
# Statistical decision theory and multiscale analyses of human brain data

D.A. Pinotsis

PII: S0165-0270(20)30335-6

DOI: https://doi.org/10.1016/j.jneumeth.2020.108912

Reference: NSM 108912

To appear in: *Journal of Neuroscience Methods*

Received Date: 25 May 2020

Revised Date: 11 August 2020

Accepted Date: 12 August 2020

Please cite this article as: { doi: https://doi.org/

# Statistical decision theory and multiscale analyses of human brain data

D. A. Pinotsis[1,2,*]

[1]Centre for Mathematical Neuroscience and Psychology and Department of Psychology, City —University of London, London EC1V 0HB, United Kingdom

[2]The Picower Institute for Learning & Memory and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
*pinotsis@mit.edu

Highlights

- Combinations of compartmental and mean field models needed in the Big Data era
- Mathematical proof of a multiscale approach for explaining M/EEG data
- M/EEG data can reveal laminar differences in neural dynamics

Abstract

Background

In the era of Big Data, large scale electrophysiological data from animal and human studies are abundant. These data contain information at multiple spatiotemporal scales. However, current approaches for the analysis of electrophysiological data often contain information at a single spatiotemporal scale only.

New method

We discuss a multiscale approach for the analysis of electrophysiological data. This is based on combining neural models that describe brain responses at different scales. It allows us to make laminar-specific inferences about neurobiological properties of cortical sources using non invasive human electrophysiology data.

Results

We provide a mathematical proof of this approach using statistical decision theory. We also consider its extensions to brain imaging studies including data from the same subjects performing different tasks. As an illustration, we show that changes in gamma oscillations between different people might originate from differences in recurrent connection strengths of inhibitory interneurons in layers 5/6.

Comparison with Existing Methods

This is a new approach that follows up on our recent work. It is different from other approaches where the scale of spatiotemporal dynamics is fixed.

Conclusions

We discussed a multiscale approach for the analysis of human MEG data. This uses a neural mass model that includes constraints informed by a compartmental model. This has two advantages. First, it allows us to find differences in cortical laminar dynamics and understand neurobiological properties like neuromodulation, excitation to inhibition balance etc. using non invasive data. Second, it also allows us to validate macroscale models by exploiting animal data.

Keywords: computational psychiatry; dynamic causal modelling; compartmental models; multiscale approaches; MEG data; statistical decision theory

*Introduction*

Recent developments in brain recording techniques allow one to record data with high spatiotemporal resolution (Jun et al., 2017). At the same time, the ability to collect human brain recordings from large numbers of subjects have revolutionised the study of neurological diseases and disorders (Braund et al., 2018; Williams et al., 2011). We can now study details at the scale of a local cortical circuit using animal models and describe differences between very large numbers of individuals at the macroscopic scale with non-invasive human electrophysiology. These developments suggest the need to develop multiscale approaches. These will allow us to connect animal and human models. So far, approaches for the analysis of brain data contain information at a single spatiotemporal scale only. We here discuss a multiscale approach for brain imaging data analysis. This is based on combining neural models that describe brain responses at different scales.

We focus on a neural mass model that can explain both animal data obtained with thin laminar probes and human MEG data. Neural masses are biophysical models describing neural population responses where ensemble activity is considered as a point process. For a general

introduction to these and other models, see (Deco et al., 2008; Moran et al., 2013). This and other nomenclature used below are defined in Table 1. Following our earlier work (Pinotsis et al., 2017), we provide a mathematical proof that the neural mass model makes similar predictions to a microscopic, compartmental model. This is based on statistical decision theory (Berger, 2013) and shows that both models can be thought of as rules belonging to the same equivalence class. We show that a Bayesian observer could not distinguish between the data predicted separately by each model. Alternatively, if both models are fitted to the same data using Bayesian inference then these fits will have the same error. This suggests the similarity of their predictions.

As an illustration, we considered laminar differences in the excitation to inhibition balance (E-I) in human MEG data reported in (Schwarzkopf et al., 2012). We asked whether cortical function changes at various depths, and focus on differences in the (E-I) balance relevant to both pathophysiology (Chen et al., 2003) and information processing in the brain (Auksztulewicz and Friston, 2015; Friston et al., 2015a; Pinotsis et al., 2014). We find that differences in the E-I balance are expressed in the recurrent connection strengths of inhibitory interneurons in layers 5/6. Although MEG does not provide direct access to laminar data, the use of a neural mass model that makes laminar predictions allowed us to disclose details about cortical function that would otherwise be accessible only by using invasive recordings. In (Pinotsis et al., 2013), we analysed the same dataset using a neural field model. Here, we used a neural mass model instead. Neural masses are a limiting case of neural fields when intrinsic delays on the cortical manifold are neglected, see also Table 1. We did not use a field model because it does not have the same number of parameters as the compartmental model considered below (see Lemma 1 in the Theory and Calculations section).

**Table 1** *Terminology used and definitions*

| **Term** | **Definition and relevant references** |
|---|---|
| *Neural mass model* | Describes the coarse grained responses of neuronal populations using few, biophysically meaningful parameters (Deco et al., 2008, Moran et al., 2013) |

| | |
|---|---|
| *Neural field model* | Similar to a neural mass but also includes explicit parameterization of the location on a cortical manifold (Deco et al., 2008, Pinotsis et al., 2014) |
| *Steady state responses* | Neuronal population responses after steady conditions have been reached and perturbations due to sensory drive have decayed. They reflect synchronous activity of many neurons whose variance is preserved over time (Moran et al., 2009, Pinotsis et al., 2013) |
| *Big data (computational psychiatry)* | Advanced data analytics methods to extract information from large datasets that can be used to understand and treat diseases (Williams et al., 2013, Rutledge et al., 2019; this term has a similar meaning in other fields like economics, meteorology etc). |

In (Pinotsis et al., 2017), we analysed data from laminar electrodes recorded from a single subject. Compared to that earlier work, our current paper includes two new contributions: the analysis of non invasive data and also of data from multiple subjects. Here we analysed MEG data using a hierarchical Bayesian approach (Parametric Empirical Bayes ; PEB) that downweighs neural model parameter estimates from subjects with less reliable data. The neural models used to explain invasive vs non invasive data are different. Although the neural circuitry is the same, the neural model that explains invasive data outputs responses at different depths, while the model that explains non invasive data sums responses across depths. In our earlier work, laminar predictions were fitted to separate electrode tips. These recorded neural activity from superficial and deep cortical layers separately (cf. Figure 7 in Pinotsis et al., 2017). These recordings corresponded to the outputs of neural populations occupying different cortical layers. Here, we used MEG data. Fitting entailed summing up output responses across all layers. The best fit was achieved via a joint optimization of different weights with which different populations contribute to the MEG signal as well as observation (lead field) parameters of the virtual electrode. Also, the analysis presented here focuses on variability in the structure and function of neural sources between different people that might account for the variability in observed brain responses.

*Materials and Methods*

*Data.* We used MEG source reconstructed responses reported in (Schwarzkopf et al., 2012). We considered visually induced oscillations between 30-80Hz from the visual cortex of 16 subjects. In that task, subjects paid attention to the centre of a screen that showed a static, high-contrast, square-wave, vertical grating. MEG data were obtained using a CTF axial gradiometer including 275 sensors, with a sampling rate of 600 Hz. Subject head movement was also recorded and data were preprocessed with the use of SPM8 routines (http://www.fil.ion.ucl.ac.uk/spm). An LCMV beamformer algorithm extracted oscillatory amplitude in an epoch between 0.5 and 1.5 s after stimulus normalised to the prestimulus epoch. This epoch was chosen because neural activity and the corresponding power increase during this epoch was stationary. This is an assumption of the DCM for steady state responses used here (Moran et al., 2009). Peak gamma responses were found in the medial occipital cortex, and at this peak location we used beamforming to obtain virtual electrode responses. Source areas were identified bilaterally. Here we analysed data from source activations in the right hemisphere only. Power spectra are shown in the Supplementary Figure. Peaks were calculated in (Schwarzkopf et al., 2012) using a multitaper spectral estimate. They were identified after fitting a Gaussian function. 3 subjects were excluded from further analysis due to poor goodness of fit. 31% (5 out of 16) of the subjects showed clear peaks (see arrows).

*Biophysical models.* We implemented a neural mass model that describes the local cortical circuit shown in Figure 1. This is based on the DCM toolbox of SPM12, https://www.fil.ion.ucl.ac.uk/spm/software/spm12/, see *spm_fx_cmc_BS.m* in https://github.com/pinotsislab/MicroMacro/. Then we fitted this model to the simulated data above using DCM and estimated its parameters. We also used the symmetric compartmental model of our earlier work (Pinotsis et al., 2016a, 2017) based on the model from (Jones et al., 2007). That model comprised 10 PNs in layers 2/3, 10 PNs in layer 5, and 10 INs in both layers. The synaptic architecture followed general tenets of cortical micro-circuitry where FF connections target the granular layer and FB connections target agranular layers, see e.g. (Pinotsis et al., 2013) for a further discussion. Modelling of single neuron morphology and physiology followed (Bush and Sejnowski, 1993), using the same parameters as in (Jones et al., 2007). Details can be found in (Pinotsis et al., 2017).

This is the compartmental model of the local microcircuit shown in Figure 1 introduced by (Bush and Sejnowski, 1993). In our variant, we changed conductance values and increased the number of inhibitory units from 3 to 10, so that the distribution of E-I neurons is uniform. To ensure that relative differences in interneuron densities were accommodated, we multiplied the maximum conductance values of the corresponding connections by a factor of 0.3. A further simplification was carried out by reducing the network to a single minicolumn, consisting only of one superficial, and one deep PN, and an IN in each layer. Thus, the number of connection parameters (synaptic strengths) in the microscopic and macroscopic models was the same. These parameters are included in Table 2 of the Results section. This satisfies the theoretical requirements suggested by statistical decision theory see the Theory and Calculations section below. It assumes symmetry constraints on horizontal connectivity (within each cortical layer) of the sort assumed in mean field models that describe aggregate activity over hundreds of neurons. A detailed description of the correspondence between the parameters of the two models is provided in the next section.

*Parameter estimation.* To fit the neural mass model *m* to simulated and MEG data we used DCM for steady state responses $g_Y(\omega)$, that is implemented in the function *spm_dcm_csd.m* of the DCM toolbox. Steady state responses are neuronal population responses after perturbations due to sensory drive have decayed. They reflect coordinated activity of many neurons under stationarity and ergodicity assumptions (Moran et al., 2009) , see also Table 1. DCM for steady state responses estimates the connection strengths of a biophysical model that reflect the E-I balance and neurotransmission described by microscale (compartmental) and macroscale (neural mass) models. The optimization of the likelihood model uses a gradient scheme on the negative Free energy, *F*. It minimizes it with respect to a variational density over model parameters $\theta$, $q(\theta) \sim \mathcal{N}(\mu, C)$. After convergence, the variational density approximates the true posterior $q(\theta) \approx p(\theta \mid g_Y(\omega), m)$ and the Free energy approximates the model evidence $F = -\ln p(y|m)$. Thus, the variational density provides posterior estimates of connection parameters.

*Parametric Empirical Bayes.* To model between-subject effects we used a hierarchical Bayesian inference approach known as Parametric Empirical Bayes (PEB), see (Friston et al.,

2015b; Pinotsis et al., 2016b; Preller et al., 2019) and the function *spm_dcm_peb_bmc.m* in the DCM toolbox. This provides an efficient scoring and averaging of large sets of (nested) models (Friston and Penny, 2011). It employs Bayesian model reduction to estimate the posterior density over hidden model parameters for a reduced model (defined in terms of a prior density) using just the posterior density estimated from a full model (with a complete set of parameters). These model parameters were used as regressors in the design matrix. This was part of a GLM with dependent variables gamma peak frequency and V1 size.

*Theory and Calculations*

Here, we present a mathematical proof of the functional equivalence between the compartmental and neural mass model used here and in (Pinotsis et al., 2016a, 2017). We also discuss extensions of our approach that could include computational models predicting data from different modalities (e.g. EEG/fMRI) or tasks (e.g. resting state, task evoked responses). Our proof relies on statistical decision theory. For a primer on this topic, see Appendix and (Berger, 2013).

Assume two Bayesian decision rules $\tilde{\delta}_1$ and $\tilde{\delta}_2$ with corresponding estimates (of the true states of the world $\vartheta$), $\tilde{a} = (\tilde{a}_1, \tilde{a}_2, ..., \tilde{a}_q)$ and $\tilde{b} = (\tilde{b}_1, \tilde{b}_2, ..., \tilde{b}_q)$. Then, the *complete class theorem* (Berger, 2013) implies that the two rules have the same expected loss, $C(\vartheta, \tilde{a})$ (see Equation A.2), that is,

$$C(\vartheta, \tilde{a}) = C(\vartheta, \tilde{b}) \tag{1}$$

We show below that this property of *Bayesian* decision rules has interesting implications for Bayesian estimates of parameters from *different* biophysical *models* that describe the *same brain network*. Consider two computational models of neural population activity, model *M1* and model *M2*. *M1* is a mesoscopic and *M2* is a microscopic model. For concreteness we here consider a compartmental and a neural mass model. They describe the cortical circuit shown

in Figure 1. Our approach however could be straightforwardly applied to other kinds of computational models describing different scales (e.g. point processes, neural fields etc).

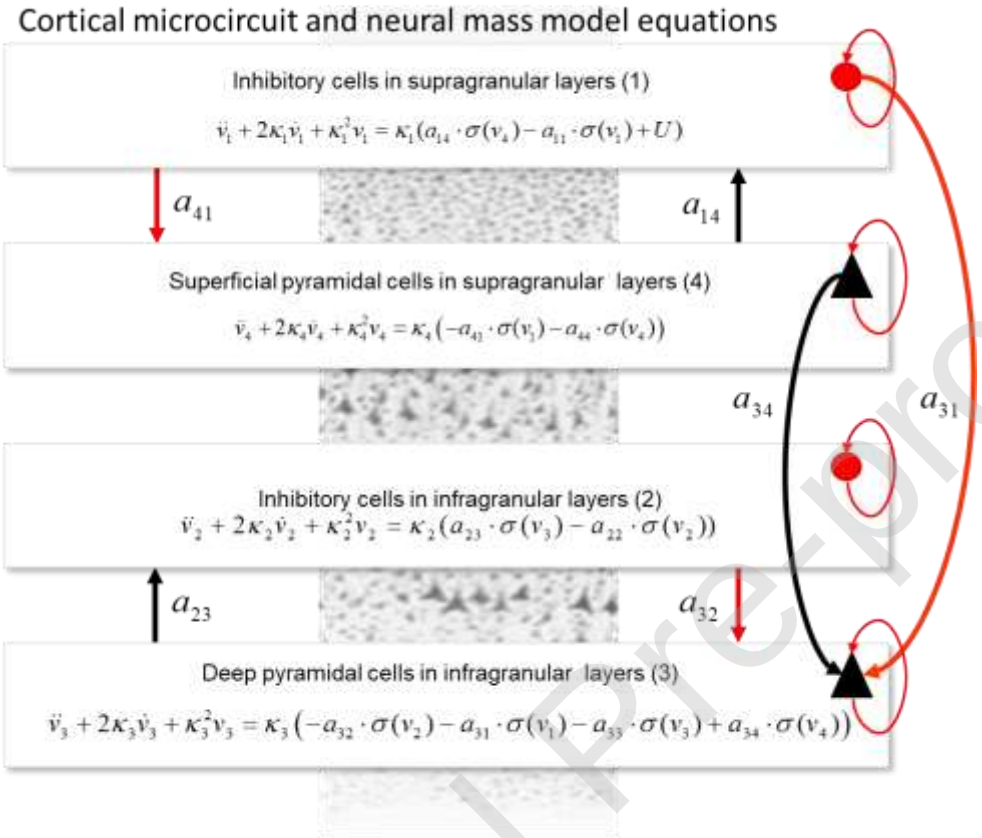


***Fig. 1. The Bush and Sejnowski microcircuit.*** *Excitatory (black) and inhibitory (red) populations occupy superficial and deep cortical layers. Firing rates within each population provide inputs to other populations and convolution of presynaptic activity produces postsynaptic depolarization. Arrows denote excitatory and inhibitory connections. All recurrent connections are inhibitory. The same microcircuit was implemented both as a mean field (Pinotsis et al., 2017) and a compartmental model (Jones et al., 2007). Evolution equations describe the flow of hidden states representing population responses.*

Let model parameters be given by $\vartheta = (\vartheta_1, \vartheta_2, ..., \vartheta_q)$ and $\phi = (\phi_1, \phi_2, ..., \phi_r)$ for models *M1* and *M2* respectively. Fitting them to brain data $o \in O$ using Bayesian inference, we obtain Bayesian

parameter estimates given by $\tilde{a}_{\vartheta} = (\tilde{a}_{\vartheta_1}, \tilde{a}_{\vartheta_2}, ..., \tilde{a}_{\vartheta_q})$ and $\tilde{a}_{\phi} = (\tilde{a}_{\phi_1}, \tilde{a}_{\phi_2}, ..., \tilde{a}_{\phi_r})$. Assume now that $o \in O$ are non invasive electrophysiology data, like the MEG data we considered here. Fitting model *M1* is possible using DCM or some other approach. On the other hand, fitting the model *M2* is hard due to large number of parameters that the compartmental model has. However, we show below that parameter estimates of model *M1 (maximum a posteriori or MAP estimates)* will be equal to a variant of model *M2,* which we call *M2'*.

*Lemma 1. Consider a variant of M2,* called *M2', with the same number of parameters as M1. Then, the MAP estimates* $\mu(\tilde{a}_{\vartheta} | \vartheta)$ and $\mu(\tilde{a}_{\varphi} | \varphi)$ *of parameters* $\vartheta$ *and* $\varphi$ *of M1 and M2' obtained by fitting the same data* $o \in O$ *are equal.*

**Proof.** First, we construct a variant of model *M2* with the same number of parameters as *M1*, *q*. Models *M1* and *M2* are described by the following general equation

$$y_X = \Gamma_X(\theta) + \varepsilon_X \tag{2}$$

where $X$ indexes the two models, $X = \{M1, M2\}$, $\Gamma_X(\theta)$ represents the differential equations of the model that generates responses $y_X$ with parameters $\theta = \{\vartheta, \phi\}$ and $\varepsilon_X$ is observation noise. Specifically, $\Gamma_X(\theta)$ is given by

$$\begin{aligned}
&\Gamma_{M1}(\phi) = \sum_q L_q(\phi) V_q(t), \quad q = 1, ..., 4 \\
&\ddot{V}_q = f_2(V_m, U, \phi) = -2\lambda_q \ddot{V}_q - \lambda_q{}^2 \ddot{V}_q + \lambda_q f_q(v_q, U, \phi) \qquad X = M1 \\
&f_q(v_q, U, \phi) = \begin{cases} a_{14} \cdot \sigma(v_4) - a_{11} \cdot \sigma(v_1) + U, & q = 1 \\ a_{23} \cdot \sigma(v_3) - a_{22} \cdot \sigma(v_2), & q = 2 \\ -a_{32} \cdot \sigma(v_2) - a_{31} \cdot \sigma(v_1) - a_{33} \cdot \sigma(v_3) + a_{34} \cdot \sigma(v_4), & q = 3 \\ -a_{41} \cdot \sigma(v_1) - a_{44} \cdot \sigma(v_4), & q = 4 \end{cases}
\end{aligned} \tag{3}$$

for the neural mass model and

$$\Gamma_{M2}(\vartheta) = \sum_m A_m l_m L_{im} J_m(t)$$
$$J_m = -\frac{\pi A_m^2}{\rho_m} \nabla_x v_m \qquad X = M2 \qquad (4)$$
$$\dot{v}_m = h(v_m, U, \vartheta) = c_m^{-1}[1/(2A_m\rho_m)\nabla_x(A_m^2 \nabla_x v_m) + U]$$

for the compartmental model. *M1* is the neural mass analogue of *M2* introduced in (Pinotsis et al., 2017, 2016a). *M2* is a well-known conductance based (microscopic) model (Bush and Sejnowski, 1993). In this model, neurons and their constituent parts (axonal arbours, soma etc) are considered as cylindrical conductors (segments) and transmembrane potentials are given by aggregates of Ohmic currents. These currents flow across the compartment, forming an RC circuit and obey Kirchhoff's law. $L_{im}$ are lead field coefficients for each compartment and sensor, $A_m, l_m$ are the cross-sectional area and the length of compartment *m* (projected in a direction perpendicular to apical dendrites). $\rho_m$, $c_m$ are the axial resistivity and membrane capacitance and $J_m(t)$ is the longitudinal current density. This model yields detailed descriptions of intracellular longitudinal currents – within the long apical dendrites of synchronized cortical pyramidal cells – that follow from cable theory. Neuronal populations are modelled as spatially organised networks with the soma of principal cells in supragranular and infragranular layers. This model captures the laminar structure of cortical columns and can characterize the cellular and circuit level processes that are measured with multi-electrode arrays or MEG. It also provides a model of neuronal morphology and how neurons are grouped together to form spatially extended networks, with precise connectivity.

The crucial difference between the two models is that *M1* operates at the mesoscale and cannot describe microscopic effects like dendritic delays or back propagation. The model predicts activity $y_X$ based on a local microcolumn depicted in Figure 1. *M2* on the other hand, operates at the microscale and assumes that predictions $y_X$ are generated by an ensemble of smaller structures called mini-columns. Below we considered ten mini-columns. These comprise the

macro-column shown in Figure 1. In this setting, activity predicted from the compartmental model *M2* (see (Jones et al., 2007)) is a simple superposition of minicolumn activities $Y_a(t,\theta)$

$$\begin{aligned} y_{M2}(t,\phi) &= \sum_{a\in\{q\}} Y_a(t,\phi) \\ Y_a(t,\phi) &= Y_a\left(\sum_{q'} A_{q'} l_{q'} L_{q'} J_{q'}(t), Q(J_k, L_k, c_{ka}, \xi)_{\substack{k\in\{q\} \\ k\neq q'}}\right) \end{aligned} \tag{5}$$

where the index $q' \subseteq q$ runs over a subset of *compartments* $q'$ that comprise each mini-column and $a = 1,...,10$ runs over the mini-columns. Each mini-column comprises the compartments of superficial PN, deep PN and superficial and deep interneurons, and $Q(J_k, L_k, c_{ka}, \xi)$ stands for the exogenous input – that depends on activity in proximate compartments indexed by $k \in \{q\}, k \neq q'$. The argument in the factor *Q* in Equation (5) above simply means that this input depends upon the current density in the adjacent mini-columns, their lead fields, anatomical parameters $\xi$ and the strength of connections $c_{ka}$.

We assumed that $c_{ka}$ are the same between any mini-column pair and that all mini-columns have the same anatomy. Thus, the number of parameters in model *M2* was reduced by a factor equal to the number of minicolumns comprising one macrocolumn. Because *M2* describes the same circuit as *M1* and all minicolumns have the same parameters, the resulting model (which we call, the *symmetric compartmental model* and denote by *M2'*) has the same number of connection parameters as the mean field model *M1*. These parameters are included in Table 2. We denote the *M2'* parameters by $\varphi = (\varphi_1, \varphi_2, ..., \varphi_q)$ and the corresponding Bayesian estimates by $\tilde{a}_\varphi = (\tilde{a}_{\varphi_1}, \tilde{a}_{\varphi_2}, ..., \tilde{a}_{\varphi_q})$. Despite the reduction in the number of parameters, predictions of laminar dynamics from models *M2* and *M2'* were very similar, see (Pinotsis et al., 2017). Activities of minicolumns lie on an invariant subspace in the full phase space of the network and have the same dynamics; see (Afraimovich et al., 2001; Breakspear et al., 2003) for a discussion of invariant subspaces. In this case, it is *not* possible to distinguish between minicolumns and describe horizontal interactions within a cortical layer. Assuming the same connection

parameters across mini-columns comes at a cost: it means that we have neglected horizontal interactions within the same cortical layer mediated by recurrent synaptic coupling. On the other hand, we have described laminar interactions with a mean field model. Following DCM for steady state (Moran et al., 2009), we here considered that the neural mass model *M2* is perturbed around a fixed point attractor state. We also considered that the compartmental model M1 operates in the same dynamical regime. In (Pinotsis et al., 2017), we found a very high correlation between the responses of the two models ($r=0.9343$, $p < 0.001$). We did not consider other non-trivial bifurcations and richer dynamics that *M1* can exhibit. To sum up, we assumed that all mini-columns in the compartmental model *M2* have the same anatomy and the connection weights $c_{ka}$ are the same between any mini-column pair.

Consider now the connection parameters of models *M1* and *M2'* as the true states of the world (in the language of statistical decision theory). Then *M1* and *M2'* can be thought of as two Bayesian decision rules $\tilde{\delta}_1$ and $\tilde{\delta}_2$ that the statistician used to infer them. In this case, the corresponding estimates of the true states $\vartheta$, are given by $\tilde{a}_\vartheta = (\tilde{a}_{\vartheta 1}, \tilde{a}_{\vartheta 2}, ..., \tilde{a}_{\vartheta q})$ and $\tilde{a}_\varphi = (\tilde{a}_{\varphi 1}, \tilde{a}_{\varphi 2}, ..., \tilde{a}_{\varphi q})$. Also, Equation (1) above implies

$$C_\vartheta(\vartheta, \tilde{a}_\vartheta) = C_\varphi(\vartheta, \tilde{a}_\varphi) \tag{6}$$

The corresponding Bayes loss satisfies (see also Equation A.2)

$$\int R(\vartheta, \tilde{a}_\vartheta)\, p(\vartheta) d\vartheta = \int R(\varphi, \tilde{a}_\varphi)\, p(\varphi) d\varphi \tag{7}$$

Assuming the same priors of the statistician about the true states, $p(\vartheta) = p(\varphi)$, we get

$$R(\vartheta, \tilde{a}_\vartheta) = R(\varphi, \tilde{a}_\varphi)$$
$$\int L(\vartheta, \tilde{a}_\vartheta)\, p(\tilde{a}_\vartheta | \vartheta) d\tilde{a}_\vartheta = \int L(\varphi, \tilde{a}_{\varphi})\, p(\tilde{a}_\varphi | \varphi) d\tilde{a}_\varphi \quad (8)$$

Considering point density probability distributions, the above Equation implies an Equality of MAP estimates $\mu(\tilde{a}_\vartheta | \vartheta) = \mu(\tilde{a}_\varphi | \varphi)$.

This concludes the proof of Lemma 1.QED.

To sum up, the MAP estimates for the *M1* and *M2'* are equal $\mu(\tilde{a}^1_\vartheta | \vartheta) = \mu(\tilde{a}^1_\varphi | \varphi)$, that is, $< p(\tilde{a}^1_\vartheta | \vartheta) >=< p(\tilde{a}^1_\varphi | \varphi) >$ and we assumed that *M1* and *M2'* describe the cortical microcircuit equally well, i.e. $\boldsymbol{p(\vartheta) = p(\varphi)}$. Motivated by these results we now prove the second Lemma that establishes the construct validity of *M1* with respect of *M2'*, i.e. that the two models predict similar data –and *M1* is a neural mass model that predicts laminar responses.

*Lemma 2. Let $o \in O$ be simulated data from M2' (the symmetric compartmental model). The Maximum Likelihood (ML) problems for models M1 and M2'*

$$\begin{aligned} p(\tilde{a}_\vartheta | \vartheta) &= \arg\min_{\tilde{a}_\vartheta} [\, f(\vartheta) + f(o)] \\ f(\vartheta) &= -KL(q(\tilde{a}_\vartheta | \vartheta) | p(\vartheta)) \\ f(o) &= \int \log p(o | \tilde{a}_\vartheta, \vartheta) q(\tilde{a}_\vartheta | \vartheta) d\vartheta \end{aligned} \quad (9)$$

*and*

$$\begin{aligned} p(\tilde{a}_\varphi | \varphi) &= \arg\min_{\tilde{a}_\varphi} [\, f(\varphi)] \\ f(\varphi) &= -KL(q(\tilde{a}_\varphi | \varphi) | p(\varphi)) \end{aligned} \quad (10)$$

*are the same.*

**Proof.** The *ML* estimation problem for a model *M* is the problem of finding the posterior distribution of model parameters $p(\tilde{a}_\vartheta|\vartheta)$ by optimizing the free energy, $F(o,\vartheta)$, see e.g. (Pinotsis et al., 2012) for details:

$$\begin{aligned} p(\tilde{a}_\vartheta|\vartheta) &= \arg\min_{\tilde{a}_\vartheta} F(o,\vartheta) \\ F(o,\vartheta) &= -KL(q(\tilde{a}_\vartheta\vartheta)|p(\vartheta)) + \int \log p(o|\tilde{a}_\vartheta,\vartheta) q(\tilde{a}_\vartheta|\vartheta) d\vartheta \end{aligned} \tag{11}$$

Here $p(o|\tilde{a}_\vartheta,\vartheta)$ is the data log-likelihood and $q(\tilde{a}_\vartheta|\vartheta)$ is the approximate posterior. The Free Energy $F(o,\vartheta)$ has two parts

$$F(o,\vartheta) = f(\vartheta) + f(o) \tag{12}$$

One depends on prior beliefs about the model parameters $p(\vartheta)$ and the other on observations $o \in O$. Thus the posterior should not move too far from our prior guess and also maximize the data log-likelihood. Then Equation (11) for model *M1* is the same as Equation (9). Using Lemma 1, we can also replace $\vartheta$ by $\varphi$ in Equation (9)

$$\begin{aligned} p(\tilde{a}_\varphi|\varphi) &= \arg\min_{\tilde{a}_\varphi} [f(\varphi) + f(o)] \\ f(\varphi) &= -KL(q(\tilde{a}_\varphi|\varphi)|p(\varphi)) \\ f(o) &= \int \log p(o|\tilde{a}_\varphi,\varphi) q(\tilde{a}_\varphi|\varphi) d\varphi \end{aligned} \tag{13}$$

To sum up, Equation (13) is equivalent to Equation (9). Because $o \in O$ are *simulated data from M2'*, $p(o|\tilde{a}_\varphi,\varphi) = 1$ and Equation (13) yields Equation (10).

This concludes the proof of Lemma 2.QED.

Since the *ML* estimation problem for the two models are the same, it follows that fitting the neural mass (*M1*) model to simulated data $o$ from the compartmental model (*M2'*), we perform the same parameter inference as if we were fitting the compartmental model itself. In the context of statistical decision theory , the free energy can be thought of as a cumulative loss function of the form of Equation (12). Then Bayesian inference is a decision problem that optimizes (12). The statistician collects observations $o \in O$ to infer parameters. Assume that we sample the data space successively $O$, obtaining brain data $o_1$ first, then $o_2, o_3$ etc. with $o_1, o_2, o_3, ... \in O.$ This is known as Bayesian belief updating (Cooray et al., 2016) and suggests that that we can assume that the loss function (12) can optimized in the same successive, optimal and rational manner after observing each data sample $o_i \in O$, *i=1,2,…*, that is $F(o, \vartheta) = \sum_i \left( f_i(\vartheta) + f_i(o) \right)$ Let us then assume that the first sample $o_1$ was *simulated data from M2'* similarly to Lemma 2 and that the second sample $o_2 \in O$ was the MEG source reconstructed data that we considered in the Results section below. Similarly to Equation (11), the inference problem for model *M1* and sample $o_2$ can be written as

$$
\begin{aligned}
p(\tilde{a}_\vartheta^2 | \vartheta) &= \arg\min_{\tilde{a}_\vartheta} \left[ f_2(\vartheta) + f_2(o^2) \right] \\
f_2(\vartheta) &= -KL(q(\tilde{a}_\vartheta^2 | \vartheta) | p(\vartheta)) \\
f_2(o^2) &= \int \log p(o^2 | \tilde{a}_\vartheta^2, \vartheta) q(\tilde{a}_\vartheta^2 | \vartheta) d\vartheta
\end{aligned}
\tag{14}
$$

where we use an upper index, *i=2*, in $\tilde{a}_\vartheta^i$ to denote the parameter estimates obtained after observing the data sample $o_i \in O$ and a lower index in $f_i(\vartheta), f_i(o)$ to denote step wise components of $F(o, \vartheta)$ as above. For simulated data, the expectation of $p(\vartheta)$, $< p(\vartheta) >= \mu(\tilde{a}_\vartheta^1 | \vartheta) = \mu(\tilde{a}_\varphi^1 | \varphi)$. It is equal to the MAP estimate obtained from the previous step (fitting $o_1$). Thus, instead of fitting the mean field model *M1* to the MEG data $o_2 \in O$ (as done in traditional DCM), we consider an extended data space comprising simulated and empirical data $o_1, o_2 \in O$ and successively fit *M1* to $o_1$ and $o_2$. The first fit ensured that model *M1* makes

the same predictions as the compartmental model *M2'* when fitted to MEG data. This corresponds to our two-step approach—where each step corresponds to a different sample. Step one: construct a micro-scale (e.g. compartmental) model that captures biophysical properties of single neurons. Test it with intracranial recordings from animals, $o_1$. Step two: tune the parameters of a neural mass model to give similar predictions as the compartmental model. Test it with human M/EEG data, $o_2$.

One could also consider a third sample $o_3$, with $o_1, o_2, o_3, ... \in O$. This could correspond to data from a different modality that model *M1* can predict and fit the model to the new dataset using the posteriors obtained from fitting $o_2$ as priors. This is the approach taken by (Jafarian et al., 2019; Wei et al., 2020) to achieve Bayesian fusion. In that work, the neural model *M1* is coupled with different observation models that predict EEG and fMRI data: $o_2$ and $o_3$ are source reconstructed EEG data and deconvolved fMRI data respectively. Similarly, one could use the approach described above to fit the model to different datasets from the same subjects. Let for example, $o_2$ and $o_3$ be resting state and task data and assume that model *M1* predicts both for different values of its input, e.g. it is a canonical microcircuit neural mass model that predicts both steady state responses $o_2$ (Pinotsis et al., 2018) and evoked potentials $o_3$ (Díez et al., 2017; Ranlund et al., 2016). This will be pursued elsewhere.

*Results*

As an illustration, we applied our approach to human MEG data from (Schwarzkopf et al., 2012). We fitted power spectra of visually induced oscillations between 30—80Hz from different people, similar to our previous work (Pinotsis et al., 2013). The corresponding gamma-band frequency correlated with the retinotopically determined surface area of central V1 (Schwarzkopf et al., 2012). Data were recorded from the visual cortex of 16 subjects while they viewed a static, high-contrast vertical grating.

We used the neural mass model shown in Figure 1 described in detail in (Pinotsis et al., 2017). This model includes two pairs of excitatory-inhibitory populations, one in superficial and the other in deep layers. Each population pair is connected with intralaminar connections. Superficial and deep populations are also connected with interlaminar connections. Arrows correspond to excitatory (black) and inhibitory (red) connections.

To fit these data, the parameters of the neural mass model had been tuned to predict similar laminar data as a compartmental model. The mathematical basis and constraints for this are described in the Theory and Calculations section. In Figure 2A, model predictions and data are shown in dashed and solid lines along with 95% confidence intervals across subjects for the data from (Schwarzkopf et al., 2012). These quantify variability across subjects and overlap for model predictions and data. In Figure 2B, percentage variance explained for individual subject fits is shown. A Welch t-test yielded non-significant differences in the accuracy of fits to the five subjects that had strong gamma peaks compared to the rest (shown with arrows in Supplementary Figure). Mean and standard deviation of the accuracy of fits for the five subjects with pronounced gamma peaks ("strong") is shown in the left bar of Figure 2B; while mean and standard deviation of fits of the remaining subjects in the right bar ("weak"). Example fits from individual subjects are shown in Figure 2C. The parameters of the tuned neural mass model were used as priors for subsequent fits using Dynamic Causal Modeling (DCM) for steady state responses (Moran et al., 2009). They were obtained by fitting the neural mass model of Figure 1 to simulated data from the compartmental model of (Jones et al., 2007). These parameters are included in Table 2.

**Table 2** ***Synaptic connectivity parameters for the microscale and macroscale models.***

| | Description | Max conductance (μS) Microscale model | Intrinsic connectivity Macroscale model (a.u) |
|---|---|---|---|
| $a_{44}$ | **SP–>SP** | **0.001** | **4.4** |
| $a_{14}$ | **SP–>SI** | **0.003** | **4.8** |
| $a_{34}$ | **SP–>DP** | **0.00025** | **23.3** |
| $a_{41}$ | **SI–>SP** | **0.015** | **3.8** |
| $a_{31}$ | **SI–>DP** | **0.0003** | **5.9** |
| $a_{11}$ | **SI–>SI** | **0.0006** | **4.2** |
| $a_{33}$ | **DP–>DP** | **0.005** | **2.2** |
| $a_{23}$ | **DP–>DI** | **0.0003** | **4.6** |
| $a_{32}$ | **DI–>DP** | **0.0075** | **6.9** |
| $a_{22}$ | **DI–>DI** | **0.0006** | **4.16** |

The fine tuning of the neural mass model based on compartmental model predictions was described in detail in (Pinotsis et al., 2017). In that earlier work, fine tuning adjusted the priors of the neural mass model to describe the difference in the timings of neural activity and delays in signal propagation between superficial vs deep populations, due e.g. to the dispersion of axons, in the same way as the microscale (compartmental) model. We also validated this model using laminar data from mice reported in (Pinto et al., 2013), see (Pinotsis et al., 2017) and also monkey data, see (Pinotsis et al., 2016a). We showed (using Bayesian model comparison (Pinotsis et al., 2013)) that the neural mass model could correctly distinguish between deep and superficial neural responses. This established its construct validity. We here used the same model considered in that earlier work after adapting it to fit MEG data. This has the advantage that its parameters had been tuned to predict superficial and deep cortical dynamics. In other words, the model makes predictions of neural responses at different cortical depths.

Fitting the fine tuned neural mass model to human MEG data his allowed us to perform laminar-specific inferences. Motivated by earlier results (Muthukumaraswamy et al., 2009; Schwarzkopf et al., 2012) we asked whether gamma peak frequency variability might relate to differences in the concentration of the inhibitory neurotransmitter GABA in the occipital lobe, measured using magnetic resonance spectroscopy (MRS). This has been shown to correlate with peak gamma- oscillation frequency. We tested this hypothesis and asked what the origin of differences in cortical inhibition between different people might be. Specifically, we asked which inhibitory connection strengths can predict V1 size[1] and gamma peak frequency. We used a recent approach known as parametric empirical Bayes (PEB; see Materials and Methods and Friston et al., 2015b; Pinotsis et al., 2016b). This employs a Savage-Dickey approximation (Verdinelli and Wasserman, 1995) to perform an exhaustive search over all combinations of model parameters that might drive differences in gamma responses or anatomy between different people. This approach considers all possible combination of connection strengths shown as arrows in Figure 1. It scores, in terms of a Bayesian GLM, how well they can predict a given phenotype; here V1 size and gamma peak frequency that were available for each subject along with MEG oscillations from the dataset from (Schwarzkopf et al., 2012).

[1] In (Schwarzkopf et al., 2012) V1 size had been measured using retinotopic mapping with fMRI: this calculated the stimulus location at which each voxel responded (Sereno et al., 1995) and these data were subsequently used to delineate V1 surface using Freesurfer (Fischl et al., 1999).

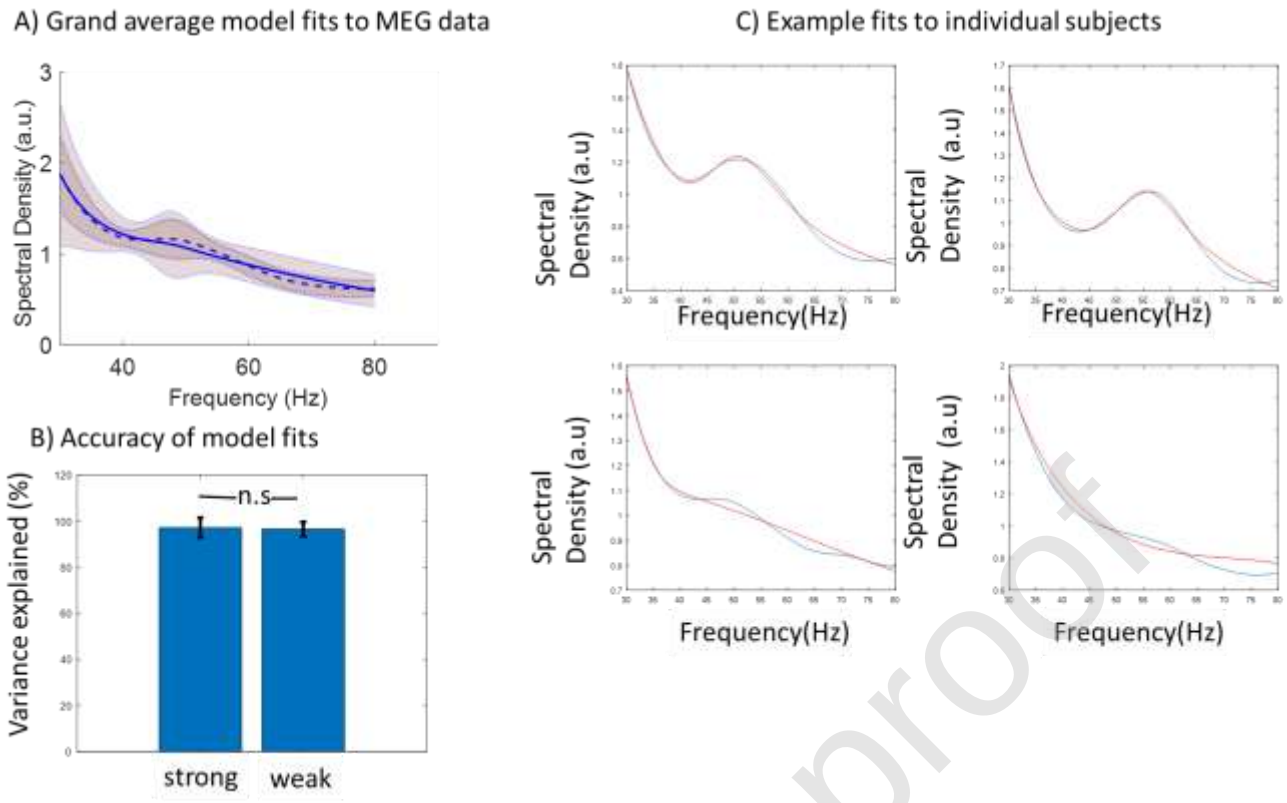


***Fig. 2.*** *A) Empirical responses (power spectra) and model fits are shown in dashed and solid lines along with 95% confidence intervals across subjects in data from (Schwarzkopf et al., 2012). B) Percentage variance explained in individual subject fits. No significant differences were found in the accuracy of fits between spectra including strong and weak peaks (see text and Supplementary Figure). C)Example* ***model fits to individual subjects.*** *Red and blue lines correspond to data and model predictions.*

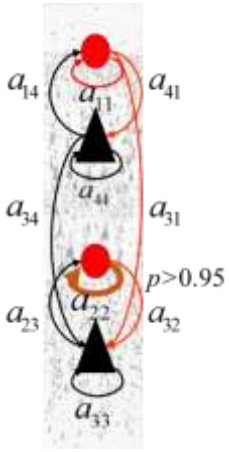


***Fig. 3. PEB analysis.*** *We scored alternative GLMs where predictors of variability in V1 included any combination of the connections (arrows) in Fig.1. We found that for the data from* (Schwarzkopf et al., 2012). *V1 size could be best predicted by the recurrent connectivity of deep inhibitory interneurons,* $a_{22}$ *(orange arrow). Evidence was very strong p>0.95.*

We found that gamma peak frequency variability was not predicted by connection changes between subjects. However, V1 size variability was. Using PEB, single subject estimates of the recurrent connection strengths in the deep inhibitory population could predict the corresponding size of V1 (Figure 3). This was in agreement with our earlier results in (Pinotsis et al., 2013) where we had also found (using a different, neural field model), that individual differences in gamma oscillations correlated with variations in the excitatory drive to GABAergic interneurons. The current results are also similar to results in (Pinotsis et al., 2016b), where we found using the same neural field model and a similar MEG dataset that differences between subjects correlated with interneuron output. Both those earlier results used a model that describes spatially distributed spectral responses over the cortical manifold (neural field), that is, *within* the same cortical layer. However, that earlier model could not make predictions of differences in cortical dynamics *between* different layers, like the model we considered here. To sum up, we here confirmed that individual variability in V1 size correlated with differences in the excitation to inhibition balance that we had found earlier. We also found a novel result: these differences suggest differences in deep inhibitory interneuron activity.

*Discussion*

The ability to process and store big datasets, known as Big Data, has revolutionized cognitive and computational neuroscience among other fields. We now have the ability to record electrophysiological data at an unprecedented spatial and temporal resolution, e.g. the first generation of Neuropixel probes could record more than 700 neurons from five brain structures (Jun et al., 2017). We can also collect and process data from thousands of healthy

and patient subjects in large scale clinical studies with psychiatric patients, like the i-SPOTD study of more than 2000 patients with depression (Braund et al., 2018; Williams et al., 2011). In brief, as a result of the Big Data revolution, there are new, unparalleled opportunities to study brain structure and function. We can study details at the scale of a local cortical circuit using animal models and describe differences between very large numbers of individuals at the macroscopic scale with non-invasive human electrophysiology. These developments suggest that multi-model, multi-scale approaches are required for the analysis of brain data. We presented such an approach here.

Specifically, we used a neural mass model to explain non-invasive human MEG data. The parameters of this model were fine tuned so that the model makes similar predictions with a compartmental modelthat describes neural activity at the microscopic scale. This allowed us to make laminar-specific inferences regarding the generators of the bulk macroscopic response. Our approach is different from some common uses of biophysical models in computational neuroscience. Often, in these applications, the scale of spatiotemporal dynamics described by the model is fixed. It depends on the available dataset and the researchers' interests, see e.g. (Bassett et al., 2018; Poldrack et al., 2018). Here, we present an alternative approach that uses *combinations* of biophysical models that explain spatiotemporal dynamics at the micro- and macro- scales (Pinotsis et al., 2016, 2017). This has two advantages. First, it allows us to find differences in cortical laminar dynamics using non invasive M/EEG data. Second, it also allows us to validate macroscale models of the sort used in Dynamic Causal Modeling (DCM) by exploiting animal data.

Our approach has two steps. Step one: construct a mesoscale (mean field) model that includes the same neuronal populations as a validated microscopic (compartmental) model that captures biophysical properties of single neurons (e.g. the geometry of the dendritic tree, kinetics and densities of ion channels, inputs from subcortical areas etc; (Bush and Sejnowski, 1993; Jones, 2016; Kopell et al., 2000) Fine tune its parameters to give similar predictions as the compartmental model. Test it using invasive, animal data, including local field potentials and multielectrode unit activity, similarly to (Pinotsis et al., 2016a; Pinotsis et al., 2017). Step two: use the same mescoscale model to explain human M/EEG data.

We established our approach by constructing a mathematical proof based on statistical decision theory. This theory studies optimal decisions in the face of uncertainty in the data (Berger, 2013). It has found numerous applications in neuroscience and psychology, including reinforcement learning (Dayan and Daw, 2008) and decision making (Trommershäuser et al., 2008). Our proof uses statistical decision theory to reformulate biophysical models as decision rules. Then it establishes model equivalence by considering models as admissible rules.

Besides its mathematical merit, the proof shows how the compartmental model constraints the particular mean field model that can be used to explain the same data. This model needs to have the same connection parameters as the compartmental model (see also Lemma 1). It can be a neural mass (that we used here) but it cannot be *any* mean field model (even if it predicted similar data), like e.g. a neural field that has an unequal number of connection parameters. The proof also dictates simplifications and adaptations of the original model by (Jones et al., 2007). In the variant we used here, activities of minicolumns were constrained to lie on an invariant subspace neglecting horizontal interactions within layer.

In brief, using the complete class theorem, our proof suggests how to construct low dimensional models making similar predictions to detailed models with large number of parameters (Lemma 1) and how to test if they are equivalent (Lemma 2). Although parameter estimation of the sort we considered here is an ill posed inverse problem (many different parameterizations might give rise to the same data features, see e.g. ( Pinotsis et al., 2013)), predicting the same data features does not mean that models are equivalent from the point of view of a Bayesian observer making decisions. They do not describe the same latent structure (true states of the nature). This is actually the essence of Bayesian model comparison where alternative models are scored in terms of how well they fit the data, see e.g. (Pinotsis et al., 2012) for a comparison of neural masses vs. fields using local field potential data.

Our approach can be used to test the biological plausibility of mean field models. These describe coarse grained neural activity at the macroscale. The approach can assess the limitations of simpler, low dimensional models in terms of reproducing predictions from more detailed e.g. compartmental or point process models. The latter describe activity at the microscale and can be validated in vitro (White et al., 2000). Validating mean field models against more detailed descriptions of brain anatomy and structure is important also for effective

connectivity studies as when we find causality, we do not want this to be just "model, not biological causality" (Mehler and Kording, 2018).

Our approach provides a way to test for biological plausibility of mean field models using multiscale data. Before fitting them to M/EEG data, one can do an analysis similar to what we discussed here by using animal data. This would correspond to step one of our approach. It would fine tune their parameters to reproduce the same predictions as earlier compartmental models. Then they could be further validated using invasive, animal data, like we did in (Pinotsis et al,, 2016a; Pinotsis et al., 2017) . Thus, while fitting the macroscale, neural mass model to MEG data here, the same temporal constraints inherited from the microscale model were preserved. Because of constraints imposed upon the parameters of the microscale model, spatial effects within the same layer were neglected (e.g. interactions between cortical columns and the dependence of gamma peak frequency on the horizontal spread of the underlying cortical source).

Our approach provides a way to implement data fusion across scales. Here, we applied the approach to explain data with different spatiotemporal resolution. It can also be used to explain multimodal datasets similar to (Jafarian et al., 2019; Wei et al., 2020) and also data from different tasks obtained with the same modality.

In all the above three cases, data fusion (across scales, modalities or tasks) rests upon constructing models that can be thought of belonging to the same equivalence class, as suggested by our proof (see Theory and Calculations). For example, for data fusion across tasks, the parameters of a neural mass model for a single subject could be estimated successively by first fitting it to data from one task (e.g. resting state) and then refining these estimates by fitting the model to data from the second task (e.g. oddball response). Importantly, the priors for the second fitting would be the posteriors of the first and the drive to the model would be adapted to describe the difference of inputs between tasks. This can be important for Computational Psychiatry where different aspects of the same pathology might be evident in various tasks.

In future work, we will also study the relationship between population activity predicted by the macroscale model and the symmetric compartmental model we used here. Spatial effects within the same cortical layer can be described by extending our approach to a different class of mean field models besides neural masses, called neural fields (Pinotsis et al., 2016b). We will also consider datasets from the same human subject recorded with different modalities or performing different tasks. In that case, parameter estimates obtained by e.g. fitting resting state activity will be used as priors to explain task induced responses. This can be important for applications as there might be complementary information in different datasets that could lead to a better understanding of the heterogeneity of neurological diseases and disorders than what afforded by current diagnostic classifications.

*Acknowledgments*

This work was supported by UKRI ES/T01279X/1. I thank Professor Earl Miller and Dr Daniel Gibson for important suggestions. I also thank and Dr Sam Schwarzkopf for informative discussions and providing data.

Author contributions. DAP: conceptualization, data curation, analysis and writing the ms.

*Appendix: Elements of statistical decision theory*

Consider a sample space $O$, that is a Borel subset of a Euclidean space, and an observation or sample of brain data $o \in O \subseteq \mathbb{R}^n$. A decision rule $\delta$ is a method that the statistician uses to infer the true states of nature $\vartheta$ based on brain data $o \in O$. In the context of statistical decision theory, there are many decision rules from which the true states can be obtained. These give rise to the space $A$, the space of possible estimates, called also the decision space. Each different rule gives a different estimate of $\vartheta$. Let $\vartheta \in \Theta \subseteq \mathbb{R}^m$, $m<n$, where $\Theta$ is called the space of all possible parameters. Here, $\vartheta = (\vartheta_1, \vartheta_2, ..., \vartheta_q)$ $k \in \{1, ..., q\}$, parametrize a biophysical model *M* and the space $\Theta$ includes parameters that give rise to observed brain dynamics. This is a generative model of brain data, that is, a mapping from biophysical parameters $\vartheta$ to observed data $o \in O$. Parameter inference can be reformulated as a problem of statistical decision theory (Bissiri et al., 2016). The decision rule $\delta$ is used to obtain an estimate of model parameters. Instead of optimizing a likelihood function, this approach entails the optimization of a loss function. After applying the decision rule, estimates of the parameters $\vartheta$ are denoted by $a \in A \subseteq \mathbb{R}^m$ and denoted by $a = a(o)$ as they depend on the data $o$. They are collectively denoted by $a = (a_1, a_2, ..., a_q)$. Thus, the decision rule is a mapping from the sample space $O$ into the decision space $A$, $\delta : O \overset{\delta}{\rightarrow} A$. To sum up, the true and estimated

model parameters are denoted by $\vartheta_k$ and $a_k$ respectively, where *k* runs over all model parameters.

We here consider a *Bayesian* decision rule and denote the corresponding estimates of model parameters by a tilde, $\tilde{a} \in A$. These are distributed according to $p(\tilde{a}|\vartheta)$, that is, the probability that the true states of nature are equal to $\tilde{a}$ *after* observing $o \in O$. The error the statistician makes when obtaining $a = a(o)$ is given by the *loss function* $L(\vartheta, a): \mathbb{R}^q \times \mathbb{R}^q \to \mathbb{R}$. This is a real valued function that quantifies the discrepancy between $\vartheta_k$ and $a_k$. The expected value of the loss function $L(\vartheta, \tilde{a})$ *after* obtaining Bayesian estimates $\tilde{a} = (\tilde{a}_1, \tilde{a}_2, ..., \tilde{a}_q)$, is called *Bayes loss* and is given by

$$R(\vartheta, \tilde{a}) = \int L(\vartheta, \tilde{a}) p(\tilde{a}|\vartheta) d\tilde{a} \tag{A.1}$$

In other words, the Bayes loss $R(\vartheta, \tilde{a})$ quantifies the expectation of the statistician about the error they will make when using Bayesian inference.

There are many ways to implement Bayesian inference, each corresponding to a different Bayesian decision rule $\tilde{\delta}$. All such rules are called admissible, that is, they *minimize* the same *expected loss* $C(\vartheta, \tilde{a}) \equiv \inf_{\delta: o \in O \to a \in A} C\vartheta$ (a This is known as the *complete class theorem* (Berger, 2013). $C(\vartheta, \tilde{a})$ is given in terms of Bayes loss $R(\vartheta, \tilde{a})$ by the following relation

$$C(\vartheta, \tilde{a}) = \int R(\vartheta, \tilde{a}) p(\vartheta) d\vartheta \tag{A.2}$$

Thus, the expected loss is just the expected value of the Bayes loss under the prior $p(\vartheta)$.

*Example (adapted from Berger, 2013).* A drug company wants to decide whether or not to market a new pain reliever. The main factor affecting its decision are the proportion of people for which the drug will prove effective $\theta_1$. This is the latent variable (true state of nature) that is typically unknown. Statistical decision amounts to evaluating costs or losses based on same sample information combined with some other, e.g. prior or complementary, information. In particular, the use of prior information –outside the experiment considered—is known as Bayesian decision theory. In this example, the prior $p(\theta_1)$ could be calculated based on the success of other similar drugs etc. Assume that the drug is effective for a percentage of users which we denote by $a$, that is, $p(\theta_1) = N_{(a,\sigma)}(\theta_1)$. It could also be more detrimental for the

company to overestimate $\theta_1$, compared to underestimating it, because of potential fines; this can be included in a loss function of a form

$$L(\theta_1, a) = \begin{cases} \theta_1 - a & |\vartheta_1 - a| \le 0, \\ 2(\theta_1 - a) & |\vartheta_1 - a| > 0 \end{cases} \tag{A.3}$$

The posterior $p(\tilde{a}|\theta_1)$ can be found using Bayes rule and iteratively sampling data points $o_1, o_2, o_3, \ldots \in O$

$$p(\tilde{a}_1|\theta_1) = \frac{p(\theta_1)}{p(O)} \prod_{l=1}^{r} p(o_l|\theta_1) \tag{A.4}$$

where the denominator is given by

$$p(O) = \int p(\theta_1') \prod_{l=1}^{r} p(o_l|\theta_1') d\theta_1' \tag{A.5}$$

To find $p(o_k|\theta_1)$, the drug company could perform tests or surveys asking people whether they feel less pain after taking the drug $p(o_l|\theta_1) = N_{(a,s)}(\theta_1)$. Then equation (A.2) furnishes the expected loss

$$\begin{aligned} C(\theta_1, \tilde{a}) &= \int_0^{\tilde{a}} (\tilde{a} - \theta_1) N_{(\tilde{a},\sigma)} d\theta_1 + 2\int_{\tilde{a}}^{1} (\theta_1 - \tilde{a}) N_{(\tilde{a},\sigma)} d\theta_1 \\ &= \left(1 - \mathrm{e}^{-\frac{(-1+\tilde{a})^2}{2\sigma^2}}\right)\sqrt{\frac{2}{\pi}}\sigma + \frac{\left(1 - \mathrm{e}^{-\frac{\tilde{a}^2}{2\sigma^2}}\right)s}{\sqrt{2\pi}} \end{aligned} \tag{A.6}$$

The above expression yields $C(\theta_1, \tilde{a})$, based on the drug company's prior expectation about proportion of people for which the drug will prove effective, $a$, and the uncertainty of this estimate, $\sigma$. Then after the new drug has entered the market this is updated to

$$C(\theta_1,\tilde{a}) = \frac{\left(1-\mathrm{e}^{-\frac{(-1+\tilde{a})^2\left(\sigma^2+s^2\right)}{2\sigma^2 s^2}}\right)\sigma s}{\pi\left(\sigma^2+s^2\right)} + \frac{\left(1-\mathrm{e}^{-\frac{\tilde{a}^2\left(\sigma^2+s^2\right)}{2\sigma^2 s^2}}\right)\sigma s}{2\pi\left(\sigma^2+s^2\right)} \tag{A.7}$$

## References


Afraimovich, V., Chazottes, J.R., Cordonet, A., 2001. Synchronization in directionally coupled systems: Some rigorous results. DISCRETE Contin. Dyn. Syst. Ser. B 1, 421–442.

Auksztulewicz, R., Friston, K., 2015. Attentional enhancement of auditory mismatch responses: a DCM/MEG study. Cereb. Cortex bhu323.

Bassett, D.S., Zurn, P., Gold, J.I., 2018. On the nature and use of models in network neuroscience. Nat. Rev. Neurosci. 19, 566–578.

Berger, J., 2013. Statistical decision theory: foundations, concepts, and methods. Springer Science & Business Media.

Bissiri, P.G., Holmes, C.C., Walker, S.G., 2016. A general framework for updating belief distributions. J. R. Stat. Soc. Ser. B Stat. Methodol. 78, 1103–1130.

Braund, T., Palmer, D., Williams, L.M., Etkin, A., Harris, A., 2018. O43. Cognitive and Emotional Biomarkers of Anxious Major Depressive Disorder: An iSPOT-D Report. Biol. Psychiatry 83, S126.

Breakspear, M., Terry, J.R., Friston, K.J., Harris, A.W.F., Williams, L.M., Brown, K., Brennan, J., Gordon, E., 2003. A disturbance of nonlinear interdependence in scalp EEG of subjects with first episode schizophrenia. Neuroimage 20, 466–478.

Bush, P.C., Sejnowski, T.J., 1993. Reduced compartmental models of neocortical pyramidal cells. J. Neurosci. Methods 46, 159–166.

Chen, Y., Nakayama, K., Levy, D., Matthysse, S., Holzman, P., 2003. Processing of global, but not local, motion direction is deficient in schizophrenia. Schizophr. Res. 61, 215–227.

Cooray, G.K., Sengupta, B., Douglas, P.K., Friston, K., 2016. Dynamic causal modelling of electrographic seizure activity using Bayesian belief updating. Neuroimage 125, 1142–1154.

Dayan, P., Daw, N.D., 2008. Decision theory, reinforcement learning, and the brain. Cogn. Affect. Behav. Neurosci. 8, 429–453.

Deco, G., Jirsa, V.K., Robinson, P.A., Breakspear, M., Friston, K., 2008. The Dynamic Brain: From Spiking Neurons to Neural Masses and Cortical Fields. Plos Comput. Biol. 4.

Díez, Á., Ranlund, S., Pinotsis, D., Calafato, S., Shaikh, M., Hall, M.-H., Walshe, M., Nevado, Á., Friston, K.J., Adams, R.A., 2017. Abnormal frontoparietal synaptic gain mediating the P 300 in patients with psychotic disorder and their unaffected relatives. Hum. Brain Mapp. 38, 3262–3276.

Fischl, B., Sereno, M.I., Dale, A.M., 1999. Cortical surface-based analysis: II: inflation, flattening, and a surface-based coordinate system. Neuroimage 9, 195–207.

Friston, K., Penny, W., 2011. Post hoc Bayesian model selection. Neuroimage 56, 2089–2099.

Friston, K.J., Bastos, A.M., Pinotsis, D., Litvak, V., 2015a. LFP and oscillations—what do they tell us? Curr. Opin. Neurobiol. 31, 1–6.

Friston, K.J., Litvak, V., Oswal, A., Razi, A., Stephan, K.E., van Wijk, B.C., Ziegler, G., Zeidman, P., 2015b. Bayesian model reduction and empirical Bayes for group (DCM) studies. NeuroImage.

Jafarian, A., Litvak, V., Cagnan, H., Friston, K.J., Zeidman, P., 2019. Neurovascular coupling: insights from multi-modal dynamic causal modelling of fMRI and MEG. ArXiv Prepr. ArXiv190307478.

Jones, S.R., 2016. When brain rhythms aren't 'rhythmic': implication for their mechanisms and meaning. Curr. Opin. Neurobiol. 40, 72–80.

Jones, S.R., Pritchett, D.L., Stufflebeam, S.M., Hämäläinen, M., Moore, C.I., 2007. Neural correlates of tactile detection: a combined magnetoencephalography and biophysically based computational modeling study. J. Neurosci. 27, 10751–10764.

Jun, J.J., Steinmetz, N.A., Siegle, J.H., Denman, D.J., Bauza, M., Barbarits, B., Lee, A.K., Anastassiou, C.A., Andrei, A., Aydın, Ç., 2017. Fully integrated silicon probes for high-density recording of neural activity. Nature 551, 232–236.

Kopell, N., Ermentrout, G.B., Whittington, M.A., Traub, R.D., 2000. Gamma rhythms and beta rhythms have different synchronization properties. Proc Natl Acad Sci U A 97, 1867–72.

Mehler, D.M.A., Kording, K.P., 2018. The lure of causal statements: Rampant mis-inference of causality in estimated connectivity. ArXiv Prepr. ArXiv181203363.

Moran, R., Pinotsis, D.A., Friston, K., 2013. Neural masses and fields in dynamic causal modeling. Front. Comput. Neurosci. 7.

Moran, R.J., Stephan, K.E., Seidenbecher, T., Pape, H.C., Dolan, R.J., Friston, K.J., 2009. Dynamic causal models of steady-state responses. Neuroimage 44, 796–811.

Muthukumaraswamy, S.D., Edden, R.A.E., Jones, D.K., Swettenham, J.B., Singh, K.D., 2009. Resting GABA concentration predicts peak gamma frequency and fMRI amplitude in response to visual stimulation in humans. Proc. Natl. Acad. Sci. 106, 8356.

Pinotsis, D., Leite, M., Friston, K., 2013. On conductance-based neural field models. Front. Comput. Neurosci. 7, 158.

Pinotsis, D.A., Brunet, N., Bastos, A., Bosman, C.A., Litvak, V., Fries, P., Friston, K.J., 2014. Contrast gain control and horizontal interactions in V1: A DCM study. NeuroImage 92, 143–155.

Pinotsis, D.A., Buschman, T.J., Miller, E.K., 2018. Working Memory Load Modulates Neuronal Coupling. Cereb. Cortex.

Pinotsis, D.A., Geerts, J.P., Pinto, L., FitzGerald, T.H., Litvak, V., Auksztulewicz, R., Friston, K.J., 2017. Linking canonical microcircuits and neuronal activity: Dynamic causal modelling of laminar recordings. Neuroimage 146, 355–366.

Pinotsis, D.A., Loonis, R., Bastos, A.M., Miller, E.K., Friston, K.J., 2016a. Bayesian modelling of induced responses and neuronal rhythms. Brain Topogr. 1–14.

Pinotsis, D.A., Moran, R.J., Friston, K.J., 2012. Dynamic causal modeling with neural fields. Neuroimage 59, 1261–1274.

Pinotsis, D.A., Perry, G., Litvak, V., Singh, K.D., Friston, K.J., 2016b. Intersubject variability and induced gamma in the visual cortex: DCM with empirical B ayes and neural fields. Hum. Brain Mapp. 37, 4597–4614.

Pinotsis, Schwarzkopf, D.S., Litvak, V., Rees, G., Barnes, G., Friston, K.J., 2013. Dynamic causal modelling of lateral interactions in the visual cortex. NeuroImage 66, 563–576.

Pinto, L., Goard, M.J., Estandian, D., Xu, M., Kwan, A.C., Lee, S.-H., Harrison, T.C., Feng, G., Dan, Y., 2013. Fast modulation of visual perception by basal forebrain cholinergic neurons. Nat. Neurosci. 16, 1857–1863.

Poldrack, R.A., Monahan, J., Imrey, P.B., Reyna, V., Raichle, M.E., Faigman, D., Buckholtz, J.W., 2018. Predicting violent behavior: What can neuroscience add? Trends Cogn. Sci. 22, 111–123.

Preller, K.H., Razi, A., Zeidman, P., Stämpfli, P., Friston, K.J., Vollenweider, F.X., 2019. Effective connectivity changes in LSD-induced altered states of consciousness in humans. Proc. Natl. Acad. Sci. 116, 2743–2748.

Ranlund, S., Adams, R.A., Díez, Á., Constante, M., Dutt, A., Hall, M.-H., Maestro Carbayo, A., McDonald, C., Petrella, S., Schulze, K., 2016. Impaired prefrontal synaptic gain in people with psychosis and their relatives during the mismatch negativity. Hum. Brain Mapp. 37, 351–365.

Rutledge, R.B., Chekroud, A.M., Huys, Q.J., 2019. Machine learning and big data in psychiatry: toward clinical applications. Curr. Opin. Neurobiol. 55, 152–159.

Schwarzkopf, D.S., Robertson, D.J., Song, C., Barnes, G.R., Rees, G., 2012. The frequency of visually induced gamma-band oscillations depends on the size of early human visual cortex. J Neurosci 32, 1507–1512.

Sereno, M.I., Dale, A.M., Reppas, J.B., Kwong, K.K., Belliveau, J.W., Brady, T.J., Rosen, B.R., Tootell, R.B., 1995. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science 268, 889–893.

Trommershäuser, J., Maloney, L.T., Landy, M.S., 2008. Decision making, movement planning and statistical decision theory. Trends Cogn. Sci. 12, 291–297.

Verdinelli, I., Wasserman, L., 1995. Computing Bayes factors using a generalization of the Savage-Dickey density ratio. J. Am. Stat. Assoc. 90, 614–618.

Wei, H., Jafarian, A., Zeidman, P., Litvak, V., Razi, A., Hu, D., Friston, K.J., 2020. Bayesian fusion and multimodal DCM for EEG and fMRI. NeuroImage 211, 116595.

White, J.A., Banks, M.I., Pearce, R.A., Kopell, N.J., 2000. Networks of interneurons with fast and slow γ-aminobutyric acid type A (GABAA) kinetics provide substrate for mixed gamma-theta rhythm. Proc. Natl. Acad. Sci. 97, 8128–8133.

Williams, L.M., Rush, A.J., Koslow, S.H., Wisniewski, S.R., Cooper, N.J., Nemeroff, C.B., Schatzberg, A.F., Gordon, E., 2011. International Study to Predict Optimized Treatment for Depression (iSPOT-D), a randomized clinical trial: rationale and protocol. Trials 12, 4.