Technical Note

Population dynamics: Variance and the sigmoid activation function

André C. Marreiros,* Jean Daunizeau, Stefan J. Kiebel, and Karl J. Friston

Wellcome Trust Centre for Neuroimaging, University College London, UK

This paper demonstrates how the sigmoid activation function of neural-mass models can be understood in terms of the variance or dispersion of neuronal states. We use this relationship to estimate the probability density on hidden neuronal states, using non-invasive electrophysiological (EEG) measures and dynamic casual modelling. The importance of implicit variance in neuronal states for neural-mass models of cortical dynamics is illustrated using both synthetic data and real EEG measurements of sensory evoked responses.
© 2008 Elsevier Inc. All rights reserved.

Keywords: Neural-mass models; Nonlinear; Modelling

Introduction

The aim of this paper is to show how the sigmoid activation function in neural-mass models can be understood in terms of the dispersion of underlying neuronal states. Furthermore, we show how this relationship can be used to estimate the probability density of neuronal states using non-invasive electrophysiological measures such as the electroencephalogram (EEG).

There is growing interest in the use of mean-field and neural-mass models as observation models for empirical neurophysiological time-series (Wilson and Cowan, 1972; Nunez, 1974; Lopes da Silva et al., 1976; Freeman, 1978, 1975; Jansen and Rit, 1995; Jirsa and Haken, 1996; Wright and Liley, 1996; Valdes et al., 1999; Steyn-Ross et al., 1999; Frank et al., 2001; David and Friston, 2003; Robinson et al., 1997, 2001; Robinson, 2005; Rodrigues et al., 2006). Models of neuronal dynamics allow one to ask mechanistic questions about how observed data are generated. These questions or hypotheses can be addressed through model selection by comparing the evidence for different models, given the same data. This endeavour is referred to as dynamic causal modelling (DCM) (Friston, 2002, 2003; Penny et al., 2004; David et al., 2006a,b; Kiebel et al., 2006). There has been considerable success in modelling fMRI, EEG, MEG and LFP data using DCM

(David et al., 2006a,b; Kiebel et al., 2006; Garrido et al., 2007; Moran et al., 2007). All these models embed key nonlinearities that characterise real neuronal interactions. The most prevalent models are called neural-mass models and are generally formulated as a convolution of inputs to a neuronal ensemble or population to produce an output. Critically, the outputs of one ensemble serve as input to another, after some static transformation. Usually, the convolution operator is linear, whereas the transformation of outputs (*e.g.*, mean depolarisation of pyramidal cells) to inputs (firing rates in presynaptic inputs) is a nonlinear sigmoidal function. This function generates the nonlinear behaviours that are critical for modelling and understanding neuronal activity. We will refer to these functions as activation or input-firing curves.

The mechanisms that cause a neuron to fire are complex (Mainen and Sejnowski, 1995; Destexhe and Paré, 1999); they depend on the state (open, closed; active, inactive) of several kinds of ion channels in the postsynaptic membrane. The configuration of these channels depends on many factors, such as the history of presynaptic inputs and the presence of certain neuromodulators. As a result, neuronal firing is often treated as a stochastic process. Random fluctuations in neuronal firing function are an important aspect of neuronal dynamics and have been the subject of much study. For example, Miller and Wang (2006) look at the temporal fluctuations in firing patterns in working memory models with persistent states. One perspective on this variability is that it is caused by *fluctuations in the threshold* of the input-firing curve of individual neurons. This is one motivation for a sigmoid activation function at the level of population dynamics; which rests on the well-known result that the average of many different threshold functions is a nonlinear sigmoid. We will show the same sigmoid function can be motivated by assuming *fluctuations in the neuronal states* (Hodgkin and Huxley, 1952). This is a more plausible assumption because variations in postsynaptic depolarisation over a population are greater than variations in firing threshold (Fricker et al 1999): in active cells, membrane potential values fluctuate by up to about 20 mV, due largely to hyperpolarisations that follow activation. In contrast, firing thresholds vary up to only 8 mV. Furthermore, empirical studies show that voltage thresholds, determined from current injection or by elevating extracellular K^+, vary little with the rate of membrane polarisation and that the "speed of transition into the inactivated states also appears to

* Corresponding author. Wellcome Trust Centre for Neuroimaging, Institute of Neurology, UCL, 12 Queen Square, London, WC1N 3BG, UK. Fax: +44 207 813 1445.
  E-mail address: a.marreiros@fil.ion.ucl.ac.uk (A.C. Marreiros).

Available online on ScienceDirect (www.sciencedirect.com).

contribute to the invariance of threshold for all but the fastest depolarisations" (Fricker et al., 1999). In short, the same mean-field model can be interpreted in terms of random fluctuations on the firing thresholds of different neurons or fluctuations in their states. The latter interpretation is probably more plausible from neurobiological point of view and endows the sigmoid function parameters with an interesting interpretation, which we exploit in this paper. It should be noted that Wilson and Cowan (1972) anticipated that the sigmoid could arise from a fixed threshold and population variance in neural states; after Eq. (1) of their seminal paper they state: "Alternatively, assume that all cells within a subpopulation have the same threshold, … but let there be a distribution of the number of afferent synapses per cell." This distribution induces variability in the afferent activity seen by any cell.

This is the first in a series of papers that addresses the importance of high-order statistics (*i.e.*, variance) in neuronal dynamics, when trying to model and understand observed neurophysiological time-series. In this paper, we focus on the origin of the sigmoid activation function, which is a ubiquitous component of many neural-mass and cortical-field models. In brief, this treatment provides an interpretation of the sigmoid function as the cumulative density on postsynaptic depolarisation over an ensemble or population of neurons. Using real EEG data we will show that population variance, in the depolarisation of neurons in somatosensory sources generating sensory evoked potentials (SEP) (Litvak et al., 2007) can be quite substantial, especially in relation to evoked changes in the mean. In a subsequent paper, we will present a mean-field model of population dynamics that covers both the mean and variance of neuronal states. A special case of this model is the neural-mass model, which assumes that the variance is fixed (David et al., 2006a,b; Kiebel et al., 2006). In a final paper, we will use these models as probabilistic generative models (*i.e.*, dynamic causal models) to show that population variance can be an important quantity, when explaining observed EEG and MEG responses.

This paper comprises three sections. In the first, we present the background and motivation for using sigmoid activation functions. These functions map mean depolarisation, within a neuronal population, to expected firing rate. We will illustrate the origins of their sigmoid form using a simple conductance-based model of a single population. We rehearse the well-known fact that threshold or Heaviside operators in the equations of motion for a single neuron lead to sigmoid activation functions, when the model is formulated in terms of mean neuronal states. We will show that the sigmoid function can be interpreted as the cumulative density function on depolarisation, within a population.

In the second section we emphasise the importance of variance or dispersion by noting that a change in variance leads to a change in the form of the sigmoid function. This changes the transfer function of the system and its input–output properties. We will illustrate this by looking at the Volterra kernels of the model and computing the modulation transfer function to show how the frequency response of a neuronal ensemble depends on population variance.

In the final section, we estimate the form of the sigmoid function using the established dynamic causal modelling technique and SEPs, following medium nerve stimulation. In this analysis, we focus on a simple DCM of brainstem and somatosensory sources, each comprising three neuronal populations. Using standard variational techniques, we invert the model to estimate the density on various parameters, including the parameters controlling the shape of the

sigmoid function. This enables us to estimate the implicit probability density function on depolarisation of neurons within each population. We conclude by discussing the implications of our results for neural-mass models, which ignore the effects of population variance on the evolution of mean activity. We use these conclusions to motivate a more general model of population dynamics that will be presented in a subsequent paper (Marreiros et al., manuscript in preparation).

Theory

In this section, we will show that the sigmoid activation function used in neural-mass models can be derived from straightforward considerations about single-neuron dynamics. To do this, we look at the relationship between variance introduced at the level of individual neurons and their population behaviour.

Saturating nonlinear activation functions can be motivated by considering neurons as binary units; *i.e.*, as being in an active or inactive state. Wilson and Cowan (1972) showed that (assuming neuronal responses rest on a threshold or Heaviside function of activity) any unimodal distribution of thresholds results in a sigmoid activation function at the population level. This can be seen easily by assuming a distribution of thresholds within a population characterised by the density, $p(w)$. For unimodal $p(w)$, the response function, which is the integral of the threshold density, will have a sigmoid form. For symmetric and unimodal distributions, the sigmoid is symmetric and monotonically increasing; for asymmetric distributions, the sigmoid loses point symmetry around the inflection point; in the case of multimodal distributions, the sigmoid becomes wiggly (monotonically increasing but with more than one inflexion point). Another motivation for saturating activation functions considers the firing rate of a neuron and assumes that its time average equals the population average (*i.e.*, activity is ergodic). The firing rate of neurons always shows saturation and hence sigmoid-like behaviour. There are two distinct types of input-firing curves, type I and type II. The former curves are continuous and represent an increasing analytic function of input. The latter has a discontinuity, where firing starts after some critical input level is reached. These transitions correspond to a bifurcation from equilibrium to a limit-cycle attractor.[1] The type of bifurcation determines the fundamental computational properties of neurons. Type I and II neuronal behaviour can be generated by the same neuronal model (Izhikevich, 2007). From these considerations, it is possible to deduce population models (Dayan and Abbott, 2001).

We will start with the following ordinary differential equation (ODE) modelling the dynamics of a single neuron from the neural-mass model for EEG/MEG (David and friston, 2003, 2006a,b; Kiebel et al., 2006; Garrido et al., in press; Moran et al., 2007); for example, the *i*-th neuron in a population of excitatory spiny stellate cells in the granular layer:

$$
\begin{aligned}
\dot{x}_1^{(i)} &= x_2^{(i)} \\
\dot{x}_2^{(i)} &= \kappa G\left(\left\langle H(x_1^{(j)} - w^{(j)})\right\rangle_j + Cu\right) - 2\kappa x_2^{(i)} - \kappa^2 x_1^{(i)}.
\end{aligned} \tag{1}
$$

---

[1] In all cases, type I cells experience a saddle-node bifurcation on the invariant circle, at threshold. Type II neurons, may have three different bifurcations; i.e., a subcritical Hopf bifurcation (most frequent), a supercritical Hopf bifurcation, or a saddle node bifurcation outside the invariant circle.

This is a fairly ubiquitous form for neuronal dynamics in many neural-mass and cortical-field models and describes neuronal dynamics in terms of two states; $\dot{x}_1^{(i)}$ which can be regarded as depolarisation and $\dot{x}_2^{(i)}$, which corresponds to a scaled current. These ordinary differential equations correspond to a convolution of input with a 'synaptic' differential alpha-function (*e.g.*, Gerstner 2001). This synaptic kernel is parameterised by G, controlling the maximum postsynaptic potential and κ, which represents a lumped rate constant. Here input has exogenous and endogenous components: exogenous input is injected current u scaled by the parameter, C. Endogenous input arise from connections with other neurons in the same population (more generally, any population). It is assumed that each neuron senses all others, so that the endogenous input is the expected firing over neurons in the population. Therefore, neural-mass models are necessarily associated with a spatial scale over which the population is deployed; *i.e.* the so-called mesoscale,[2] from a few hundred to a few thousand neurons.

The firing of one neuron is assumed to be a Heaviside function of its depolarisation that is parameterised by some neuron-specific threshold, $w^{(i)}$. We can write this in terms of the cumulative density over the states and thresholds of the population

$$\left\langle H(x_1^{(j)} - w^{(j)}) \right\rangle_j = \iint H(x_1 - w)p(x_1,w)\mathrm{d}x_1\mathrm{d}w = \iint_{x_1>w} p(x_1,w)\mathrm{d}x_1\mathrm{d}w. \tag{2}$$

This expression can be simplified, if we assume the states have a large variability in relation to the thresholds (see Fricker et al., 1999) and replace the density on the thresholds, $p(w)$ with a point mass at its mode, w. Under this assumption, the input from other neurons can be expressed as a function of the sufficient statistics[3] of the population's states; for example, if we assume a Gaussian density $p(x_1) = N(x_1 : \mu_1, \sigma_1^2)$ we can write

$$\left\langle H(x_1^{(j)} - w) \right\rangle_j = \int_w^\infty p(x_1)\mathrm{d}x_1 = S(\mu_1 - w) \Rightarrow p(x_1) = S'(x_1 - \mu_1). \tag{3}$$

Where $S(\cdot)$ is the sigmoid cumulative density of a zero-mean normal distribution with variance σ_1^2 (*c.f.*, Freeman, 1975). Eq. (3) is quite critical because it links the motion of a single neuron to the population density and therefore couples microscopic and mesoscopic dynamics. Finally, we can summarise the population dynamics in terms of the sufficient statistics of the states to give a mean-field model $\dot{\mu} = f(\mu, \mu)$ by taking the expectation of Eq. (1)

$$\dot{x}_1^{(i)} = x_2^{(i)}$$
$$\dot{x}_2^{(i)} = \kappa G(S(\mu_1 - w) + Cu) - 2\kappa x_2^{(i)} - \kappa^2 x_1^{(i)} \Rightarrow$$

$$\dot{\mu}_1 = \mu_2$$
$$\dot{\mu}_2 = \kappa G(S(\mu_1 - w) + Cu) - 2\kappa\mu_2 - \kappa^2\mu_1. \tag{4}$$

---

[2] Different descriptions pertain to at least three levels of organization. At the lowest level we have single neurons and synapses (microscale) and at the highest, anatomically distinct brain regions and inter-regional pathways (macroscale). Between these lies the level of neuronal groups or populations (mesoscale) (Sporns et al., 2005).
[3] The quantities that specify a probability density; *e.g.*, the mean and variance.
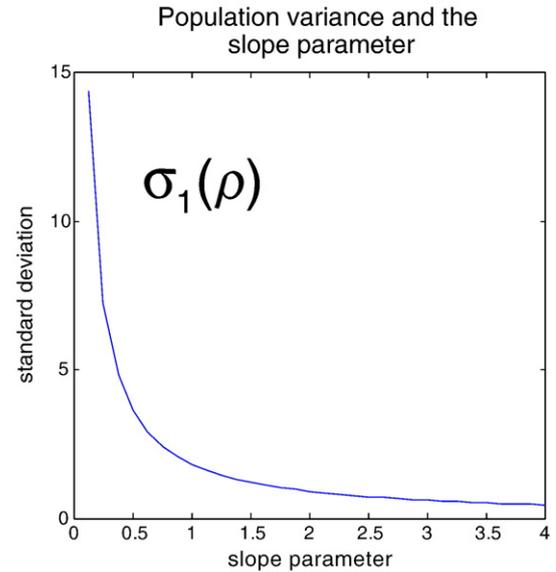

Fig. 1. Relationship between the sigmoid slope ρ and the population variance, expressed as the standard deviation.

We can do this easily because the equations of motion are linear in the states (note the sigmoid is not a function of the states). The ensuing mean-field model has exactly the same form as the neural-mass model we use in dynamic causal modelling of electromagnetic observations (David et al., 2006a,b). It basically describes the evolution of mean states that are observed directly or indirectly. In these neural-mass models the sigmoid has a fixed form[4]

$$S(\mu_i - w) = \frac{1}{1 + \exp(-\rho(\mu_i - w))} \tag{5}$$

where ρ is a parameter that determines its slope (*c.f.*, voltage-sensitivity). It is this function that endows the model with nonlinear behaviour and biological plausibility. However, this form assumes that the variance of the states is fixed, because the sigmoid encodes the density on neuronal states (see Eq. (3)). In the particular parameterisation of Eq. (5), the slope-parameter corresponds roughly to the inverse variance or precision of $p(x_i)$; more precisely

$$\sigma_i^2(\rho) = \int (x_i - \mu_i)^2 p(x_i)\mathrm{d}x_i$$
$$p(x_i) = S'(x_i - \mu_i) = \frac{\rho\exp(-\rho(x_i - \mu_i))}{(1 + \exp(-\rho(x_i - \mu_i)))^2}. \tag{6}$$

Fig. 1 shows the implicit standard deviation over neural states as a function of the slope-parameter, ρ. Heuristically, a high voltage-sensitivity or gain corresponds to a tighter distribution of voltages around the mean, so that near-threshold increases in the mean cause a greater proportion of neurons to fire and an increased sensitivity to changes in the mean.

This analysis is based on the assumption that variations in threshold are small, in relation to variability in neuronal states

---

[4] By fixed we mean constant over time. Note that we ignore a constant term here that can be absorbed into exogenous input.

themselves. Clearly, in the real brain, threshold variance is not zero; in the Appendix we show that if we allow for variance on the thresholds (as suggested by our reviewers), the standard deviation in Fig. 1 becomes an upper bound on the population variability of the states. In the next section, we look at how the dynamics of a population can change profoundly when the inverse variance (*i.e.*, gain) changes.

Kernels, transfer functions and the sigmoid

In this section, we illustrate the effect of changing the slope-parameter (*i.e.*, variance of the underlying neuronal states) on the input–output behaviour of neuronal populations. We will start with a time-domain characterisation, in terms of convolution kernels and conclude with a frequency-domain characterisation, in terms of transfer functions. We will see that the effects of changing the implicit variance are mediated largely by first-order effects and can be quite profound.

Nonlinear analysis and Volterra kernels

The input–output behaviour of population responses can be characterised in terms of a Volterra series. These series are a functional expansion of a population's input that produces its outputs (where the outputs from one population constitute the inputs to another). The existence of this expansion suggests that the history of inputs and the Volterra kernels represent a complete and sufficient specification of population dynamics (Friston et al., 2003b). The theory states that, under fairly general conditions, the

output y of a nonlinear dynamic system can be expressed in terms of an infinite sum of integral operators

$$y(t) = \sum_i \int \dots \int k_i(\sigma_1, \dots \sigma_i) u(t - \sigma_1) u(t - \sigma_i) d\sigma_1 \dots d\sigma_i \quad (7a)$$

where the i-th order kernel is

$$k_i(\sigma_1, \dots \sigma_i) = \frac{\partial y^i(t)}{\partial u(t - \sigma_1) \dots \partial u(t - \sigma_i)}. \quad (7b)$$

Volterra kernels represent the causal input–output characteristics of a system and can be regarded as generalised impulse response functions (*i.e.*, the response to an impulse or spike). The first-order kernel $\kappa_1(\sigma_1) = \partial y(t) / \partial u(t - \sigma_1)$ encodes the response evoked by a change in input at $t - \sigma_1$. In other words, it is a time-dependent measure of driving efficacy. Similarly the second-order kernel $\kappa_2(\sigma_1, \sigma_2) = \partial_y{}^2(t) / \partial u(t - \sigma_1) \partial u(t - \sigma_2)$ reflects the modulatory influence of the input at $t - \sigma_1$ on the response evoked by input at $t - \sigma_2$; and so on for higher orders.

Volterra series have been described as a 'power series with memory' and are generally thought of as a high-order or nonlinear convolution of inputs to provide an output. Essentially, the kernels are a re-parameterisation of the system that encodes the input–output properties directly, in terms of impulse response functions. In what follows, we computed the first and second-order kernels (*i.e.*, impulse response functions) of the neural-mass models, using different slope-parameters. This enabled us to see whether the changes in population variance are expressed primarily in first or second-order effects.
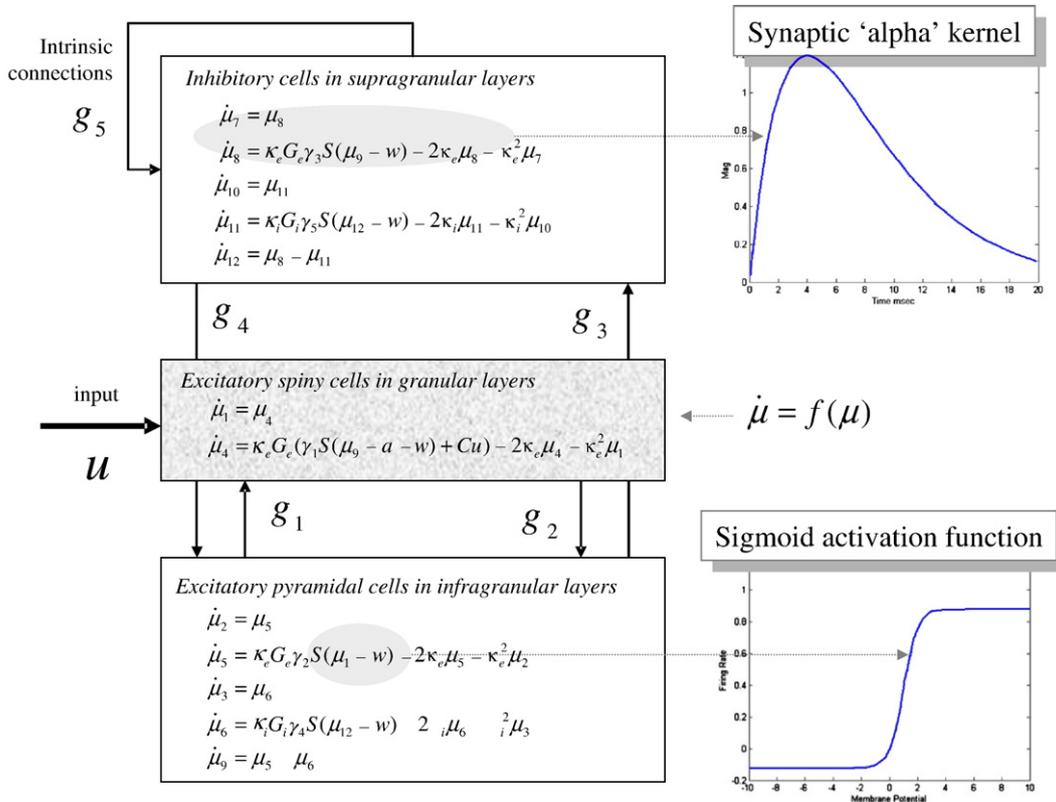

Fig. 2. Schematic of the neural-mass model used to model a single source (Moran et al., 2007).
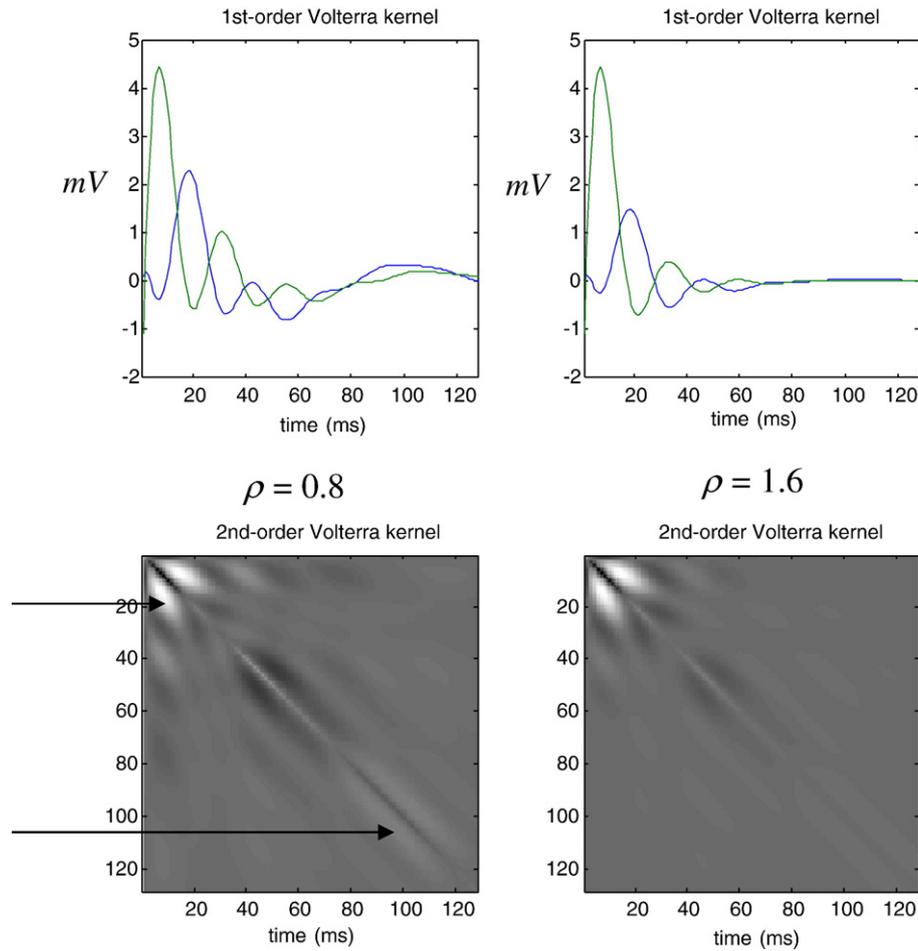
Volterra kernels and the slope parameter

Fig. 3. Upper panels: The first-order Volterra kernels for the depolarisation of pyramidal (blue) and spiny stellate (green) populations, for two different values of ρ (left: 0.8, right: 1.6). There is a difference between the waveform, which is marked for the pyramidal cells. Lower panels: The corresponding second-order Volterra kernels in image format. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The specific neural-mass model we used has been presented in detail by Moran et al. (2007). This model uses intrinsic coupling parameters, γ_i, between three subpopulations within any source of observed electromagnetic activity. Each source comprises an inhibitory subpopulation in the supragranular layer and excitatory pyramidal (output) population in an infra-granular layer. Both these populations are connected to an excitatory spiny (input) population in the granular layer. This model differs from the model used by David and friston, (2003) in two ways: (i) the inhibitory subpopulation has recurrent self-connections and (ii) spike-rate adaptation is included to mediate slow neuronal dynamics. The equations of motions for a three-population source are shown in Fig. 2; these all have the form of Eq. (4).

The first-order kernels or response functions for the depolarisation of the two excitatory populations are shown in Fig. 3 (upper panels) and the second-order kernels for the excitatory pyramidal cell population are shown in the lower panels, for two values of the slope-parameter; $\rho = 0.8$ and $\rho = 1.6$. The other parameters were chosen such that the system was dynamically stable; see Moran et al. (2007) and Table 1. The kernels were computed as described in the appendix of Friston et al. (2000).

The first-order responses exhibit a more complicated response for the smaller value of ρ; with pronounced peaks at about 10 ms and 20 ms for the stellate and pyramidal populations respectively. Both responses resemble damped fast oscillations in the gamma range (about 40 Hz). In addition, there appears to be a slower dynamic, with late peaks at about 100ms. This is lost with larger values of ρ (right panels); furthermore, the pyramidal response is attenuated and more heavily damped. The second-order kernels have two pronounced off-diagonal wing-shaped positivities that do not differ markedly for the two values of ρ. These high-order kernels tell us about nonlinear or modulatory interactions among

Table 1
Model parameters

Parameter	Physiological interpretation	Value
$H_{e,i}$	Maximum postsynaptic potentials	8 mV, 32 mV
$\tau_{e,i} = 1/\kappa_{e,i}$	Postsynaptic rate constants	4 ms, 16 ms
$\tau_a = 1/\kappa_{e,i}$	Adaptation rate constant	512 ms
$\gamma_{1,2,3,4,5}$	Intrinsic connectivity	128, 128, 64, 64, 4
w	Threshold	1.8

inputs and speak to asynchronous coupling. For example, the peaks in the second-order kernel at 10 ms and 20 ms (upper arrow) mean that the response to an input 10 ms in the past is positively modulated by an input 20 ms ago (and *vice versa*). The long-term memory of the population dynamics is expressed in positive asynchronous interactions (lower arrow) around 100 ms. These second-order effects correspond to interactions between inputs at different times, in terms of producing changes in the output. They can be construed as input effects that interact nonlinearly with intrinsic states, which 'remember' the inputs. In the present context, these effects are due to, and only to, the nonlinear form of the sigmoid function, which is mandated by the fact it is a cumulative probability density function. This is an important observation, which means, under the models considered here, population dynamics must, necessarily exhibit nonlinear responses.

The effect of changing the gain or slope-parameter is much more evident in the first-order, relative to the second-order kernels. This suggests population variance does not, in itself, change the nonlinear properties of the population dynamics, compared to linear effects. The reason that the slope-parameter has quantitatively more marked effects on the first-order kernel is that our neural-mass model is only weakly nonlinear; it does not involve any interactions among the states, apart from those mediated by the sigmoid activation function. We can use this to motivate a focus on linear effects using linear systems theory in the frequency domain.

Linear analysis and transfer functions

An alternative characterisation of generalised kernels is in terms of their Fourier transforms, which furnish generalised transfer functions. A transfer function allows one to compute the frequency or spectral response of a population given the spectral characteristics of its inputs. We have presented a linear transfer function analysis of this neural-mass model previously (Moran et al., 2007). Our model is linearised by replacing the sigmoid function with a first-order expansion around $\mu_i = 0$ to give

$$S(\mu_i) = S'(-w)\mu_i. \tag{8}$$

This assumes small perturbations of neuronal states around steady-state. Linearising the model in this way allows us to evaluate the transfer function

$$H(s) = C(sI - A)^{-1}B \tag{9}$$

where the state matrices, $A = \partial f / \partial x$ and $B = \partial f / \partial u$ are simply the derivatives of the equations of motion (*i.e.*, Eq. (4)) with respect to the states and inputs respectively. The frequency response for steady-state input oscillations at ω radians per second, obtains by evaluating the transfer function at $s = j\omega$ (where $j\omega$ represents the axis of the complex *s*-plane corresponding to steady-state frequency responses). When the system is driven by exogenous input with spectrum, $U(j\omega)$, the output is the frequency profile of the stimulus modulated by the transfer function

$$|Y(j\omega)| = |H(j\omega)||U(j\omega)|. \tag{10}$$

In brief, the transfer function, $H(s)$, filters or shapes the frequency spectra of the input, $U(s)$ to produce the observed spectral response, $Y(s)$. The transfer function $H(s)$ represents a normalized model of the systems input–output properties and embodies the steady-state behaviour of the system. Eq. (9) results from one of the most useful

properties of the Laplace transform, which enables differentiation to be cast as a multiplication. One benefit of this is that convolution in the time domain can be replaced by multiplication in the s-domain. This reduces the computational complexity of the calculations required to analyse the system.

We examined the effects of the slope-parameter on the transfer function by computing $|H(j\omega)|$ for different values of $\rho = \frac{1}{16}, \frac{2}{16}, \ldots, 2$. $|H(j\omega)|$ corresponds to the spectral response under white noise input (see Eq. (10)). Fig. 4 shows the spectral response is greatest at about, $\rho = 0.8$ when it exhibits a bimodal frequency distribution; with a pronounced alpha peak (~12 Hz) and a broader gamma peak (~40 Hz). As ρ increases or decreases from this value the alpha component is lost, leading to broad-band responses expressed maximally in the gamma range. This is an interesting result, which suggests that the population's spectral responses are quite sensitive to changes in the dispersion of states, particularly with respect to the relative amount of alpha and gamma power. Having said this, these results

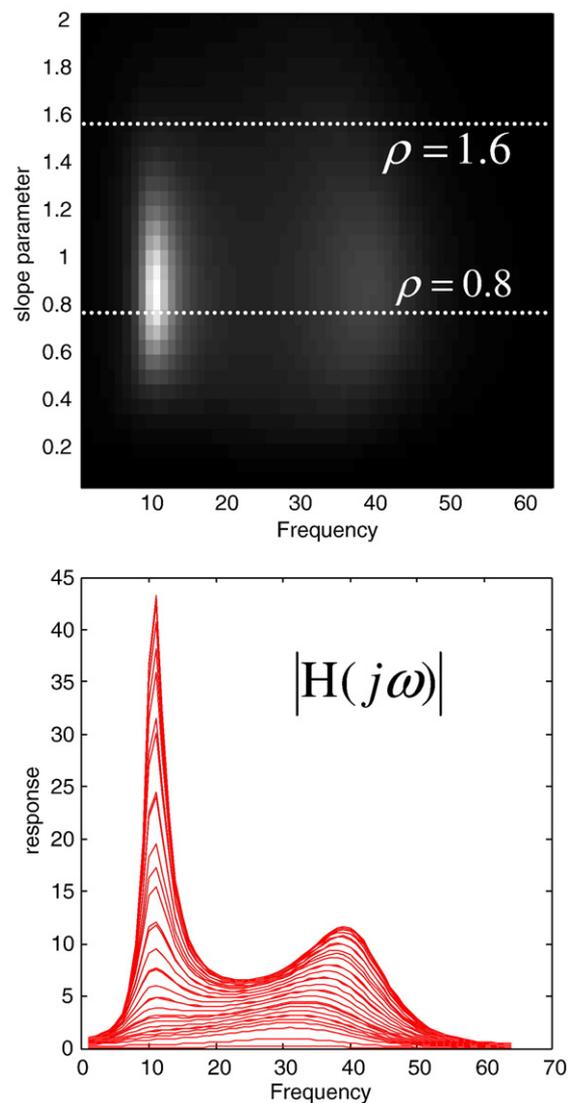Fig. 4. Upper panel: Image of the transfer function magnitude $H(s)$ where ρ is varied from a sixteenth to two. Lower panel: Plot of the same data over frequencies.

should not be generalised because they only hold for the values of the other model parameters we used. These values were chosen to highlight the dependency on the slope-parameter.

To illustrate the change in the response properties caused by a change in ρ, we computed the response of the excitatory populations to an input spike embedded in white noise process (where the amplitude of the noise was one sixteenth of the spike). Using exactly the same input, the responses were integrated for the two values of ρ above: $\rho = 0.8$ which maximises the frequency response and a larger value, $\rho = 1.6$. Fig. 5, shows the ensuing depolarisation of pyramidal and spiny cells and corresponding time–frequency plots. For the smaller ρ (large population variance), the output is relatively enduring with a preponderance of alpha power. For the larger value (small population variance), the output is more transient and embraces higher frequencies. We will return to this distinction in an empirical setting in the next section, where we try to estimate the slope-parameters and implicit population variance using real data.

Estimating population variance with DCM

In this final section, we exploit the interpretation of the sigmoid as a cumulative density on the states, specifically the depolarisa-

tion. This interpretation renders the derivative of the sigmoid a probability density on the voltage: recall from the first section

$$\int_{w}^{\infty} p(x_i)\mathrm{d}x_i = S(\mu_i - w) \Rightarrow p(x_i) = S'(x_i - \mu_i). \tag{11}$$

This means we can use estimates of the slope-parameter, which specifies S', to infer the underlying variance of depolarisation in neuronal populations (or an upper bound; see Appendix). In what follows, we estimate the slope-parameter using EEG data and Dynamic Causal Modelling. We present two analyses. The first addressed the question: "are changes in the mean depolarisation small or large relative to the dispersion of voltages?" We answered this by evaluating the evoked changes in mean depolarisation in somatosensory sources generating SEPs and then comparing the amplitude of these perturbations with the implicit variance. The second analysis tried to establish whether population variance is stable over time. This issue has profound implications for neural-mass models that assume variance does not change with time.
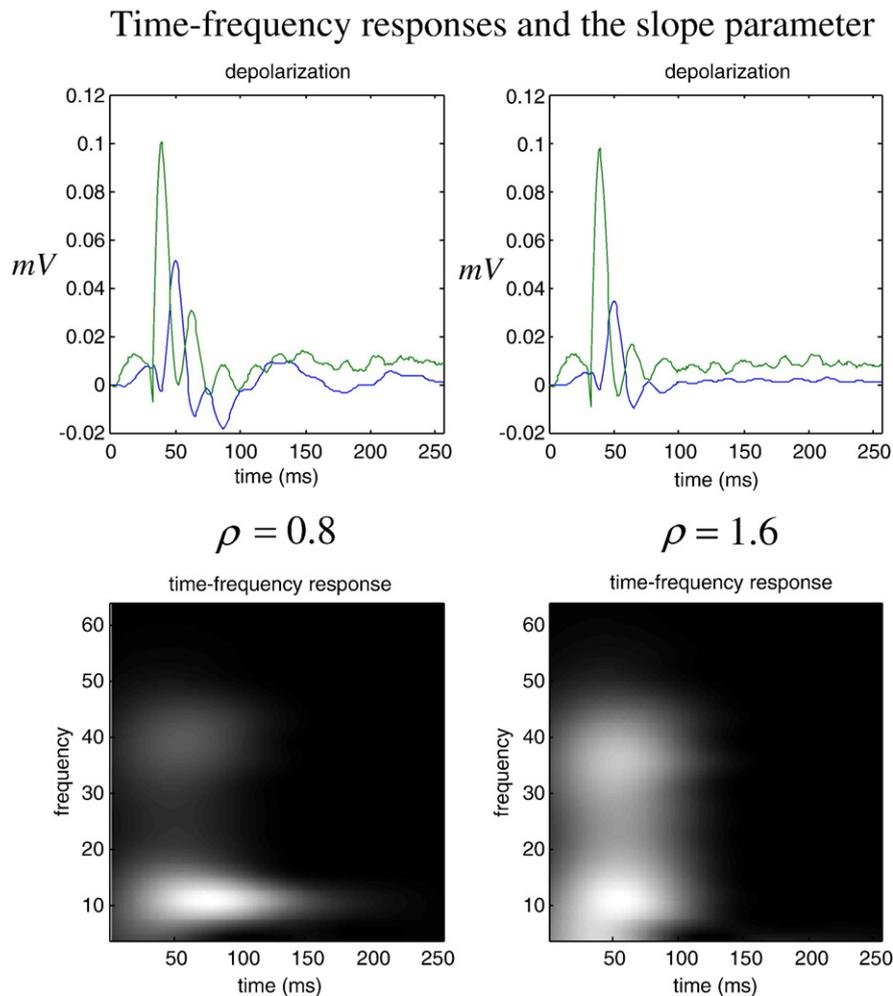

Fig. 5. Upper panels: Integrated response to a noisy spike input, for two different values of ρ (left: 0.8, right: 1.6). The response of the excitatory pyramidal (output) population is shown in blue, and the response of the spiny stellate in green. Lower panels: The respective time–frequency responses for the two ρ cases. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Analysis of somatosensory responses

We analysed data from a study of long-term potentiation (LTP) reported in Litvak et al. (2007). LTP is a long-lasting modification of synaptic efficacy and is believed to represent a physiological substrate of learning and memory (Bliss and Lomo, 1973; Martin et al., 2000; Malenka and Bear, 2004; Cooke and Bliss, 2006). Litvak et al. used paired associative stimulation (PAS), which involved repetitive magnetic cortical stimulation timed to interact with median nerve (MN) stimulation-induced peripheral signals from the hand. The PAS paradigm has been shown to induce long-lasting changes in MN somatosensory evoked potentials (MN-SSEP; Wolters et al., 2005) as measured by single-channel recordings from the scalp region overlying somatosensory cortex. The generators of MN-SSEPs evoked by compound nerve stimulation have been studied extensively with both invasive and non-invasive methods in humans and in animal models (for a review see Allison et al., 1991). Litvak et al. (2007) characterised the topographical distribution of PAS-induced excitability changes as a function of the timing and composition of afferent (MN) somatosensory stimulation, with respect to transcranial magnetic stimulation (TMS).

In this work, we analysed the SEP data from one subject, following MN stimulation (*i.e.*, in the absence of magnetic stimulation), with DCM. The network architecture was based on reports in published literature (Buchner et al., 1995; Ravazzani et al., 1995, Litvak et al., 2007). We modelled the somatosensory system with four equivalent current dipoles or sources, each comprising three neuronal subpopulations as described in the previous section. Exogenous input was modelled with a gamma function (with free parameters), peaking shortly after MN stimulation. In this model, exogenous input was delivered to the brainstem source (BS), which accounts for early responses in the medulla. In Brodmann area (BA) 3b of S1, we deployed three sources, given previous work showing distinct tangential and radial dipoles. We employed a third source to account for any other activity. These sources received endogenous input from the BS source, via extrinsic connections to the stellate cells.

We inverted the resulting DCM using a variational scheme (David et al., 2006a) and scalp data from 12 ms to 100 ms, following MN stimulation. This inversion used standard variational techniques, which rest on a Bayesian expectation maximisation (EM) algorithm under a Laplace approximation to the true posterior. This provided the posterior densities of the models parameters, which included the synaptic parameters of each population, the extrinsic connection strengths, the parameters of the gamma input function and the spatial parameters of the dipoles (for details see David et al., 2006a,b; Kiebel et al., 2006). The resulting posterior means of dipole locations and moments are shown in Fig. 6 (upper panel).

In terms of the temporal pattern of responses, the MN-SSEP has been studied extensively (Allison et al., 1991). A P14 component is generated subcortically, then a N20–P30 complex at the sensorimotor cortex (BA 3b) exhibits a typical 'tangential source pattern'.
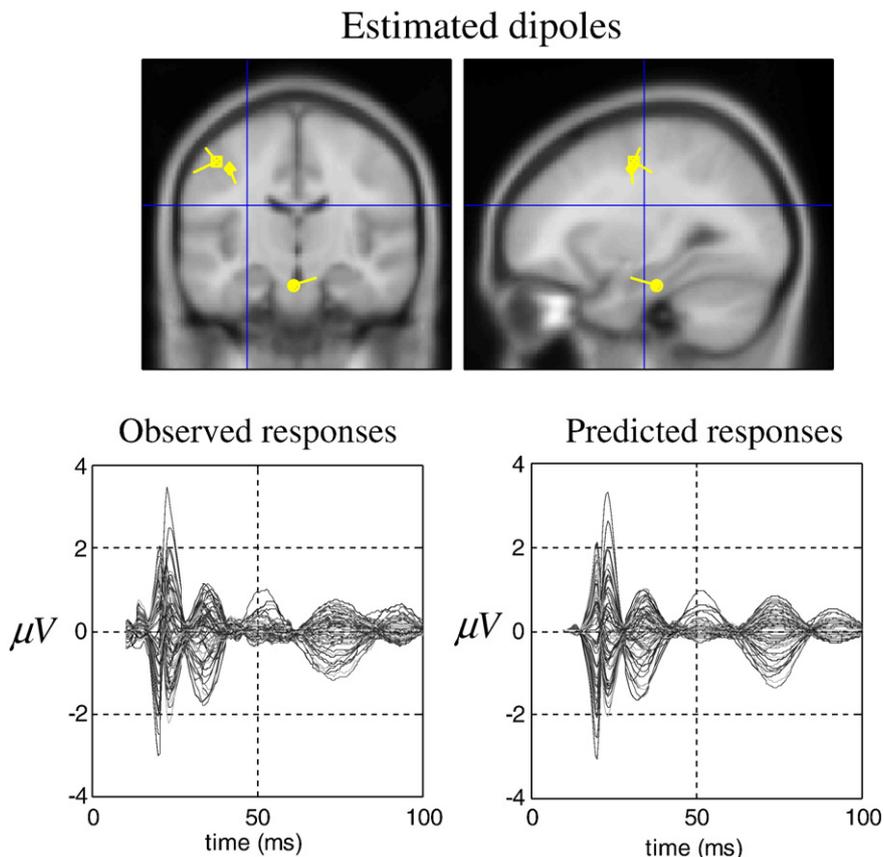

Fig. 6. Upper panels: Source locations estimated with DCM: Orthogonal slices showing the brainstem dipole (BS) and the left primary somatosensory cortex (S1) source (consisting of three dipoles: tangential, radial and orthogonal). Lower panels: The left graph shows the observed MN-SSEP in channel space. The right graph demonstrates the goodness of fit of the DCM using the same format.

This is followed by a P25–N35 complex with a 'radial source pattern'. The remainder of the SEP can be explained by an 'orthogonal source pattern' originating from the hand representation in S1 (Litvak et al., 2007). These successive response components were reproduced precisely by the DCM. The accuracy of the DCM can be appreciated by comparing the observed data with predicted responses in Fig. 6 (lower panels).

Using Eq. (6) and the maximum *a posteriori* estimate of the slope-parameter, we evaluated the implicit variance of depolarisation $\sigma_i^2(\rho)$ within each neuronal population (see Eq. (6) and Fig. 1). This variance can be combined with the time-dependent mean depolarisation $\mu_i(t)$ of any population, estimated by the DCM, to reconstruct the implicit density on population depolarisation over peristimulus time. Fig. 7 shows this density in terms of its mean and 90% confidence intervals for the first S1 pyramidal population. This quantitative analysis is quite revealing; it shows that evoked changes in the mean depolarisation are small in relation to the dispersion. This means that only a small proportion of neurons are driven above threshold, even during peak responses. For example, using the estimated threshold, *w*, during peak responses only about 12% of neurons would be above threshold and contribute to the output of the population. In short, this sort of result suggests that communication among different populations is mediated by a relatively small faction of available neurons and that small changes in mean depolarisation are sufficient to cause large changes in firing rates, because depolarisation is dispersed over large ranges.

Epilogue

The preceding analysis assumes that the variance is fixed over peristimulus time. Indeed neural-mass models in general assume a fixed variance because they assume a fixed form for the sigmoid activation function. Neural-mass models are obliged to make this assumption because their state variables allow only changes in mean states, not changes in variance or higher-order statistics of neuronal activity. Is this assumption sensible?

In our next paper on population variance we will compare mean-field models that cover both the mean and variance as time-varying
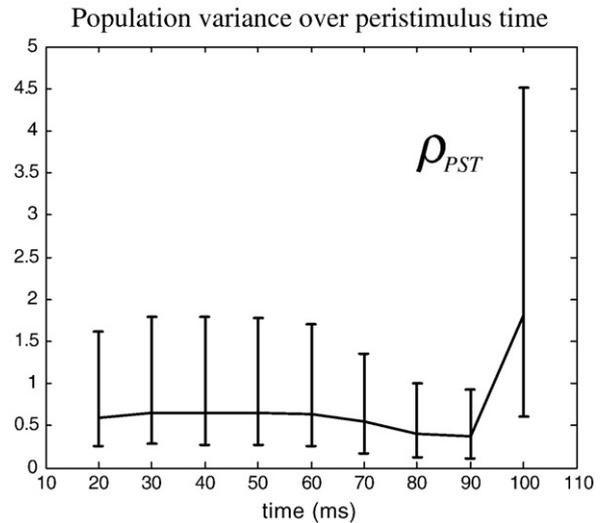
Fig. 8. Change in the conditional estimates of ρ (mean and 90% confidence intervals) as a function of the peristimulus time-window used for model inversion.

quantities. Under the neural-mass model considered here, we cannot test formally for changes in variance. However, we can provide anecdotal evidence for changes in variance by estimating the slope-parameters over different time-windows of the data. If the variance does not change with time, then the estimate of population variance should not change with the time-window used to estimate it. Fig. 8 show estimates of ρ (with 90% confidence intervals)[5] that obtain using different time-windows of the MN-SSEP data. For example, the estimate, ρ_{80} was obtained using the time period from 10 to 80 ms. It can be seen immediately that the slope-parameter and implicit variance changes markedly with the time-window analysed.

The results in Fig. 8 should not be over interpreted because there are many factors that can lead to differences in the conditional density when the data change; not least a differential shrinkage to the prior expectation. However, this instability in the conditional estimates speaks to the potential importance of modelling population variance as a dynamic quantity.

Conclusion

In this paper, our focus was on how the sigmoid activation function, linking mean population depolarisation to expected firing rate can be understood in terms of the variance or dispersion of neuronal states. We showed that the slope-parameter ρ models formally the effects of variance (to a first approximation) on neuronal interactions. Specifically, we saw that the sigmoid function can be interpreted as a cumulative density function on depolarisation, within a population. Then, we looked at how the dynamics of a population can change profoundly when the variance (slope-parameter) changes. In particular, we examined how the input–output properties of populations depend on ρ, in terms of first (driving) and second (modulatory) order convolution kernels and corresponding transfer functions.
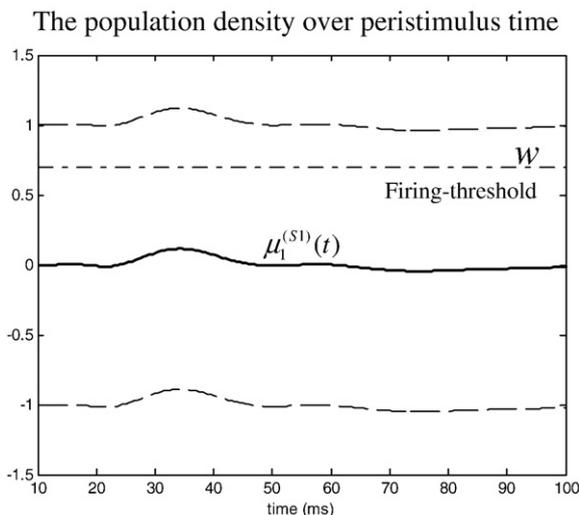

Fig. 7. S1 source (pyramidal population) mean depolarisation (solid line) as estimated by DCM. The variance is depicted with 90% confidence intervals (dashed lines); *i.e.*, $\pm 1.641 \times \sigma_i^2(\rho)$.

---

[5] Note that these confidence intervals are not symmetric about the mean. This is because we actually estimate ln ρ, under Gaussian shrinkage priors. Under the Laplace assumption (David et al., 2006a) this means the condition density $q(\rho)$ has a log-normal form.

We used real EEG data to show that population variance, in the depolarisation of neurons from somatosensory sources generating SEPs, can be quite substantial. Using DCM, we estimated the SEP parameter density controlling the shape of the sigmoid function. This allowed us to quantify the population variance in relation to the evolution of mean activity of neural-masses. The quantitative results of this analysis suggested that only a small proportion of neurons are actually firing at any time, even during the peak of evoked responses.

We have seen that the dynamics of neuronal populations can be captured qualitatively via a system of coupled differential equations, which describe the evolution of the average firing rate of each population. To accommodate stochastic models of neural activity, one could solve the associated Fokker–Planck equation for the probability distribution of activities in the different neuronal populations. This can be a difficult computational task, in the context of a large number of states and populations (*e.g.*, Harrison et al., 2005). Rodriguez and Tuckwell (1996, 1998) presented an alternative approach for noisy systems using the method of moments (MM). This entails the derivation of deterministic ordinary differential equations (ODE) for the first and second-order moments of the population density. The resulting reduced system lends itself to both analytical and numerical solution, as compared with the original Langevin formulation.

Hasegawa (2003a) proposed a semi-analytical mean-field approximation, in which equations of motions for moments were derived for a FitzHugh–Nagumo ensemble. In Hasegawa (2003b), the original stochastic differential equations were replaced by deterministic ODEs by applying the method of moments (Rodriguez and Tuckwell, 1996). This approach was applied to an ensemble of Hodgkin–Huxley neurons, for which effects of noise, coupling strength, and ensemble size have been investigated. In Deco and Martí (2007), the MM was extended to cover bimodal densities on the state variables; such that a reduced system of deterministic ODEs could be derived to characterise regimes of multistability. We will use MM in our next paper, where we derive the ODEs of the sufficient statistics of integrate-and-fire ensembles of distributed neuronal sources. These ODEs will form the basis of dynamical causal models of empirical EEG and LFP data.

The insights from these studies and this paper motivate a more general model of population dynamics that will be presented in the next paper on this subject. In that paper, we will compare DCMs based on density-dynamics with those based on neural-mass models. Modelling the interactions between mean neuronal states (*e.g.*, depolarisation) and their dispersion or variance over each population may provide a better and more principled model of real data. In brief, these models allow us to ask if the variance of neuronal states in a population affects the mean (or *vice versa*) using the evidence or marginal likelihood of the data under different models. Moreover, we can see if observed responses are best explained by mean firing rates, or some mixture of the mean and higher-order moments. This will allow one to adjudicate between models that include high-order statistics of neuronal states in EEG time-series models. In a final paper, we will use these models as probabilistic generative models (*i.e.*, dynamic causal models) to show that population variance is an important quantity, when explaining observed EEG and MEG responses.

Acknowledgments

Software note

Matlab demonstration and modelling routines referred to in this paper are available as academic freeware as part of the SPM software from http://www.fil.ion.ucl.ac.uk/spm (neural models toolbox).

Appendix

The predicted firing rate of a population is the expectation of the step or Heaviside function of depolarisation, over both the states and the threshold probability density functions (Eq. (2) in the main text):

$$\left\langle H(x_1^{(j)} - w^{(j)}) \right\rangle_{p(w)p(x_1)} = \int\int H(x_1 - w)p(x_1, w)dx_1 dw. \qquad \text{A.1}$$

Assuming that x_1 and w are independent and normally distributed; $z = x_1 - w$ has a Gaussian distribution; $p(z) = N(\mu_z, \Sigma_z) = N(\mu_x - \mu_w, \Sigma_x + \Sigma_w)$ and Eq. (A.1) can be written as:

$$\left\langle H(x_1^{(j)} - w^{(j)}) \right\rangle_{p(w)p(x_1)} = \int H(z)dz$$
$$= \int_{z \geq 0} p(z)dz \approx \frac{1}{1 + \exp\left(\rho\left(\sum_z\right)(\mu_x - \mu_w)\right)}. \qquad \text{A.2}$$

This means the expected firing rate remains a function of the sufficient statistics of the population and retains the same form as Eq. (3). Furthermore, it shows that for any given value of the slope-parameter, $\rho(\Sigma_z)$ the implicit variance

$$\rho\left(\sum_z\right)^{-1} = \sum_z(\rho) = \sum_x + \sum_w \geq \sum_x \qquad \text{A.3}$$

is always greater than the population variance on neuronal states. This means we always overestimate the proportion of supra-threshold neurons that contribute to the firing because $\Sigma_z(\rho)$ is an overestimate of the population variance. In other words, the 12% estimate from Fig. 7 is an upper bound on the actual proportion of firing neurons.

References

Allison, T., McGarthy, G., Wood, C.C., Jones, S.J., 1991. Potentials evoked in human and monkey cerebral cortex by stimulation of the median nerve. A review of scalp and intracranial recordings. Brain 114, 2465–2503.

Bliss, T.V., Lomo, T., 1973. Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. J. Physiol. (Lond.) 232, 331–356.

Buchner, H., Adams, L., Muller, A., Ludwig, I., Knepper, A., Thron, A., Niemann, K., Scherg, M., 1995. Somatotopy of human hand somatosensory cortex revealed by dipole source analysis of early somatosensory evoked potentials and 3D-NMR tomography. Electroencephalogr. Clin. Neurophysiol. 96, 121–134.

Cooke, S.F., Bliss, T.V., 2006. Plasticity in the human central nervous system. Brain 129, 1659–1673.

David, O., Friston, K.J., 2003. A neural-mass model for MEG/EEG: coupling and neuronal dynamics. NeuroImage 20, 1743–1755.

David, O., Kiebel, S., Harrison, L.M., Mattout, J., Kilner, J.M., Friston, K.J., 2006a. Dynamic causal modelling of evoked responses in EEG and MEG. NeuroImage 20, 1255–1272 (Feb).

David, O., Kilner, J.M., Friston, K.J., 2006b. Mechanisms of evoked and induced responses in MEG/EEG. NeuroImage 31 (4), 1580–1591 (Jul 15).

Dayan, P., Abbott, L.F., 2001. Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems. MIT Press. ISBN0-262-04199-5.

Deco, G., Martí, D., 2007. Extended method of moments for deterministic analysis of stochastic multistable neurodynamical systems. Phys. Rev., E 75, 031913.

Destexhe, A., Paré, D., 1999. Impact of network activity on the integrative properties of neocortical pyramidal neurons in vivo. J. Neurophysiol. 81 (4), 1531–1547 (Apr).

Kiebel, S.J., David, O., Friston, K.J., 2006. Dynamic causal modelling of evoked responses in EEG/MEG with lead field parameterization. NeuroImage 30 (4), 1273–1284 (May).

Frank, T.D., Daffertshofer, A., Beek, P.J., 2001. Multivariate Ornstein–Uhlenbeck processes with mean-field dependent coefficients: application to postural sway. Phys. Rev., E Stat. Nonlin. Soft. Matter Phys. 63, 011905.

Freeman, W.J., 1975. Mass Action in the Nervous System. Academic Press, New York.

Freeman, W.J., 1978. Models of the dynamics of neural populations. Electroencephalogr. Clin. Neurophysiol., Suppl. 9–18.

Fricker, D., Verheugen, J.H.S., Miles, R., 1999. Cell-attached measurements of the firing threshold of rat hippocampal neurones. J. Physiol. 517.3, 791–804.

Friston, K.J., 2002. Bayesian estimation of dynamical systems: an application to fMRI. NeuroImage 16 (4), 513–530.

Friston, K.J., 2003. Volterra kernels and connectivity, In: Frackowiak, R.S.J., Friston, K.J., Frith, C., Dolan, R., Friston, K.J., Price, C.J., Zeki, S., Ashburner, J., Penny, W.D. (Eds.), Human Brain Function, 2nd edition. Academic Press.

Friston, K.J., Mechelli, A., Turner, R., Price, C.J., 2000. Nonlinear responses in fMRI: the Balloon model, Volterra kernels, and other hemodynamics. NeuroImage 12 (4), 466–477 (Oct).

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

Garrido, M.I., Kilner, J.M., Kiebel, S.J., Stephan, K.E., Friston, K.J., 2007. Dynamic causal modelling of evoked potentials: a reproducibility study. NeuroImage 36, 571–580.

Garrido, M.I., Kilner, J.M., Kiebel, S.J., Friston, K.J., in press. Evoked brain responses are generated by feedback loops. NeuroImage.

Gerstner, W., 2001. A framework for spiking neuron models: the spike response model. In: Moss, F., Gielen, S. (Eds.), The Handbook of Biological Physics, vol. 4, pp. 469–516.

Harrison, L.M., David, O., Friston, K.J., 2005. Stochastic models of neuronal dynamics. Philos. Trans. R. Soc. Lond., B Biol. Sci. 360 (1457), 1075–1091.

Hasegawa, H., 2003a. Dynamical mean-field theory of spiking neuron ensembles: response to a single spike with independent noises. Phys. Rev., E 67, 041903.

Hasegawa, H., 2003b. Dynamical mean-field theory of spiking neuron ensembles: response to a single spike with independent noises. Phys. Rev., E 68, 041909.

Hodgkin, A.L., Huxley, A.F., 1952. A quantitative description of membrane current and its application to conduction and excitation in nerve. J. Physiol. 117, 500–544.

Izhikevich, E., 2007. Dynamical Systems in Neuroscience. The Geometry of Excitability and Bursting. MIT Press.

Jansen, B.H., 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.K., Haken, H., 1996. Field theory of electromagnetic brain activity. Phys. Rev. Lett. 77 (5), 960–963 (Jul 29).

Litvak, V., Zeller, D., Oostenveld, R., Maris, E., Cohen, A., Schramm, A., Gentner, R., Zaaroor, M., Pratt, H., Classen, J., 2007. LTP-like changes induced by paired associative stimulation of the primary somatosensory cortex in humans: source analysis and associated changes in behaviour. Eur. J. Neurosci. 25 (9), 2862–2874 (May).

Lopes da Silva, F.H., van Rotterdam, A., Barts, P., van Heusden, E., Burr, W., 1976. Models of neuronal populations: the basic mechanisms of rhythmicity. Prog. Brain Res. 45, 281–308.

Mainen, Z.F., Sejnowski, T.J., 1995. Reliability of spike timing in neocortical neurons. Science 268, 1503–1506.

Malenka, R.C., Bear, M.F., 2004. LTP and LTD: an embarrassment of riches. Neuron 44, 5–21.

Martin, S.J., Grimwood, P.D., Morris, R.G., 2000. Synaptic plasticity and memory: an evaluation of the hypothesis. Annu. Rev. Neurosci. 23, 649–711.

Miller, P., Wang, X.-J., 2006. Power-law neuronal fluctuations in a recurrent network model of parametric working memory. J. Neurophysiol. 95 (2), 1099–1114.

Moran, R.J., Kiebel, S.J., Stephan, K.E., Reilly, R.B., Daunizeau, J., Friston, K.J., 2007. A Neural-mass Model of spectral responses in electrophysiology. NeuroImage 37 (3), 706–720.

Nunez, P., 1974. The brain wave equation: a model for the EEG. Math. Biosci. 21, 279.

Penny, W.D., Stephan, K.E., Mechelli, A., Friston, K.J., 2004. Comparing dynamic causal models. NeuroImage 22, 1157–1172.

Ravazzani, P., Tognola, G., Grandori, F., Budai, R., Locatelli, T., Cursi, M., Di Benedetto, G., Comi, G., 1995. Temporal segmentation and multiple source analysis of short-latency median nerve SEPs. J. Med. Eng. Technol. 19, 70–76.

Robinson, P.A., Rennie, C.J., Wright, J.J., 1997. Propagation and stability of waves of electrical activity in the cerebral cortex. Phys. Rev., E 56 (1), 826–840.

Robinson, P.A., Rennie, C.J., Wright, J.J., Bahramali, H., Gordon, E., Rowe, D.L., 2001. Prediction of electroencephalographic spectra from neurophysiology. Phys. Rev., E 63, 021903.1–021903.18.

Robinson, P.A., 2005. Propagator theory of brain dynamics. Phys. Rev., E Stat. Nonlin. Soft. Matter Phys. 72 (1), 011904.1–011904.13.

Rodriguez, R., Tuckwell, H.C., 1996. A dynamical system for the approximate moments of nonlinear stochastic models of spiking neurons and networks. Phys. Rev., E 54, 5585.

Rodriguez, R., Tuckwell, H.C., 1998. Noisy spiking neurons and networks: useful approximations for firing probabilities and global behavior. Biosystems 48 (1–3), 187–194 (Sep–Dec).

Rodrigues, S., Terry, J.R., Breakspear, M., 2006. On the genesis of spikewave activity in a mean-field model of human thalamic and corticothalamic dynamics. Phys. Lett., A 355, 352–357.

Sporns, O., Tononi, G., Kotter, R., 2005. A structural description of the human brain. Plos Comp. Biol. 1 (4), e42.

Steyn-Ross, M.L., Steyn-Ross, D.A., Sleigh, J.W., Liley, D.T., 1999. Theoretical electroencephalogram stationary spectrum for a white-noise-driven cortex: evidence for a general anesthetic-induced phase transition. Phys. Rev. E. Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics 60, 7299–7311.

Valdes, P.A., Jimenez, J.C., Riera, J., Biscay, R., Ozaki, T., 1999. Nonlinear EEG analysis based on a neural-mass model. Biol. Cybern. 81, 415–424.

Wilson, H.R., Cowan, J.D., 1972. Excitatory and inhibitory interactions in localized populations of model neurons. Biophys. J. 12, 1–24.

Wolters, A., Schmidt, A., Schramm, A., Zeller, D., Naumann, M., Kunesch, E., Benecke, R., Reiners, K., Classen, J., 2005. Timing-dependent plasticity in human primary somatosensory cortex. J. Physiol. (Lond.) 565, 1039–1052.

Wright, J.J., Liley, D.T.J., 1996. Dynamics of the brain at global and microscopic scales: neural networks and the EEG. Behav. Brain Sci. 19, 285–320.