

RESEARCH ARTICLE

Reliability of dynamic causal modelling of resting-state magnetoencephalography

Amirhossein Jafarian^{1,2}  | Melek Karadag Assem^{1,2} | Ece Kocagoncu^{1,2} | Juliette H. Lanskey^{1,2} | Rebecca Williams^{1,2} | Yun-Ju Cheng³ | Andrew J. Quinn^{4,5} | Jemma Pitt⁶ | Vanessa Raymont⁶ | Stephen Lowe⁷ | Krish D. Singh⁸ | Mark Woolrich⁴ | Anna C. Nobre^{6,9} | Richard N. Henson¹  | Karl J. Friston¹⁰  | James B. Rowe^{1,2}

¹MRC Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, UK

²Department of Clinical Neurosciences and Cambridge University Hospitals NHS Foundation Trust, Cambridge Biomedical Campus, Cambridge, UK

³Lilly Corporate Center, Indianapolis, Indiana, USA

⁴Oxford Centre for Human Brain Activity, Wellcome Centre for Integrative Neuroimaging, Department of Psychiatry, University of Oxford, Oxford, UK

⁵Department of Psychology, University of Birmingham, Birmingham, UK

⁶Department of Psychiatry, University of Oxford, Oxford, UK

⁷Lilly Centre for Clinical Pharmacology, Singapore, Singapore

⁸Cardiff University Brain Research Imaging Centre, School of Psychology, Cardiff University, Cardiff, UK

⁹Department of Psychology and Center for Neurocognition and Behavior, Wu Tsai Institute, Yale University, New Haven, Connecticut, USA

¹⁰Wellcome Centre for Human Neuroimaging, University College London, London, UK

Correspondence

Amirhossein Jafarian, MRC Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, UK.
Email: aj631@cam.ac.uk

Funding information

NIHR Cambridge Biomedical Research Centre, University of Cambridge, Grant/Award Number: NIHR203312; BRC-1215-20014; Canada-UK Artificial Intelligence Initiative, Grant/Award Number: ES/T01279X/1; Medical Research Council, Grant/Award Numbers: M C_UU_00030/14, MR/T033371/1; Wellcome Trust, Grant/Award Numbers: 205103/Z/16/Z, 220258

Abstract

This study assesses the reliability of resting-state dynamic causal modelling (DCM) of magnetoencephalography (MEG) under conductance-based canonical microcircuit models, in terms of both posterior parameter estimates and model evidence. We use resting-state MEG data from two sessions, acquired 2 weeks apart, from a cohort with high between-subject variance arising from Alzheimer's disease. Our focus is not on the effect of disease, but on the reliability of the methods (as within-subject between-session agreement), which is crucial for future studies of disease progression and drug intervention. To assess the reliability of first-level DCMs, we compare model evidence associated with the covariance among subject-specific free energies (i.e., the 'quality' of the models) with versus without interclass correlations. We then used parametric empirical Bayes (PEB) to investigate the differences between the inferred DCM parameter probability distributions at the between subject level. Specifically, we examined the evidence for or against parameter differences (i) within-subject, within-session, and between-epochs; (ii) within-subject between-session; and (iii) within-site between-subjects, accommodating the conditional dependency among parameter estimates. We show that for data acquired close in time, and under

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Author(s). *Human Brain Mapping* published by Wiley Periodicals LLC.

similar circumstances, more than 95% of inferred DCM parameters are unlikely to differ, speaking to mutual predictability over sessions. Using PEB, we show a reciprocal relationship between a conventional definition of ‘reliability’ and the conditional dependency among inferred model parameters. Our analyses confirm the reliability and reproducibility of the conductance-based DCMs for resting-state neurophysiological data. In this respect, the implicit generative modelling is suitable for interventional and longitudinal studies of neurological and psychiatric disorders.

KEYWORDS

dynamic causal modelling, magnetoencephalography, parametric empirical Bayes, reliability

1 | INTRODUCTION

Dynamic causal modelling (DCM) has been used widely in translational neuroscience to elucidate the underlying causes of physiological observations including electro/magnetoencephalography (MEG) (Adams et al., 2022; Adams, Hughes, et al., 2021; Gilbert et al., 2016; Jafarian, Hughes, et al., 2023; Shaw et al., 2017; Shaw et al., 2021). However, the use of imaging and analytical methods to reveal the effects of disease progression and treatment intervention rests on reliability. This article addresses the reliability of DCM, using variational Bayesian inversion of biologically informed models of neuroimaging data that can be used to characterise the neural mechanisms of cognition and the effect of disease or drugs. We test whether inferences from DCMs of resting-state MEG data are reliable, that is, predict both the results across trials within the same session, and across separate sessions from the same participants.

DCM uses the variational Bayesian inversion of biologically motivated dynamical systems from neuroimaging data, to provide posterior estimates of unknown parameters (e.g., synaptic physiology) from a given model, and the model evidence (Friston et al., 2007; Friston et al., 2008). To compare alternative hypotheses, represented by alternative models, one uses differences in the free energy bound on [log-] model evidence, akin to log Bayes factors (Friston, 2011; Friston et al., 2011; Jafarian et al., 2019; Kass & Raftery, 1995). Bayesian model reduction (BMR) can be used for post hoc calculation of model evidence (and posterior parameter estimates) under nested, or alternative priors. Removing redundant parameters can improve model evidence by reducing model complexity. BMR is not only computationally efficient, but also eludes local minima during model inversion (Friston & Penny, 2011). At the group level, a hierarchical Bayesian inversion known as parametric empirical Bayes (PEB) accommodates multiple first-level (single subject) models and constrains physiological parameters according to empirical priors quantifying between subject effects (Friston et al., 2015; Friston et al., 2016; Litvak et al., 2015). PEB leverages BMR for the fast and efficient calculation of posterior estimates for each subject under group constraints. The implicit revision of empirical priors for each subject render's local minima less likely, because they are informed by the group mean. This effect is

reflected quantitatively in terms of the improved (free energy bound on) model evidence at the between subjective level.

In this article, we assess the reliability of these methods in DCM, in terms of between trials and across sessions as measures of their reliability.

A classical statistical approach to measure reliability is the correlation between measures from the same participants under matched conditions (Bartko, 1966; Fisher, 1992). For two groups of data, $(x_1(k), x_2(k))$ ($n = 1, \dots, k$), the reliability can be defined as the modified Pearson correlation r , as follows:

$$r = \frac{1}{nS_2} \sum_{k=1}^n \left(x_1(k) - \frac{1}{2n} \sum_{k=1}^n (x_1(k) + x_2(k)) \right) \left(x_2(k) - \frac{1}{2n} \sum_{k=1}^n (x_1(k) + x_2(k)) \right)$$

$$S_2 = \frac{1}{2n} \left(\sum_{k=1}^n \left(x_1(k) - \frac{1}{2n} \sum_{k=1}^n (x_1(k) + x_2(k)) \right)^2 + \sum_{k=1}^n \left(x_2(k) - \frac{1}{2n} \sum_{k=1}^n (x_1(k) + x_2(k)) \right)^2 \right) \quad (1)$$

Conventionally, a general linear model or one-way analysis of variance (ANOVA) with random effects is used to quantify the reliability between measurements. If y_{ij} is the i th measurements in the j th group, ANOVA can be used to estimate the unknown group mean μ , unknown j th group random effect α_j and normally distributed random effect ϵ_{ij} in the following general linear model:

$$y_{ij} = \mu + \alpha_j + \epsilon_{ij} \quad (2)$$

The between-group reliability is defined as $\frac{\sigma_\alpha}{\sigma_\alpha + \sigma_\epsilon}$, where the variance of group mean is σ_α and random effect variance is denoted by σ_ϵ . However, as shown by Box and Tiao (1973) and Wang and Sun (2014), this point estimate of reliability is not always robust. Alternative reliability estimates using, for example, MCMC, can be limited by high computational burden (Mulder & Fox, 2019).

The frequentist approach to reliability considers only the expected values of parameters, that is, the maximum likelihood estimates, but not posterior variance or covariance. In ideal settings, frequentist estimates of DCM reliability may be sufficient. However,

with complex models—with high posterior covariance among parameters—the reliability of the expected value of particular parameters can be very low (Frässle & Stephan, 2022; Rowe, 2010; Rowe et al., 2010; Schuylar et al., 2010). The reliability of DCM parameters has also been examined using split sampling over evoked potentials (Adams et al., 2022) and resting-state MEG (Jafarian, Hughes, et al., 2023). However, the frequentist approach to reliability is ill-suited to complex nonlinear dynamic systems with posterior covariance among model parameters. We therefore quantified reliability in terms of the contribution of subject or session effects using PEB; effectively, comparing between subject or session models with and without between session or subject effects. This can be interpreted as an assessment of reliability between probability distribution of parameters, in the sense that in the absence of between subject or session effects, the posterior estimates from one subject or session are similar the estimates from another.

This article has three aims: (i) to use variational Bayes to test the reliability of evidence estimates (i.e., free energy) under DCMs inverted from MEG data; (ii) to leverage the PEB to test reliability, in terms of inference about effective connectivity (i.e., posterior probability distribution of parameters) and (iii) test the influence of covariance among posterior parameters from DCM on the reliability of parameter estimates. Reliability refers to the ability to consistently reproduce a given result (and that instruments or tools used to interrogate data do so in a consistent, reproducible way). Our proposed method based on PEB (for ‘group DCM’) considers the reliability of ‘inferences’ about synaptic physiology, both within-session and between-session. This precludes the use of classic metrics of reliability (e.g., correlation, or intraclass correlation coefficient [ICC]), because these classical statistics can only be used with singular measurements (i.e., without uncertainty), as opposed to the probability distributions that underlie the inference (of synaptic physiology) using DCM. Reliability plays a crucial role in assessing reproducibility, generalisability, and predictive validity, whether in the context of parameter estimation or model-based inferences.

We use repeated measures of resting-state MEG data collected from participants in the ‘*New therapeutics in Alzheimer's disease*’ study (Lanskey et al., 2022). Our focus was not on modelling the effect of disease, but on the reliability of the ensuing estimates of synaptic efficacy. We briefly describe the participants and data, collected at baseline and 2 weeks later at rest. These data were acquired in a task-free or resting-state, which is suited for longitudinal studies of patients with progressive diseases. In the context of Alzheimer's disease, we consider the oscillatory dynamics of the default mode network (comprising bilateral angular gyri, medial prefrontal cortex [MPFC], and precuneus). Although medial temporal cortex is a priori associated with Alzheimer's disease, MEG is relatively insensitive to activity in this region (e.g., Hillebrand & Barnes, 2002; Piastra et al., 2021). Default mode network connectivity has been studied extensively in Alzheimer's disease and its treatment (Greicius et al., 2004; Lorenzi et al., 2011). We performed first-level DCM to estimate synaptic parameters in the default mode regions, from the cross-spectral density (CSD) of the MEG. We then consider the reliability of model

TABLE 1 Acronyms.

Acronyms	Description
AMPA	α -Amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid
BMR	Bayesian model reduction
CMM	Conductance microcircuit model
DCM	Dynamic causal modelling
FT	Fourier transform
fMRI	Functional magnetic resonance imaging
GABA	Gamma-aminobutyric acid
GLU	Glutamate
MEG	Magnetoencephalography
NMDA	N-methyl-D-aspartate receptors
PEB	Parametric empirical Bayes
PSDs	Power spectral density that was derived from MEG data
ss, ss, inh, dp	Superficial pyramidal cells, spiny stellate excitatory neurons, interneurons, deep pyramidal cells

TABLE 2 Glossary of variables and expressions in the conductance-based model (CMM-NMDA).

Variable	Description
u	Exogenous input
V	Mean depolarisation of a neuronal population
$\sigma(V)$	The neuronal firing rate—A sigmoid squashing function of depolarisation
L	Lead field vector mapping from (neuronal) states to measured (electrophysiological) responses
$g_x(\omega), g_o(\omega), g_y(\omega)$	Spectral density of (neuronal) state fluctuations, observation noise and measurement, respectively
$\nabla_x f$	System Jacobian or derivative of system flow with respect to (neuronal) states
$k(t) = FT[K(\omega)]$	First-order kernel mapping from inputs to responses; c.f., an impulse response function of time. This is the Fourier transform of the transfer function
$K(\omega) = FT[k(t)]$	The frequency transfer function modulates the power of endogenous neuronal fluctuations to produce a (cross-spectral density) response. This is the Fourier transform of the kernel

evidence estimates, using a general linear model approach and the reliability of the DCM parameters using PEB. Finally, we discuss the potential applications and limitations of the foregoing analyses. A glossary of acronyms and variables used in this article are provided in Tables 1, 2 and 3.

2 | MATERIALS AND METHODS

We tested the reliability of probability distribution of DCM parameters in terms of the evidence for between session (and subject) effects, inferred from repeated-measures data with conductance-based dynamic causal models.

TABLE 3 Parameters of the neuronal CMM-NMDA model (see also Figure 1). The (i,j) element in the matrix associated with parametrisation of intrinsic connection H , means connections that originate from population j and target population i in a region (here elements 1–4 correspond to ss, sp, inh, and dp layers, respectively). The (i,j) element in the matrix associated with parametrisation of extrinsic connections A and AN , means connections that originate from population j in a region and project to population i in a distal region (here elements 1–4 correspond to ss, sp, inh, and dp layers, respectively).

Description	Parameterisation	Prior
κ Rate constants of ion channels, AMPA, GABA, and NMDA, respectively	$\exp(\theta_\kappa) \cdot \kappa$ $\kappa = [4,16,100]$	$p(\theta_\kappa) = N(0,1/16)$
C Membrane capacitance of ss, sp, inh, and dp populations, respectively	$\exp(\theta_c) \cdot C$ $C = [128\ 128\ 256\ 32]/1000$	$p(\theta_c) = N(0,1/16)$
H Intrinsic connections	$\exp(\theta_H) \cdot H$ $H = \begin{bmatrix} 8 & 0 & 2 & 0 \\ 4 & 8 & 8 & 0 \\ 4 & 0 & 3 & 2 & 2 \\ 0 & 4 & 8 & 1 & 2 & 8 \end{bmatrix}$	$p(\theta_H) = N(0,1/32)$
A Extrinsic forward connection	$\exp(\theta_A) \cdot A$ $A = [1\ 0; 0\ 1; 0\ 2; 0\ 0]/8$	$p(\theta_A) = N(0,1/8)$
AN Extrinsic backward connection	$\exp(\theta_{AN}) \cdot AN$ $AN = [1\ 0; 0\ 1; 0\ 2; 0\ 0]/8$	$p(\theta_A) = N(0,1/8)$
L Sensor gain	L	$p(L) = N(1,64)$
J Contribution of spiny stellate population (J_{ss}) and deep pyramidal (J_{dp}) to observation data	J_* , ($* = dp, ss$)	$p(J_{ss}) = N(0,1/16)$ $p(J_{dp}) = N(0,1/16)$
a Endogenous random fluctuation with transfer function $\frac{a_1}{\omega^2}$	$\exp(a)$	$p(a_{1,2}) = N(0,1/128)$
d Structural cosine coefficients of endogenous random fluctuation	$\exp(d)$	$p(d_{1,2,3,4}) = N(0,1/128)$
b Common sensor noise with transfer function $\frac{b_1}{\omega^2}$	$\exp(b)$	$p(b_{1,2}) = N(0,1/128)$
c Specific sensor noise with transfer function $\frac{c_1}{\omega^2}$	$\exp(c)$	$p(c_{1,2}) = N(0,1/128)$
f Scaling some frequencies as model of data filtration	$\exp(f)$	$p(f_{1,2}) = N(0,1/128)$
D Neuronal delay between regions and within layers	$\exp(D) \cdot D$ $D = [2,16]$	$p(D) = N(0,1/64)$

2.1 | Participants

The study received ethical approval from the Cambridge 2 Research Ethics Committee. Participants provided written informed consent. Participants met clinical diagnostic criteria for symptomatic Alzheimer's disease, including mild cognitive impairment, with positive amyloid biomarkers (Lanskey et al., 2022). Principal participants (fourteen) undertook two MEG scans in a resting-state with eyes open, in two sessions that were 2 weeks apart—at a similar time of day—and without any medication changes. Their average age was 74.8 (standard deviation ± 7.33), mini-mental state examination score was 26.1/30 (standard deviation ± 3.05) and Addenbrookes Cognitive Examination (revised) score 78.5/100 (standard deviation ± 10.5). A second set of participants underwent baseline MEG, with closely matched age 73 (± 5.33) and cognition.

2.1.1 | Resting-state MEG data

Then, 5 min of resting-state MEG data (with eyes open) were collected using an Elekta Vector View system with 204 planar

gradiometers and 102 magnetometers. MEG data were recorded continuously with 1000 Hz sampling rate. Participants' horizontal and vertical eye movements were recorded using bipolar electrooculogram and electro-cardiogram electrodes. Five head position indicator coils were placed on an EEG cap to track the head position. Three fiducial points (nasion, left, and right pre-auricular) and >100 head shape points were digitised using Polhemus digitisation.

The Elekta Neuromag toolbox (Elekta Oy), with MaxFilter v2.2.12 was used for the detection and interpolation of bad sensors, signal space separation to remove external noise from the data and head movement correction. Then we pre-processed the data by downsampling to 500 Hz, band-pass filtering between 0.1 and 100 Hz, and applying a notch filter between 48–52 Hz and 98–102 Hz. We then performed artefact rejection/removal using ICA, with EOG data. We epoched the data into 1000 ms segments and repeated each epoch's artefact rejection and removal.

Using T1-weighted structural MRI (3 T Siemens, TR = 2300 ms, TE = 2.91 ms, resolution 1 mm), we performed DICOM conversion to NII and inverse-normalised the canonical mesh, size 2. We co-registered the MRI to the mesh using three fiducials and head shape points to create the forward model for MEG with the single shell

boundary element model method. We used ‘COH’ source inversion (Litvak et al., 2011) for extracting four default mode network sources/regions in left and right angular gyri (LAG [49 −63 33], RAG [−46 −66 30]), MPFC [−1 54 27], and Precuneus (PCC) [0 −55 32]. We used the induced source inversion option over 1000 ms epochs, frequency range 0.1–100 Hz, with fusion across magnetometers and gradiometers. We extracted principal components of power spectral densities over trials, as data features for the DCM of CSD.

We use three sets of resting-state eyes open MEG data for the reliability study: (i) split sampled baseline data where each individual patients is divided into odd and even epochs, (ii) baseline versus 2 weeks data, and (iii) data from baseline and a second set of participants at baseline.

2.2 | DCM of resting-states MEG data

We use DCM for CSD, SPM12-version 8163 (Friston et al., 2012; Moran et al., 2007; Moran et al., 2011) for inferring parameters and the marginal likelihood of the conductance-based biophysically canonical microcircuit model (*cmn-nmda* model in SPM12) from spectral features of MEG data. DCM for CSD explains the frequency content of MEG data in terms of a local linear perturbation (due to endogenous neuronal fluctuations) around the fixed point of a nonlinear model of canonical neuronal circuitry (Basar et al., 2012; Haken, 1977).

The conductance-based model describes the electrical activity of a cortical source based on the interactions of four neuronal

populations: excitatory spiny stellate cells, superficial pyramidal cells, inhibitory interneurons, and deep pyramidal cells, as shown in Figure 1. Each cortical source is connected to other regions via forward connections (that originate from the superficial pyramidal population and project to excitatory spiny stellate and deep pyramidal cells of other regions) and backward connections (that originate from deep pyramidal cells and project to superficial pyramidal and inhibitory populations in the target source). Each population is modelled by a Morris–Lecar model (driven by random endogenous fluctuations) (Moran et al., 2013) as follows:

$$\frac{dV}{dt} = \frac{1}{C} \left[g_L(V_L - V) + g_{AMPA}(V_{AMPA} - V) + g_{GABA}(V_{GABA} - V) + g_{NMDA} m(V)(V_{NMDA} - V) \right] + u$$

$$\frac{dg_*}{dt} = \frac{1}{\tau_*} \left(\sum_{k=sp,inh,dp,ss} S_k \sigma_k - g_* \right) + u, * = [L, AMPA, GABA, NMDA] \quad (3)$$

In Equation 3, the membrane potential of each population is denoted by V , the conductance of ion channels are g_L , g_{NMDA} , g_{AMPA} , and g_{GABA} . The input to the model is which represents random endogenous fluctuations. Constant parameters are C as the membrane capacitance, L as a passive leak current with a fixed conductance, and τ_* as ion channel receptor time constants. V_L , V_{NMDA} , V_{AMPA} , and V_{GABA} denote the reversal equilibrium potentials of the ion channels. The dynamics of depolarisation is equipped with activity-dependent

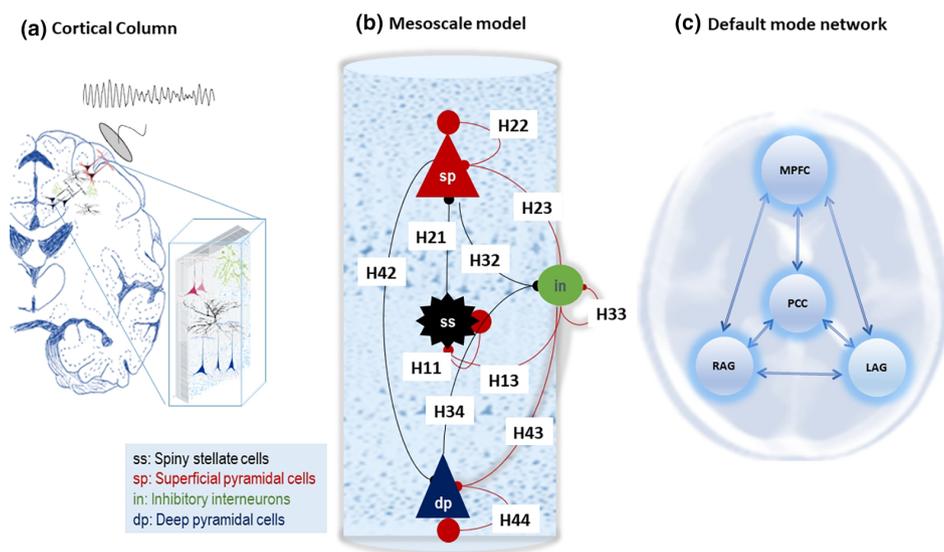


FIGURE 1 Mesoscale model of MEG data. Panel (a) illustrates a cortical column as the origin of electrical brain activity as recorded by neuroimaging modalities such as MEG. Panel (b) illustrates the laminar specific conductance based model with superficial (sp) and deep pyramidal (dp) cells in the top and bottom layers, respectively, excitatory interneurons (spiny stellate cells, ss) situated in layer four, and inhibitory interneurons that are distributed across all layers and modelled using one population. The dynamics of ion compartments with a population are governed by the Morris–Lecar model. This model explains the dynamics of different ion currents: NMDA, AMAP, and GABA and passive ion current and membrane capacitance as explained in Equation 1. Panel (c) shows the fully connected default mode network which contains MPFC, PCC, RAG, and LAG sources.

magnesium channels, which are modelled as $m(V) = \frac{1}{1+0.2 \exp(-\alpha_{\text{NMDA}} V)}$. Afferent presynaptic firings from a population k are denoted by σ_k , which is scaled by connectivity S_k which can be, in nature, excitatory AMPA, NMDA or inhibitory GABA afferent intrinsic H_k or forward and backward extrinsic AMPA (denoted by A) and NMDA (denoted by AN) connections. Glossary of variables and prior for unknown parameters are reported in Tables 2 and 3, respectively.

The generative model of MEG data (denoted by y) can be written as a partially observed dynamical system as follows (please see Table 2 for definition of notations):

$$\begin{aligned} \dot{x} &= f_\theta(x, u) \\ y &= L_s(\Gamma_J(x)) + e \end{aligned} \quad (4)$$

In Equation 4, θ denotes a lump representation of all unknown parameters (i.e., H_k , τ), x is a vector of biological states in the model, $f_\theta(x, u)$ is a function that is the concatenated version of the right-hand sides of Equation 1 over all populations. In Equation 4, u is the endogenous (structured pink) noise with a cross-spectrum, $g_u(\omega, \theta) = FT(E[u(t), u(t-\tau)])$. In the second line of Equation 4, L_s is a parameterisation of lead field projections and sensor gain, and $\Gamma_J(x)$ is an operator (parameterised by unknown parameters J) that links the activity of populations (e.g., weighted sum—by J parameters—of different population responses) to MEG data (through lead field projection).

Resting-state dynamics can be modelled as a response to endogenous random fluctuations. In this DCM, a linearised neuronal model (around a stable equilibrium point) is used to generate spectral content of the MEG data. The spectral response of the neuronal model, $g_x(\omega)$, can be modelled as follows:

$$g_x(\omega) = K(\omega, \theta) \cdot g_u(\omega, \theta) \cdot K(\omega, \theta)^T + g_o(\omega, \theta) \quad (5)$$

In Equation 5, $g_o(\omega, \theta)$ represents the spectrum of the observation noise, which is a sum of common and source-specific noise and $K(\omega, \theta) = FT(\exp \tau \cdot \nabla_x f(x, \theta))$ (∇_x is the Jacobian) is the transfer function (input to output) of the neural model (parameterised by J). The spectral response in sensor space can also be generated by the inclusion of a forward electromagnetic model into Equation 4, which is denoted by $L.M$ (L is the sensor gain and M is the head model), as follows:

$$g_y(\omega) = L \cdot M \cdot g_x(\omega, \theta) \cdot M^T \cdot L^T + \epsilon \quad (6)$$

In Equation 6, $g_y(\omega)$ is the cross spectra of the MEG data, and $\epsilon \sim N(0, \sigma^2)$ is a random effect (with unknown covariance). Because we perform the DCM in the source space, the forward electromagnetic model reduces to a scaling parameter.

The laminar interpretation of the conductance-based model (e.g., superficial, deep geometry) is supported by (i) specification of priors for intrinsic connections (Table 1), (ii) the equation of the observer for MEG data, and (iii) endogenous inputs to the model which targets spiny stellate cells in layer four. This model

parametrisation supports inferences about the laminar basis of degenerative brain neurological disorders (Adams et al., 2022; Adams, Hughes, et al., 2021; Adams, Pinotsis, et al., 2021; Shaw et al., 2021).

The unknown parameters in the DCM are specified as log-scale values $\theta = \theta_0 \exp(\hat{\theta})$ where θ_0 are biologically informed scaling constants for the parameter, and $\hat{\theta}$ has Gaussian distribution, with prior normal density $\hat{\theta}_0 \sim N(0, \Sigma_\theta)$ of zero mean and covariance Σ_θ . This expresses Bayesian beliefs about the range over which parameters can vary and constrains the posterior density over parameters accordingly. The prior distribution for parameters assures the stability and plausibility of the model, and we refer to them as ‘micro-priors’ hereafter.

2.3 | Parametric empirical Bayes

DCM at the between session or subject level implements hierarchical variational Bayesian inversion of data under empirical priors at the second level (e.g., age, disease severity, etc) on the first level (e.g., synaptic) parameters. At the first level, the neuronal model is optimised to fit each individual's data. At the second level, the estimation of some parameters is constrained by group estimates to improve group model evidence. These parameters are those that are equipped with random effects. For these parameters, there are micro- and macro-priors that constrain the accompanying posterior estimates at the first and second levels, respectively.

The micro priors are taken from the physiological literature (Friston et al., 2003; Friston et al., 2015; Friston & Penny, 2011), while macro priors impose constraints on parameters based on information about the population from which the data are drawn (i.e., an empirical prior, such as age or disease severity) (Friston et al., 2016). Due to leveraging BMR, assessing the impact of macro-scale constraints on synaptic parameters does not require re-estimation of the first level DCMs (Friston et al., 2016; Friston, Preller, et al., 2019; Zeidman et al., 2019).

2.4 | Reliability of DCM

2.4.1 | Reliability of fixed quantities without uncertainty

This section assesses the reliability between fixed effects (i.e., measurements without uncertainty or random effects), such as the free energy of a model. We denote two sets of measurements of similar phenomena by column vectors k_1 and k_2 , each of which has dimension $n \times 1$ (n number of, e.g., participants) and specify the following linear model:

$$k_1 = k_2 \beta + \epsilon \quad (7)$$

In Equation 7, β represents the intercepts of the model between the two measurements. The reliability between measurements can be

specified as a covariance component of the noise term $\epsilon \sim N(0, \Sigma)$, with compound symmetry as follows (Jelenkowska, 1998; Jelenkowska, 1999).

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \cdots & \rho \\ \vdots & \ddots & \vdots \\ \rho & \cdots & 1 \end{bmatrix} \quad (8)$$

The covariance matrix elements are defined as $[\Sigma_{ii}] = \sigma^2$ and $[\Sigma_{ij}] = \rho\sigma^2$ ($i \neq j$). The parameter ρ is the inter-class correlation coefficient. The compound symmetry structure of a covariance matrix assumes that all off-diagonal elements of the matrix are equal. This means that the covariance between any two variables or measurements is the same across all pairs. This is a common assumption in statistical analyses, particularly in the context of repeated-measures or within-subject designs. This form of covariance is relevant for measuring reliability between measurements and can be motivated in several ways. First, the compound symmetry applies the Homogeneity assumption, with equal covariances between all pairs of variables, meaning that the relationships among variables or measurements are similar across the board. This simplifies the modelling process and reduces the number of (covariance) parameters that need to be estimated. Second, there is interchangeability of observations arising from repeated measurements on the same subjects. That is, any pair of measurements are equally related to each other regardless of when they were taken. The compound symmetry assumption reflects this interchangeability by setting all off-diagonal elements equal. Third, the compound symmetry form simplifies the covariance matrix, aiding the estimation and interpretation of the models. With fewer parameters to estimate, analyses become more efficient: this is especially relevant when dealing with limited data. In the context of our reliability analysis, the compound symmetry implies that the reliability of measurements remains consistent across different conditions or time points.

To test reliability, one can invert the general linear model in Equation 7 with and without ρ in the covariance matrix and use Bayesian model comparison to compare the respective free energies to assess the evidence for an interclass correlation. We use variational Laplace (Friston et al., 2007) as implemented in SPM12 (e.g., Friston, Diedrichsen, et al., 2019) to estimate the covariance components, with and without the interclass correlation.

2.4.2 | Reliability of inferred DCM parameters

To assess the reliability of parameters with random effects, we used a parametric empirical Bayesian approach to assess the contribution of random effects to conditionally dependent parameter estimates. Mathematically, let a column vector of model parameters at the first level DCM, over cohort, is $\theta_{np \times 1}^{(1)}$ (n number of participants and p is the number of parameters for each participant). Then, the generative model of the PEB is given by (Friston et al., 2016):

$$\begin{aligned} y^{(i)} &= L_{\theta^{(1)}}(\Gamma(x)) + \epsilon^{(i)} \quad i = 1, \dots, n \\ \theta^{(1)} &= (X \otimes I)\theta^{(2)} + \epsilon^{(2)} \quad \theta^{(1)} = [\theta^{i,1}, \dots, \theta^{i,n}] \end{aligned} \quad (9)$$

The first line of Equation 9 is the generative model of each DCM with unknown parameters inferred from neuroimaging data at the first level. The second line models (macro-level) empirical priors that constrain parameter estimates from the first-level DCM. In the second line of Equation 9, $X \in R^{n \times r}$ is the design matrix with $r \geq 1$ covariant. The first column of X is equal to one and reflects the group mean; generally, the rest of the column can be defined based on empirical data. The symbol \otimes is the Kronecker product, and I is the $p \times p$ identity matrix. The random effects have a Gaussian distribution and at the second line of equation (6A), $\epsilon^{(2)} \sim N(0, \Pi^{(2)})$ (where $\Pi^{(2)}$ is the precision matrix or inverse of covariance). The precision matrix is parameterised with a single (hyper-precision) parameter, γ , as follows (Friston et al., 2016):

$$\Pi^{(2)} = I_S \otimes (Q_0 + e^{-\gamma} Q_1) \quad (10)$$

In Equation 10, $Q_0 \in R^{p \times p}$ is the lower bound on the precision, defined with a small positive value. The (hyper)parameter, γ , scales a precision matrix $Q_1 \in R^{p \times p}$, which is (by default) 16 times the prior precision of the group mean (Zeidman et al., 2019): this hyper prior ensures that random effects are small compared to prior uncertainty about the parameter in question. The objective of group DCMs is to maximise the second level free energy under the PEB constraint. The inversion of group DCM (aka PEB) starts with the inversion of each data, followed by assessing the effect of empirical prior and adjusting parameters so that the free energy is maximised. As part of group DCMs inversion, BMR is used to re-evaluate first-level posteriors under updated second-level parameters (Friston et al., 2016; Litvak et al., 2015). This greatly expedites inversion of group data. It is common practice to ask whether there is any difference between two cohorts. This question can be addressed using what is known as a ‘PEB of PEBs’. This analysis entails inverting second-level models for separate cohorts that are then combined in a third level (PEB) analysis, to identify shared characteristics and differences. The aim of a ‘PEB of PEB’ analysis is to elucidate the similarities and differences between the models representing distinct cohorts, providing insights into the underlying difference between the populations under study. Motivated by the classical definition of reliability, that is, that there are no between-session or between-subject effects, we evaluated the evidence for PEB models with and without these effects. This constitute the predictive validity equivalent, which is useful in the context of inferred parameters (i.e., multivariate probably distribution). We perform PEB estimation for each parameter separately; in which synaptic rate constants (T), intrinsic synaptic gain (H), extrinsic connections (A and AN), state to observation parameters (L and J), and physiological inputs (a and d) are constrained by their group average (a single column matrix of ones). We employed the PEB of PEB approach, utilizing a matrix with the first column containing ones (constant term) and the second column containing zeros and ones,

indicating each group, respectively. This method was used to evaluate the evidence for between-session or between-subject effects.

3 | RESULTS

The power spectra of the MEG data from each source and subject are shown in the supplementary material. In addition, in Figure 2, we have shown a sample of DCM fit in terms of predicted versus observed response.

Comparison between the baseline and 2-week spectra show that they are not identical. These differences are not attributable to the progression of the disease over such a short interval but may rise from other factors, including, for example, psychological states (e.g., fatigue), recording noise, or movements.

3.1 | Reliability of DCM

We tested the reliability of the model's free energy estimates based on test and re-test data: as split-sample data from one session, or two sessions acquired in the same participants, or of two similar groups of participants, as summarised in Figure 3. The free energy estimates were highly reliable for within-session split-sample data, and within-

subject between session data. For the within-subject models, the free energy was higher for the model with compound symmetry—that is, with a high interclass correlation—with a difference of 20 (equivalent to a Bayes factor of $\sim 5 \times 10^8$). On the other hand, there was no evidence for an interclass correlation between the free energy of models of data acquired from different people, even if demographically and clinically matched. This is expected: even though most of the neuronal parameters of the biophysical models that generate MEG data were similar for matched adults, differences in signal-to-noise and other non-neuronal factors can have a profound effect on the free energy estimates of model evidence (a.k.a., marginal likelihood).

To test for between session and subject effects, we used PEB of PEBs to assess the reliability of 156 physiological parameters; including synaptic rate constants (T), intrinsic synaptic gains (H), extrinsic forward and backward AMPA connections (denoted by A) and extrinsic forward and backward NMDA connections (denoted by AN), state to observation parameters (L and J), and physiological inputs (a and d). For the within-session split-sample analysis, the PEB of PEBs revealed that only 4 of 156 parameters differed between odd and even trials (Figure 4a). For the within-subject between-session analysis, the PEB of PEBs showed that only four of 156 parameters differed between sessions (Figure 4b). Finally, there was fair agreement across the two separate groups of similar participants, with nine of 156 showing evidence of between subject effects on parameter estimates (Figure 4c).

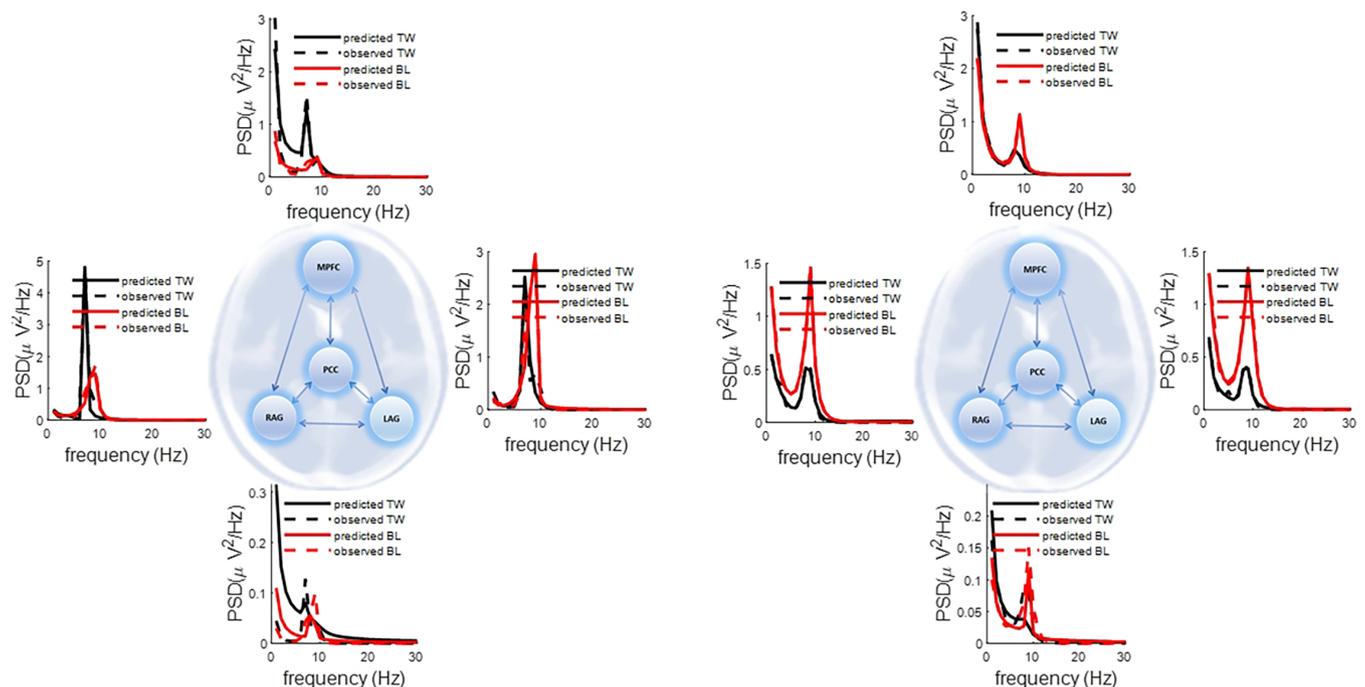


FIGURE 2 This graphics illustrate observed spectral densities (PSD) and their associated predicted response by DCM at baseline (BL) and after 2 weeks (TW) in the eyes open condition, over the four default mode network sources. The differences between baseline and 2 weeks later data are not attributable to the progression of the disease but may be linked to plasticity, psychological effects, differential fatigue, measurement noise or movement etc. Predicted DCM responses for the four-node default model network DCMs suggest that the neuronal model replicates regions' PSDs data. Similar graphics associated with other subject's predicted and observed responses for their baseline and 2 weeks apart data are given in supplementary Figures 1S–7S.

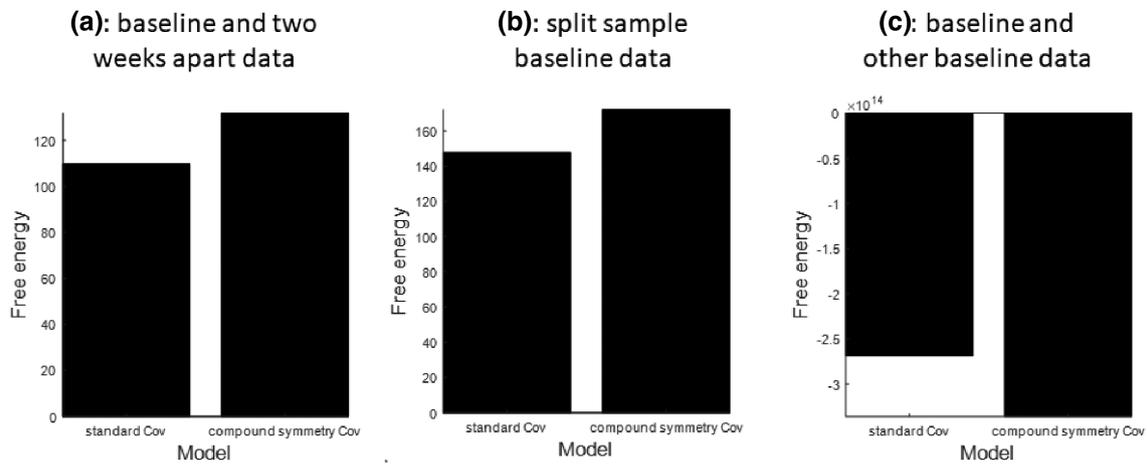


FIGURE 3 Reliability of free energies of the first level multichannel DCMs of the default mode network. There is greater evidence for the model with compound symmetry structure than the model without, when fitting data for different trials within-subjects (panel a) or different sessions within-subjects (panel b; baseline vs. 2-week data), but greater evidence for the model without compound symmetry when fitting data from different subjects, as expected (panel c; baseline data).

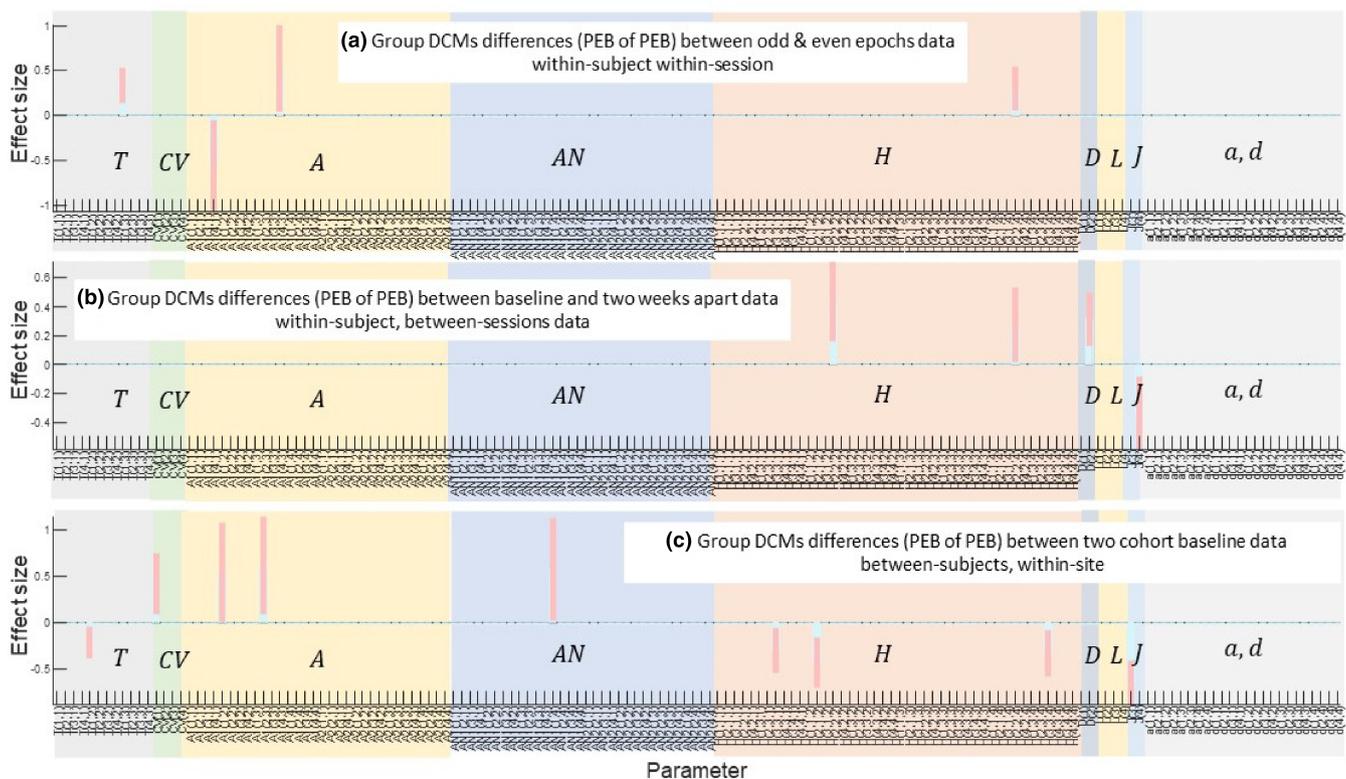


FIGURE 4 Reliability using the PEB of PEB approach for 156 inferred parameters of the fully connected default model network DCMs for (a) split sample data within-subject within-session, (b) within-subject between-sessions, 2 weeks apart and (c) between-subjects, within-site. Each plot illustrates the expectation of parameters (blue bar) and their 95% confidence interval (pink bar) for parameters for which there was evidence of a difference. Note that >95% of parameters do not differ within-session, or between-sessions. Time constants are denoted by T and CV are membrane capacitances, A and AN are between-sources forward and backward extrinsic AMPA and NMDA connections, respectively, H are the intrinsic within-regions connections, D are within and between regions delays, L and J are sensor gains and states that contributed to local field potentials, respectively, and a and d are the parameters of random neuronal fluctuations. The list of all 156 parameters in x axis are provided in the supplementary information.

3.2 | Reliability and the correlation of parameters

To quantify the relationship between the reliability of the posterior parameters estimates and their correlation, we examined the expected values of parameters after first level DCM and their re-estimation by the PEB, that is, second level DCM to implement empirical priors or constraints. In Table 4, we report number of intrinsic, extrinsic forward and backward synaptic parameters that have correlations larger than 0.5 between their first-level DCM estimates and after re-estimation by PEB under empirical (second level) constraints. This confirms that after implementing empirical priors, their reliability improves. The PEB re-estimates first-level DCM parameters to improve model evidence by reducing model complexity and implicitly resolving conditional dependencies. In effect, inferred parameters become more conditionally independent, and thereby more reliable in the classical sense.

4 | DISCUSSION

We assessed the reliability of DCM using MEG data acquired within-session, between-session, and between-subject. In all three cases, the group level conductance-based DCM estimates, obtained from resting-state MEG power spectra, are highly reliable in terms of the inferred neuronal parameters' distributions. To compare fixed quantities (e.g., the free energy estimates of model evidence), we used variational Laplace to estimate the covariance components due to intra-class correlations. To assess agreement between parameters with random effects, we used 'PEB of PEBs' to test for between-session and between-subject effects, finding excellent generalisation over split-sample and test-retest analyses. These indicate a high reliability of DCM of electrophysiological observations. In addition, we confirmed the reciprocal relationship between reliability and conditional dependency (covariance) between parameter estimates. In other words, as the covariance among parameters reduces, the reliability of their expectations increases.

The frequentist correlations of singular parameters, as typically measured by the 'traditional' ICC approach, are not well-suited to the context of complex models—with high posterior covariances—such as DCM. This is primarily due to the conditional dependency between parameters. However, alternative metrics such as the free energies

remain consistent between sessions, despite the presence of conditional dependency among inferred parameters. This consistency indicates that model comparison is reliable, thereby providing a useful means to specify and test different hypotheses related to session differences (e.g., a drug intervention, or disease progression). Moreover, we focus on the underlying meaning of reliability and employ the 'PEB of PEB' approach to determine between-session effects. This is especially important with the use of subject-specific biomarkers as priors in DCM, for example, in translational modelling and precision medicine. The posterior covariance between parameters is a major contributor to poor reliability (something that has been observed before but is explicitly investigated here). By improving the free energy of—that is, evidence for—the model (for instance, by using PEB as reported in Table 4), one may observe some improvement in the reliability of parameters (one can intuit this as an improvement in model evidence via reducing model complexity and, consequently, reducing the posterior correlations among parameters). By refining and providing better priors (e.g., through, e.g., MRS or PET data), the free energy of a model improves. However, it is important to note that even in the case of multimodal DCM (Jafarian et al., 2020) or pathology-enriched DCM (e.g., Jafarian, Assem, et al., 2023) the classic reliability formulation would still not be applicable to the inferred parameters simply because the parameters are probability distributions, whereas classic reliability tests were defined for repeated singular measurements without considering uncertainty.

A key motivation for this analysis was to address the reliability of inferences from complex biophysical models in which neuronal generators are nonlinear, and parameter estimates are correlated (Litvak et al., 2019). In previous studies, individual DCM connectivity parameters have proven unreliable, except for very simple models, despite the high reliability of model evidence and model comparison for hypothesis testing (Jafarian et al., 2022; Rowe, 2010; Rowe et al., 2010; Schuyler et al., 2010). Poor reliability was attributed to their covariance, and the use of model comparison based on model evidence was recommended for hypothesis testing.

Parameter inference is sensitive to changes in the data (which may arise in repeated-measures or between-subject data). The translational neuronal modelling of neurological disorders presents a significant challenge, necessitating a reliable platform for hypothesis testing. DCM offers such a platform, but its utility rests on reliability for repeated measures and agreement between sessions. By utilizing

TABLE 4 Reliability of parameters before and after using PEB for split sampled and follow up data.

Inferred DCM parameters	Number of parameters with correlation score > 0.5 in DCMs of split sample baseline data		Number of parameters with correlation score > 0.5 in DCMs of 2 weeks apart data and baseline data	
	First level DCM	Re-estimated first level DCM by PEB	First level DCM	Re-estimated first level DCM by PEB
Intrinsic connections H	15	19	8	10
Extrinsic forward connection A	8	9	7	8
Intrinsic backward connections AN	9	10	11	12

the resting-state data, we invert neuronal models of the default mode network with moderate complexity. We demonstrate that consistent conclusions can be drawn from either dataset because free energies from these data remain consistent across datasets. This complements earlier reports of the reliability of evoked responses, spectral power and functional connectivity (Colclough et al., 2016; Kumar et al., 2022; Lew et al., 2021; Marquetand et al., 2019). Indeed, the use of evoked responses rather than resting state, may increase reliability further, although resting-state data facilitates large scale clinical studies. The reliability we demonstrate is adequate for a modelling platform for experimental medicine, improving our understanding and treatment of neurological disorders. However, using PEB (Friston et al., 2016; Friston & Penny, 2011; Litvak et al., 2015; Litvak et al., 2019) one can effectively identify the similarity of underlying neuronal dynamics between split-sample and test-re-test data for ~95% of parameters; consistent with the data being drawn from the same distribution, and in our study the absence of disease progression.

Parameter estimates under nonlinear models are accompanied by a degree of co-linearity that is, changing one parameter is equivalent to changing others. In other words, different combinations of parameters in the generative model could give rise to very similar data. The set of parameters can be considered as a manifold, which is biologically plausible (Prinz et al., 2004) but challenging when testing hypotheses based on particular parameter values. Historically, solutions to this problem have included reducing model complexity (Stephan et al., 2007), by re-parameterisation and defining priors for the generative models for functional MRI. However, such re-parameterisation is not straightforward, and may not be possible for generative models of M/EEG. This is because of the inherent complexity of the models of M/EEG cortical generators (Penny, 2012). To address this problem, we leveraged BMR and PEB to find nested models with lower complexity that can better capture the underlying dynamics.

The use of PEB can be seen as an extension of BMR to cohort studies. PEB is well suited to address whether model evidence from cohort data can be improved by replacing non-informative (or weakly informative) prior parameters in DCM by empirical priors; that is, empirical constraints (Adams et al., 2022; Adams, Pinotsis, et al., 2021; Friston et al., 2022; Jafarian et al., 2021; Jafarian, Hughes, et al., 2023). In PEB, a reduced model with higher model evidence is sought at the group level by constraining parameters using group information to enhance cohort model evidence. PEB re-estimates first-level DCMs where all parameters are informed by the group information. If two sets of data are sampled from the same distribution, as expected within-session, then PEB of PEB should not identify differences in the parameters of the generative model. The improvement of reliability after the application of PEB is partly due to the application of BMR during group DCM inversion, which serves to reduce the complexity and, implicitly, the posterior correlation between parameters.

The reliability of the probability distributions of parameters and model evidence from DCMs is just one aspect of their validation. It is not expected that the inferred parameters are identical, given the

likely differences between the data acquired close together and in 'similar' conditions. Although disease progression is unlikely to a meaningful degree in 2 weeks, other differences like fatigue, anxiety, motion or scanner noise, may arise. These differences are likely to occur in addition to longitudinal or interventional 'repeated-measures' studies. The effect of such cofounds for translational neuroscience is a matter of degree, and approaches to reduce the effect of such cofounds include (i) using larger data samples and (ii) inclusion of informative priors on individual differences (Jafarian, Hughes, et al., 2023; Stephan et al., 2009; Zeidman et al., 2019). It is anticipated that DCM holds promise for experimental medicine studies, particularly in estimating within-subject differences following a drug intervention or disease progression, as well as in guiding treatment decisions through model comparison. Previous studies, such as those investigating the effects of deep brain stimulation or drug interventions in Parkinson's disease and psychotic disorders (Jin et al., 2023), have demonstrated this potential application. However, some prior applications of DCM have relied on the expectation of parameters, using posterior expected values as dependent variables in repeated measures ANOVA. This approach ignores posterior uncertainty about parameter estimates, which may be essential for understanding of the underlying generators of data. One practical strategy—to address this limitation—is through BMR, which allows for the assessment of how alternative priors, such as treatment or time effects, predict and elucidate the effects of interventions on individual patients. For example, recent single-subject work by Friston, Preller, et al. (2019) used this approach to shed light on the mechanisms of neurovascular coupling.

In summary, we have demonstrated reliable inferences based on the posterior probabilities and reliable relative model evidences, both within-session and between-sessions. The use of group inversion improves the free energy of first-level DCMs and improves reliability of individual parameters. Such DCMs, including the canonical micro-circuit model of MEG, provide a sufficiently reliable modelling platform to consider for use in longitudinal or interventional studies.

AUTHOR CONTRIBUTIONS

Amirhossein Jafarian: Conceptualization, methodology, software, validation, formal analysis, writing original draft, data curation, visualization. **Melek Karadag Assem:** Data acquisition, pre-processing of MEG data, review and editing. **Ece Kocagoncu:** Data acquisition, pre-processing of MEG data, review and editing. **Juliette H. Lanskey:** Data acquisition, review. **Rebecca Williams:** Review. **Andrew J. Quinn:** Data acquisition, review and editing. **Jemma Pitt:** Data acquisition, review. **Stephen Lowe:** Review. **Vanessa Raymont:** Review. **Krish D. Singh:** Review. **Mark Woolrich:** Review. **Anna C. Nobre:** Discussion, review and editing. **Richard N. Henson:** Discussion, review and editing. **Karl J. Friston:** Conceptualization, methodology, review and editing. **James B. Rowe:** Conceptualization, methodology, review and editing, funding acquisition.

ACKNOWLEDGEMENTS

This work has been funded by the Wellcome Trust (220258); the Medical Research Council (M C_UU_00030/14; MR/T033371/1) and

Cambridge Centre for Parkinson-plus; the Holt fellowship; and the NIHR Cambridge Biomedical Research Centre (NIHR203312: BRC-1215-20014). KJF is supported by funding for the Wellcome Centre for Human Neuroimaging (Ref: 205103/Z/16/Z) and a Canada-UK Artificial Intelligence Initiative (Ref: ES/T01279X/1). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. For the purpose of open access, the authors have applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

CONFLICT OF INTEREST STATEMENT

The authors declare that there is no competing interests.

DATA AVAILABILITY STATEMENT

The SPM12 software (<https://www.fil.ion.ucl.ac.uk/spm/>) was used in this research. The code associated with this paper is available at <https://github.com/NIMG-22-2183/HM-reliability-of-DCM>. Minor changes were applied to the `spm_ar_reml` code to accommodate specifying the compound symmetry structure for the covariance of the random effect. Anonymised imaging data can be requested from the senior author, but a data transfer agreement is likely to be required under the terms of consent and data protection and the need to prevent patient identification.

ORCID

Amirhossein Jafarian  <https://orcid.org/0000-0001-9547-7941>

Richard N. Henson  <https://orcid.org/0000-0002-0712-2639>

Karl J. Friston  <https://orcid.org/0000-0001-7984-8909>

REFERENCES

- Adams, N. E., Hughes, L. E., Rouse, M. A., Phillips, H. N., Shaw, A. D., Murley, A. G., Cope, T. E., Bevan-Jones, W. R., Passamonti, L., & Street, D. (2021). GABAergic cortical network physiology in fronto-temporal lobar degeneration. *Brain*, *144*, 2135–2145.
- Adams, N. E., Jafarian, A., Perry, A., Rouse, M., Shaw, A. D., Murley, A. G., Cope, T. E., Bevan-Jones, W. R., Passamonti, L., & Street, D. (2022). Neurophysiological consequences of synapse loss in progressive supranuclear palsy. *medRxiv*.
- Adams, R., Pinotsis, D., Tsirlis, K., Unruh, L., Mahajan, A., Horas, A., Convertino, L., Summerfelt, A., Sampath, H., & Du, X. (2021). Computational modelling of EEG and fMRI paradigms indicates a consistent loss of pyramidal cell synaptic gain in schizophrenia. *Biological Psychiatry*, *89*, S95.
- Bartko, J. J. (1966). The intraclass correlation coefficient as a measure of reliability. *Psychological Reports*, *19*, 3–11.
- Basar, E., Flohr, H., Haken, H., & Mandell, A. (2012). Synergetics of the brain: Proceedings of the International Symposium on Synergetics at Schloß Elmau, Bavaria, May 2–7, 1983, Springer Science & Business Media.
- Box, G., & Tiao, G. (1973). *Bayesian inference in statistical analyses*. Addison-Wesley-5 Longman.
- Coldclough, G. L., Woolrich, M. W., Tewarie, P., Brookes, M. J., Quinn, A. J., & Smith, S. M. (2016). How reliable are MEG resting-state connectivity metrics? *NeuroImage*, *138*, 284–293.
- Fisher, R. A. (1992). Statistical methods for research workers. In *Breakthroughs in statistics*. Springer.
- Frässle, S., & Stephan, K. E. (2022). Test-retest reliability of regression dynamic causal modeling. *Network Neuroscience*, *6*, 135–160.
- Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J., & Penny, W. (2007). Variational free energy and the Laplace approximation. *NeuroImage*, *34*, 220–234.
- Friston, K., & Penny, W. (2011). Post hoc Bayesian model selection. *NeuroImage*, *56*, 2089–2099.
- Friston, K., Zeidman, P., & Litvak, V. (2015). Empirical Bayes for DCM: A group inversion scheme. *Frontiers in Systems Neuroscience*, *9*, 164.
- Friston, K. J. (2011). Functional and effective connectivity: A review. *Brain Connectivity*, *1*, 13–36.
- Friston, K. J., Bastos, A., Litvak, V., Stephan, K. E., Fries, P., & Moran, R. J. (2012). DCM for complex-valued data: Cross-spectra, coherence and phase-delays. *NeuroImage*, *59*, 439–455.
- Friston, K. J., Diedrichsen, J., Holmes, E., & Zeidman, P. (2019). Variational representational similarity analysis. *NeuroImage*, *201*, 115986.
- Friston, K. J., Flandin, G., & Razi, A. (2022). Dynamic causal modelling of COVID-19 and its mitigations. *Scientific Reports*, *12*, 12419.
- Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *NeuroImage*, *19*, 1273–1302.
- Friston, K. J., Li, B., Daunizeau, J., & Stephan, K. E. (2011). Network discovery with DCM. *NeuroImage*, *56*, 1202–1221.
- 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.
- Friston, K. J., Preller, K. H., Mathys, C., Cagnan, H., Heinzle, J., Razi, A., & Zeidman, P. (2019). Dynamic causal modelling revisited. *NeuroImage*, *199*, 730–744.
- Friston, K. J., Trujillo-Barreto, N., & Daunizeau, J. (2008). DEM: A variational treatment of dynamic systems. *NeuroImage*, *41*, 849–885.
- Gilbert, J. R., Symmonds, M., Hanna, M. G., Dolan, R. J., Friston, K. J., & Moran, R. J. (2016). Profiling neuronal ion channelopathies with non-invasive brain imaging and dynamic causal models: Case studies of single gene mutations. *NeuroImage*, *124*, 43–53.
- Greicius, M. D., Srivastava, G., Reiss, A. L., & Menon, V. (2004). Default-mode network activity distinguishes Alzheimer's disease from healthy aging: Evidence from functional MRI. *Proceedings of the National Academy of Sciences*, *101*, 4637–4642.
- Haken, H. (1977). Synergetics. *Physics Bulletin*, *28*, 412–414.
- Hillebrand, A., & Barnes, G. R. (2002). A quantitative assessment of the sensitivity of whole-head MEG to activity in the adult human cortex. *NeuroImage*, *16*, 638–650.
- Jafarian, A., Assem, M. K., ECE Kocagoncu, E., Lanskey, J. H., Williams, R., Cheng, Y.-J., Quinn, A. J., Pitt, J., Raymont, V., & Lowe, S. (2023). Reliability of dynamic causal modelling of resting state magnetoencephalography. *bioRxiv*, 2023.10.16.562379.
- Jafarian, A., Hughes, L., Adams, N. E., Lanskey, J., Naessens, M., Rouse, M. A., Murley, A. G., Friston, K., & Rowe, J. (2022). Neurochemistry-enriched dynamic causal models of magnetoencephalography, using magnetic resonance spectroscopy. *bioRxiv*.
- Jafarian, A., Hughes, L. E., Adams, N. E., Lanskey, J., Naessens, M., Rouse, M. A., Murley, A. G., Friston, K. J., & Rowe, J. B. (2023). Neurochemistry-enriched dynamic causal models of magnetoencephalography, using magnetic resonance spectroscopy. *NeuroImage*, *276*, 120193.
- Jafarian, A., Litvak, V., Cagnan, H., Friston, K. J., & Zeidman, P. (2020). Comparing dynamic causal models of neurovascular coupling with fMRI and EEG/MEG. *NeuroImage*, *216*, 116734.
- Jafarian, A., Zeidman, P., Litvak, V., & Friston, K. (2019). Structure learning in coupled dynamical systems and dynamic causal modelling. *Philosophical Transactions of the Royal Society A*, *377*, 20190048.
- Jafarian, A., Zeidman, P., Wykes, R. C., Walker, M., & Friston, K. J. (2021). Adiabatic dynamic causal modelling. *NeuroImage*, *238*, 118243.

- Jelenkowska, T. H. (1998). Bayesian estimation of the intraclass correlation coefficients in the mixed linear model. *Applications of Mathematics*, 43, 103–110.
- Jelenkowska, T. H. (1999). Bayesian estimation of the intraclass correlation coefficients in the multivariate one-way model. *Listy Biometryczne-Biometrical Letters*, 36, 61–67.
- Jin, J., Zeidman, P., Friston, K. J., & Kotov, R. (2023). Inferring trajectories of psychotic disorders using dynamic causal modeling. *Computational Psychiatry*, 7, 60–75.
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90, 773–795.
- Kumar, W. S., Manikandan, K., Murty, D. V., Ramesh, R. G., Purokayastha, S., Javali, M., Rao, N. P., & Ray, S. (2022). Stimulus-induced narrowband gamma oscillations are test-retest reliable in human EEG. *Cerebral Cortex Communications*, 3, tgab066.
- Lanskey, J. H., Kocagoncu, E., Quinn, A. J., Cheng, Y.-J., Karadag, M., Pitt, J., Lowe, S., Perkinton, M., Raymont, V., & Singh, K. D. (2022). New therapeutics in Alzheimer's disease longitudinal cohort study (NTAD): Study protocol. *BMJ Open*, 12, e055135.
- Lew, B. J., Fitzgerald, E. E., Ott, L. R., Penhale, S. H., & Wilson, T. W. (2021). Three-year reliability of MEG resting-state oscillatory power. *NeuroImage*, 243, 118516.
- Litvak, V., Garrido, M., Zeidman, P., & Friston, K. (2015). Empirical Bayes for group (DCM) studies: A reproducibility study. *Frontiers in Human Neuroscience*, 9, 670.
- Litvak, V., Jafarian, A., Zeidman, P., Tibon, R., Henson, R. N., & Friston, K. (2019). There's no such thing as a 'true' model: The challenge of assessing face validity. 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC). IEEE, 4403–4408.
- Litvak, V., Mattout, J., Kiebel, S., Phillips, C., Henson, R., Kilner, J., Barnes, G., Oostenveld, R., Daunizeau, J., & Flandin, G. (2011). EEG and MEG data analysis in SPM8. *Computational Intelligence and Neuroscience*, 2011, 1–32.
- Lorenzi, M., Beltramello, A., Mercuri, N. B., Canu, E., Zoccatelli, G., Pizzini, F. B., Alessandrini, F., Cotelli, M., Rosini, S., & Costardi, D. (2011). Effect of memantine on resting state default mode network activity in Alzheimer's disease. *Drugs & Aging*, 28, 205–217.
- Marquetand, J., Vannoni, S., Carboni, M., Li Hegner, Y., Stier, C., Braun, C., & Focke, N. K. (2019). Reliability of magnetoencephalography and high-density electroencephalography resting-state functional connectivity metrics. *Brain Connectivity*, 9, 539–553.
- Moran, R. J., Kiebel, S. J., Stephan, K. E., Reilly, R., Daunizeau, J., & Friston, K. J. (2007). A neural mass model of spectral responses in electrophysiology. *NeuroImage*, 37, 706–720.
- Moran, R. J., Pinotsis, D. A., & Friston, K. J. (2013). Neural masses and fields in dynamic causal modeling. *Frontiers in Computational Neuroscience*, 7, 57.
- Moran, R. J., Stephan, K. E., Dolan, R. J., & Friston, K. J. (2011). Consistent spectral predictors for dynamic causal models of steady-state responses. *NeuroImage*, 55, 1694–1708.
- Mulder, J., & Fox, J.-P. (2019). Bayes factor testing of multiple intraclass correlations. *Bayesian Analysis*, 14, 521–552.
- Penny, W. D. (2012). Comparing dynamic causal models using Aic, BIC and free energy. *NeuroImage*, 59, 319–330.
- Piastra, M. C., Nüßing, A., Vorwerk, J., Clerc, M., Engwer, C., & Wolters, C. H. (2021). A comprehensive study on electroencephalography and magnetoencephalography sensitivity to cortical and subcortical sources. *Human Brain Mapping*, 42, 978–992.
- Prinz, A. A., Bucher, D., & Marder, E. (2004). Similar network activity from disparate circuit parameters. *Nature Neuroscience*, 7, 1345–1352.
- Rowe, J. B. (2010). Connectivity analysis is essential to understand neurological disorders. *Frontiers in Systems Neuroscience*, 4, 144.
- Rowe, J. B., Hughes, L. E., Barker, R. A., & Owen, A. M. (2010). Dynamic causal modelling of effective connectivity from fMRI: Are results reproducible and sensitive to Parkinson's disease and its treatment? *NeuroImage*, 52, 1015–1026.
- Schuyler, B., Ollinger, J. M., Oakes, T. R., Johnstone, T., & Davidson, R. J. (2010). Dynamic causal modeling applied to fMRI data shows high reliability. *NeuroImage*, 49, 603–611.
- Shaw, A., Moran, R. J., Muthukumaraswamy, S. D., Brealy, J., Linden, D., Friston, K. J., & Singh, K. D. (2017). Neurophysiologically-informed markers of individual variability and pharmacological manipulation of human cortical gamma. *NeuroImage*, 161, 19–31.
- Shaw, A. D., Hughes, L. E., Moran, R., Coyle-Gilchrist, I., Rittman, T., & Rowe, J. B. (2021). In vivo assay of cortical microcircuitry in fronto-temporal dementia: A platform for experimental medicine studies. *Cerebral Cortex*, 31, 1837–1847.
- Stephan, K. E., Tittgemeyer, M., Knösche, T. R., Moran, R. J., & Friston, K. J. (2009). Tractography-based priors for dynamic causal models. *NeuroImage*, 47, 1628–1638.
- Stephan, K. E., Weiskopf, N., Drysdale, P. M., Robinson, P. A., & Friston, K. J. (2007). Comparing hemodynamic models with DCM. *NeuroImage*, 38, 387–401.
- Wang, M., & Sun, X. (2014). Bayes factor consistency for one-way random effects model. In *Communications in statistics-theory and methods* (Vol. 43, pp. 5072–5090). Taylor & Francis.
- Zeidman, P., Jafarian, A., Seghier, M. L., Litvak, V., Cagnan, H., Price, C. J., & Friston, K. J. (2019). A guide to group effective connectivity analysis, part 2: Second level analysis with PEB. *NeuroImage*, 200, 12–25.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Jafarian, A., Assem, M. K., Kocagoncu, E., Lanskey, J. H., Williams, R., Cheng, Y.-J., Quinn, A. J., Pitt, J., Raymont, V., Lowe, S., Singh, K. D., Woolrich, M., Nobre, A. C., Henson, R. N., Friston, K. J., & Rowe, J. B. (2024). Reliability of dynamic causal modelling of resting-state magnetoencephalography. *Human Brain Mapping*, 45(10), e26782. <https://doi.org/10.1002/hbm.26782>