – in line with previous observations from functional (Davies-Thompson and Andrews, 2012) and effective connectivity studies (Frässle et al., 2016c).

In an additional analysis step, we assessed the effective connectivity in the same model (i.e., model A) using DCM10 (as implemented in SPM8, version R4290) in order to compare results from rDCM and VBL qualitatively. Note that a quantitative match of parameter estimates by rDCM and VBL cannot be expected since the generative models of these two frameworks are quite different. Three major differences are worth reiterating: First, while the hemodynamic model in classical DCM is nonlinear and contains region-specific parameters, rDCM models the hemodynamic response as a fixed, linear convolution of neuronal states. Second, rDCM uses a mean field approximation (or partial independence assumption) that ignores potential dependencies amongst parameters affecting different regions. Third, in contrast to the log-normal prior on noise variance in classical DCM, rDCM uses a Gamma prior on noise precision. These differences in likelihood functions and priors between rDCM and VBL translate into quantitatively different posterior estimates (see Eq. (15)).

Qualitatively, however, parameter estimates by VBL were very similar to rDCM (Fig. 5C, right). In brief, all six driving inputs were excitatory (although this was not significant for the face-specific driving input to left FFA when correcting for multiple comparisons). Similarly, the intra-hemispheric forward connections in both hemispheres and the inter-hemispheric connections among bilateral OFA were excitatory. One difference between VBL and rDCM parameter estimates concerned the inter-hemispheric connections among bilateral FFA, which were inhibitory for VBL. These inhibitory effects, however, did not reach significance for the connection from right to left FFA ($t_{(19)}$=-1.43, p=0.17), and only at an uncorrected statistical threshold for the connection from left to right FFA ($t_{(19)}$=-2.79, p=0.01).

Bayesian model selection

In order to evaluate the Bayesian model selection (BMS) performance of rDCM in an empirical setting, we constructed a second model (model B; Fig. 5B, bottom) and inverted this model under rDCM and VBL, respectively. Importantly, in this alternative model, the visual baseline driving inputs were permuted compared to the original model (model A). In contradiction to neuroanatomy, model B thus proposed that visual stimuli presented in the periphery entered ipsilateral V1 (i.e., LVF influenced left V1, RVF influenced right V1). Hence, we expected both rDCM and VBL to select model A as the winning model.

We used random effects BMS (Stephan et al., 2009a; as implemented in SPM12, version: R6685) to compare the two alternative models based on their negative free energies. For both, rDCM and VBL, model A was the decisive winning model with a protected exceedance probability of 1.00 in either case (Rigoux et al., 2014). This indicates that rDCM not only yields BMS results comparable to VBL (with equally high confidence), but also selects the expected and biologically more plausible model amongst two competing hypotheses.

Computational burden

We evaluated the run-time of rDCM and VBL for both models A and B. Again, we would like to highlight that the reported values should be interpreted in a comparative, not absolute, manner as they depend on the specific hardware and software settings. Consistent with our previous observations in the context of simulations, we found model inversion under rDCM to be on average three orders of magnitude faster than VBL (Table 2). More precisely, rDCM was highly efficient taking less than half a second per model, whereas the time required for model inversion under VBL was on the order of 15 min.

Regression DCM for large-scale networks

Model parameter estimation

In a final step, we used simulations to evaluate the utility of rDCM for inferring effective connectivity in a large (whole-brain) network. We chose a network comprising 66 brain regions and 300 free connectivity parameters (Fig. 6A), where model structure

Table 2
Computational burden of rDCM and VBL quantified in terms of the approximate run-times (in s) required for estimating a single linear DCM (i.e., model A or model B) of the intra- and inter-hemispheric connectivity in the core face perception network. The run-times required for inverting the model are given as the mean ± standard deviation of the 20 subjects, as well as the range.

		Computational burden (*S*)	
		Mean ± std	Range
rDCM	Model A	0.24 ± 0.03	0.21–0.28
	Model B	0.26 ± 0.04	0.21–0.32
VBL	Model A	827.3 ± 233.4	519.7–1422.2
	Model B	973.3 ± 371.7	491.6–1706.3

was based on the structural connectome provided by the diffusion-weighted imaging work by Hagmann et al. (2008). As above, we computed the root mean squared error (RMSE) and the number of sign errors (SE) to quantify the accuracy of parameter recovery.

Consistent with our previous findings from the synthetic and empirical dataset, we observed a dependence of both the RMSE and SE on SNR and TR, indicating that parameter estimation improved with increasing data quality (i.e., higher SNR) and sampling rates (i.e., shorter TR). For a realistic setting of fMRI data (TR = 1 s, SNR = 3), the RMSE was 0.29 ± 0.01 (Fig. 6B). In the case of high SNR data (SNR = 100) and ultra-fast data acquisition (TR = 0.1 s), the RMSE was 0.09 ± 0.02. While the RMSE for this case of "ideal data" is larger than in the simulations using the much smaller six-region network above, errors are still in an acceptable range, indicating a promising scalability of rDCM.

It is worth highlighting that even for "ideal data", one would not expect rDCM (nor VBL or any other model inversion method) to exactly recover the "true" parameter values that were used to simulate the data. This is because Bayesian methods optimize the posterior rather than the likelihood, and the influence of the prior exerts a bias on model parameter recovery whenever the prior mean does not coincide exactly with the parameter values used for data generation. This can produce counterintuitive results: even when both prior and likelihood have means of the same sign, parameter dependencies (which arise from the mathematical form of the likelihood function) can lead to a posterior mean of the opposite sign (see Supplementary Fig. S1 for a graphical visualization). For pronounced parameter interdependencies (which are unavoidable in large-scale models with hundreds of free parameters), this sign flipping can occur even when prior mean and data mean (likelihood) are identical, provided their means are not too far away from zero.

Given these considerations, it was unsurprising to find that rDCM of the whole-brain model suffered from considerably more sign errors than the small six-region DCMs described above (Fig. 6C). Again, this was a function of SNR and TR: While only 4.0 ± 1.9 sign errors occurred for high quality data (TR = 0.1 s, SNR = 100), we observed 45.7 ± 4.6 sign errors in more realistic settings (TR = 1 s, SNR = 3), corresponding to an error rate of $15.2 \pm 1.5\%$.

As illustrated by Supplementary Fig. S1, the likelihood of sign flipping having occurred is smaller for parameters whose posterior mean deviates strongly from zero. In a second analysis, we therefore restricted the evaluation of the sign errors to those connections for which zero was not within the 95% Bayesian credible interval of the posterior density. In this way, we asked whether estimates of connections that provided sufficiently large evidence for an effect could be trusted (in terms of revealing the correct direction of influence). When focusing on these connections, sign errors were considerably reduced (Fig. 6D). Specifically, even for (relatively) slow image acquisitions and noisy data (TR = 2 s; SNR = 1), the number of sign errors was in an acceptable range (19.5 ± 5.1, corresponding to an error rate of $6.5 \pm 1.7\%$). For more realistic image acquisition and SNR settings (TR = 1 s; SNR = 3), the number of sign errors reduced to 12.4 ± 3.3 (error rate of $4.1 \pm 1.1\%$). Ultimately, when approaching ideal data (TR = 0.1 s; SNR = 100), hardly any sign error was observed (2.2 ± 1.1, corresponding to an error rate of $0.7 \pm 0.4\%$). This suggests that parameter estimates representing a non-trivial effect size (i.e., the 95% Bayesian credible interval not containing zero) correctly indicate the direction of influences, rendering rDCM a meaningful tool for inferring effective connectivity patterns in large (whole-brain) networks.

Computational burden

We evaluated the run-time of rDCM for the whole-brain DCM for all possible combinations of SNR and TR. Again, the reported values should be interpreted in a qualitative, not absolute, manner as they depend on the specific hardware and software settings. We found model inversion to be extremely efficient even for such a large number of brain regions and free parameters. More specifically, estimation of the model using rDCM took on average 2–3 s (Table 3), suggesting that our approach scales easily with the number of brain regions (data points) and, thus, makes inference on the effective connectivity in large (whole-brain) networks computationally feasible.

Discussion

In this paper, we have introduced regression DCM (rDCM) for functional magnetic resonance imaging (fMRI) data as a novel variant of DCM that enables computationally highly efficient analyses of effective connectivity in large-scale brain networks. This development rests on reformulating a linear DCM in the time domain as a special case of Bayesian linear regression (Bishop, 2006) in the frequency domain, together with a highly efficient VB inference scheme. Using synthetic and empirical data, we first demonstrated the face validity of rDCM for small six-region networks before providing a simulation-based proof-of-principle for using rDCM to infer effective connectivity in a large network consisting of 66 brain regions, with a realistic human structural connectome and 300 free parameters to be estimated.

Our initial simulations using a six-region network (a typical size of conventional DCMs) indicated that, as expected, the accuracy of rDCM – with regard to both parameter estimation and model comparison – varies as a function of the signal-to-noise ratio (SNR) and the repetition time (TR) of fMRI data. Overall, our results demonstrated reasonable performance with regard to parameter recovery and model selection accuracy but also highlighted the importance of sufficiently high SNR (3 or higher) and

Table 3
Computational burden of rDCM quantified in terms of the approximate run-times (in s) required for estimating the large (whole-brain) DCM consisting of 66 brain regions, with a realistic human structural connectome and 300 free parameters. Run-times are shown for various combinations of the signal-to-noise ratio (SNR) and the repetition time (TR) of the synthetic fMRI data. The run-times required for inverting the model are given as the mean ± standard deviation of the 20 simulations, as well as the range.

SNR	Computational burden (s)									
	TR = 2 s		TR = 1 s		TR = 0.5 s		TR = 0.25 s		TR = 0.1 s	
	Mean ± std	Range	Mean ± std	Range	Mean ± std	Range	Mean ± std	Range	Mean ± std	Range
1	1.7 ± 0.3	1.2–2.3	1.8 ± 0.4	1.3–2.7	1.4 ± 0.2	1.1–1.7	1.8 ± 0.3	1.2–2.4	3.1 ± 0.2	2.8–3.4
3	1.6 ± 0.4	1.1–2.7	1.7 ± 0.6	1.2–3.1	1.4 ± 0.3	1.1–1.8	1.7 ± 0.3	1.4–2.4	3.2 ± 0.2	2.8–3.5
5	1.6 ± 0.4	1.2–2.8	1.8 ± 0.4	1.4–3.0	1.4 ± 0.2	1.1–1.9	1.8 ± 0.4	1.3–2.5	3.2 ± 0.2	2.9–3.7
10	1.7 ± 0.4	1.2–2.6	2.4 ± 0.3	1.9–3.0	1.5 ± 0.2	1.2–1.7	1.8 ± 0.3	1.3–2.4	3.0 ± 0.3	2.8–3.9
100	1.8 ± 0.4	1.5–2.6	2.0 ± 0.3	1.5–2.9	1.5 ± 0.1	1.3–1.7	1.6 ± 0.2	1.3–2.0	3.2 ± 0.3	2.8–3.8

fast data acquisition (TR < 2 s) for veridical inference. In situations where these conditions are not met (e.g., for subcortical regions with inherently low SNR), the current formulation of rDCM might not give reliable results. Our simulations suggest that the early version of rDCM reported in this paper is particularly promising when exploiting sophisticated scanner hardware and/or acquisition sequences that boost SNR and reduce TR. Fortunately, the development trends in fMRI move in the required direction. For example, high-field MRI (7 T and beyond) allows for considerably higher SNRs (Duyn, 2012; Redpath, 1998) and 7 T MR scanners are now becoming widely available. Similarly, the application of ultrafast inverse imaging and multiband EPI techniques enable very high sampling rates (with TRs far below one second) with whole-brain coverage (Lin et al., 2012; Moeller et al., 2010; Xu et al., 2013). Alternatively, even conventional data acquisition focused on regions of interest and using only a few slices enables TRs with only a few hundred milliseconds (for a previous DCM example, see Kasess et al. (2008)). Taking advantage of such methodological advancements may help to further exploit the full potential of rDCM for inferring effective connectivity from fMRI data. However, whether the benefits of short TRs for rDCM translate from simulations to real world datasets needs to be examined by empirical validation studies; for example, by testing whether the accuracy of connectivity-based decoding of diagnostic status (cf. Brodersen et al. (2011)) is improved by short TRs.

Having established the validity of rDCM for six-region networks, we then provided a proof-of-principle that rDCM is suitable for inferring effective connectivity in a whole-brain network comprising 66 nodes, with connectivity according to the human structural connectome reported in Hagmann et al. (2008) and 300 free parameters to estimate. These analyses suggested that rDCM can adequately recover connectivity parameters in large networks whose size is an order of magnitude larger than currently established DCM applications. Importantly, our run-time analyses suggest, that the approach scales easily and can be applied to much larger networks, provided that enough data are available. Specifically, run-time analyses did not indicate a significant increase in computation time when increasing the number of data points. Even for the shortest TR, corresponding to roughly 13,500 data points (per brain region), run-time was still only on the order of 2–3 s. The striking efficiency of rDCM rests on the fact that – due to the algebraic form of the likelihood function in the frequency domain – the computationally most expensive operation on each iteration is essentially the inversion of an $N \times N$ covariance matrix (whereas, in VBL, it is the computation of an $N \times N$ Hessian, in addition to integrating the state equation).

These findings suggest that rDCM has promising potential for the exploration of effective connectivity patterns in large (whole-brain) networks. We presently see four main application domains. First, given the computational efficiency of rDCM – which only requires few seconds for the inversion of whole-brain DCMs including an estimate of the negative free energy as an approximation to the log model evidence – it may serve as a useful tool for network discovery (Biswal et al., 2010; Friston et al., 2011). Second, it may enable the application of graph theoretical approaches to effective connectivity patterns from large networks, which have so far been restricted to structural and functional connectivity (for a comprehensive review, see Bullmore and Sporns (2009), Rubinov and Sporns (2010)). Extending graph theory to effective connectivity estimates in networks of non-trivial size opens up exciting new possibilities for studying the functional integration of the human brain. Specifically, given the inherently directed network of the human brain, graph-theoretical measure such as small-worldness, node degree, path length, centrality of nodes, or modularity will only provide an accurate view on the network topology underlying brain dynamics when accommodating the directionality of edges. Third, regardless of whether brain-wide effective connectivity estimates are used by themselves or undergo further (e.g., graph-theoretical) processing, rDCM may serve useful for computational phenotyping of patients with diseases for which global dysconnectivity is suspected, such as schizophrenia (Bullmore et al., 1997; Friston and Frith, 1995; Pettersson-Yeo et al., 2011; Stephan et al., 2006). Finally, due to its high computational efficiency, rDCM would be ideally suited for initializing the starting values of the VBL algorithm for standard DCM analyses of effective connectivity. This could be achieved by running rDCM repeatedly from multiple starting points (either defined as a grid in parameter space or randomly chosen) and using the model with the highest evidence to provide a starting point for subsequent VBL under conventional DCM. This might not only considerably speed up model inversion under the current DCM framework but potentially also prevent the algorithm from getting stuck in local extrema, against which local optimization schemes like VBL are vulnerable (Daunizeau et al., 2011a).

So far, there has only been one study that extended DCM for fMRI to larger networks (Seghier and Friston, 2013). This approach used functional connectivity measures to provide prior constraints that bounded the effective number of free parameters by essentially replacing the number of nodes with a (lower) number of modes (the principal components of the functional connectivity matrix). Seghier and Friston (2013) described their approach by an application to a network consisting of 20 regions, and it remains to be tested whether this approach also generalizes to a larger number of network nodes as required for whole-brain connectomics under commonly used parcellation schemes (e.g., Deco et al., 2013b; Hagmann et al., 2008; Honey et al., 2009).

Investigating connectivity in large-scale networks by means of mathematical models has been addressed from a different methodological angle using biophysical network models (BNM; Deco et al., 2013a, 2013b; Honey et al., 2007, 2009; Sanz-Leon et al., 2015; Woolrich and Stephan, 2013). BNMs of fMRI data typically consist of up to 10^3 network nodes, where each node is represented by a neural mass or mean-field model of local neuronal populations. Nodes are linked by long-range connections which are typically based on anatomical knowledge from human diffusion-weighted imaging data or from tract tracing studies in the macaque monkey. The resulting network dynamics are then fed into an observation model to predict fMRI data. The complexity of BNMs, however, has so far prevented estimating the strengths of individual connections. Existing applications have typically focused on simulations under fixed parameters (Deco et al., 2013a; Honey et al., 2007) or used a simplified model allowing for the estimation of a global scaling parameter (Deco et al., 2013b). Critically, such a single parameter has an indiscriminative effect on all connections, which cannot capture the selective changes in subsets of long-range connections evoked by cognitive processes. The ability of rDCM to estimate the strengths of individual connections in large-scale networks may represent a starting point for further convergence between DCMs and BNMs, as has been predicted repeatedly in the recent past (Deco and Kringelbach, 2014; Stephan et al., 2015).

Notably, whole-brain connectivity analyses face a number of potential pitfalls and challenges, regardless whether structural, functional or effective connectivity measures are obtained (Fornito et al., 2013; Kelly et al., 2012). Among others, this includes the correct identification of nodes and edges, the highly complex and dynamic structure of noise in fMRI, and the development of rigorous statistical frameworks for the analysis of whole-brain graphs. For instance, macroscopic criteria for parcellating the brain into functionally meaningful and biologically valid nodes are only just emerging (Glasser et al., 2016). Valid node identification is critical for accurate mapping of inter-regional connectivity (Smith

Please cite this article as: Frässle, S., et al., (2017), http://dx.doi.org/10.1016/j.neuroimage.2017.02.090

et al., 2011) and ill-defined nodes can have profound impact on the inferred organization of the brain as, for instance, derived from graph-theoretical measures (Fornito et al., 2010). Similarly, statistical frameworks need to be refined to address key challenges in the analysis of whole-brain connectomes – including the multiple comparison problem, graph thresholding, and the interpretation of topological measures (Fornito et al., 2013). A systematic and thorough assessment of these issues will therefore be of major importance for whole-brain connectivity analyses in the near future.

We would like to emphasize that the current implementation of rDCM only represents a starting point of development and is subject to three major limitations when compared to the original DCM framework. First, due to the replacement of the hemodynamic forward model with a fixed hemodynamic response function, rDCM does not presently capture the variability in the BOLD signal across brain regions and individuals (Aguirre et al., 1998; Handwerker et al., 2004). Critically, accounting for the inter-regional variability is crucial to avoid confounds when inferring effective connectivity from fMRI data (David et al., 2008; Valdes-Sosa et al., 2011). In forthcoming work, we will improve rDCM by replacing the fixed HRF with a basis set of hemodynamic response functions (i.e., a canonical HRF, and its temporal and dispersion derivatives; Friston et al., 1998), which can capture variations in the latency and duration of hemodynamic responses. This basis set is almost identical to the principal components of variation with respect to the hemodynamic model used in DCM, conferring biophysical validity to the set (Friston et al., 2000). An alternative approach to account for inter-regional variability in hemodynamic responses would be to use a linearized version of the hemodynamic model in DCM as described in previous work (Stephan et al., 2007).

Second, in its current implementation, rDCM ignores possible interdependencies between connections that affect different regions. This is because we assumed the measurement noise to be an independent random vector with noise precision τ and neglected its dependency on the endogenous connectivity parameters. We will improve this approximation by introducing appropriate covariance components (similar to standard approaches to non-sphericity correction in conventional GLM analyses; Friston et al., 2002) and augment our inference scheme to estimate these additional hyperparameters.

Third, rDCM is restricted to linear models (i.e., DCMs that include an A and C matrix) and thus cannot account for modulatory influences by experimental manipulations (B matrix). A bilinear extension to the rDCM approach is challenging in that the bilinear term of the neuronal state equation in the time domain induces a convolution of the experimental input and the noise term in the frequency domain. This leads to a non-trivial structure of the error covariance matrix. Addressing these three major limitations in forthcoming extensions will further improve the utility of rDCM for inferring effective connectivity among neuronal populations from fMRI data.

Apart from addressing the limitations mentioned above, there are two further important extensions that we will present in forthcoming work. First, while the present model is designed to work with experimentally controlled perturbations (the driving inputs in rDCM) and hence task-related fMRI, it is possible to extend the model to describe the "resting state", i.e., unconstrained cognition in the absence of external pertubations. A second important advance concerns the introduction of sparsity constraints to our approach (see Lomakina (2016)). This sparse rDCM (srDCM; Frässle et al., in preparation) is of likely importance for the analysis of large-scale networks – both because a complete description of larger networks complicates interpretability, but also because the available measurements may not offer sufficient information (i.e., number of data points per parameter) to allow for precise estimation of all connectivity parameters. The formulation of rDCM as a linear regression problem makes this extension of rDCM straightforward, as we can exploit well-established methods for sparse linear regression and feature selection such as LASSO (Tibshirani, 1996), elastic net regularization (Zou and Hastie, 2005), or Spike-and-Slab priors for Bayesian linear regression (Hernandez-Lobato et al., 2013). The ensuing automatic pruning of connections could be further informed by including information about the strength or likelihood of anatomical connections (cf. anatomically informed priors in DCM; Stephan et al., 2009b). These methods all implement some form of sparsity hyperpriors on the parameters; either implicitly or explicitly. An alternative to these bespoke models of sparsity would be to use Bayesian model reduction (Friston et al., 2016) where parameters are removed (and thus connectivity graphs are reduced) using an efficient scoring of models. In our context, this may represent an efficient alternative way to introduce sparsity – based upon the posterior densities furnished by rDCM. In summary, augmenting the current rDCM approach with the ability to impose sparsity constraints may result in a powerful tool for automatic "pruning" of whole-brain graphs to the most essential connections.

Finally, we would like to emphasize that the analysis of effective connectivity in whole-brain networks will not only prove valuable for studying the neural basis of cognitive processes in the healthy human brain, but may also contribute to a deeper understanding of pathophysiology of psychiatric and neurological disorders. A translational neuromodeling approach to neuroimaging data has potential for establishing novel diagnostic and predictive tools, enabling the emergence of Computational Psychiatry and Computational Neurology (Deco and Kringelbach, 2014; Friston et al., 2014a, 2014b; Huys et al., 2011; Maia and Frank, 2011; Montague et al., 2012; Stephan and Mathys, 2014; Stephan et al., 2015). Here, one important goal concerns the stratification of patients from heterogeneous spectrum diseases into mechanistically more well-defined subgroups that have predictive validity for individual treatment responses. Despite some encouraging first successes of insights into heterogeneous spectrum diseases based on connectivity inferred from fMRI data (e.g., Anticevic et al., 2015; Brodersen et al., 2014; Dima et al., 2009; Yang et al., 2014), considerable challenges remain that have so far prevented the successful transition to clinical applications (Stephan et al., 2015). The computational approach introduced in this paper may serve useful in this regard since rDCM represents a step towards a practical and computationally extremely efficient approach to obtaining directed estimates of individual connections in networks of non-trivial size, rendering computational phenotyping of whole-brain dynamics a feasible endeavour.

Software note

A MATLAB implementation of the rDCM approach introduced in this paper will be made available as open source code in a future release of the TAPAS Toolbox (www.translationalneuromodeling.org/software).

Acknowledgements

We thank Jakob Heinzle and Lars Kasper for helpful advice. This work was supported by the ETH Zurich Postdoctoral Fellowship Program and the Marie Curie Actions for People COFUND Program (both to SF), as well as the René and Susanne Braginsky Foundation and the University of Zurich (KES).

Please cite this article as: Frässle, S., et al., (2017), http://dx.doi.org/10.1016/j.neuroimage.2017.02.090

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.neuroimage.2017.02.090.

References

Aguirre, G.K., Zarahn, E., D'esposito, M., 1998. The variability of human, BOLD hemodynamic responses. Neuroimage 8, 360–369.

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.

Bishop, C.M., 2006. Pattern Recognition and Machine Learning. Springer, New York, p. 105 (12, 13, 47).

Biswal, B.B., Mennes, M., Zuo, X.N., Gohel, S., Kelly, C., Smith, S.M., Beckmann, C.F., Adelstein, J.S., Buckner, R.L., Colcombe, S., Dogonowski, A.M., Ernst, M., Fair, D., Hampson, M., Hoptman, M.J., Hyde, J.S., Kiviniemi, V.J., Kötter, R., Li, S.J., Lin, C.P., Lowe, M.J., Mackay, C., Madden, D.J., Madsen, K.H., Margulies, D.S., Mayberg, H. S., McMahon, K., Monk, C.S., Mostofsky, S.H., Nagel, B.J., Pekar, J.J., Peltier, S.J., Petersen, S.E., Riedl, V., Rombouts, S.A., Rypma, B., Schlaggar, B.L., Schmidt, S., Seidler, R.D., Siegle, G.J., Sorg, C., Teng, G.J., Veijola, J., Villringer, A., Walter, M., Wang, L., Weng, X.C., Whitfield-Gabrieli, S., Williamson, P., Windischberger, C., Zang, Y.F., Zhang, H.Y., Castellanos, F.X., Milham, M.P., 2010. Toward discovery science of human brain function. Proc. Natl. Acad. Sci. USA 107, 4734–4739.

Q6 Bracewell, R., 1999. The Fourier Transform and its Applications, 3rd Edition McGraw-Hill Science/Engineering/Math.

Brodersen, K.H., Schofield, T.M., Leff, A.P., Ong, C.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.

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.

Buckner, R.L., Krienen, F.M., Yeo, B.T., 2013. Opportunities and limitations of intrinsic functional connectivity MRI. Nat. Neurosci. 16, 832–837.

Bullmore, E.T., Sporns, O., 2009. Complex brain networks: graph theoretical analysis of structural and functional systems. Nat. Rev. Neurosci. 10, 186–198.

Bullmore, E.T., Frangou, S., Murray, R.M., 1997. The dysplastic net hypothesis: an integration of developmental and dysconnectivity theories of schizophrenia. Schizophr. Res. 28, 143–156.

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.

Catani, M., Thiebaut de Schotten, M., 2008. A diffusion tensor imaging tractography atlas for virtual in vivo dissections. Cortex 44, 1105–1132.

Daunizeau, J., Friston, K., Kiebel, S., 2009. Variational Bayesian identification and prediction of stochastic nonlinear dynamic causal models. Physica D-Nonlinear Phenom. 238, 2089–2118.

Daunizeau, J., David, O., Stephan, K., 2011a. Dynamic causal modelling: a critical review of the biophysical and statistical foundations. Neuroimage 58, 312–322.

Daunizeau, J., Preuschoff, K., Friston, K., Stephan, K., 2011b. Optimizing experimental design for comparing models of brain function. PLoS Comput. Biol. 7.

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, 2683–2697.

Davies-Thompson, J., Andrews, T.J., 2012. Intra- and interhemispheric connectivity between face-selective regions in the human brain. J. Neurophysiol. 108, 3087–3095.

Deco, G., Kringelbach, M.L., 2014. Great expectations: using whole-brain computational connectomics for understanding neuropsychiatric disorders. Neuron 84, 892–905.

Deco, G., Jirsa, V.K., McIntosh, A.R., 2013a. Resting brains never rest: computational insights into potential cognitive architectures. Trends Neurosci. 36, 268–274.

Deco, G., Ponce-Alvarez, A., Mantini, D., Romani, G.L., Hagmann, P., Corbetta, M., 2013b. Resting-state functional connectivity emerges from structurally and dynamically shaped slow linear fluctuations. J. Neurosci. 33, 11239–11252.

Dima, D., Roiser, J.P., Dietrich, D.E., Bonnemann, C., Lanfermann, H., Emrich, H.M., Dillo, W., 2009. Understanding why patients with schizophrenia do not perceive the hollow-mask illusion using dynamic causal modelling. Neuroimage 46, 1180–1186.

Duyn, J.H., 2012. The future of ultra-high field MRI and fMRI for study of the human brain. Neuroimage 62, 1241–1248.

Fornito, A., Zalesky, A., Bullmore, E., 2010. Network scaling effects in graph analytic studies of human resting-state fMRI data. Front. Syst. Neurosci. 4, 1–16.

Fornito, A., Zalesky, A., Breakspear, M., 2013. Graph analysis of the human connectome: promise, progress, and pitfalls. NeuroImage 80, 426–444.

Fornito, A., Zalesky, A., Breakspear, M., 2015. The connectomics of brain disorders. Nat. Rev. Neurosci. 16, 159–172.

Frässle, S., Krach, S., Paulus, F.M., Jansen, A., 2016a. Handedness is related to neural mechanisms underlying hemispheric lateralization of face processing. Sci. Rep. 6, 27153.

Frässle, S., Paulus, F.M., Krach, S., Jansen, A., 2016b. Test-retest reliability of effective connectivity in the face perception network. Hum. Brain Mapp. 37, 730–744.

Frässle, S., Paulus, F.M., Krach, S., Schweinberger, S.R., Stephan, K.E., Jansen, A., 2016c. Mechanisms of hemispheric lateralization: asymmetric interhemispheric recruitment in the face perception network. Neuroimage 124, 977–988.

Friston, K., Moran, R., Seth, A., 2013. Analysing connectivity with Granger causality and dynamic causal modelling. Curr. Opin. Neurobiol. 23, 172–178.

Friston, K.J., 2002. Beyond phrenology: what can neuroimaging tell us about distributed circuitry? Annu. Rev. Neurosci. 25, 221–250.

Friston, K.J., 2011. Functional and effective connectivity: a review. Brain Connect. 1, 13–36.

Friston, K.J., Frith, C.D., 1995. Schizophrenia: a disconnection syndrome? Clin. Neurosci. 3, 89–97.

Friston, K.J., Harrison, L., Penny, W., 2003. Dynamic causal modelling. Neuroimage 19, 1273–1302.

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.

Friston, K.J., Li, B., Daunizeau, J., Stephan, K., 2011. Network discovery with DCM. Neuroimage 56, 1202–1221.

Friston, K.J., Kahan, J., Biswal, B., Razi, A., 2014a. A DCM for resting state fMRI. Q7 Neuroimage 94, 396–407.

Friston, K.J., Stephan, K.E., Montague, R., Dolan, R.J., 2014b. Computational psychiatry: the brain as a phantastic organ. Lancet Psychiatry 1, 148–158.

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

Friston, K.J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M.D., Turner, R., 1998. Event-related fMRI: characterizing differential responses. Neuroimage 7, 30–40.

Friston, K.J., Glaser, D.E., Henson, R.N., Kiebel, S., Phillips, C., Ashburner, J., 2002. Classical and Bayesian inference in neuroimaging: applications. Neuroimage 16, 484–512.

Friston, K.J., Holmes, A., Poline, J., Grasby, P., Williams, S., Frackowiak, R., Turner, R., 1995. Analysis of fMRI time-series revisited. Neuroimage 2, 45–53.

Friston, K.J., Litvak, V., Oswal, A., Razi, A., Stephan, K.E., van Wijk, B.C., Ziegler, G., Zeidman, P., 2016. Bayesian model reduction and empirical Bayes for group (DCM) studies. Neuroimage 128, 413–431.

Glasser, M.F., Coalson, T.S., Robinson, E.C., Hacker, C.D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C.F., Jenkinson, M., Smith, S.M., Van Essen, D.C., 2016. A multi-modal parcellation of human cerebral cortex. Nature 536, 171–178.

Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C.J., Wedeen, V.J., Sporns, O., 2008. Mapping the structural core of human cerebral cortex. PLoS Biol. 6, e159.

Handwerker, D.A., Ollinger, J.M., D'Esposito, M., 2004. Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. Neuroimage 21, 1639–1651.

Havlicek, M., Roebroeck, A., Friston, K., Gardumi, A., Ivanov, D., Uludag, K., 2015. Physiologically informed dynamic causal modeling of fMRI data. Neuroimage 122, 355–372.

Hernandez-Lobato, D., Hernandez-Lobato, J., Dupont, P., 2013. Generalized spike-and-slab priors for Bayesian group feature selection using expectation propagation. J. Mach. Learn. Res. 14, 1891–1945.

Honey, C.J., Kötter, R., Breakspear, M., Sporns, O., 2007. Network structure of cerebral cortex shapes functional connectivity on multiple time scales. Proc. Natl. Acad. Sci. USA 104, 10240–10245.

Honey, C.J., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J.P., Meuli, R., Hagmann, P., 2009. Predicting human resting-state functional connectivity from structural connectivity. Proc. Natl. Acad. Sci. USA 106, 2035–2040.

Huys, Q., Moutoussis, M., Williams, J., 2011. Are computational models of any use to psychiatry? Neural Netw. 24, 544–551.

Jirsa, V.K., Proix, T., Perdikis, D., Woodman, M.M., Wang, H., Gonzalez-Martinez, J., Bernard, C., Bénar, C., Guye, M., Chauvel, P., Bartolomei, F., 2016. The virtual epileptic patient: individualized whole-brain models of epilepsy spread. Neuroimage 145, 377–388.

Kanwisher, N., McDermott, J., Chun, M., 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. J. Neurosci. 17, 4302–4311.

Kasess, C.H., Windischberger, C., Cunnington, R., Lanzenberger, R., Pezawas, L., Moser, E., 2008. The suppressive influence of SMA on M1 in motor imagery revealed by fMRI and dynamic causal modeling. Neuroimage 40, 828–837.

Kelly, C., Biswal, B.B., Craddock, R.C., Castellanos, F.X., Milham, M.P., 2012. Characterizing variation in the functional connectome: promise and pitfalls. Trends Cogn. Sci. 16, 181–188.

Lin, F.H., Tsai, K.W., Chu, Y.H., Witzel, T., Nummenmaa, A., Raij, T., Ahveninen, J., Kuo, W.J., Belliveau, J.W., 2012. Ultrafast inverse imaging techniques for fMRI. Neuroimage 62, 699–705.

Lomakina, E.I., 2016. Machine Learning in Neuroimaging: Methodological Investigations and Applications to fMRI (Ph.D. thesis). ETH Zurich, http://dx.doi.org/10.3929/ethz-a-010639985.

Mackay, D.J.C., 1992. A practical Bayesian framework for backpropagation networks. Neural Comput. 4, 448–472.

Mackay, D.J.C., 2003. Information Theory, Inference and Learning Algorithms. Cambridge University Press, Cambridge.

Maia, T., Frank, M., 2011. From reinforcement learning models to psychiatric and neurological disorders. Nat. Neurosci. 14, 154–162.

McIntosh, A.R., 1998. Understanding neural interactions in learning and memory

using functional neuroimaging. Ann. N. Y. Acad. Sci., 556–571.
Moeller, S., Yacoub, E., Olman, C.A., Auerbach, E., Strupp, J., Harel, N., Uğurbil, K., 2010. Multiband multislice GE-EPI at 7 T, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. Magn. Reson. Med. 63, 1144–1153.
Montague, P., Dolan, R., Friston, K., Dayan, P., 2012. Computational psychiatry. Trends Cogn. Sci. 16, 72–80.
Oppenheim, A., Schafer, R., Buck, J., 1999. Discrete-time Signal Processing. Prentice-Hall, Inc.
Park, H., Kim, J., Lee, S., Seok, J., Chun, J., Kim, D., Lee, J., 2008. Corpus callosal connection mapping using cortical gray matter parcellation and DT-MRI. Hum. Brain Mapp. 29, 503–516.
Passingham, R.E., Stephan, K.E., Kötter, R., 2002. The anatomical basis of functional localization in the cortex. Nat. Rev. Neurosci. 3, 606–616.
Penny, W.D., Stephan, K.E., Mechelli, A., Friston, K.J., 2004. Comparing dynamic causal models. Neuroimage 22, 1157–1172.
Pettersson-Yeo, W., Allen, P., Benetti, S., McGuire, P., Mechelli, A., 2011. Dysconnectivity in schizophrenia: where are we now? Neurosci. Biobehav. Rev. 35, 1110–1124.
Puce, A., Allison, T., Asgari, M., Gore, J., McCarthy, G., 1996. Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. J. Neurosci. 16, 5205–5215.
Redpath, T.W., 1998. Signal-to-noise ratio in MRI. Br. J. Radiol. 71, 704–707.
Rigoux, L., Stephan, K.E., Friston, K.J., Daunizeau, J., 2014. Bayesian model selection for group studies – revisited. Neuroimage 84, 971–985.
Roebroeck, A., Formisano, E., Goebel, R., 2005. Mapping directed influence over the brain using Granger causality and fMRI. Neuroimage 25, 230–242.
Rubinov, M., Sporns, O., 2010. Complex network measures of brain connectivity: uses and interpretations. Neuroimage 52, 1059–1069.
Sanz-Leon, P., Knock, S.A., Spiegler, A., Jirsa, V.K., 2015. Mathematical framework for large-scale brain network modeling in the virtual brain. Neuroimage 111, 385–430.
Seghier, M.L., Friston, K.J., 2013. Network discovery with large DCMs. Neuroimage 68, 181–191.
Smith, S.M., 2012. The future of FMRI connectivity. Neuroimage 62, 1257–1266.
Smith, S.M., Miller, K.L., Salimi-Khorshidi, G., Webster, M., Beckmann, C.F., Nichols, T.E., Ramsey, J.D., Woolrich, M.W., 2011. Network modelling methods for fMRI. Neuroimage 54, 875–891.
Sporns, O., Tononi, G., Kötter, R., 2005. The human connectome: a structural description of the human brain. PLoS Comput. Biol. 1, e42.
Stephan, K.E., Roebroeck, A., 2012. A short history of causal modeling of fMRI data. Neuroimage 62, 856–863.
Stephan, K.E., Mathys, C., 2014. Computational approaches to psychiatry. Curr. Opin. Neurobiol. 25, 85–92.
Stephan, K.E., Baldeweg, T., Friston, K., 2006. Synaptic plasticity and dysconnection in schizophrenia. Biol. Psychiatry 59, 929–939.
Stephan, K.E., Iglesias, S., Heinzle, J., Diaconescu, A.O., 2015. Translational perspectives for computational neuroimaging. Neuron 87, 716–732.
Stephan, K.E., Weiskopf, N., Drysdale, P.M., Robinson, P.A., Friston, K.J., 2007. Comparing hemodynamic models with DCM. Neuroimage 38, 387–401.
Stephan, K.E., Penny, W., Daunizeau, J., Moran, R., Friston, K., 2009a. Bayesian model selection for group studies. Neuroimage 46, 1004–1017.
Stephan, K.E., Tittgemeyer, M., Knösche, T.R., Moran, R.J., Friston, K.J., 2009b. Tractography-based priors for dynamic causal models. Neuroimage 47, 1628–1638.
Tibshirani, R., 1996. Regression shrinkage and selection via the Lasso. J. R. Stat. Soc. Ser. B-Methodol. 58, 267–288.
Valdes-Sosa, P.A., Roebroeck, A., Daunizeau, J., Friston, K., 2011. Effective connectivity: influence, causality and biophysical modeling. Neuroimage 58, 339–361.
Van Essen, D.C., Newsome, W.T., Bixby, J.L., 1982. The pattern of interhemispheric connections and its relationship to extrastriate visual areas in the macaque monkey. J. Neurosci. 2, 265–283.
Welvaert, M., Rosseel, Y., 2013. On the definition of signal-to-noise ratio and contrast-to-noise ratio for FMRI data. PLoS One 8, e77089.
Woolrich, M.W., Stephan, K.E., 2013. Biophysical network models and the human connectome. Neuroimage 80, 330–338.
Xia, M., Wang, J., He, Y., 2013. BrainNet viewer: a network visualization tool for human brain connectomics. PLoS One 8, e68910.
Xu, J., Moeller, S., Auerbach, E.J., Strupp, J., Smith, S.M., Feinberg, D.A., Yacoub, E., Uğurbil, K., 2013. Evaluation of slice accelerations using multiband echo planar imaging at 3 T. Neuroimage 83, 991–1001.
Yang, G.J., Murray, J.D., Repovs, G., Cole, M.W., Savic, A., Glasser, M.F., Pittenger, C., Krystal, J.H., Wang, X.J., Pearlson, G.D., Glahn, D.C., Anticevic, A., 2014. Altered global brain signal in schizophrenia. Proc. Natl. Acad. Sci. USA 111, 7438–7443.
Zeki, S., 1970. Interhemispheric connections of prestriate cortex in monkey. Brain Res. 19, 63–75.
Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. J. R. Stat. Soc. Ser. B-Stat. Methodol. 67, 301–320.

Please cite this article as: Frässle, S., et al., (2017), http://dx.doi.org/10.1016/j.neuroimage.2017.02.090