



# An electrophysiological validation of stochastic DCM for fMRI

J. Daunizeau<sup>1,2\*</sup>, L. Lemieux<sup>3</sup>, A. E. Vaudano<sup>4</sup>, K. J. Friston<sup>2</sup> and K. E. Stephan<sup>2,5,6</sup>

<sup>1</sup> Motivation, Brain and Behaviour Group, Brain and Spine Institute, Paris, France

<sup>2</sup> Wellcome Trust Centre for Neuroimaging, University College London, London, UK

<sup>3</sup> Institute of Neurology, University College London, London, UK

<sup>4</sup> Department of Biomedical science, Metabolism, and Neuroscience, University of Modena and Reggio Emilia, NOCSE Hospital, Modena, Italy

<sup>5</sup> Translational Neuromodeling Unit, Institute for Biomedical Engineering, University of Zurich and ETH Zurich, Zurich, Switzerland

<sup>6</sup> Laboratory for Social and Neural Systems Research, Department of Economics, University of Zurich, Zurich, Switzerland

## Edited by:

Viktor Jirsa, Movement Science  
Institute CNRS, France

## Reviewed by:

Ingo Bojak, University of  
Birmingham, UK

Petra Ritter, Charité University  
Medicine Berlin, Germany

## \*Correspondence:

J. Daunizeau, Motivation, Brain and  
Behaviour Group, Brain and Spine  
Institute, 47, bvd de l'Hoptial,  
75013 Paris, France.  
e-mail: jean.daunizeau@gmail.com

In this note, we assess the predictive validity of stochastic dynamic causal modeling (sDCM) of functional magnetic resonance imaging (fMRI) data, in terms of its ability to explain changes in the frequency spectrum of concurrently acquired electroencephalography (EEG) signal. We first revisit the heuristic model proposed in Kilner et al. (2005), which suggests that fMRI activation is associated with a frequency modulation of the EEG signal (rather than an amplitude modulation within frequency bands). We propose a quantitative derivation of the underlying idea, based upon a neural field formulation of cortical activity. In brief, dense lateral connections induce a separation of time scales, whereby fast (and high spatial frequency) modes are enslaved by slow (low spatial frequency) modes. This slaving effect is such that the frequency spectrum of fast modes (which dominate EEG signals) is controlled by the amplitude of slow modes (which dominate fMRI signals). We then use conjoint empirical EEG-fMRI data—acquired in epilepsy patients—to demonstrate the electrophysiological underpinning of neural fluctuations inferred from sDCM for fMRI.

**Keywords:** dynamic causal modeling, neural noise, EEG, fMRI, effective connectivity, neural field, separation of time scales

## INTRODUCTION

In addition to localizing brain regions that encode specific sensory, motor, or cognitive processes, contemporary neuroimaging research tries to understand how information is transmitted within brain networks (Sporns, 2010). Ultimately, the ambition is to understand the nature of the information that neuronal populations pass on to each other. This speaks to the notion of *functional integration*, which views cognitive function as resulting from information exchange within brain networks (Tononi et al., 1994). This means one has to understand the directed influence or *effective connectivity* that brain systems (e.g., cortical areas, neuronal populations, or single neurons) exert on each other. Analyzing effective connectivity rest on models that formalize assumptions about how neuronal systems are wired and how they respond to different stimuli. These models are then used to interpret brain responses measured using, e.g., functional magnetic resonance imaging (fMRI) or magneto-/electroencephalography (M/EEG). We refer to Valdes-Sosa et al. (2011) for a comprehensive review on effective connectivity, and its relationship with influence and causality.

In the context of M/EEG, detailed biophysical knowledge about the activity of neural ensembles has led the community to develop and validate dynamical models that can describe macro-scale dynamics in great detail (cf. Deco et al., 2008; Coombes, 2010; Bressloff, 2012). Most of these models are inspired by

approaches in statistical physics based on the notion of a *mean field*, i.e., the idea that interactions among ensembles of neurons can be captured by summary statistics (i.e., moments of the relevant distribution). Dynamic causal modeling [DCM, see Daunizeau et al. (2011) for a recent review] for M/EEG embeds these models into a formal (Bayesian) statistical framework that allows for parameter estimation and model comparison when analyzing evoked or induced responses. In this context, the approach has proven successful in exploiting the realism of these biophysical models to capture experimental effects of cognitive manipulations in terms of network plasticity (e.g., Garrido et al., 2007 or Moran et al., 2011).

Even more established is the use of DCM for fMRI data, for which the approach was originally proposed by Friston et al. (2003), building on previous advances in the biophysics of hemodynamic processes (Buxton et al., 1998). Interestingly, fMRI paradigms elicit neural responses that unfold on a much slower time scale (seconds) than that typical of electrophysiological measurements (milliseconds). In contrast to hemodynamic processes, relatively little is known about the biophysical underpinning of these neuronal responses, which has precluded the use of detailed (realistic) models of neural dynamics in DCM for fMRI. Instead, the slow (and widespread) dynamics of interacting brain regions are captured by phenomenological models that are much simpler than those employed for M/EEG data. Even though these models

serve as a reference when evaluating fMRI effective connectivity methods (Smith et al., 2011), their simplicity may impose limitations on the ensuing inferences (but see, for example, David et al., 2008 or Brodersen et al., 2011). This is why we have previously proposed variational Bayesian approaches that can deal with some of the (unavoidably simplifying) model assumptions by introducing random fluctuations of physiological states (Friston et al., 2008, 2010; Daunizeau et al., 2009b). Initial applications of this “stochastic” version of DCM [stochastic dynamic causal modeling (sDCM)] have established some face validity in the context of fMRI data analysis. For example, Friston et al. (2011) showed how sDCM can be used in combination with *post-hoc* model comparison (Friston and Penny, 2011) to explore large model spaces, and Li et al. (2011) examined the smoothness of the underlying neural fluctuations. In Daunizeau et al. (2012), we provide a comprehensive comparison of deterministic and stochastic DCM, in terms of parameter estimation and model comparison. In particular, we showed how accounting for random effects on the system’s dynamics can improve network identification by exploiting the decorrelation of neural time series induced by neural noise.

However, sDCM for fMRI has its own methodological challenges and still lacks solid experimental validation. On the one hand, the ability of sDCM for fMRI to discriminate between the respective contributions of neuronal noise and measurement noise in the observed fMRI signal depends on modeling assumptions (Roebroeck et al., 2011; Daunizeau et al., 2012). On the other hand, it is yet unclear how to formally relate the neural states of DCM for fMRI to electrophysiological measures of activity. In this note, we use EEG data to provide empirical evidence for the predictive validity of sDCM for fMRI data. We therefore place this work in the context of the ongoing debate regarding the hemodynamic correlate of the EEG [see Rosa et al. (2010a) for a recent critical review of this literature]. Here, we appeal to neural field theory (Wilson and Cowan, 1972, 1973; Lopes da Silva et al., 1973; Nunez, 1974; Amari, 1975, 1977; Freeman, 1975; Jirsa and Haken, 1996) to revisit the heuristic hemodynamic correlate of EEG signals originally proposed in Kilner et al. (2005). In essence, this hypothesis suggests that the fMRI signal is correlated with the frequency modulation of the EEG (where larger fMRI signals reflect a shift of the EEG spectrum toward higher frequencies). This is based on the idea that slow dynamics underpinning fMRI responses reflect the instantaneous frequency of oscillating modes of activity (Cabral et al., 2011; Friston et al., 2011). Here, we show how slow macroscopic modes of activity that emerge from intrinsic (within region) connections may modulate the frequency spectrum of fast dynamical modes that contribute to the EEG signal. Guided by this notion, we obtain estimates of neural dynamics from fMRI data using sDCM and ask whether these estimates predict the spectral behavior of concurrent EEG data—in terms of the temporal changes in its center frequency (above and beyond physiological confounds). This enables us to establish the predictive validity of sDCM for fMRI data, in relation to electrophysiological responses.

This paper comprises three sections. In the first, we present the neural field model that motivates a separation of time scales

and the ensuing heuristic hemodynamic correlate of the EEG. The second section is an empirical demonstration that serves to evaluate the predictive validity of stochastic DCM for fMRI for the ensuing electrophysiological correlates. We close with a discussion of the implications of our findings in the final section.

## MODEL: REVISITING THE HEURISTIC HEMODYNAMIC CORRELATE OF EEG

In this section, we first introduce a neural field model of a cortical region, and show how its spatiotemporal pattern of activity can be decomposed into canonical dynamic modes that have distinct time scales. In brief, slow macroscopic neural states control the rate of change of fast components. This motivates both the separation of time scales that underlies the generative model of sDCM for fMRI (Friston et al., 2011) and the heuristic hemodynamic correlate of EEG (Kilner et al., 2005). In brief, this section serves as a partial theoretical derivation of sDCM for fMRI, from the perspective of biologically-informed (neural field) models of macroscopic electrophysiological activity.

### NEURAL FIELDS AND SEPARATION OF TIME SCALES

We start with a description of the dynamics of single neurons within an ensemble, which we describe in terms of the probability density function  $p(\zeta)$  of neuronal post-synaptic membrane potentials (PSP)  $\zeta$ . We assume that neurons fire an action potential if their PSP surpasses a depolarization threshold  $\chi$ . This means that  $\langle H(\zeta - \chi) \rangle$  represents the average (mean field) firing rate over the ensemble, where  $H$  is the Heaviside step function and  $\langle \cdot \rangle$  is the expectation under the ensemble density. Using the central limit theorem, one can show that the average firing rate over the ensemble is determined mainly by the first two moments of the PSP ensemble density  $p(\zeta)$  (Liley et al., 2002; Marreiros et al., 2008):

$$\begin{aligned} \langle H(\zeta - \chi) \rangle &= \int_{\chi}^{\infty} p(\zeta) d\zeta \approx S(\eta) = \frac{1}{1 + \exp(-\rho(\eta - \chi))} \\ \eta &= \langle \zeta \rangle \\ \rho^{-2} &= \frac{3}{\pi^2} \langle (\zeta - \eta)^2 \rangle \end{aligned} \quad (1)$$

where  $S$  is a sigmoid function and  $\rho$  is proportional to the inverse of the standard deviation (dispersion) of depolarization within the ensemble. This is a statistical (mesoscopic) summary of neuronal ensemble activity, corresponding roughly to a cortical macrocolumn (about 100,000 neurons and 1 mm<sup>2</sup> of cortex). For our modeling purposes, this summary statistics is now taken to the macroscale, where each ensemble or column has a position  $\mathbf{r}$  on the 2D-cortical manifold. At this macroscopic scale, states like depolarization can be regarded as a continuum or *field* (Jirsa and Haken, 1996; Liley et al., 2002; Coombes et al., 2007; Daunizeau et al., 2009a; Pinotsis et al., 2012), which is a function of space and time:  $\eta(t) \rightarrow \eta(\mathbf{r}, t)$ . This means we can describe the local average input and output activity of neurons using the depolarization field  $\eta(\mathbf{r}, t)$  and the sigmoid function. The spatiotemporal dynamics of the neural field essentially depends

on how local ensembles influence each other—as described below.

We now consider a region of the cortical manifold that is defined by its continuous domain  $D$ , over which the neural field  $\eta(\mathbf{r}, t)$  is defined. The field is considered to result from the temporal convolution of pre-synaptic input  $\mu(\mathbf{r}, t)$  as follows:

$$\begin{aligned}\eta(\mathbf{r}, t) &= \int_{-\infty}^{\infty} J(t-t')\mu(\mathbf{r}, t)dt' \\ &= J \otimes \mu(\mathbf{r}, t) \\ J(t) &= \begin{cases} \gamma \frac{t}{\tau} \exp\left(-\frac{t}{\tau}\right) & t \geq 0 \\ 0 & t < 0 \end{cases},\end{aligned}\quad (2)$$

where  $J$  is the synaptic impulse response function,  $\tau$  is a time-constant, and  $\gamma$  controls the maximum PSP following an action potential. The maximum PSP and time-constant summarize the microscopic properties of the ensemble (Jansen and Rit, 1995). Because action potentials propagate with a finite conduction velocity, the mean rate of arrival of pre-synaptic impulses can be expressed as a time-delayed integral (spatiotemporal convolution) of local firing rates:

$$\begin{aligned}\mu(\mathbf{r}, t) &= \int_{-\infty}^{\infty} \int_D G(|\mathbf{r}-\mathbf{r}'|, t-t')S(\eta(\mathbf{r}', t'))d\mathbf{r}'dt' \\ &= G \otimes S \circ \eta(\mathbf{r}, t),\end{aligned}\quad (3)$$

where  $D$  is the domain, over which the neural field is defined, and  $G$  is a homogeneous Green function that acts as a spatiotemporal convolution operator. Let  $r = |\mathbf{r}-\mathbf{r}'|$  be the distance between two points with positions  $\mathbf{r}$  and  $\mathbf{r}'$ . Below, we follow Bojak and Liley

(2010) and parameterize the Green function as follows:

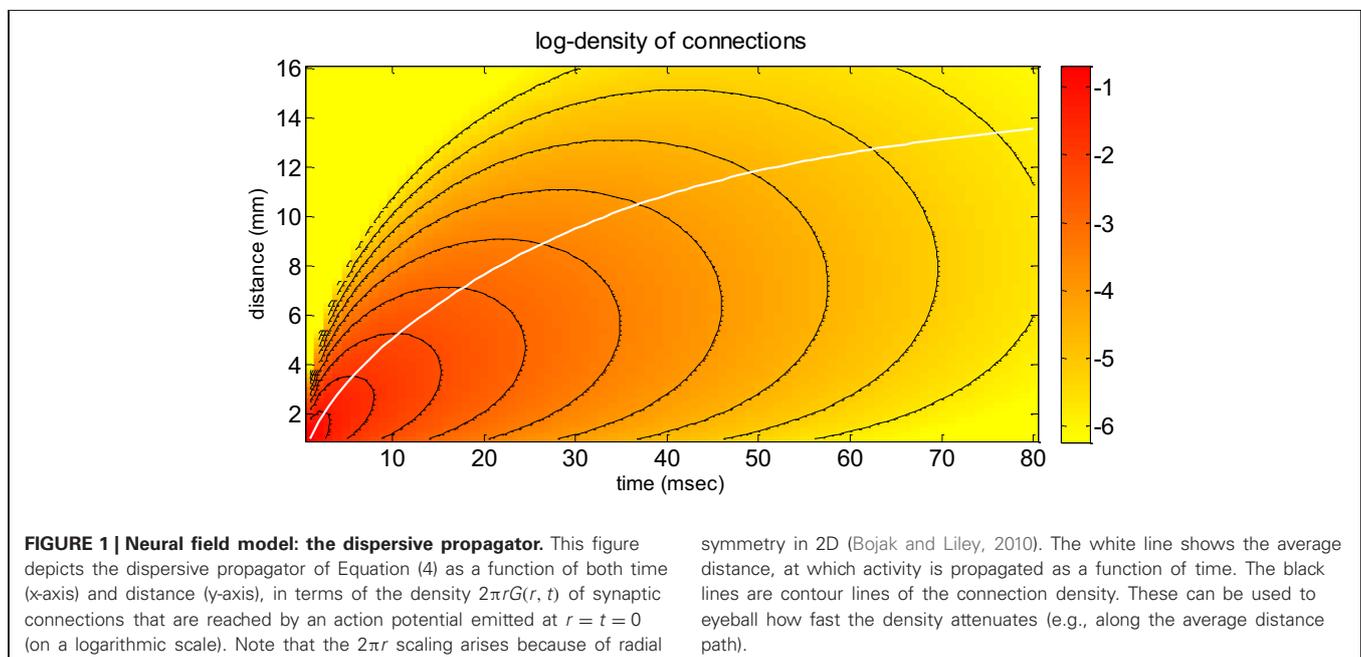
$$G(r, t) = \frac{G^0}{4\pi\sigma^2} t^{-1} \exp\left(-\frac{r^2 + \bar{v}^2 t^2}{2\sigma\bar{v}t}\right) H(t), \quad (4)$$

where the Heaviside step function endows Equation (3) with temporal causality and  $G^0$  is the total number of synaptic connections. This *dispersive propagator* assumes a Gaussian falloff in the density of synaptic connections and accounts for a certain variability in velocities ( $\bar{v}$  is the mode of the velocity distribution at “large distance,” i.e., when  $r/\sigma \rightarrow \infty$ ). The spatial scale of lateral connectivity is controlled by the spatial dispersion  $\sigma$ . **Figure 1** depicts the dependence of the dispersive propagator with respect to space and time (see also Liley et al., 2012—**Figure 3**).

In addition to its plausible form, this homogeneous Green function has proven very useful, in that it has an analytical Fourier transform. Applying the Fourier transform back and forth to Equations (2) and (3) yields the following partial differential equation (see Bojak and Liley, 2010):

$$\begin{aligned}\left(\frac{\bar{v}}{2\sigma} + \frac{\partial}{\partial t} - \frac{\bar{v}\sigma}{2}\nabla^2\right)\mu(\mathbf{r}, t) &= \frac{G^0\bar{v}}{2\sigma}S(\eta(\mathbf{r}, t)) \\ \left(\tau^2\frac{\partial^2}{\partial t^2} + 2\tau\frac{\partial}{\partial t} + 1\right)\eta(\mathbf{r}, t) &= \tau\gamma\mu(\mathbf{r}, t).\end{aligned}\quad (5)$$

The first partial differential equation describes the diffusion of action potentials of spikes (through the Laplacian operator) and derives from the lateral connectivity kernel in the Green function. The second ordinary differential equation derives from Equation (2) and describes synaptic transmission; i.e., how propagated pre-synaptic firing is accumulated locally to produce depolarization.



The solution to Equation (5) can be decomposed into canonical dynamical modes, by projecting the neural field on the set of eigenvectors  $w^{(k)}(\mathbf{r})$  of the Laplacian operator defined over the domain  $D$  (see Daunizeau et al., 2009a). These spatial modes solve the Laplacian eigenvalue problem, i.e.,  $\nabla^2 w^{(k)}(\mathbf{r}) = \lambda_k w^{(k)}(\mathbf{r})$ , where the distribution of eigenvalues  $\lambda$  is such that:  $\lambda_k < \lambda_0 = 0$ . They form a complete and orthonormal basis function set for the neural field, allowing for an eigen-decomposition of the field, as follows:

$$\begin{aligned}\eta(\mathbf{r}, t) &= \sum_k z_1^{(k)}(t) w^{(k)}(\mathbf{r}) \\ \mu(\mathbf{r}, t) &= \sum_k z_2^{(k)}(t) w^{(k)}(\mathbf{r}),\end{aligned}\quad (6)$$

here, the dynamics of the spatial modes are defined as:

$$\begin{aligned}z_1^{(k)}(t) &= \int_D w^{(k)}(\mathbf{r}) \eta(\mathbf{r}, t) d\mathbf{r} \\ z_2^{(k)}(t) &= \int_D w^{(k)}(\mathbf{r}) \mu(\mathbf{r}, t) d\mathbf{r}.\end{aligned}\quad (7)$$

Furthermore, it can be shown that the dynamics of the spatial modes (Equation 5) can be approximated as third-order ordinary differential equations that are coupled through the fundamental mode  $k = 0$  (see Appendix 1):

$$\begin{aligned}\dot{\mathbf{z}}^{(k)} &= \begin{bmatrix} \dot{z}_1^{(k)} \\ \dot{z}_2^{(k)} \\ \dot{z}_3^{(k)} \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 & 1 \\ c(z_1^{(0)}) \frac{\bar{v}}{2\sigma} (\sigma^2 \lambda_k - 1) & 0 & 0 \\ -\frac{1}{\tau^2} & \frac{\gamma}{\tau} & -\frac{2}{\tau} \end{bmatrix} \mathbf{z}^{(k)} + O(z^2) \\ c(z_1^{(0)}) &\equiv \frac{\rho G^0 \bar{v}}{2\sigma} S(z_1^{(0)}) \left(1 - S(z_1^{(0)})\right)\end{aligned}\quad (8)$$

here,  $z_3 \equiv \dot{z}_1$ , and  $c(z_1^{(0)})$  is a non-linear function of the fundamental mode  $z_1^{(0)}$  that mediates the coupling between modes describing firing rate and depolarization. Roughly speaking, each mode decays at a rate that is proportional to  $\sigma \bar{v} \lambda_k$ . As a consequence, high propagation velocities in large cortical regions dissipate the spatial modes quickly, with the exception of the fundamental mode, which has a zero eigenvalue. More generally, smooth eigenmodes (low spatial scales) will be associated with slow dynamics (see, e.g., Schultze-Kraft et al., 2011). Note that the natural time scale of the fundamental mode can be very slow, if the conduction velocity  $\bar{v}$  becomes small compared to the dispersion of lateral connections  $\sigma$  (i.e., if  $\bar{v}/\sigma \rightarrow 0$ ). In this case, the fundamental mode  $z^{(0)}$  will decay so slowly that it will predominate over other stable (higher spatial frequency) modes  $z^{(k)}$ , which disappear as soon as they are created. This means that Equation (8) effectively instantiates a separation of time scales, where the slowest mode “enslaves” the fast modes (Haken, 1983)

through the coupling function  $c(z_1^{(0)})$ . This slaving effect is such that the frequency profile of fast modes ( $k > 0$ ) is controlled by the amplitude of slow modes ( $k = 0$ ).

### MODELING LOCAL ACTIVATION IN TERMS OF EEG FREQUENCY MODULATION

Equation (8) expresses the Jacobian of eigenmode dynamics as a function of the fundamental mode of activity  $z_1^{(0)}$  within a given brain region, over which the neural field is defined. This means that slow fluctuations in the fundamental mode  $z_1^{(0)}$  should be associated with changes in the EEG frequency spectrum  $P(\omega)$ , which derives from the Laplace transform of Equation (8) (see Appendix 2):

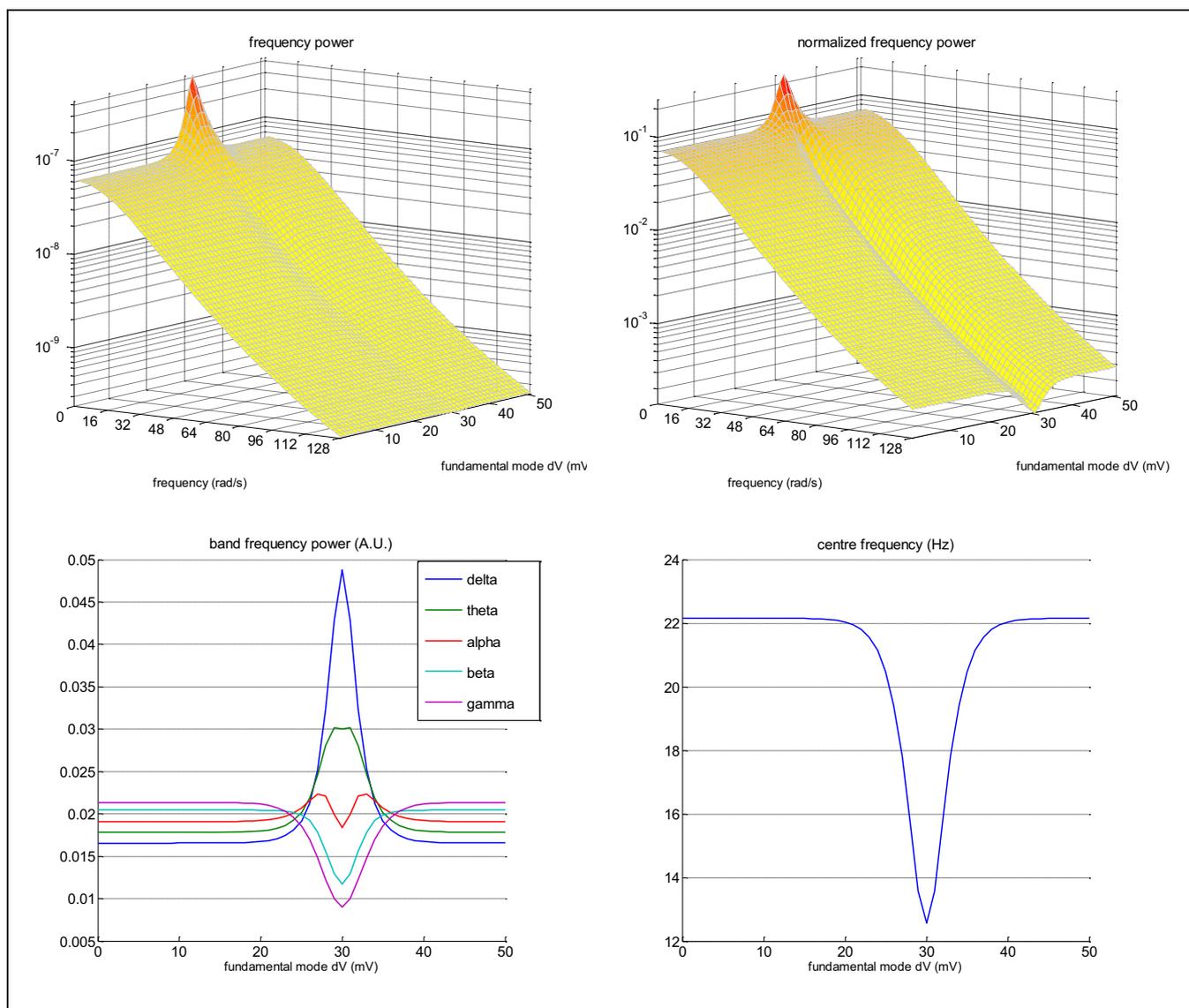
$$P(\omega) \propto \left| \sum_{k \leq k_c} \frac{\left(i\omega + \frac{\bar{v}}{2\sigma} (1 - \sigma^2 \lambda_k)\right) \left(i\omega + \frac{2}{\tau} + 1\right) + \frac{\gamma}{\tau}}{\left(i\omega \left(i\omega + \frac{2}{\tau}\right) + \frac{1}{\tau^2}\right) - \frac{\gamma}{\tau} c(z_1^{(0)})} \right|^2 \quad (9)$$

where  $k_c$  is the cut-off order induced by the skull's low electrical conductivity. Here, the summation is over the fast modes (up to  $k_c = 16$ ), and we have assumed that EEG responses reflect post-synaptic potentials, i.e.,  $z_1^{(k)}$ . Similarly to Kilner et al. (2005), let us now define the EEG center frequency  $\bar{\omega}$  as the first moment of the (normalized) frequency power spectrum  $\bar{P}(\omega)$ :

$$\begin{aligned}\bar{P}(\omega) &= \frac{P(\omega)}{\int_{\omega_0}^{\infty} P(\omega) d\omega} \\ \bar{\omega} &\equiv \int_{\omega_0}^{\infty} \omega \bar{P}(\omega) d\omega\end{aligned}\quad (10)$$

By construction, normalized frequency power  $\bar{P}(\omega)$  is the proportion of total power attributed to any given frequency of interest. Thus, one can think of it in terms of a probability density function over frequencies, whose first-order moment would be the center frequency. Intuitively, an increase in the center frequency  $\bar{\omega}$  signals a shift of frequency power toward higher frequencies. Note that the center frequency is not the main peak in the frequency spectrum. Instead, variations of  $\bar{\omega}$  over time quantify frequency modulations of the EEG signal. The key idea here is that the EEG center frequency  $\bar{\omega}$  is a function of the slow eigenmode dynamics  $z_1^{(0)}$ , and its modulation over time thus follows the slow eigenmode dynamics  $z_1^{(0)}$ , i.e.,  $\bar{\omega} = \bar{\omega}(z_1^{(0)})$ .

**Figure 2** depicts the frequency modulation of the EEG signal induced by changes in the slow eigenmode  $z_1^{(0)}$ . The frequency power spectrum  $P(\omega)$  given in Equation (9) was evaluated using the parameter values given in **Table 1** below. One can see that when the slow eigenmode tends toward the action-potential firing threshold ( $z_1^{(0)} \rightarrow \chi$ ), there is both a power increase in the low frequencies and a decrease in the high frequencies (cf. increase in the delta/theta/alpha band, and decrease in the beta/gamma



**FIGURE 2 | Neural field model: EEG frequency modulation.** This figure shows the effect of changes in the mean membrane depolarization fundamental eigenmode  $z_1^{(0)}$  on the frequency content of the EEG signal. **Upper-left:** EEG frequency spectrum  $P(\omega)$  (z-axis) as a function of frequency  $\omega$  (x-axis) and eigenmode amplitude  $z_1^{(0)}$  (y-axis). **Upper-right:** Normalized frequency spectrum  $\bar{P}(\omega)$  (cf. Equation 10,

same format). **Lower-left:** Magnitude of standard EEG frequency bands (y-axis; blue: delta, green: theta, red: alpha, magenta: beta, violet: gamma) as a function of the eigenmode amplitude  $z_1^{(0)}$  (x-axis).

**Lower-right:** Centre frequency  $\bar{\omega}$  (y-axis) as a function of the eigenmode amplitude  $z_1^{(0)}$  (x-axis). Note that here, the action potential firing threshold  $\chi$  is 30 mV.

band). This shift toward lower frequencies can be seen most clearly on the normalized frequency power  $\bar{P}(\omega)$ . As a consequence, the center frequency  $\bar{\omega}$  increases as the slow eigenmode gets further away from the action potential firing threshold. In our example, this induces a change of about 10 Hz in the EEG center frequency. Note that the precise numerical value of the center frequency  $\bar{\omega}$  (as well as its susceptibility to the eigenmode  $z_1^{(0)}$ ) depends upon the neural field's parameters (cf. Table 1). In fact, using more realistic neural fields models (accounting for, e.g., excitatory and inhibitory interneurons, see Pinotsis et al., 2012) would profoundly change the shape of the frequency power spectrum. However, the qualitative effect of neural

activation upon the center frequency is qualitatively invariant to such changes. We will comment on this in the “Discussion” section.

#### THE LINK BETWEEN EEG TIME-FREQUENCY RESPONSE AND fMRI TEMPORAL RESPONSE

In this paper, we assume that the slow macroscopic dynamics  $x^{(n)}$  underlying fMRI data in a given region of interest follow the fundamental eigenmode  $z_1^{(0)}$  of its membrane depolarization cortical field. The motivations for this assumption include the following:

- Its relative simplicity.

**Table 1 | Neural field parameters.**

Constant	Physical meaning	Value
$\gamma$	Maximum post-synaptic impulse response	8 mV
$\tau$	Post-synaptic impulse response decay rate	4 ms
$G^0$	Total number of synaptic connections	2000
$\sigma$	Spatial decay rate of synaptic connections	3 mm
$\rho$	Inverse of the within-ensemble PSP standard deviation	0.54 mV <sup>-1</sup>
$\chi$	Action potential firing threshold (w.r.t. PSP steady-state)	30 mV
$\bar{v}$	Average conduction velocity	3 m/s

*These parameter values were taken from Pinotsis et al. (2012).*

- Sensitivity to input versus output spiking activity: there is a consensus now on the fact that mean membrane depolarization (as opposed to e.g., firing rate) drives hemodynamic changes [see Logothetis and Wandell (2004) for a review on the relationship between local field potentials and the BOLD signal].
- Energy budget: both the slow time scale and the low spatial frequency of the fundamental eigenmode make it likely to be the main driver of induced BOLD changes, when compared with faster (and higher spatial frequency) eigenmodes. This is because blood flow increases in local vascular trees are induced by coherent spatio-temporal summation of neural activity (see e.g., Daunizeau et al., 2010; Rosa et al., 2010a,b; Schmuell, 2010).
- It relates to theoretical accounts on critical slowing of high-dimensional dynamical systems, which has recently been used to motivate the rate of change of neural states in stochastic DCM for fMRI (Friston et al., 2010).

We will further comment on this point in the “Discussion” section.

In the previous section, we have described how a separation of time scales could arise from local interactions at the level of a cortical region. It turns out that this result can be generalized to a set of coupled local neural fields, in that the fundamental mode mediates the extrinsic (between-region) coupling. In other words, the influence different neural fields exert on each other expresses itself through the coupling of their respective fundamental modes, which locally drive higher order modes (as in Equation 8). Deriving the precise dynamical properties of coupled neural fields is beyond the scope of the present study. However, one can invoke the slaving effect when considering coupled (distal) brain regions, the network properties of which shape the temporal dynamics of local fundamental eigenmodes  $\mathbf{x}_1^{(0)}$ . This means that long-range connections control the frequency profile of brain regions responses (see e.g., Bojak et al., 2011), through their effect on their respective fundamental eigenmodes  $\mathbf{x}_1^{(0)}$ . Since we did not consider coupled neural fields in the above treatment, we replace their (unknown) evolution function with a first-order Taylor expansion on slow neural states, yielding the

following Langevin equation for stochastic DCM (see Daunizeau et al., 2012):

$$\dot{\mathbf{x}}^{(n)} \approx \mathbf{A}\mathbf{x}^{(n)} + \mathbf{v} \quad (11)$$

where  $\boldsymbol{\theta}^{(n)} = \{\mathbf{A}_{ij}\}$  are neuronal parameters that measure extrinsic coupling strengths. Here, we have neglected non-linear interaction terms that act as gating factors on the network connections  $\mathbf{x}^{(n)}$  (cf. Stephan et al., 2008). The influence of the fast modes on the motion of slow modes  $\mathbf{x}^{(n)}$  is expressed via fluctuations  $\mathbf{v}$ , which are comparatively much faster. This allows us, in the context of sDCM for fMRI, to treat  $\mathbf{v}$  as stochastic (neural) noise and place priors  $p(\mathbf{v}|m)$  on its generalized motion (see below).

In contradistinction, typical evoked EEG responses are driven by a mixture of slow and fast modes. Note that the relative contribution of the slow and fast modes is imbalanced (toward fast modes), simply because the summation span all eigenmodes up to cut-off order  $k_c$ . In fact, typical evoked EEG responses disappear within a second or so of peristimulus time; i.e., their limit frequency has an order of magnitude similar to the fMRI sampling rate (about 1 Hz).

Taken together, we expect to see concurrent changes in the slow neural states  $\mathbf{x}^{(n)}$  that drive BOLD and in the EEG power spectrum  $P(\omega)$ . However, from the above section, we know that the underlying relationship is non-linear and depends upon intrinsic properties of local neural fields, such as the size of active brain regions and the spatial decay rate of synaptic connections (cf. Equation 9). In addition, the respective contribution of each region to the scalp EEG power spectrum is virtually unknown, as this would require solving the so-called EEG inverse problem (see Rosa et al., 2010a,b). These issues conspire with the unavoidable simplifications of our model to make the prediction of the full EEG frequency spectrum practically irrelevant. In the following, we will thus focus on the EEG center frequency  $\bar{\omega}$ , which captures global changes in the frequency spectrum. For the sake of simplicity, in the remainder of this work, we will invoke a first-order Taylor expansion of the following form:

$$\bar{\omega}(\mathbf{x}^{(n)}) = \bar{\omega}(0) + \left. \frac{\partial \bar{\omega}}{\partial \mathbf{x}^{(n)}} \right|_0 \mathbf{x}^{(n)} + O(x^2) \quad (12)$$

where the gradient  $\left. \frac{\partial \bar{\omega}}{\partial \mathbf{x}^{(n)}} \right|_0$  measures the susceptibility of the center frequency to the slow neural states. This quantity will be directly estimated from experimental data (along with the intercept  $\bar{\omega}(0)$ , which acts as a confound). Testing for Equation (12) effectively bypasses the dependency of the EEG power spectrum to the respective contribution each local neural field included in the network, as well as their biophysical properties, and only retains the qualitative prediction of the model.

### Summary

In this section, we have used the neural field formulation of local neural activity to demonstrate an emergent separation of temporal scales, in which the slow fluctuations of modes or patterns of depolarization enslave faster modes. Crucially, the amplitude of slow modes (which we associate with fMRI signals) and modulate

the frequencies of fast modes (which we associate with EEG signals). This dynamical behavior is based upon a plausible mean field approximation to macroscopic neuronal activity on the cortical surface and, within the setting, provides a formal verification of the heuristic in Kilner et al. (2005), linking fMRI signals to changes in the center frequency of electrophysiological signals. In the next section, we will use stochastic DCM to obtain estimates of macroscopic (slow) neural fluctuations from fMRI data. We then test whether these estimates can predict fluctuations in concurrently measured EEG data—in terms of its center frequency, as suggested by the neural field treatment above (Equations 9–12).

## EMPIRICAL DATA: RESTING STATE AND GENERALIZED SPIKE AND WAVE (GSW) ACTIVITY

In this section, we conduct an sDCM analysis of data from patients with epilepsy who exhibited generalized spike and wave (GSW) activity, while undergoing concurrent EEG/fMRI measurements in the scanner. We first describe the EEG/fMRI acquisition and conventional activation analysis. We then describe the sDCM analysis of fMRI data and ensuing validation using EEG data.

### EEG/fMRI ACQUISITION AND ACTIVATION ANALYSIS

These data were part of a previous neuroimaging study of idiopathic generalized epilepsy (Hamandi et al., 2006; Vaudano et al., 2009), whose acquisition protocol and data pre-processing are briefly summarized here.

Ten-channel EEG (10–20 system) was recorded using MR-compatible equipment, along with bipolar electrocardiogram. After filtering and artifact correction (Krakow et al., 2000), the onsets and offsets of GSW activity were identified by two experts (see Vaudano et al., 2009 for details). Seven hundred T2\*-weighted single-shot gradient echo echo-planar images (TE = 40 ms, TR = 3 s, 21 interleaved axial slices of 5 mm thickness, FOV = 24 × 24 cm<sup>2</sup>, 64 × 64 matrix) were acquired over a 35 min session with a 1.5 Tesla MRI scanner (Horizon EchoSpeed, General Electric). Patients were asked to rest with their eyes closed and to keep still. fMRI data were pre-processed using SPM8 (<http://www.fil.ion.ucl.ac.uk/spm/>). EPI time series were realigned and spatially smoothed with an 8 mm FWHM isotropic Gaussian kernel and spatially normalized to the standard anatomical space. For each patient, a general linear model (GLM) was constructed to test for the presence of regional GSW-related BOLD changes. Periods of GSW activity were modeled as blocks, beginning at GSW onset and terminating at their offset. The GSW regressors were then convolved with the canonical hemodynamic response function (plus temporal and dispersion derivatives) before inclusion in the GLM. In addition, we included slow drifts (Fourier basis function set), as well as motion-related effects (head and eye movements), cardiac-related effects (see Liston et al., 2006) and scan-nulling regressors (modeling inter-scan motion events larger than 0.2 mm, cf. Lemieux et al., 2007) as confounding factors.

On average, the GLM design matrices contained about 100 confounds regressors. Significant positive and negative GSW-related BOLD responses were identified by means of an *F*-contrast

on the GSW regressors for the nine patients included in this study. SPMs were thresholded at  $p < 0.05$  (FWE whole-brain corrected) to define three regions of interest (ROIs), which were implicated in the initiation and termination of GSW discharges in all patients: thalamus, prefrontal cortex (PFC), and precuneus. A summary time series was derived for each ROI by computing the first eigenvariate of all suprathreshold voxel time series within a 10 mm of the ROI centers. The time series were adjusted for all confounding effects included in the GLM analysis.

## VALIDATING THE sDCM MACROSCOPIC NEURAL STATES ESTIMATES WITH EEG: METHODS

In addition to the neural evolution function given in Equation (11), DCM for fMRI requires the specification of an additional set of hemodynamic states that couple neural dynamics to observed BOLD signal changes:

$$\dot{\mathbf{x}}^{(h)} = \begin{bmatrix} x^{(n)} - \kappa_s x_1^{(h)} - \kappa_f (e^{x_2^{(h)}} - 1) \\ x_1^{(h)} e^{-x_2^{(h)}} \\ \frac{1}{\tau_0} (e^{x_2^{(h)}} - e^{x_3^{(h)}/\alpha}) e^{-x_3^{(h)}} \\ \frac{1}{\tau_0} \left( \frac{e^{x_2^{(h)}} - x_4^{(h)} (1 - (1 - E_0) e^{-x_2^{(h)}})}{E_0} - e^{(1-\alpha)x_3^{(h)}/\alpha} \right) \end{bmatrix} \quad (13)$$

here,  $x^{(n)}$  represents regional neural activity, whose dynamics is given in Equation (11). Equation (13) expresses changes in hemodynamic states  $\mathbf{x}^{(h)}$ , as a response to a neural perturbation  $x^{(n)}$  and as a function of hemodynamic parameters  $\boldsymbol{\theta}^{(h)} = \{\kappa_s, \kappa_f, \alpha, \tau_0, E_0, \varepsilon_0\}$  (for details, see the appendix in Stephan et al., 2008). Finally, one has to specify the observation mapping from hemodynamic states  $\mathbf{x}^{(h)}$  to observed local BOLD changes  $y$  (Stephan et al., 2007):

$$y = V_0 \left( 4.3 v_0 E_0 TE \left( 1 - e^{x_4^{(h)}} \right) + \varepsilon_0 r_0 E_0 TE \left( 1 - e^{x_4^{(h)} - x_3^{(h)}} \right) + (1 - \varepsilon_0) \left( 1 - e^{x_3^{(h)}} \right) \right) + \varepsilon, \quad (14)$$

where  $\varepsilon$  is an additive measurement noise, and the parameters are defined in **Table 2** below.

In brief, the predicted data  $\mathbf{y}$  depend non-linearly on the unknown model variables  $\boldsymbol{\Theta} = \{\mathbf{x}, \boldsymbol{\theta}\}$ , through Equations (11), (13), and (14). This model (as well as statistical assumptions about measurement noise  $\boldsymbol{\varepsilon}$ ) is encoded in the likelihood function  $p(\mathbf{y} | \boldsymbol{\Theta}, m)$ . Priors  $p(\boldsymbol{\Theta} | m)$  specify our assumptions about the magnitude of state noise, evolution parameters and observation parameters—where the prior means of the hemodynamic parameters of Equations (13) and (14) are given in **Table 2**:

Inverting the generative model  $m$  means (1) approximating the conditional density  $p(\boldsymbol{\Theta} | \mathbf{y}, m)$  of unknown variables  $\boldsymbol{\Theta}$  given the set of sampled measurements  $\mathbf{y}$  and (2) quantifying the model evidence  $p(\mathbf{y} | m)$ . Non-linearities in the generative model eschew exact analytical solutions to this problem, which is finessed using variational approaches that rest on optimizing a

**Table 2 | Prior means of hemodynamic parameters and acquisition-related constants for fMRI at 1.5 T.**

Constant	Physical meaning	Value
$\kappa_s$	Vasodilatory signal decay rate	0.65 Hz
$\kappa_f$	Vasodilatory signal feedback rate	0.41 Hz
$\tau_0$	Mean transit time	2 s
$\alpha$	Vessel stiffness	0.32
$E_0$	Oxygen extraction fraction at rest	0.34
$V_0$	Venous volume fraction at rest	4
$\nu_0$	Frequency offset	40.3 Hz
$TE$	EPI echo time	0.04 s
$r_0$	Intravascular relaxation rate	25 Hz
$\varepsilon_0$	Ratio of intra- and extra-vascular signals	1

Note that  $\theta^{(h)} = \{\kappa_s, \kappa_f, \alpha, \tau_0, E_0, \varepsilon_0\}$  are estimated from the data (i.e., have a positive prior variance), whereas the remaining parameters are treated as constants.

free-energy lower bound  $F(q)$  to the model evidence, with respect to an approximate conditional density  $q(\Theta)$ :

$$F(q) = \langle \ln p(\Theta | m) + \ln p(\mathbf{y} | \Theta, m) - \ln q(\Theta) \rangle_q \quad (15)$$

$$= \ln p(\mathbf{y} | m) - D_{\text{KL}}(q(\Theta); p(\Theta | \mathbf{y}, m)),$$

where  $D_{\text{KL}}$  is the Kullback-Leibler divergence and the expectation  $\langle \cdot \rangle_q$  is taken under  $q$ . From Equation (15), maximizing the functional  $F(q)$  with respect to  $q$  minimizes the Kullback-Leibler divergence between  $q(\Theta)$  and the exact posterior  $p(\Theta | \mathbf{y}, m)$ . This decomposition is complete in the sense that if  $q(\Theta) = p(\Theta | \mathbf{y}, m)$ , then  $F(q) = \ln p(\mathbf{y} | m)$ . Typically, the iterative maximization of free energy proceeds under the Laplace approximation, where the approximate posterior  $q(\Theta) \approx p(\Theta | \mathbf{y}, m)$  is assumed to have a Gaussian form (see Friston et al., 2007). We refer to Daunizeau et al. (2009b) for details about the application of the variational Bayesian approach to stochastic DCM.

Following model inversion as described above, inference on hidden states is based on the conditional density  $q(\mathbf{x})$  and thus depends upon the generative model  $m$ . Usually, in DCM, inference on states or parameters is preceded by model selection. However, our focus was on the validity of sDCM for fMRI with regard to predicting the spectral behavior of concurrently acquired EEG data—based on the idea that slow macroscopic neural states  $\mathbf{x}^{(n)}$  control the EEG frequency modulation. We therefore endowed the network with full reciprocal connectivity (no zero entry in matrix  $\mathbf{A}$  of Equation 11) and simply applied the above variational Bayesian approach to derive the first moment  $\hat{\mathbf{x}} = \langle \mathbf{x}^{(n)} \rangle_q$  of the conditional density  $q(\mathbf{x})$ . In the following,  $\hat{\mathbf{x}}$  thus refers to the sDCM estimate of the slow neural states, given fMRI time series  $\mathbf{y}$  (and under model  $m$ ).

To assess the relationship between EEG frequency modulation and neural states as estimated by sDCM, we constructed a GLM  $H_1$ , whose dependent variable was the trajectory  $\bar{\omega}$  of the EEG center frequency sampled at each fMRI time sample and whose independent variables were the neural states estimated by sDCM

plus confounds (cf. Equation 12):

$$H_1 : \bar{\omega} = [\hat{\mathbf{x}} \ \mathbf{X}_0] \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\beta}_0 \end{bmatrix} + \mathbf{e} \quad (16)$$

where  $\hat{\mathbf{x}}$  are the sDCM conditional estimates,  $\mathbf{X}_0$  contains the confounds (see below),  $\boldsymbol{\beta}$  and  $\boldsymbol{\beta}_0$  are unknown regression coefficients, and  $\mathbf{e}$  are i.i.d. residuals (with zero mean and unknown variance). Parameters  $\boldsymbol{\beta}$  effectively capture the susceptibility of the EEG center frequency with respect to the slow neural states (cf. gradient  $\partial \bar{\omega} / \partial \mathbf{x}^{(n)}|_0$  in Equation 12). They lump together the contributions of the multiple regions that constitute the network as well as biophysical properties of the local neural fields. Parameters  $\boldsymbol{\beta}_0$  weight the contribution of the confounding factors to changes in the EEG center frequency. We then tested whether the sDCM neural states predicted EEG frequency modulation above and beyond what could be explained using the confounds. To this end, we used Bayesian model comparison at the group level (Stephan et al., 2009) to quantify the evidence in favor of the full model  $H_1$ , relative to a reduced model  $H_0$  that only included the confounds:

$$H_0 : \bar{\omega} = \mathbf{X}_0 \boldsymbol{\beta}_0 + \mathbf{e} \quad (17)$$

The assessment of sDCM predictive validity thus depends on the definition of the confounds, i.e., uncontrolled sources of variations in the EEG center frequency. In order to investigate the robustness of the statistical relationship between the slow eigenmode dynamics  $\hat{\mathbf{x}}$  and the EEG center frequency trajectory  $\bar{\omega}$ , we have defined two different sets of confounds:

- *Slow drifts*: this assumes that the EEG center frequency undergoes slow and unspecific variations due to e.g., skin conductance changes, residual eye blinks artifacts, etc. ... Under this assumption, we set up  $\mathbf{X}_0$  similarly to the minimal confounds of a typical fMRI activation analysis, i.e., a Fourier basis function set (16 first harmonics; see Figure 6).
- *Full confounds*: this generalizes the intuition that confounds on the EEG center frequency are similar to those included in the fMRI activation analysis. For example, cardiac and respiratory effects, as well as body movements, could be modeled directly, as potential sources of (nuisance) variations. Under this assumption, we set up  $\mathbf{X}_0$  identically to the full confounding matrix used for each subject fMRI activation analysis (cf. section “EEG/fMRI Acquisition and Activation Analysis” above).

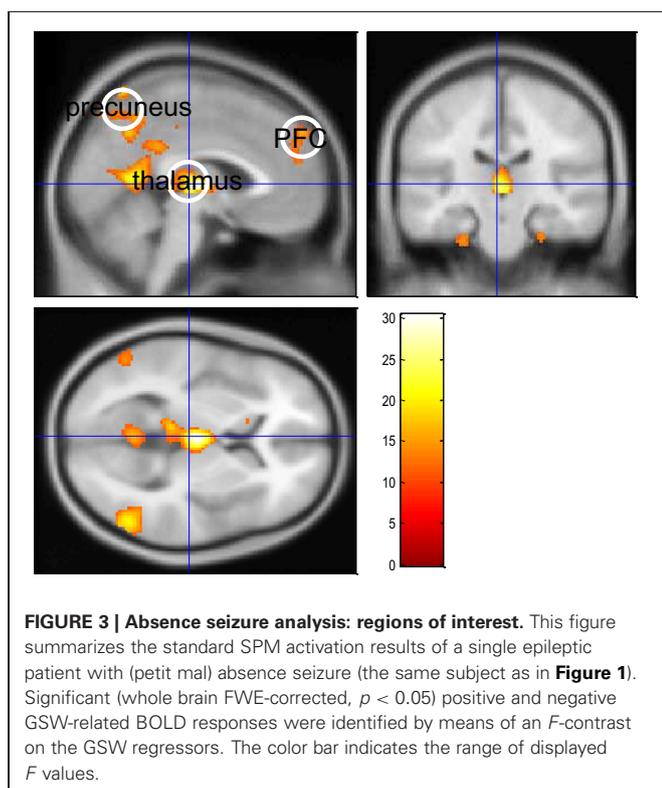
In brief, testing for the contribution of sDCM neural estimates  $\hat{\mathbf{x}}$  on variations of the center frequency  $\bar{\omega}$ , above and beyond the full confounds is the most conservative of the two alternatives. However, it will be the least sensitive of both models as well; recall that  $\boldsymbol{\beta}$  has only  $n_{\text{ROI}}$  entries, where  $n_{\text{ROI}} = 3$  is the number of ROIs included in the sDCM analysis (to be compared with the dimension of  $\boldsymbol{\beta}_0$ ). In other words, these two sets of confounds can be motivated on principled grounds. We will thus perform a statistical comparison between  $H_0$  and  $H_1$  that marginalizes over the two sets of confounds.

### VALIDATING NEURAL STATES ESTIMATE OBTAINED BY sDCM AGAINST EEG

**Figure 3** shows the SPM of significant GSW-related BOLD changes (above and beyond physiological confounds), for the first patient of the study ( $F$ -test,  $p < 0.05$ , whole-brain corrected). This was used to define the three ROIs included in the sDCM analysis (**Figure 4**).

One can see that most of the variance of the BOLD response is explained by the stochastic DCM. Focusing on peri-GSW periods, one can see that the conditional density  $q(\mathbf{x})$  shows high temporal variability, with increased uncertainty during head motion. This is expected, as scan-nulling confound regressors effectively prevent sDCM from deriving any information from fMRI data about hidden states during these periods. It is interesting to note that the network connectivity coefficients (elements of the  $\mathbf{A}$  matrix in Equation 11) are small (about 0.02 Hz). This means that the average coupling of these regions (over the entire recording) is rather weak. Importantly, with the (partial) exception of the inhibitory intrinsic self-connections, the posterior correlation matrix (**Figure 4**) shows that most system parameters are identifiable and not confounded by hemodynamic parameter estimation. Generally, **Figure 4** summarizes the motivation for this work—is the temporal variability of sDCM estimates of neuronal activity an artifact of measurement noise, or does it have an electrophysiological underpinning?

The cross-validation of neural state estimates by sDCM from fMRI data against EEG data requires the extraction of the time-dependent center frequency  $\bar{\omega}$  for each channel of each subject, according to Equation (10). This is depicted in



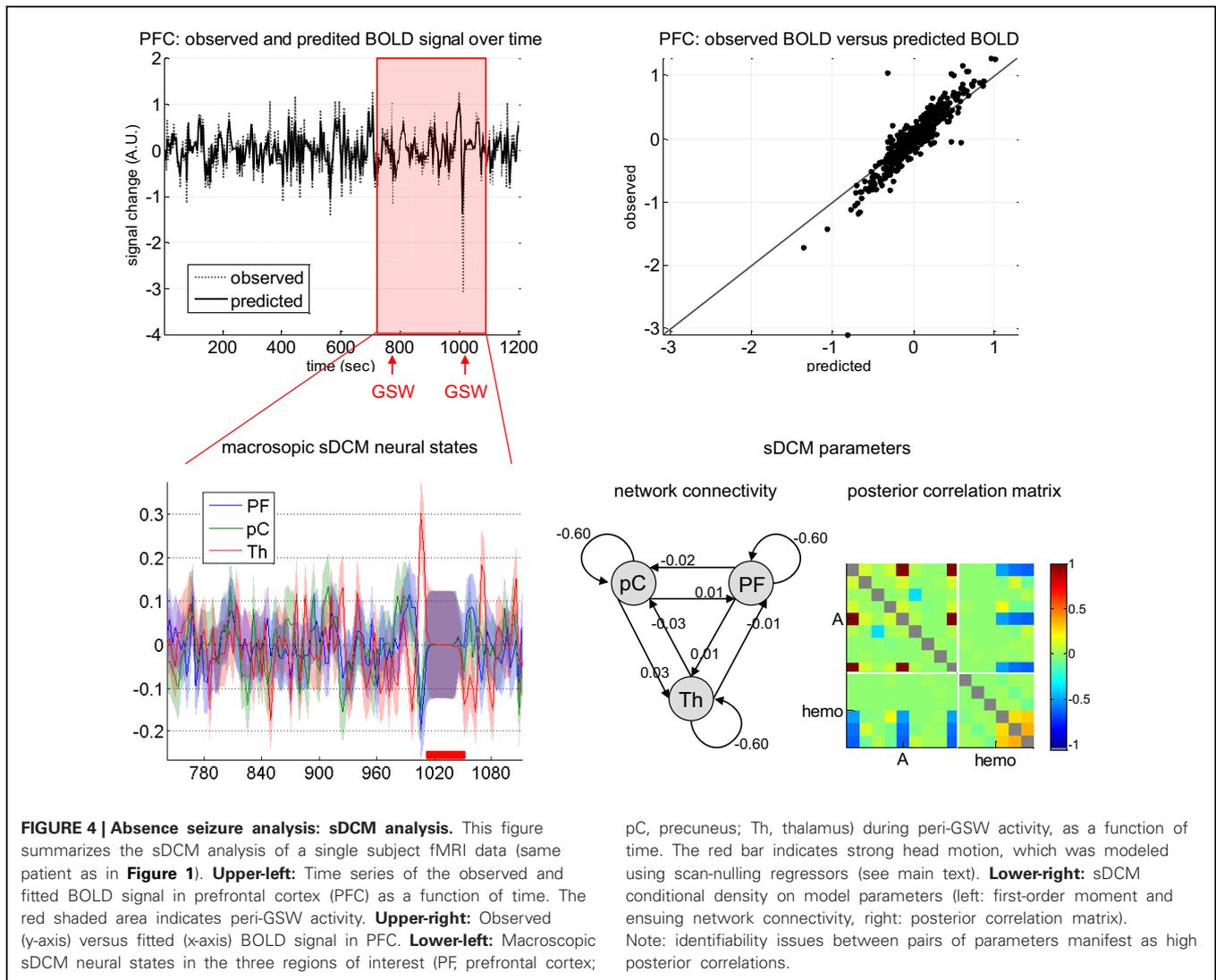
**Figure 5**, for the EEG channel C4 of the same patient in **Figures 3, 4**.

Recall that comparisons between the experimental and theoretical profiles of the frequency spectrum show differences that prevent meaningful quantitative fitting of Equation (9). For example, the model cannot account for peaks in the alpha band that appear in occipital channels (at least for the neural field parameter values of **Table 1**; results not shown). However, it is reassuring to see that the variations of the EEG frequency profile over time appear as small perturbations around a largely invariant average power spectrum. More precisely, the appearance of GSW activity does not dramatically distort its shape, but rather induces frequency modulations of small amplitude. This is important, since this affords face validity to the qualitative prediction of our neural field model. Interestingly, there is a clear change in the normalized frequency power beginning at the onset of the second GSW crisis. This leads to a marked decrease in the EEG center frequency (the first GSW event seems to be associated with a similar but weaker effect). We will briefly comment on GSW-related frequency modulations in the discussion.

**Figure 6** summarizes the relationship between slow macroscopic sDCM states and the EEG frequency modulation of channel C4 for the same patients as in previous figures. Even though one can see that the center frequency  $\bar{\omega}$  is very well fitted by the model (upper row of **Figure 6**: coefficient of determination:  $R^2 = 0.59$ ), one might think that most of the frequency modulation could be explained by confounds. For example, eyeballing the trajectory of the observed EEG center frequency reveals clear low frequency trends that are likely to be captured by slow drifts confounds. To disclose the specific contribution of neural state estimates, we have plotted the data and model prediction after removing the confounds (adjusted data). One can see that, when considering slow drifts confounds, sDCM neuronal estimates explain EEG frequency modulations of about 10 Hz. This is a similar order of magnitude than center frequency variations captured by slow drifts (cf. parameter estimates on lower-right panel of **Figure 6**). Note that the corresponding correlation coefficient is about 0.49 (adjusted  $R^2 = 0.24$ ), which indicates a reasonably good fit (above and beyond confounds).

**Figure 7** summarizes the result of Bayesian model comparison, in terms of the log-Bayes factor  $\text{LBF} = \log p(\bar{\omega} | H_1) - \log p(\bar{\omega} | H_0)$ , across all EEG channels for the same patient as before. One can see that all EEG channels show strong evidence in favor of the full model ( $H_1$ ), whereas there is evidence against  $H_1$  only for EOG and ECG channels. This suggests that the neuronal fluctuations estimated by stochastic DCM predict the EEG frequency modulation beyond slow drifts confounds. For completeness, we also show the results of classical inference ( $F$ -test with significance level  $p = 0.05$ , full model  $H_1$  against the null  $H_0$ ). As this test was repeated independently for each of the 30 channels, we applied a Bonferroni correction for multiple comparisons across channels. One can see that the profile of  $F$ -statistics and log-Bayes factors (LBFs) across channels is very similar.

Recall that the results summarized in **Figures 6, 7** are only shown to illustrate the analysis. In fact, it is important to note



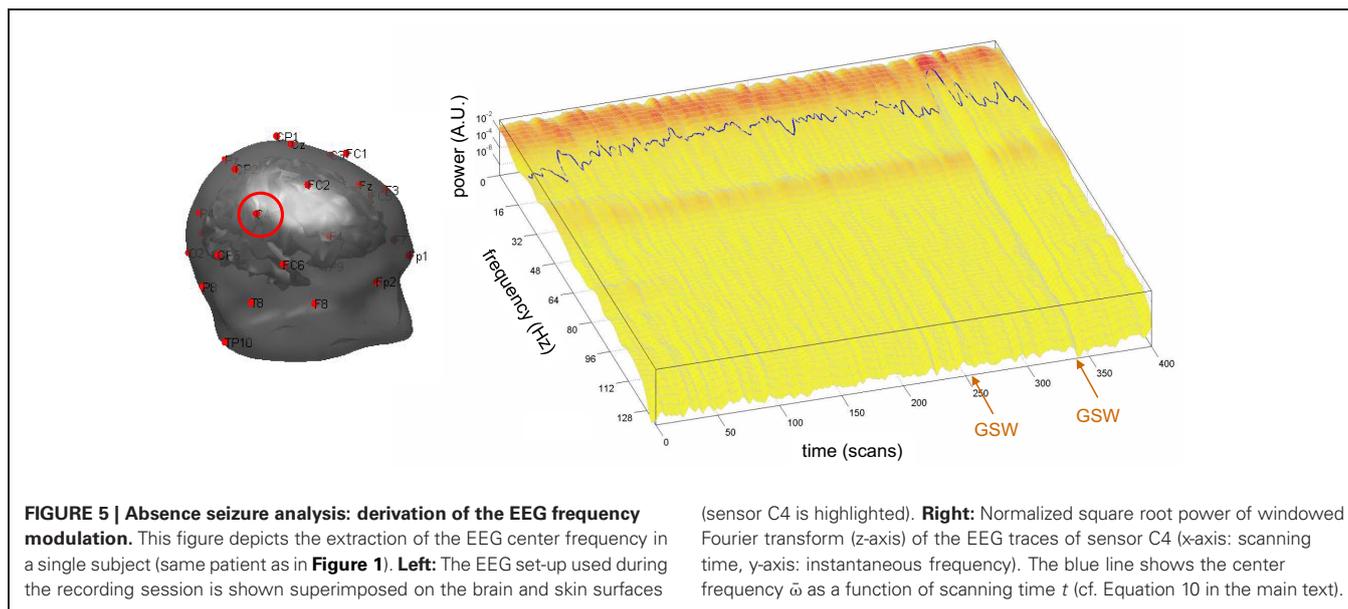
that they (1) depend upon the definition of the confounds (here slow drifts) and (2) do not capture inter-individual variability. The question of whether the accuracy with which the neural states inferred by sDCM are able to predict the frequency modulations of the EEG signal (above and beyond confounds) generalizes at the group level is addressed below.

This Bayesian model comparison for the single patient shown in **Figure 7** was then repeated for all subjects, for both sets of confounds (namely: slow drifts and full confounds). First, we pooled evidence over EEG channels for each subjects. This was simply done by summing the log- model evidences over EEG channels (within-subjects fixed effect. The ensuing distribution of LBFs across subjects—shown in **Table 3** below—indicates that most subjects show very strong evidence in favor of  $H_1$ , irrespective of the particular set of confounds.

To confirm this, we ran a random-effect analysis (Stephan et al., 2009) to assess the prevalence of  $H_1$  and  $H_0$  at the group level. Note that the two sets of confounds (“slow drifts” and “full confounds”) induce the four following models:

- $H_1^{(\text{drifts})}$  :  $H_1$  with slow drifts
- $H_1^{(\text{full})}$  :  $H_1$  with full confounds
- $H_0^{(\text{drifts})}$  :  $H_0$  with slow drifts
- $H_0^{(\text{full})}$  :  $H_0$  with full confounds.

The comparison of interest is in terms of whether or not adding the sDCM slow neural states adds anything to the prediction of the EEG center frequency, above and beyond confounds. Thus, we applied group-level family inference (Penny et al., 2010) to evaluate the evidence in favor or against  $H_1$ , irrespective of the definition of the confounds. We have thus partitioned the set of four models above into two families  $\bar{H}_1$  and  $\bar{H}_0$  that gather both sorts of confounds, as follows:  $\bar{H}_1 = \{H_1^{(\text{drifts})}, H_1^{(\text{full})}\}$  and  $\bar{H}_0 = \{H_0^{(\text{drifts})}, H_0^{(\text{full})}\}$ . Deriving the relative evidence of  $\bar{H}_1$  against  $\bar{H}_0$  effectively marginalizes over the different confounds. First of all, we checked that there was a random effect at all, i.e.,: a difference



between the respective prevalence of all models within the population. We found that the posterior probability of all models being equally frequent was  $p = 0.03$ . This Bayesian omnibus test indicates that the distribution of Bayes factors shown in **Table 3** is unlikely to be due to chance. **Figure 8** summarizes the group-level family inference results. One can see that the estimated prevalence of family  $\bar{H}_1$  is about 0.82. In addition, the exceedance probability of  $\bar{H}_1$  (versus  $\bar{H}_0$ ) was 0.991. In other words, we can be 99% confident that sDCM estimates of neuronal fluctuations predict concurrent EEG frequency modulation above and beyond physiological confounds (in more than half the population).

For completeness, we have also performed bayesian group-level model comparison, conditional on the definition of confounds. The results match those of the family-inference analysis above. In brief, the exceedance probability of  $H_1^{(\text{drifts})}$  (versus  $H_0^{(\text{drifts})}$ ) was 0.997, and the exceedance probability of  $H_1^{(\text{full})}$  (versus  $H_0^{(\text{full})}$ ) was 0.990. However, the bayesian omnibus test for these analyses was not as conclusive (slow drifts:  $p = 0.11$ , full confounds:  $p = 0.14$ ). Taken together, this means that there is statistical evidence in favor of  $H_1$  (versus  $H_0$ ), irrespective of the definition of the confounds. This concludes our quantitative assessment of the predictive validity of sDCM for fMRI with regard to concurrently obtained EEG measures.

## DISCUSSION

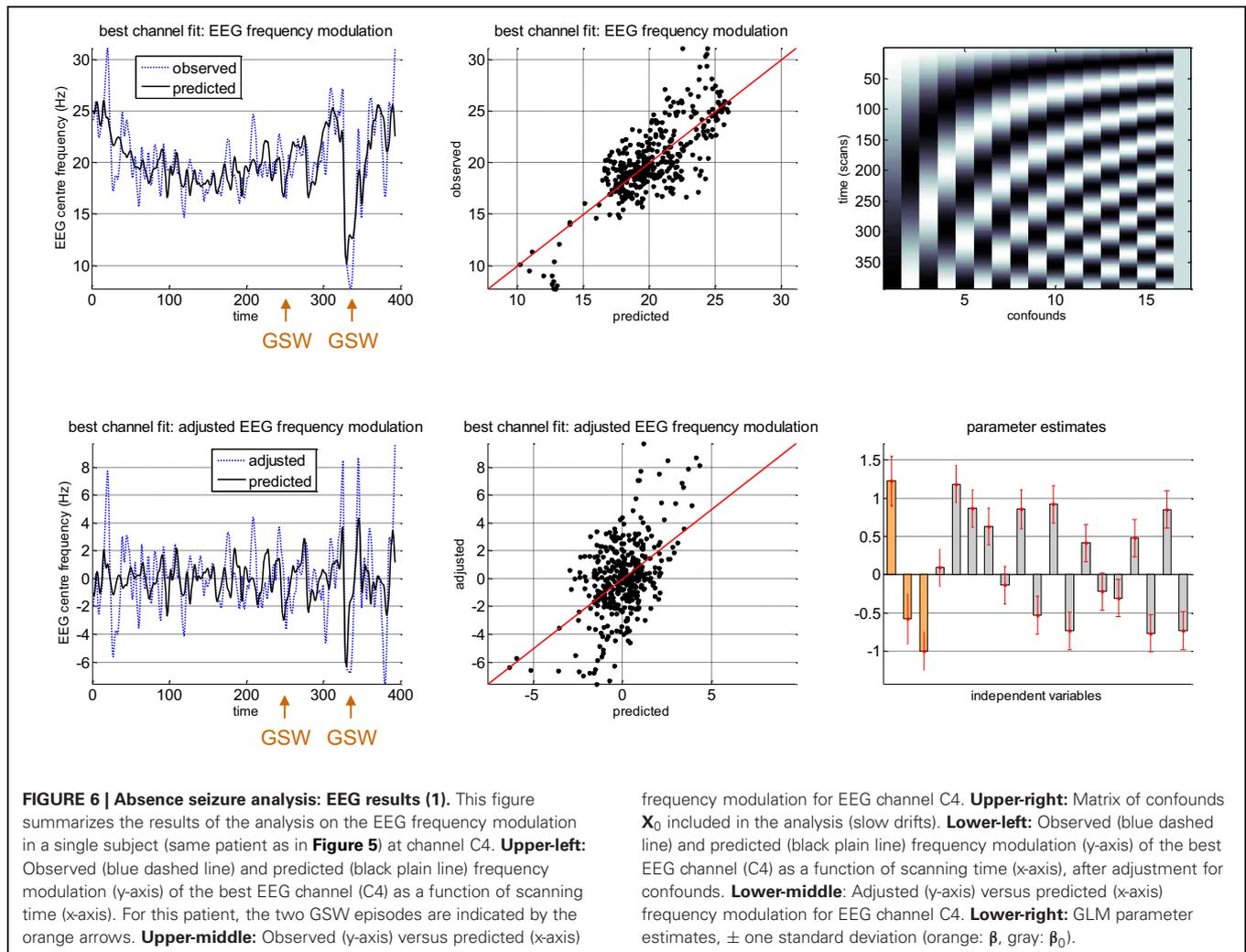
Introducing random neural fluctuations in DCM for fMRI was originally motivated by the need to estimate hidden states in the absence of experimental perturbations. In this note, we have assessed the predictive validity of sDCM estimates of slow macroscopic neural states with respect to EEG frequency modulation. We have revisited the heuristic hemodynamic correlate of the EEG proposed in Kilner et al. (2005), using a neural field formulation to show how slow modes of local electrophysiological activity

(which we assume drive changes in the fMRI signal) enslave the frequency spectrum of fast modes that contribute to the EEG signal. From this theoretical perspective, we tested whether slow neural estimates obtained by sDCM could explain concurrent EEG frequency modulation above and beyond physiological confounds.

To our knowledge, this work is the first attempt to provide empirical evidence that neural fluctuations inferred using stochastic DCM from fMRI time series have an electrophysiological underpinning. From an experimental perspective, this endorses the use of sDCM for studying spontaneous fluctuations that are beyond experimental control. Examples of this include—but are not limited to—multistable perception (Hesslmann et al., 2010) and epileptic events (e.g., absence seizures, cf. Vaudano et al., 2009). In this context, the added-value of sDCM is that it allows for comparing quantitative hypotheses about the propagation and/or the emergence of these phenomena within brain networks.

The acute reader might question the relevance of the modeling part of our paper, in the sense that quantitative predictions of the model were not used in the subsequent data analysis. However, we believe this model may contribute to the debate regarding the hemodynamic correlate of the EEG. First, it bears a number of interesting differences, with respect to the heuristic model of Kilner et al. (2005):

- Here, the main prediction, namely that fMRI activation is associated with a frequency modulation of the EEG signal, only relies on the assumption that BOLD changes are mostly driven by the fundamental (slow and smooth) eigenmode of the membrane potential field. This differential sensitivity of EEG and fMRI with respect to higher-order canonical eigenmodes bypasses most modeling assumptions of the heuristic model.

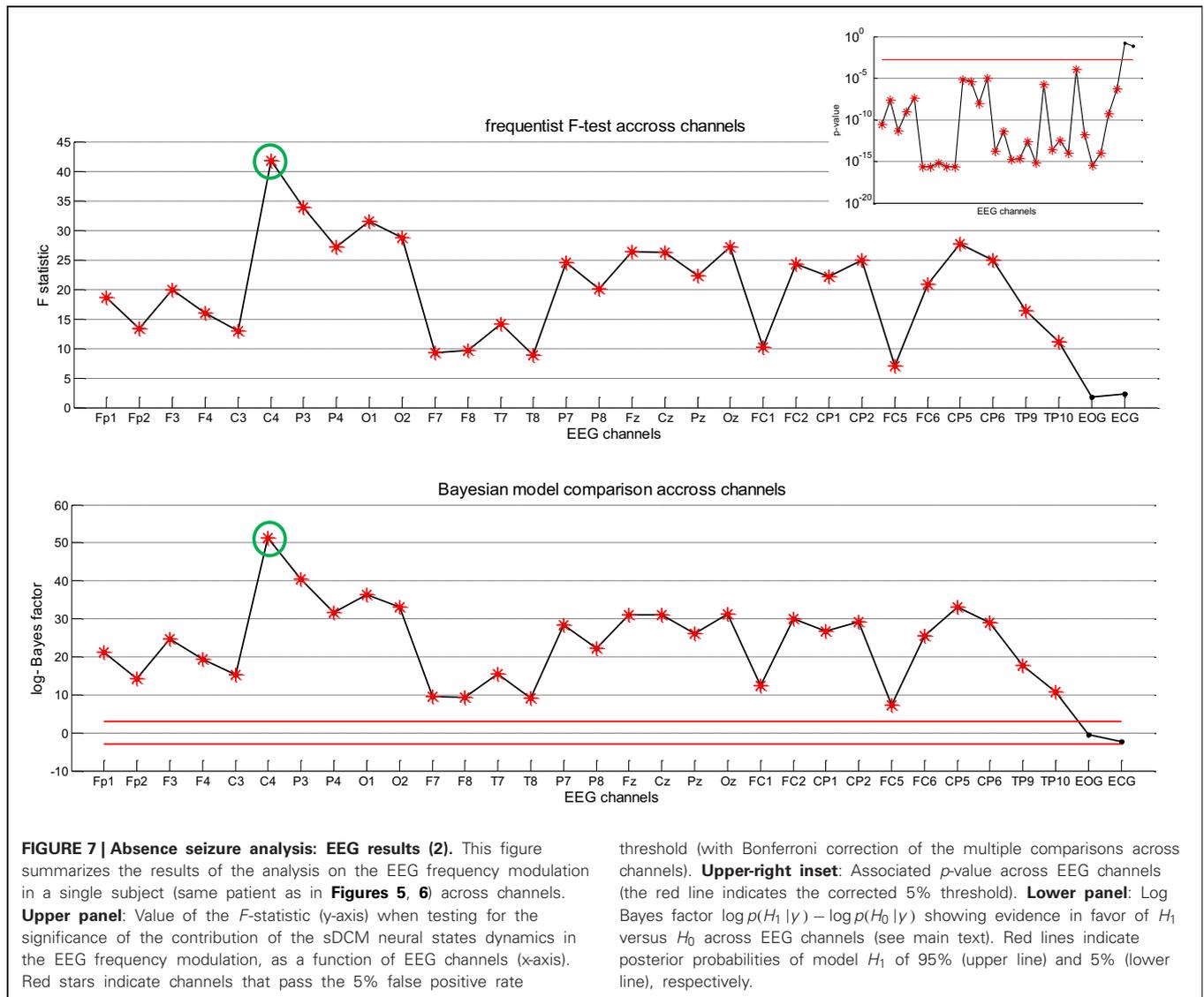


- Here, the relationship between activation and the EEG center frequency is not monotonic. More precisely, the center frequency decreases as one approaches the action potential threshold, either from hyperpolarized states (activation) or from depolarized states (de-activation). In brief, this form of critical slowing is qualitatively equivalent to the heuristic model, above the action potential threshold.

Second, we believe that the above qualitative theoretical prediction is very robust to model assumptions. First, we checked that the effect persisted in the context of our model, when changing the parameters that control the biophysical properties of the neural field (results not shown). Second, the effect is a byproduct of local propagation of activity over the field. In brief, smooth local patterns emerge in densely connected brain regions, the—slow—dynamics of which are shaped by effective connectivity between remote regions. Each local eigenmode is then expected to control the acceleration or slowing down of the response of its respective brain region. In contradistinction to e.g., quantitative predictions about frequency power spectra, we anticipate

this qualitative prediction to hold, irrespective of models' specificities.

Let us now consider the (modest) theoretical contribution of this work in relation to the separation of time scales. We believe that—if properly extended—established biophysical models of fast electrophysiological responses may be a good starting point for understanding slow dynamical modes at macroscopic spatial scales that, presumably, drive fMRI responses. The neural field treatment we have developed in this note is a first step toward a mechanistic understanding of the macroscopic dynamics that emerge at the time scale of seconds or minutes. However, this model is limited in many ways. First, we did not account for extrinsic connectivity between regions (coupled neural fields). This is known to strongly impact on the stability of spatiotemporal brain dynamics (Knock et al., 2009), and will thus be critical for predicting the dynamical repertoire of slow macroscopic modes of activity (cf. point above). For example, Bojak et al. (2011) show that increasing the long-range coupling between remote regions strongly modulates the frequency response of the target region, in terms of the relative power in low- and high- frequency bands. Second, we did not separate the



**Table 3 | Log-Bayes factors (LBF) of  $H_1$  versus  $H_0$ , for all subjects, and both sets of confounds.**

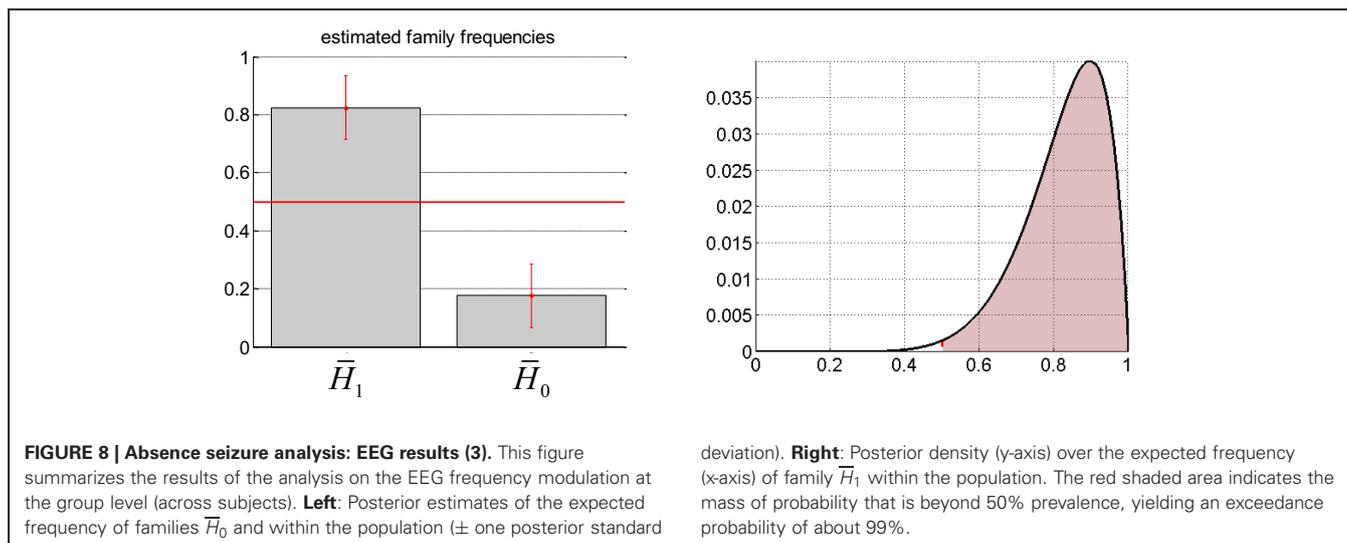
Subject	LBF ( $H_1$ against slow drifts)	LBF ( $H_1$ against full confounds)
1	670.9	198.6
2	204.3	103.9
3	51.4	-4.8
4	92.9	23.3
5	17.9	185.4
6	4.4	35.3
7	-1.8	65.0
8	1.2	28.6
9	101.4	37.2

A positive LBF indicates evidence in favor of  $H_1$  ( $|LBF| > 3$ : significant effect at the subject-level, Kass and Raftery, 1995). Highlighted cells depict the two subjects that show relative evidence in favor of  $H_0$ .

threshold (with Bonferroni correction of the multiple comparisons across channels). **Upper-right inset:** Associated *p*-value across EEG channels (the red line indicates the corrected 5% threshold). **Lower panel:** Log Bayes factor  $\log p(H_1 | y) - \log p(H_0 | y)$  showing evidence in favor of  $H_1$  versus  $H_0$  across EEG channels (see main text). Red lines indicate posterior probabilities of model  $H_1$  of 95% (upper line) and 5% (lower line), respectively.

respective contributions of excitatory and inhibitory subpopulations (Pinotsis et al., 2012). Of particular importance here is the fact that inhibitory-excitatory networks can possess stable limit cycles (Seung et al., 1995). This means that local neural fields can behave as oscillators or resonators. This relates to the work of Cabral et al. (2011) who examined the role of local network oscillations in the emergence of slow temporal synchrony between remote brain areas. The critical thing here is that the time it takes for one resonator to influence another resonator can be much slower than the local resonance frequencies. In addition, owing to the autonomous nature of local dynamics, the effect of between-regions coupling might be better understood in terms of changes in the resonance frequency of the target region. Taken together, these considerations suggest that challenging theoretical work still needs to be undertaken to establish the dynamical repertoire of coupled slow macroscopic modes of brain activity.

We take the fact that the sDCM neural states contribution was significant for all EEG channels (above and beyond slow



drifts) as a compelling validation of the conditional estimates of neuronal activity from sDCM. Importantly, this also provides some empirical evidence in favor of the fMRI to EEG mapping itself. This supports the results of Rosa et al. (2010b), which is, to our knowledge, the only experimental study that tested predictions of the heuristic model of Kilner et al. (2005). In this study, the authors experimentally controlled the frequency content of the EEG signal, and showed that the BOLD response in a group subjects was better explained by frequency modulation than by amplitude modulation of the EEG. More specifically, Rosa et al. (2010b) showed that to correctly model BOLD variations it was important to use a normalized frequency spectrum, as in Equation (10). Here, we approach this issue from the other side and use fMRI to predict the EEG frequency spectrum. Note that this is only an indirect validation of the neurovascular coupling model, as we did not compare it with other quantitative scenarios. Among these is the idea that BOLD changes correlate with amplitude modulations within frequency bands of interest in the EEG spectrum. Typically, the BOLD signal is considered a filtered version of the EEG alpha power (see e.g., De Munck et al., 2008). More generally, this idea has been applied to each and every EEG frequency band with relative success [see Laufs et al. (2008) and references therein]. As an illustration, Lachaux et al. (2007) have found that EEG gamma band modulations co-localize with BOLD variations. However, although power fluctuations of different EEG frequency bands are known to be mutually highly correlated, multi frequency models are usually not used in such studies (De Munck et al., 2009). In addition to compromising the specificity of the results, this prevents a direct comparison with our and related work (cf. Rosa et al., 2010b).

Another limitation of our data analysis is that we neglected the unavoidable non-linearity implicit in the relationship  $\bar{\omega} = \bar{\omega}(z_1^{(0)})$  (cf. Equations 9–12 and Figure 2) when testing for the contribution of slow neural states to modulations of the EEG center frequency. This means we may have not taken full advantage of our neural field formulation. However, this does not detract from

our quantitative assessment of the predictive validity of sDCM for fMRI data.

When examining the GSW-related EEG frequency modulations, we noted that the frequency spectrum did not change markedly during GSW activity. This might seem surprising, if one thinks of GSW as sudden bursts of activity within the network, associated with clearly recognizable 3 Hz transient waves of EEG activity (Destexhe and Sejnowski, 2001). Closer inspection of the EEG frequency profile confirmed the appearance of a peak at 3 Hz during GSW-activity (not shown). However, many other (faster) processes actually contribute to the overall frequency spectrum, which typically drags the center frequency toward 20 Hz. This is the main reason why we had to include as many confounding factors as possible, when assessing the specific contribution of the sDCM predictors. Importantly, physiological confounds (cardiac activity, head movements, etc.) span most of the slow frequencies in the EEG signal, which potentially masks interesting slow neural processes. This means that slow frequency EEG signals with a neural origin are confounded with physiological noise. Note that there are notorious exceptions to this detectability issue, for example readiness potentials (Jahanshahi and Hallett, 2003) that are thought to reflect the preparation of motor activity and have a typical time scale of a few seconds. Nevertheless, our results indicate that it may be possible to detect the contribution of slow neural processes to the scalp EEG signal, through their indirect effect on the frequency modulation of fast modes.

In conclusion, we have provided empirical evidence supporting the predictive validity of sDCM for fMRI data. In future work, we will examine the empirical and theoretical relationships that exist between slow and fast dynamical consequences of effective connectivity as can be accessed via fMRI and EEG, respectively.

## ACKNOWLEDGMENTS

The authors would like to thank Dr. Dimitris Pinotsis for useful discussions. K. J. Friston was funded by the Wellcome Trust. J. Daunizeau was funded by an ERC grant.

## REFERENCES

- Amari, S.-I. (1975). Homogeneous nets of neuron-like elements. *Biol. Cybern.* 17, 211–220.
- Amari, S.-I. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biol. Cybern.* 27, 77–87.
- Bojak, I., and Liley, D. T. (2010). Axonal velocity distributions in neural field equations. *PLoS Comp. Biol.* 6:e1000653. doi: 10.1371/journal.pcbi.1000653
- Bojak, I., Oostendorp, T. F., Reid, A. T., and Kotter, R. (2011). Towards a model-based integration of co-registered electroencephalography/functional magnetic resonance imaging data with realistic neural population meshes. *Philos. Transact. A Math. Phys. Eng. Sci.* 369, 3785–3801.
- Bressloff, P. C. (2012). Spatiotemporal dynamics of continuum neural fields. *J. Phys. A Math. Theor.* 45:033001. doi: 10.1088/1751-8113/45/3/033001
- Brodersen, K. H., Schofield, T. M., Leff, A. P., Ong, C. S., Lomakina, E. I., Buhmann, J. M., et al. (2011). Generative embedding for model-based classification of fMRI data. *PLoS Comput. Biol.* 7:e1002079. doi: 10.1371/journal.pcbi.1002079
- Buxton, R. B., Wong, E. C., and Franck, L. R. (1998). Dynamics of blood flow and oxygenation changes during brain activation: the Balloon model. *Magn. Res. Med.* 39, 855–864.
- Cabral, J., Hugues, E., Sporns, O., and Deco, G. (2011). Role of local network oscillations in resting-state functional connectivity. *Neuroimage* 57, 130–139.
- Coombes, S. (2010). Large-scale neural dynamics: simple and complex. *Neuroimage* 52, 731–739.
- Coombes, S., Venkov, N. A., Shiau, L., Bojak, I., Liley, D. T. J., and Laing, C. R. (2007). Modeling electrocortical activity through local approximations of integral neural field equations. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 76, 1539–1547.
- Daunizeau, J., David, O., and Stephan, K. E. (2011). Dynamic Causal Modelling: a critical review of the biophysical and statistical foundations. *Neuroimage* 58, 312–322.
- Daunizeau, J., Kiebel, S. J., and Friston, K. J. (2009a). Dynamic causal modeling of distributed electromagnetic responses. *Neuroimage* 47, 590–601.
- Daunizeau, J., Kiebel, S. J., and Friston, K. J. (2009b). Variational Bayesian identification and prediction of stochastic dynamic causal models. *Physica D* 289, 2089–2118.
- Daunizeau, J., Laufs, H., and Friston, K. J. (2010). “EEG-fMRI information fusion: biophysics and data analysis,” in *EEG-fMRI- Physiology, Technique and Applications*, eds C. Mulert and L. Lemieux (Berlin, Heidelberg: Springer), 511–526. ISBN: 978-3-540-87918-3.
- Daunizeau, J., Stephan, K. E., and Friston, K. J. (2012). Stochastic dynamic causal modelling of fMRI data: should we care about neural noise? *Neuroimage* 62, 464–481.
- David, O., Guillemain, I., Baillet, S., Rey, S., Deransart, C., Segebarth, C., et al. (2008). Identifying neural drivers with functional MRI: an electrophysiological validation. *PLoS Biol.* 6:e315. doi: 10.1371/journal.pbio.0060315
- Deco, G., Jirsa, V. K., Robinson, P., Breakspear, M., and Friston, K. (2008). The dynamic brain: from spiking neurons to neural masses and cortical fields. *PLoS Comp. Biol.* 4:e1000092. doi: 10.1371/journal.pcbi.1000092
- De Munck, J. C. (1988). The potential distribution in a layered spheroidal volume conductor. *J. Appl. Phys.* 64, 464–470.
- De Munck, J. C., Gonçalves, S. I., Faes, T. J., Kuijter, J. P., Pouwels, P. J., Heethaar, R. M., et al. (2008). A study of the brain's resting state based on alpha band power, heart rate and fMRI. *Neuroimage* 42, 112–121.
- De Munck, J. C., Gonçalves, S. I., Mammoliti, R., Heethaar, R. M., and Lopes da Silva, F. H. (2009). Interactions between different EEG frequency bands and their effect on alpha-fMRI correlations. *Neuroimage* 47, 69–76.
- Destexhe, A., and Sejnowski, T. J. (2001). *Thalamocortical Assemblies*. Oxford, UK: Oxford University Press.
- Freeman, W. J. (1975). *Mass Action in the Nervous System: Examination of the Neurophysiological Basis of Adaptive Behavior through the EEG*, 1st Edn. New York, NY: Academic Press.
- Friston, K. J., Harrison, L., and Penny, W. D. (2003). Dynamic causal modelling. *Neuroimage* 19, 1273–1302.
- Friston, K. J., Li, B., Daunizeau, J., and Stephan, K. E. (2011). Network discovery with DCM. *Neuroimage* 56, 1202–1221.
- Friston, K. J., Mattout, J., Trujillo-Barreto, N., Ashburner, J., and Peeny, W. (2007). Variational free energy and the Laplace approximation. *Neuroimage* 34, 220–234.
- Friston, K. J., and Penny, W. (2011). Post-hoc Bayesian model comparison. *Neuroimage* 56, 2089–2099.
- Friston, K. J., Stephan, K. E., Li, B., and Daunizeau, J. (2010). Generalised filtering. *Math. Probl. Eng.* 2010:621670. doi: 10.1155/2010/621670
- Friston, K. J., Trujillo-Barreto, N. J., and Daunizeau, J. (2008). DEM: a variational treatment of dynamical systems. *Neuroimage* 42, 849–885.
- Garrido, M. I., Kilner, J., Kiebel, S. J., and Friston, K. J. (2007). Evoked brain responses are generated by feedback loops. *Proc. Natl. Acad. Sci. U.S.A.* 104, 20961–20966.
- Haken, H. (1983). *Synergetics, an Introduction: Nonequilibrium Phase Transitions and Self-Organization in Physics, Chemistry, and Biology*, 3rd Rev. edn. New York, NY: Springer-Verlag.
- Hamandi, K., Salek-Haddadi, A., Laufs, H., Liston, A., Friston, K., Fish, D. R., et al. (2006). EEG-fMRI of idiopathic and secondarily generalized epilepsy. *Neuroimage* 31, 1700–1710.
- Hesselmann, G., Sadaghiani, S., Friston, K. J., and Kleinschmidt, A. (2010). Predictive coding or evidence accumulation? False inference and neuronal fluctuations. *PLoS ONE* 5:e9926. doi: 10.1371/journal.pone.0009926
- Jahanshahi, M., and Hallett, M. (eds.). (2003). *The Bereitschaftspotential, Movement-Related Cortical Potentials. Springer Series*. New York, NY: Kluwer Academic / Plenum Publishers. ISBN: 0306474077.
- Jansen, B. H., and Rit, V. G. (1995). Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biol. Cybern.* 73, 357–366.
- Jirsa, V., and Haken, H. (1996). Field theory of electromagnetic brain activity. *Phys. Rev. Lett.* 77, 960–963.
- Kass, R. E., and Raftery, A. E. (1995). Bayes factors. *J. Am. Stat. Assoc.* 90, 773–795.
- Kilner, J. M., Mattout, J., Henson, R., and Friston, K. J. (2005). Hemodynamic correlates of EEG: a heuristic. *Neuroimage* 28, 280–286.
- Knock, S. A., McIntosh, A. R., Sporns, O., Kötter, R., Hagmann, P., and Jirsa, V. K. (2009). The effects of physiologically plausible connectivity structure on local and global dynamics in large scale brain models. *J. Neurosci. Methods* 183, 86–94.
- Krakov, K., Allen, P. J., Lemieux, L., and Fish, D. R. (2000). Methodology: EEG-correlated fMRI. *Adv. Neurol.* 83, 187–201.
- Lachaux, J. P., Fonlupt, P., Kahane, P., Minotti, L., Ho mann, D., Bertrand, O., et al. (2007). Relationship between task-related gamma oscillations and BOLD signal: new insights from combined fMRI and intracranial EEG. *Hum. Brain Mapp.* 28, 1368–1375.
- Laufs, H., Daunizeau, J., Carmichael, D. W., and Kleinschmidt, A. (2008). Recent advances in recording electrophysiological data simultaneously with magnetic resonance imaging. *Neuroimage* 40, 515–528.
- Lemieux, L., Salek-Haddadi, A., Lund, T. E., Laufs, H., and Carmichael, D. (2007). Modelling large motion events in fMRI studies of patients with epilepsy. *Magn. Reson. Imaging* 25, 894–901.
- Li, B., Daunizeau, J., Stephan, K. E., Penny, W., and Friston, K. (2011). Stochastic DCM and generalized filtering. *Neuroimage* 58, 442–457.
- Liley, D. T. J., Cadusch, P. J., and Dafilis, M. P. (2002). A spatially continuous mean field theory of electrocortical activity. *Network* 13, 67–113.
- Liley, D. T. J., Foster, B. L., and Bojak, I. (2012). “Cooperative populations of neurons: mean field models of mesoscopic brain activity,” in *Computational Systems Neurobiology*, ed N. Le Novère (Netherlands: Springer), 315–362.
- Liston, A. D., Lund, T. E., Salek-Haddadi, A., Hamandi, K., Friston, K. J., and Lemieux, L. (2006). Modelling cardiac signal as a confound in EEG-fMRI and its application in focal epilepsy study. *Neuroimage* 30, 827–834.
- Logothetis, N. K., and Wandell, B. A. (2004). Interpreting the BOLD signal. *Annu. Rev. Physiol.* 66, 735–769.
- Lopes da Silva, F. H., Hoeks, A., Smits, H., and Zetterberg, L. H. (1973). Model of brain rhythmic activity: the alpha-rhythm of the thalamus. *Kybernetik* 15, 27–37.
- Marreiros, A. C., Kiebel, S. J., Daunizeau, J., Harrison, L. M., and Friston, K. J. (2008). Population dynamics under the Laplace assumption. *Neuroimage* 44, 701–714.
- Moran, R. J., Jung, F., Kumagai, T., Endepols, H., Graf, R., Dolan, R. J., et al. (2011). Dynamic causal models and physiological inference: a validation study using isoflurane anaesthesia in rodents. *PLoS ONE* 6:e22790. doi: 10.1371/journal.pone.0022790

- Nunez, P. L. (1974). The brain wave equation: a model for the EEG. *Math. Biosci.* 21, 279–297.
- Penny, W., Joao, M., Flandin, G., Daunizeau, J., Stephan, K. E., Friston, K. J., et al. (2010). Comparing families of dynamic causal models. *PLoS Comp. Biol.* 6:e1000709. doi: 10.1371/journal.pcbi.1000709
- Pinotsis, D. A., Moran, R., and Friston, K. J. (2012). Dynamic causal modelling with neural fields. *Neuroimage* 59, 1261–1274.
- Roebroek, A., Formisano, E., and Goebel, R. (2011). The identification of interacting networks in the brain using fMRI: model selection, causality and deconvolution. *Neuroimage* 58, 296–302.
- Rosa, M. J., Daunizeau, J., and Friston, K. J. (2010a). EEG-fMRI integration: a critical review of biophysical modelling and data analysis approaches. *J. Integr. Neurosci.* 9, 453–476.
- Rosa, M. J., Kilner, J., Blankenburg, F., Josephs, O., and Penny, W. (2010b). Estimating the transfer function from neuronal activity to BOLD using simultaneous EEG-fMRI. *Neuroimage* 49, 1496–1509.
- Schmuel, A. (2010). “Locally measured neuronal correlates of functional MRI signals,” in *EEG-fMRI- Physiology, Technique and Applications*, eds C. Mulert and L. Lemieux (Berlin, Heidelberg: Springer), 63–82. ISBN: 978-3-540-87918-3
- Schultze-Kraft, M., Becker, R., Breakspear, M., and Ritter, P. (2011). Exploiting the potential of three dimensional spatial wavelet analysis to explore nesting of temporal oscillations and spatial variance in simultaneous EEG-fMRI data. *Progr. Biophys. Mol. Biol.* 105, 67–79.
- Seung, H. S., Richardson, T. J., Lagarias, J. C., and Hopfield, J. J. (1995). “Minimax and Hamiltonian dynamics of excitatory-inhibitory networks,” in *NIPS '97 Proceedings of the 1997 Conference on Advances in Neural Information Processing Systems, Vol. 10* (Cambridge, MA: MIT Press), 329–335.
- Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., et al. (2011). Network modelling methods for fMRI. *Neuroimage* 54, 875–891.
- Sporns, O. (2010). *Networks of the Brain*. Cambridge, MA: MIT Press. ISBN: 9780262014694.
- Srinivasan, R., Tucker, D. M., and Murias, M. (1998). Estimating the spatial Nyquist of the human EEG. *Behav. Res. Methods Instrum. Comput.* 30, 8–19.
- Stephan, K. E., Kasper, L., Harrison, L., Daunizeau, J., Den Ouden, H., Breakspear, M., et al. (2008). Nonlinear dynamic causal models for fMRI. *Neuroimage* 42, 649–662.
- Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., and Friston, K. J. (2009). Bayesian model selection for group studies. *Neuroimage* 46, 1004–1017.
- Stephan, K. E., Weiskopf, N., Drysdale, P. M., Robinson, P. A., and Friston, K. J. (2007). Comparing hemodynamic models with DCM. *Neuroimage* 38, 387–401.
- Tononi, G., Sporns, O., and Edelman, G. M. (1994). A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proc. Natl. Acad. Sci. U.S.A.* 91, 5033–5037.
- Valdes-Sosa, P., Roebroek, A., Daunizeau, J., and Friston, K. (2011). Effective connectivity: influence, causality and biophysical modelling. *Neuroimage* 58, 339–361.
- Vaudano, A. E., Laufs, H., Kiebel, S. J., Carmichael, D. W., Hamandi, K., Guye, M., et al. (2009). Causal hierarchy within the thalamo-cortical network in spike and wave discharges. *PLoS ONE* 4:e6475. doi: 10.1371/journal.pone.0006475
- Von Holdt, R. E. (1965). Rational power of power series. *Am. Math. Mon.* 72, 740–743.
- Wilson, H. R., and Cowan, J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neuron. *Biophys. J.* 12, 1–24.
- Wilson, H. R., and Cowan, J. D. (1973). A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik* 13, 55–80.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 26 March 2012; accepted: 31 December 2012; published online: 18 January 2013.

Citation: Daunizeau J, Lemieux L, Vaudano AE, Friston KJ and Stephan KE (2013) An electrophysiological validation of stochastic DCM for fMRI. *Front. Comput. Neurosci.* 6:103. doi: 10.3389/fncom.2012.00103

Copyright © 2013 Daunizeau, Lemieux, Vaudano, Friston and Stephan. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.

## APPENDIX 1: APPROXIMATING THE LOCAL NEURAL FIELDS WITH A SPATIAL EIGENMODES EXPANSION

Let  $\nabla^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2$  be the 2D-Laplacian operator defined on a bounded (square) 2D domain of size  $\|D\|$ , where  $\{x, y\} = \mathbf{r}$  are the two cartesian coordinates on the domain. Recall that eigenvectors  $w^{(k, l)}(\mathbf{r})$  of  $\nabla^2$  are such that:

$$\begin{aligned} w^{(k, l)}(\mathbf{r}) &= e^{i\frac{\pi}{\|D\|}(kx + ly)} \\ &= w^{(k)}(x)w^{(l)}(y) \\ &= [w^{(1)}(x)]^k [w^{(1)}(y)]^l \end{aligned} \quad (\text{A1})$$

where we have used index pairs  $(k, l)$  for spatial frequencies along the x- and y-axis, respectively. The associated eigenvalues are given by:  $\lambda_{k, l} = -(k^2 + l^2) \pi^2/\|D\|^2$ . By convention,  $w^{(k)}$  refer to the eigenfunctions of the 1D-Laplacian, and its complex conjugate will be denoted as  $w^{(-k)}$ . Now, projecting Equation (5) onto the (complex conjugate) spatial eigenmodes yields the eigendynamics of the neural field (Daunizeau et al., 2009a):

$$\begin{aligned} \tau^2 \ddot{z}_1^{(k, l)}(t) &= -2\tau \dot{z}_1^{(k, l)}(t) - z_1^{(k, l)}(t) + \tau \gamma z_2^{(k, l)}(t) \\ z_2^{(k, l)}(t) &= \frac{\bar{v}}{2\sigma} \left( (\sigma^2 \lambda_{kl} - 1) z_2^{(k, l)}(t) + G^0 \underbrace{\int_D w^{(-k, -l)}(\mathbf{r}) S(\eta(\mathbf{r}, t)) d\mathbf{r}}_{\varsigma_{k, l}(t)} \right) \end{aligned} \quad (\text{A2})$$

where the projections  $\varsigma_{k, l}(t)$  of the firing rate form a complete basis function set.

Together with a truncated eigenmode expansion, Equation (A2) can be used to approximate the solution to Equation (5) of the main text. Such a truncation basically acts as a low-pass filter, neglecting the transient effects of higher spatial frequency modes. This can be motivated by noting that (1) the eigenmodes  $w^{(k, l)}(\mathbf{r})$  of the Laplacian operator are of increasing spatial frequency as the order  $(k, l)$  increases and (2) the associated eigenvalues  $\lambda_{kl}$  are negative (except for  $\lambda_{00} = 0$ ) with increasing magnitude. This implies that the dynamics of high-order spatial modes will be strongly damped and thus quickly disappear in the absence of external forcing.

We now focus on deriving an approximation to the projections  $\varsigma_{k, l}(t)$  of the firing rate.

First, let us Taylor-expand the sigmoidal mapping of PSP activity around zero:

$$S(\eta(\mathbf{r}, t)) = \sum_{m=0}^{\infty} \frac{1}{m!} S_0^{[m]} \eta^m(\mathbf{r}, t), \quad (\text{A3})$$

where  $S_0^{[m]} = \left. \frac{\partial^m S}{\partial x^m} \right|_0$  is the  $m$ th derivative of the sigmoidal mapping evaluated at  $\eta = 0$  and  $\eta^m(\mathbf{r}, t)$  is the  $m$ th power of mean PSP  $\eta(\mathbf{r}, t)$ . Assuming that Equation (A1) holds for the eigenvectors of the Laplacian operator defined onto the cortical domain  $D$ , Equation (6) becomes:

$$\begin{aligned} \eta(\mathbf{r}, t) &= \sum_k \sum_l z_1^{(k, l)}(t) w^{(k, l)}(\mathbf{r}) \\ &= \sum_k \tilde{z}_1^{(k)}(t, y) [w^{(1)}(x)]^k \\ \tilde{z}_1^{(k)}(t, y) &\equiv \sum_l z_1^{(k, l)}(t) [w^{(1)}(y)]^l \end{aligned} \quad (\text{A4})$$

where  $\tilde{z}_1^{(k)}(t, y)$  is a dummy eigenmode dynamics, that is still a function of space (more precisely, of the second coordinate  $y$ ). Inserting Equation (A4) into (A3) yields the following power series of power series in  $w^{(1)}$ :

$$\begin{aligned} S(\eta(\mathbf{r}, t)) &= \sum_{m=0}^{\infty} \frac{1}{m!} S_0^{[m]} \left( \sum_k \tilde{z}_1^{(k)}(t, y) [w^{(1)}(x)]^k \right)^m \\ &= \sum_{m=0}^{\infty} \frac{1}{m!} S_0^{[m]} \sum_k A_k(t, y, m) w^{(k)}(x) \end{aligned} \quad (\text{A5})$$

where the coefficients  $A_k(t, y, m)$  obey the following recursive equation (Von Holdt, 1965):

$$\begin{cases} A_k(t, y, m) = \frac{1}{k\tilde{z}_1^{(0)}(t)} \sum_{i=1}^k ((m+1)i - k) \tilde{z}_1^{(i)}(t) A_{k-i}(t, y, m) \\ A_0(t, y, m) = [\tilde{z}_1^{(0)}(t)]^m \end{cases} \quad (\text{A6})$$

The contribution  $\Upsilon_k(t, y)$  of the  $k$ th spatial mode to the expansion in Equation (A5) is given by:

$$\begin{aligned} S(\eta(\mathbf{r}, t)) &= \sum_k \Upsilon_k(t, y) w^{(k)}(x) \\ \Upsilon_k(t, y) &= \sum_{m=0}^{\infty} \frac{1}{m!} S_0^{[m]} A_k(t, y, m) \end{aligned} \quad (\text{A7})$$

These can be analytically derived from the recursive relationship given in Equation (A6). They are given in **Table A1** below, up to order  $k = 3$ .

It can be seen from **Table A1** that, except for the fundamental mode, the spatial mode contribution  $\Upsilon_k$  is of the form:

$$\begin{aligned} \Upsilon_k &\propto \sum_{j=1}^k \mathcal{O}([\tilde{z}_1]^j) S^{[j]}(\tilde{z}_1^{(0)}) \\ &= \sum_{j=1}^k \mathcal{O}([\tilde{z}_1]^j) \rho^j S(\tilde{z}_1^{(0)}) \prod_{m=1}^j (1 - mS(\tilde{z}_1^{(0)})), \\ &\approx \sum_{j=1}^k \mathcal{O}([\tilde{z}_1]^{2j+1}) \end{aligned} \quad (\text{A8})$$

where  $\mathcal{O}(x^j)$  is a polynomial of order  $j$  in  $x$  and  $\rho$  is the sigmoidal slope. The second line in Equation (A8) follows from the following property of the sigmoid function:  $S^{[j]}(x) = \rho^j S(x) \prod_{m=1}^j (1 - mS(x))$  and the last line would follow from a first-order Taylor expansion of the sigmoidal mapping. This means that the lower order term in each spatial mode contribution  $\Upsilon_k(t, y)$  is already a polynomial of order three in the dummy eigenmodes dynamics  $\tilde{z}_1^{(k)}(t, y)$ . We will therefore truncate the contribution to its first term, i.e.,:  $\Upsilon_0(t, y) = S(\tilde{z}_1^{(0)}(t, y))$  for  $k = 0$ , and  $\Upsilon_k(t, y) \approx \tilde{z}_1^{(k)}(t, y) S'(\tilde{z}_1^{(0)}(t, y))$  for  $k > 0$ . This yields the following approximation to the firing rate field:

$$S(\eta(\mathbf{r}, t)) \approx S(\tilde{z}_1^{(0)}(t, y)) w^{(0)}(x) + \sum_{k \geq 0} \rho S'(\tilde{z}_1^{(0)}(t, y)) \tilde{z}_1^{(k)}(t, y) w^{(k)}(x) \quad (\text{A9})$$

**Table A1 | This table gives both the coefficients  $A_k(t, y, m)$  (left) and the contribution of the  $k$ th spatial mode to the expansion in Equation (A4), as given by  $\Upsilon_k(t, y)$  (right) up to  $k = 3$ .**

	$A_k(t, y, m)$	$\Upsilon_k(t, y)$
$k = 0$	$[\tilde{z}_1^{(0)}]^m$	$S(\tilde{z}_1^{(0)})$
$k = 1$	$\frac{1}{1!} m \tilde{z}_1^{(1)} [\tilde{z}_1^{(0)}]^{m-1}$	$\frac{1}{1!} \tilde{z}_1^{(1)} S'(\tilde{z}_1^{(0)})$
$k = 2$	$m \tilde{z}_1^{(2)} [\tilde{z}_1^{(0)}]^{m-1} + \frac{m!}{2!(m-2)!} [\tilde{z}_1^{(1)}]^2 [\tilde{z}_1^{(0)}]^{m-2}$	$\tilde{z}_1^{(2)} S'(\tilde{z}_1^{(0)}) + \frac{1}{2!} [\tilde{z}_1^{(1)}]^2 S''(\tilde{z}_1^{(0)})$
$k = 3$	$m \tilde{z}_1^{(3)} [\tilde{z}_1^{(0)}]^{m-1} + \frac{m!}{(m-2)!} \tilde{z}_1^{(1)} \tilde{z}_1^{(2)} [\tilde{z}_1^{(0)}]^{m-2} + \frac{m!}{3!(m-3)!} [\tilde{z}_1^{(1)}]^3 [\tilde{z}_1^{(0)}]^{m-3}$	$\tilde{z}_1^{(3)} S'(\tilde{z}_1^{(0)}) + \tilde{z}_1^{(1)} \tilde{z}_1^{(2)} S''(\tilde{z}_1^{(0)}) + \frac{1}{3!} [\tilde{z}_1^{(1)}]^3 S^{[3]}(\tilde{z}_1^{(0)})$

Equation (A9) can now be used to approximate the integral with respect to  $x$  in Equation (A2):

$$\begin{aligned} s_{k,l}(t) &= \int w^{(-l)}(y) \left[ \int w^{(-k)}(x) S(\eta(\mathbf{r}, t)) dx \right] dy \\ &\approx \delta_k \int w^{(-l)}(y) S(\tilde{z}_1^{(0)}(t, y)) dy + (1 - \delta_k) \rho \int w^{(-l)}(y) S'(\tilde{z}_1^{(0)}(t, y)) \tilde{z}_1^{(k)}(t, y) dy \\ &= \delta_k \int w^{(-l)}(y) S(\tilde{z}_1^{(0)}(t, y)) dy + (1 - \delta_k) \rho \sum_{l'} z_1^{k,l}(t) \int w^{(l'-l)}(y) S'(\tilde{z}_1^{(0)}(t, y)) dy \end{aligned} \quad (\text{A10})$$

where  $\delta_k$  is a Dirac delta, which is zero except when  $k = 0$ .

Similarly, the sigmoidal terms in the integrands of Equation (A10) can be approximated using power series of power series, as follows:

$$\begin{aligned} S(\tilde{z}_1^{(0)}(t, y)) &\approx S(z_1^{(0,0)}(t)) w^{(0)}(y) + \sum_{l \geq 0} \rho S'(z_1^{(0,0)}(t)) z_1^{(0,l)}(t) w^{(l)}(y) \\ S'(\tilde{z}_1^{(0)}(t, y)) &\approx S'(z_1^{(0,0)}(t)) w^{(0)}(y) + \sum_{l \geq 0} \rho S''(z_1^{(0,0)}(t)) z_1^{(0,l)}(t) w^{(l)}(y) \end{aligned} \quad (\text{A11})$$

Again, inserting Equation (A11) into (A10), and ignoring high-order terms<sup>1</sup> yields:

$$s_{k,l}(t) = \delta_k \delta_l S(z_1^{(0,0)}(t)) + (1 - \delta_k \delta_l) \rho S'(z_1^{(0,0)}(t)) z_1^{(k,l)}(t) \quad (\text{A12})$$

Inserting Equation (A12) in (A2) yields:

$$\dot{z}_2^{(k,l)}(t) = \begin{cases} \frac{\bar{v}}{2\sigma} \left( G^0 S(z_1^{(0,0)}(t)) - z_2^{(0,0)}(t) \right) & \text{if } k = l = 0 \\ \frac{\bar{v}}{2\sigma} \left( (\sigma^2 \lambda_{kl} - 1) z_2^{(k,l)}(t) + \rho G^0 S'(z_1^{(0,0)}(t)) z_1^{(k,l)}(t) \right) & \text{otherwise} \end{cases} \quad (\text{A13})$$

The second line of Equation (A13) is Equation (8) of the main text, having re-ordered the eigenmodes index pairs  $(k, l) \rightarrow k$ , for the sake of notational simplicity.

The critical properties of this derivation can be summarized as follows:

First, the fundamental eigenmode response will be slower than higher-order eigenmodes (because of smaller dampening terms). Second, the response of fast eigenmodes is enslaved by the slow eigenmode. Third, the fast eigenmodes do not feedback onto the slow eigenmodes. Taken together, we can think of the neural field as a feedforward system: inputs to the field activate the slow eigenmode  $z_1^{(0,0)}(t)$ , which controls the fast modes  $z_1^{(k,l)}(t)$ . This is further discussed in the main text of this document.

## APPENDIX 2: DERIVING THE FREQUENCY SPECTRUM OF EIGENMODES

Recall that the eigenmode dynamics obey the following ODE (cf. Equation 8 in the main text):

$$\begin{aligned} \dot{\mathbf{z}}^{(k)} &= \mathbf{A} \mathbf{z}^{(k)} + \mathbf{v} \\ \mathbf{A} &= \begin{bmatrix} 0 & 0 & 1 \\ c \left( z_1^{(0)} \right) & -\frac{\bar{v}}{2\sigma} (1 - \sigma^2 \lambda_k) & 0 \\ -\frac{1}{\tau} & \frac{\gamma}{\tau} & -\frac{2}{\tau} \end{bmatrix} \end{aligned} \quad (\text{A14})$$

where  $\mathbf{v} = O(z^2)$  are high-order error terms that act as a perturbation to the deterministic evolution of eigenmodes. As a simplifying assumption, we consider that  $\mathbf{v}$  behaves as i.i.d. white noise, the Fourier transform  $F[z^{(k)}]$  of Equation (A8) writes:

$$\begin{aligned} F[\mathbf{z}^{(k)}] &= (2i\pi\omega \mathbf{I}_3 - \mathbf{A})^{-1} F[\mathbf{v}] \\ &\propto (2i\pi\omega \mathbf{I}_3 - \mathbf{A})^{-1} \mathbf{I}_3 \end{aligned} \quad (\text{A15})$$

where  $\mathbf{I}_3$  is a  $3 \times 1$  vector of ones, and  $\mathbf{I}_3$  is the  $3 \times 3$  identity matrix.

<sup>1</sup>These are  $O(z^2)$  terms that derive from the second term in the right-hand side of the second line of Equation (A11).

Note that the inverse translated Jacobian writes:

$$(2i\pi\omega\mathbf{I}_3 - \mathbf{A})^{-1} = \frac{1}{|2i\pi\omega\mathbf{I}_3 - \mathbf{A}|} \begin{bmatrix} \xi_k(\omega) \left(i\omega + \frac{2}{\tau}\right) & \frac{\gamma}{\tau} & \xi_k(\omega) \\ c(z_1^{(0)}) \left(i\omega + \frac{2}{\tau}\right) & i\omega \left(i\omega + \frac{2}{\tau}\right) + \frac{1}{\tau^2} & c(z_1^{(0)}) \\ \frac{\gamma}{\tau} c(z_1^{(0)}) - \xi_k(\omega) \frac{1}{\tau^2} & i\omega \frac{\gamma}{\tau} & i\omega \xi_k(\omega) \end{bmatrix} \quad (\text{A16})$$

$$\xi_k(\omega) = i\omega + \frac{\bar{\nu}}{2\sigma} (1 - \sigma^2 \lambda_k)$$

where  $\xi_k(\omega)$  is the only term that depends upon the mode order  $k$ , and the determinant of the translated Jacobian is given by:

$$\begin{aligned} |2i\pi\omega\mathbf{I}_3 - \mathbf{A}| &= i\omega \xi_k(\omega) \left(i\omega + \frac{2}{\tau}\right) - \frac{\gamma}{\tau} c(z_1^{(0)}) + \xi_k(\omega) \frac{1}{\tau^2} \\ &= \xi_k(\omega) \left(i\omega \left(i\omega + \frac{2}{\tau}\right) + \frac{1}{\tau^2}\right) - \frac{\gamma}{\tau} c(z_1^{(0)}) \end{aligned} \quad (\text{A17})$$

where  $\xi_k(\omega)$  is given in Equation (A12) (with appropriate eigenmodes index reordering  $(k, l) \rightarrow k$ ). The explicit Fourier transform of eigenmode dynamics now derives from inserting Equation (A16) into (A15):

$$F[\mathbf{z}^{(k)}] \propto \frac{1}{|2i\pi\omega\mathbf{I}_3 - \mathbf{A}|} \begin{bmatrix} \xi_k(\omega) \left(i\omega + \frac{2}{\tau} + 1\right) + \frac{\gamma}{\tau} \\ \left(i\omega + \frac{2}{\tau}\right) \left(c(z_1^{(0)}) + i\omega\right) + \frac{1}{\tau^2} + c(z_1^{(0)}) \\ \xi_k(\omega) \left(i\omega - \frac{1}{\tau^2}\right) + \frac{\gamma}{\tau} \left(i\omega + c(z_1^{(0)})\right) \end{bmatrix} \quad (\text{A18})$$

Now, we are interested in the total frequency power of EEG signal sampled at some position  $r'$  on the scalp that is induced by mean membrane depolarization:

$$\begin{aligned} P(\omega, \mathbf{r}') &\equiv \left| \int_D L(\mathbf{r}, \mathbf{r}') F[\eta(\mathbf{r}, t)] d\mathbf{r} \right|^2 \\ &= \left| \sum_k \alpha^{(k)}(\mathbf{r}') F[z_1^{(k)}] \right|^2 \\ \alpha^{(k)}(\mathbf{r}') &= \int_D L(\mathbf{r}, \mathbf{r}') w^{(k)}(\mathbf{r}) d\mathbf{r} \end{aligned} \quad (\text{A19})$$

where the second line derives from the eigendecomposition of the neural field (cf. Equation 6),  $L(\mathbf{r}, \mathbf{r}')$  is a spatial convolution operator that describes the quasi-static volume propagation of electromagnetic fields through head tissues, and  $\alpha^{(k)}(\mathbf{r}')$  captures the contribution of the  $k$ th eigenmode at position  $\mathbf{r}'$  on the scalp. It is known that the geometrical properties of electrical conductivity within head tissues eventually shape the EEG susceptibility to different cortical spatial scales. For the sake of simplicity, we will assume the skull acts as a perfect low-pass filter (see e.g., Srinivasan et al., 1998), which induces a cut-off order  $k_c$ , i.e.,:

$$\alpha^{(k)}(\mathbf{r}') = \begin{cases} \alpha^{(0)}(\mathbf{r}') & \text{if } k \leq k_c \\ 0 & \text{if } k > k_c \end{cases} \quad (\text{A20})$$

Equation (A20) basically neglects the continuous loss of sensitivity as one increases the spatial frequency of (cortical) membrane depolarization fields. Inserting Equation (A20) into (A19) yields Equation (9) of the main text, i.e.,:

$$\begin{aligned} P(\omega, \mathbf{r}') &= \alpha^{(0)}(\mathbf{r}')^2 \left| \sum_{k \leq k_c} F[z_1^{(k)}] \right|^2 \\ &= \alpha^{(0)}(\mathbf{r}')^2 \left| \sum_{k \leq k_c} \frac{\xi_k(\omega) \left(i\omega + \frac{2}{\tau} + 1\right) + \frac{\gamma}{\tau}}{\xi_k(\omega) \left(i\omega \left(i\omega + \frac{2}{\tau}\right) + \frac{1}{\tau^2}\right) - \frac{\gamma}{\tau} c(z_1^{(0)})} \right|^2 \end{aligned} \quad (\text{A21})$$

where  $\alpha^{(0)}(\mathbf{r}')$  is a scaling factor that will be different for different EEG sensors.

Equation (A21) is a valid approximation of the EEG power spectrum at position  $\mathbf{r}'$  on the scalp generated by one local brain region. The generalization to a network of remote brain regions requires the specification of the relative contribution of each region, which depends on their relative distance with EEG scalp sensors. In principle, this can be derived from knowledge of the forward operator  $L(\mathbf{r}, \mathbf{r}')$  (see e.g., De Munck, 1988). This is beyond the scope of the present study.

Lastly, note that the numerical values of the Laplacian's eigenvalues are approximated from the 2D-Euclidean case, as follows:  $\lambda_{k,l} = -(k^2 + l^2) \pi^2 / \|D\|^2$ . The effective frequency cut-off  $k_c$  is such that  $k^2 + l^2 \leq k_c^2$ . In practice, we use  $k_c^2 = 16$  and  $\|D\| = 1$  cm.