

Bilinear dynamical systems

W. Penny^{1,*}, Z. Ghahramani² and K. Friston¹

¹*Wellcome Department of Imaging Neuroscience, and* ²*Gatsby Computational Neuroscience Unit, University College, London WC1N 3BG, UK*

In this paper, we propose the use of bilinear dynamical systems (BDS)s for model-based deconvolution of fMRI time-series. The importance of this work lies in being able to deconvolve haemodynamic time-series, in an informed way, to disclose the underlying neuronal activity. Being able to estimate neuronal responses in a particular brain region is fundamental for many models of functional integration and connectivity in the brain. BDSs comprise a stochastic bilinear neurodynamical model specified in discrete time, and a set of linear convolution kernels for the haemodynamics. We derive an expectation-maximization (EM) algorithm for parameter estimation, in which fMRI time-series are deconvolved in an E-step and model parameters are updated in an M-Step. We report preliminary results that focus on the assumed stochastic nature of the neurodynamic model and compare the method to Wiener deconvolution.

Keywords: functional magnetic resonance imaging; deconvolution; connectivity; expectation-maximization; dynamical; embedding

1. INTRODUCTION

Imaging neuroscientists have at their disposal a variety of imaging techniques for investigating human brain function (Frackowiak *et al.* 2003). The electroencephalogram (EEG) records electrical activity from electrodes placed on the scalp, the magnetoencephalogram (MEG) records the magnetic field from sensors placed just above the head, and functional magnetic resonance imaging (fMRI) records changes in magnetic resonances owing to variations in blood oxygenation. However, as the goal of brain imaging is to obtain information about the neuronal networks that support perceptual inference and cognition, one must first transform measurements from imaging devices into estimates of intracerebral electrical activity. Brain imaging methodologists are therefore faced with an inverse problem, ‘how can one make inferences about intracerebral neuronal processes given extracerebral/vascular measurements?’

Though not often expressed in this terminology, we argue that this problem is best formulated as a model-based spatio-temporal deconvolution problem. For EEG and MEG, the required deconvolution is primarily spatial, and for fMRI it is primarily temporal. Although one can make minimal assumptions about the source signals by applying ‘blind’ deconvolution methods (McKeown *et al.* 1998; Makeig *et al.* 2002), knowledge of the underlying physical processes can be used to great effect. This information can be implemented in a forward model that is inverted during deconvolution. In EEG/MEG, forward models make use of Maxwell’s equations governing propagation of electromagnetic fields (Baillet *et al.* 2001), and in fMRI, forward models comprise haemodynamic

processes as described by ‘Balloon’ models (Buxton *et al.* 1998; Friston 2002; Riera *et al.* 2004; Stephan *et al.* 2004).

In this paper, we propose a new state-space method (Haykin 1996) for model-based deconvolution of fMRI. The importance of this work lies in being able to deconvolve haemodynamic time-series, in an informed way, to disclose the underlying neuronal activity. Such estimates are required by many models of functional integration and connectivity in the brain. Typically, the influence of an experimental manipulation on the coupling between two regions is tested using a statistical model of the interaction between the experimental factor and neuronal activity in the source region. These interaction terms rest on being able to deconvolve the fMRI time-series. The procedures described below enable the construction of precise and informed interaction terms that can then be used to detect context-sensitive coupling among brain regions. The interaction terms are variously known as psychophysiological interactions (Friston *et al.* 1997), moderator variables in structural equation models (Buchel & Friston 1997) and bilinear inputs in dynamic causal models (Friston *et al.* 2003).

The model we propose is called a bilinear dynamical system (BDS) and comprises the following elements. Experimental manipulation (input) causes changes in neuronal activation (state), which in turn cause changes in fMRI signal (output). Experimental inputs fall into two classes: (i) driving inputs, which directly excite neurodynamics and (ii) modulatory inputs, which change the neurodynamics. These modulatory inputs typically correspond to instructional or attention set, and therefore allow neurodynamic changes to be directly attributed to changes in experimental context. Readers of our earlier papers on dynamic causal modelling (DCM; Friston *et al.* 2003; Penny *et al.*

* Author for correspondence (wpenny@fil.ion.ucl.ac.uk).

One contribution of 21 to a Theme Issue ‘Multimodal neuroimaging of brain connectivity’.

2004) may well be experiencing a sense of ‘déjà vu’. This is no accident because BDS employs the same concepts of driving and modulatory inputs, neuronal states and outputs.

A key difference to DCM, however, is that BDS uses a stochastic neurodynamic model. This is biologically more realistic as regional dynamics are unlikely to be under full experimental control. Moreover, it has been established, across a variety of spatial scales, that neurodynamics can be meaningfully treated as stochastic processes, from Poisson processes describing spike generation (Dayan & Abbott 2001) to statistical mechanical treatments of activity in cortical macro-columns (Ingber 1995). Stochastic components could also be considered as the consequence of local or global deterministic nonlinear dynamics of a possibly chaotic nature (see e.g. Breakspear & Terry 2002). From another perspective, stochastic components could be considered as reflecting input from remote regions (David & Friston 2003).

In BDS, neurodynamics cause changes in fMRI signals via haemodynamic models specified in terms of a set of basis functions. Although more detailed differential equation models exist (Buxton *et al.* 1998; Friston 2002; Riera *et al.* 2004), the relationship between neuronal activation and fMRI signals can be formulated as a first-order convolution with a kernel expansion using basis functions (typically two or three). This kernel is the haemodynamic response function. This approach has enjoyed a decade of empirical success in the guise of the general linear model (GLM; Friston *et al.* 1995; Henson 2003). Moreover, this means that the state-output relation can be described by a linear convolution.

The BDS model may be viewed as a marriage of two formulations, the input-state relation being a stochastic DCM and the state-output relation being a GLM.

BDS has a similar form to models used in signal processing and machine learning. In particular, BDS is almost identical to a linear dynamical system (LDS; Roweis & Ghahramani 1999) with inputs. The only difference is that the inputs can change the linear dynamics in a bilinear fashion—hence the name ‘bilinear’ dynamical systems. In such models, the hidden variable, which in our case corresponds to the neuronal time-series, can be estimated using Kalman smoothing. Furthermore, the parameters of the model can be estimated using an expectation-maximization (EM) algorithm (described in Ghahramani & Hinton 1996), which needs only minor modification to accommodate the bilinear term.

The structure of the paper is as follows. In §2 we define BDS and describe how it can be applied to deconvolve fMRI time-series. We describe a novel EM algorithm for estimation of model parameters. In §3 we present preliminary results on applying the models to synthetic data and data from an fMRI experiment.

(a) Notation

We denote matrices and vectors with bold upper case and bold lower case letters, respectively. All vectors are column vectors. \mathbf{X}^T denotes the transpose of \mathbf{X} , and \mathbf{I}_k is the k -dimensional identity matrix. Brackets are used to denote entries in matrices/vectors. For example, $\mathbf{A}(i)$

denotes the i th row of \mathbf{A} , $\mathbf{A}(i, j)$ denotes the scalar entry in the i th row and j th column, and $\mathbf{a}(i)$ denotes the i th entry in \mathbf{a} . We define \mathbf{z}_N as a vector with a 1 as the first entry, followed by $N-1$ zeros and $\mathbf{0}_{I, \mathcal{J}}$ as the zero matrix of dimension $I \times \mathcal{J}$.

2. BILINEAR DYNAMICAL SYSTEMS

Although, in principle, we could define models for activity in a network of regions, in this paper we define a BDS model for activity in only a single region. BDS is an input-state-output model where the states correspond to ‘neuronal’ activations. Neuronal activity is defined mathematically below, and can be thought of as that component of the local field potential to which fMRI is most sensitive (Logothetis *et al.* 2001). BDS is defined, for a single region, with the following state-space equations where n indexes time

$$s_n = (a + \mathbf{b}^T \mathbf{u}_n) s_{n-1} + \mathbf{d}^T \mathbf{v}_n + w_n \quad (2.1)$$

$$\mathbf{x}_n = [s_n, s_{n-1}, s_{n-2}, \dots, s_{n-L+1}]^T \quad (2.2)$$

$$y_n = \beta^T \Phi \mathbf{x}_n + e_n \quad (2.3)$$

The first equation describes the stochastic neurodynamic model. Driving inputs \mathbf{v}_n cause an increase in neuronal activity s_n (a scalar) that decays with a time constant determined by $(a + \mathbf{b}^T \mathbf{u}_n)$. This time constant is determined by the value of the ‘intrinsic connection’, a , and ‘modulatory inputs’, \mathbf{u}_n . Driving inputs typically correspond to the presentation of experimental stimuli, and modulatory inputs typically correspond to instructional or attentional set. The strengths of the driving and modulatory effects are determined by the vectors of driving connections, \mathbf{d} , and modulatory connections, \mathbf{b} . The driving connections are also known as ‘input efficacies’. Neuronal activity is also governed by zero-mean additive Gaussian noise, w_n , having variance σ_w^2 .

Neuronal activation then gives rise to fMRI time-series according to the second and third equations. The second equation defines an embedding process (Weigend & Gershenfeld 1994) in which neuronal activity over the last L time-steps is placed in the ‘buffer’, \mathbf{x}_n . The vector \mathbf{x}_n is referred to as an ‘embedded’ time-series, where L is the embedding dimension. Neuronal activation, now described by \mathbf{x}_n , then gives rise to fMRI time-series, y_n , via a linear convolution process described in the third equation. Equation (2.3) comprises a signal term $\beta^T \Phi \mathbf{x}_n$ and a zero-mean Gaussian error term e_n with variance σ_e^2 . The h th row of matrix Φ defines the h th convolution kernel (i.e. basis of the haemodynamics response function) and $\beta(h)$ is the corresponding regression coefficient. Figure 1 shows a set of convolution kernels that have been used widely for the analysis of fMRI data in the context of the GLM. Linear combinations of these functions have been found, empirically, to account for most subject-to-subject and voxel-to-voxel variations in haemodynamic response (Henson *et al.* 2001; Henson 2003).

The model can perhaps be better understood by looking at the neuronal and haemodynamic time-series that it generates. An example is shown in figure 2. The model we have proposed, BDS, can be viewed as

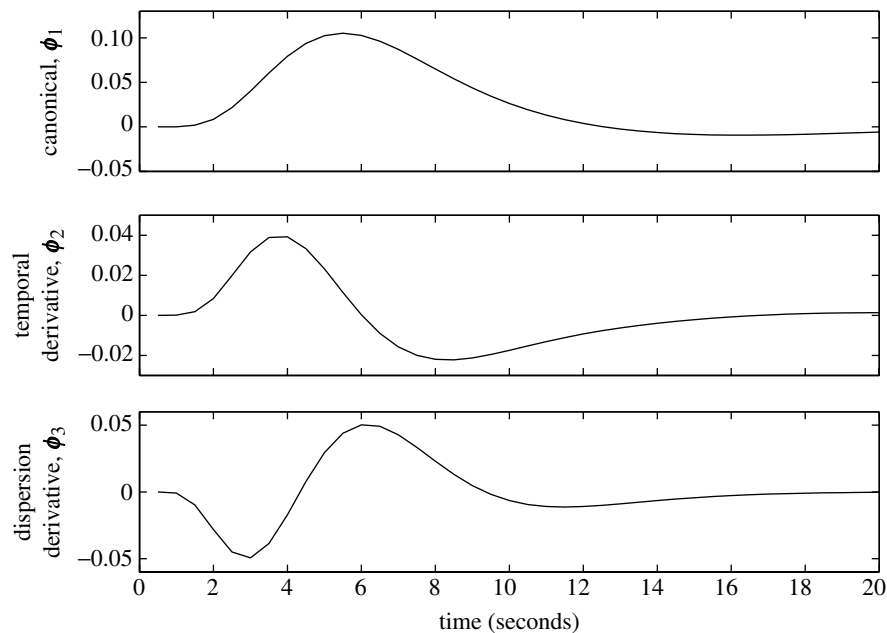


Figure 1. Haemodynamic basis functions. Top panel: canonical function; middle panel: derivative of canonical with respect to time; bottom panel: derivative of canonical with respect to dispersion. Linear combinations of these functions have been found empirically to account for most subject-to-subject and voxel-to-voxel variations in haemodynamic responses (Henson *et al.* 2001; Henson 2003).

a combination of two models: (i) a stochastic, discrete-time version of DCM to describe neurodynamics and (ii) the GLM to describe haemodynamics. GLMs are well established in the neuroimaging literature (Frackowiak *et al.* 2003), and DCMs are becoming so. Consequently, the model choices we have made are informed by precedents established for GLMs and DCMs. For example, from previous work using GLMs (Henson *et al.* 2001), we know that the embedding dimension should be chosen to span a 20–30 s period. With regard to neurodynamics, the discrete bilinear form affords analytic tractability, while allowing non-linear interactions between exogenous input and states. The splitting of exogenous inputs into driving and modulatory terms is also prompted by DCM. The motivation for this is that, for data collected from controlled experiments, one wishes to relate changes in connectivity to experimental manipulation. This distinction is echoed by the separate neurobiological mechanisms underlying driving versus modulatory activity (Penny *et al.* 2004).

The modulatory coefficients in equation (2.1), \mathbf{b} , are also known as ‘bilinear’ terms. This is because the dependent variable, s_n , is linearly dependent on the product of s_{n-1} and \mathbf{u}_n . That \mathbf{u}_n and s_{n-1} combine in multiplicative fashion endows the model with ‘non-linear’ dynamics that can be understood as a non-stationary linear system that changes according to experimental context. Importantly, because \mathbf{u}_n is known, it is straightforward to estimate how the dynamics are changing. We emphasize that the term ‘bilinear’ relates to interactions between a state and an input rather than among the states themselves.

(a) Relation to biophysical models

In the previous section, we interpreted $(a + \mathbf{b}^T \mathbf{u}_n)$ as the time constant of decaying neuronal activity. However, our BDS model has a much more general relationship

to underlying biophysical models of neuronal dynamics. Consider some biologically plausible model of neuronal activity that is formulated in continuous time $\dot{x} = f(x, \mathbf{u}, \theta)$ and allows for arbitrarily complicated and nonlinear effects of the state and exogenous inputs. The parameters of this biophysical model θ (cf. neural mass models; Valdes *et al.* 1999) can be related directly to biophysical processes. In our discrete-time formulation, under local linearity assumptions, $a = \exp(\Delta t \partial f / \partial x)$ and $\mathbf{b}(i) = \partial a / \partial \mathbf{u}(i) = \Delta t a \partial^2 f / \partial x \partial \mathbf{u}(i)$, where Δt is the sampling interval and i indexes the input. This means that our input-dependent autoregression coefficients can be regarded as a lumped representation of the underlying model parameters. This re-parametrization, in terms of a state-space model, precludes an explicit estimation of the original parameters. However, this does not cause concern, because we are not interested in the parameters *per se*; we only require the conditional estimates of the states. The transformation of a dynamic formulation into a discrete BDS is a useful perspective, because priors on the biophysical parameters can be used to specify priors on the autoregression coefficients, should they be needed.

(b) Relation to GLMs

For models with a single input, BDS reduces to a GLM if $s_n = v_n$, that is, if the neuronal activity is synonymous with the experimental manipulation (driving inputs). This highlights a key assumption of GLMs, that the dynamics of neuronal transients can be ignored.

For models with multiple inputs, the relation between BDS and GLMs is more complex. Ignoring bilinear and noise terms, BDS can be more parsimonious than the GLM, in the sense of having fewer parameters. Say, for example, we are modelling haemodynamics with H basis functions. Then for M (driving) input variables, the GLM has HM

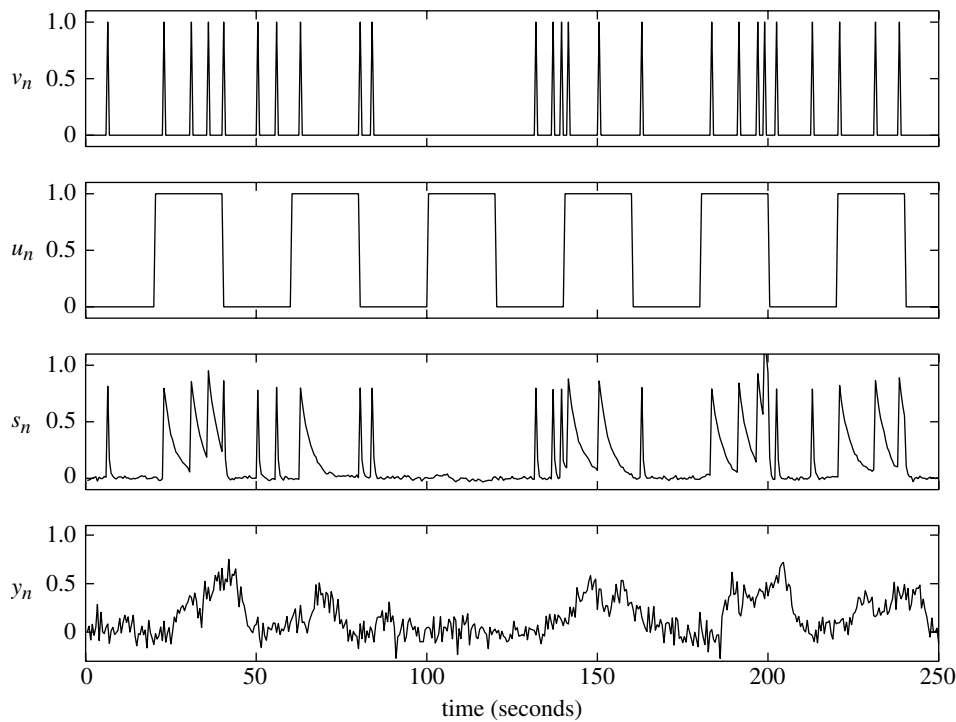


Figure 2. Time-series from exemplar BDS model. Driving input v_n (top panel) causes an increase in neuronal activity s_n (third panel). This activity decays with a time constant that changes according to a modulatory input u_n (second panel). In this instance, the modulatory coefficient, b , is negative. Driving inputs typically correspond to the presentation of experimental stimuli and modulatory inputs typically correspond to instructional or attentional set. Neuronal activity then gives rise to the fMRI time-series y_n (bottom panel) by convolving neuronal activity with a set of haemodynamic kernels (shown in figure 1).

parameters, whereas in BDS, there are $M+H+1$ parameters. For $H=3$, for example, BDS has fewer coefficients if M is 3 or more.

The reason for this reduction is that, in BDS, the inputs affect only neurodynamics. And the strength of this effect, ignoring modulatory terms, is captured in a single parameter, the input efficacy. The relation between neurodynamics and haemodynamics is independent of which input excited neuronal activity. This is obviously more consistent with physiology, where experimental manipulation only affects fMRI by changing neuronal activity.

(c) Deconvolution

Deconvolution consists of estimating a neuronal time-series, s_n , given an fMRI time-series, y_n . It is possible to perform model-based deconvolution using BDS and a modified Kalman-filtering algorithm. The modification is necessary because, owing to the modulatory terms, the state transition matrix is time dependent. Also, the BDS model must be reformulated so that the output is an instantaneous function of the state. This section describes standard Kalman filtering and the modifications required for BDS.

In the Kalman-filtering approach, deconvolution progresses via iterative computation of the probability density $p(s_n | \mathbf{y}_1^n)$, where $\mathbf{y}_1^n = \{y_1, y_2, \dots, y_n\}$. That is, our estimate of neuronal activity at time-step n is based on all the fMRI data up to time n (but not on future values).

The Kalman-filtering algorithm can be split into two steps that are applied recursively. The first step is a time update, where the density $p(s_{n-1} | \mathbf{y}_1^{n-1})$ is updated to

$p(s_n | \mathbf{y}_1^{n-1})$. In our model this will take into account the effects of inputs u_n and v_n and the natural decay of neuronal activity. The second step is a *measurement update*, where the density $p(s_n | \mathbf{y}_1^{n-1})$ is updated to $p(s_n | \mathbf{y}_1^n)$, thereby taking into account the 'new' fMRI measurement y_n . The two steps together take us from time point $n-1$ to time point n , and are applied recursively to deconvolve the entire time-series.

A complication arises, however, because the Kalman filtering updates require that the output be an instantaneous function of the hidden state. That is, that y_n depend only on s_n and not on s_{n-1} , s_{n-2} , etc., which is clearly not the case. But Kalman filtering can proceed if we use the embedded state variables \mathbf{x}_n . This will lead to an estimate of the multivariate density $p(\mathbf{x}_n | \mathbf{y}_1^n)$. Our desired density, $p(s_n | \mathbf{y}_1^n)$, is then just the first entry in $p(\mathbf{x}_n | \mathbf{y}_1^n)$. While this is conceptually simple, a good deal of book-keeping is required to translate the BDS neurodynamical model into an 'embedded space'. This is described in detail in appendices A and B.

A second complication arises in that, in the standard Kalman update formulae (Ghahramani & Hinton 1996), the inputs do not change the intrinsic dynamics. To account for this, we introduce a time-dependent state-transition matrix that embodies changes in intrinsic dynamics owing to modulatory inputs. This is also detailed in appendices A and B.

It is also possible to perform deconvolution based on Kalman smoothing rather than Kalman filtering. Kalman smoothing estimates neuronal activity, by computing the density $p(s_n | \mathbf{y}_1^N)$. It is therefore based on the whole fMRI record, i.e. past *and* future values

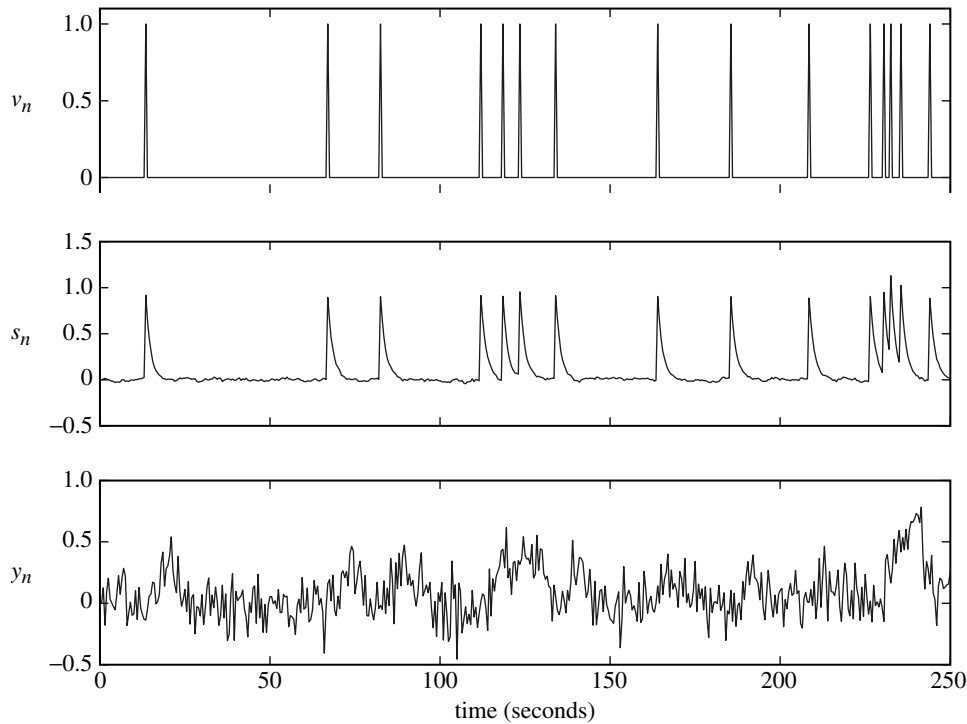


Figure 3. Low neuronal noise data. Top panel: driving input v_n ; middle panel: neuronal activity s_n ; bottom panel: fMRI data y_n .

and, as we shall see, provides more accurate deconvolutions in high-noise environments.

(d) Estimation

To run the deconvolution algorithms described in the previous section, we must know the parameters of the model, e.g. β , a , \mathbf{b} and \mathbf{d} . If these parameters are unknown, as will nearly always be the case, they can be estimated using maximum-likelihood methods.

For BDS, this can be implemented using an EM algorithm, which we derive in appendices A and B. The E-step uses a Kalman-smoothing algorithm, which contains only minor modifications to the algorithm presented in Ghahramani & Hinton (1996). The M-step contains updates derived by considering estimation of the neurodynamic parameters (a , \mathbf{b} , \mathbf{d}) as a constrained linear regression problem (Rao & Toutenberg 1995). The neurodynamic parameters are updated as shown in equation (B 14), and the haemodynamic parameters, β , are updated using equation (B 17). It is also possible to update the initial state estimates as described in Ghahramani & Hinton (1996), but this was not implemented for the empirical work in this paper.

It is also possible to implement maximum-likelihood parameter estimation by computing the likelihood using Kalman filtering (as shown in Appendix C), and then maximizing this using standard Matlab optimization algorithms such as pseudo-Newton or Simplex methods (Press *et al.* 1992). However, preliminary simulations showed EM to be faster than these other maximum-likelihood methods.

(e) Initialization

The EM estimation algorithm is initialized as follows. In the limit of zero neuronal noise (ZNN), the neuronal

activity is given by

$$s_n = (a + \mathbf{b}^T \mathbf{u}_n) s_{n-1} + \mathbf{d}^T \mathbf{v}_n, \quad (2.4)$$

and the predicted haemodynamic activity is

$$\hat{y}_n = \beta^T \Phi \mathbf{x}_n. \quad (2.5)$$

We initially set a random number between 0 and 1, and \mathbf{b} such that $\mathbf{b}^T \mathbf{u}_n$ is between 0 and $1 - a$ for all n . The values of \mathbf{d} are also set to random numbers between 0 and 1. The first β coefficient is fixed at unity, and others, if there are any, are initialized to random numbers between 0 and 1. We then run a pseudo-Newton gradient descent optimization (the function `fminunc` in Matlab) to minimize the discrepancy between the observed fMRI time-series y_n and the estimated values \hat{y}_n under the ZNN assumption. Occasionally, parameter estimates result in an unstable model. We therefore repeat these initialization steps until a stable model is returned. These parameters are then used as a starting point for EM.

3. RESULTS

(a) Simulations

We generate simulated data to demonstrate various properties of the deconvolution algorithm. These simulations are similar to the first set of simulations in Gitelman *et al.* (2003). We also compare our results to those obtained with Wiener deconvolution (Glover 1999), which makes minimal assumptions about the source (that it has a flat spectral density; Press *et al.* 1992; Gitelman *et al.* 2003).

A 250 s time-series of input events with sampling period $\Delta t = 0.5$ s was generated from a Poisson distribution, with a mean interval between events of 12 s. Events less than 2 s apart were then removed. The half-life of neuronal transients was fixed to 1 s by setting $a = 0.71$, the input efficacy was set to $d = 0.9$

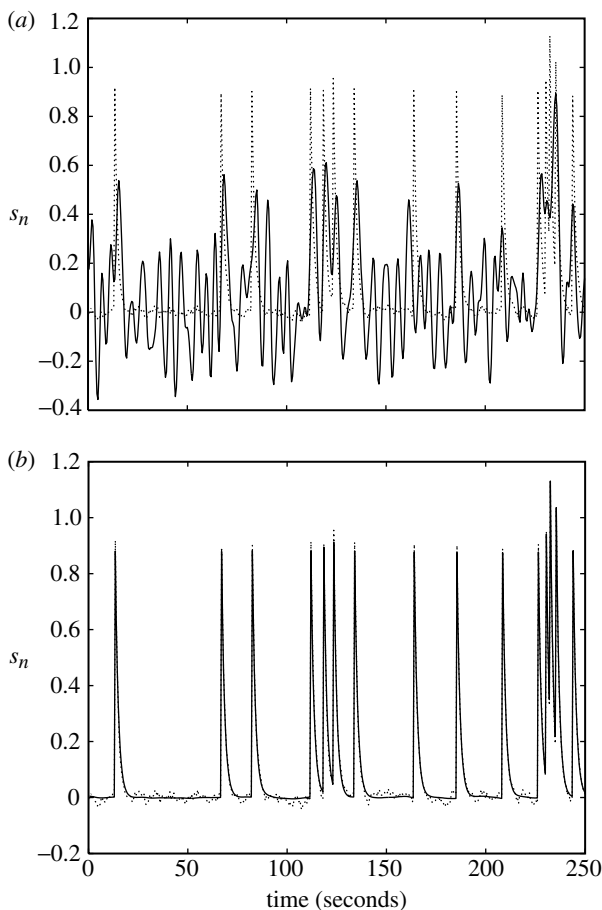


Figure 4. Deconvolving low neuronal noise data. (a) Wiener deconvolution (solid line), (b) BDS Kalman-smoothing deconvolution (solid line) using estimated parameters $a=0.72$ and $d=0.88$. The simulated (true) neuronal activity is shown as a dotted line in both plots.

and the neuronal noise variance was set to $\sigma_w^2 = 0.0001$. There were no modulatory coefficients, and a single haemodynamic basis function (the ‘canonical’, shown in figure 1) was used.

We then added observation noise to achieve a signal-to-noise ratio (SNR) of unity (this is equivalent to the 100% noise case in Gitelman *et al.* 2003). This was achieved using $\sigma_e^2 = 0.015$. The generated time-series is shown in figure 3.

The parameters were estimated using EM to be $a=0.72$ and $d=0.88$. Wiener and BDS Kalman-smoothing deconvolutions (the latter obtained using estimated parameters) are shown in figure 4. Wiener deconvolution used the known value of σ_e^2 , and BDS Kalman smoothing used the known values of σ_e^2 and σ_w^2 . The quality of the BDS deconvolution is considerably better than that using the method described in Gitelman *et al.* (2003); see, for example, fig. 3D in Gitelman *et al.* (2003). This is because information about the paradigm (e.g. experimental inputs) has been used. For Wiener deconvolution, the correlation with the generated neuronal time-series is $r=0.520$ and for BDS Kalman smoothing it is $r=0.998$.

We then ran a second simulation with the neuronal noise variance set to a high value, $\sigma_w^2 = 0.03$. These data are shown in the top panel of figure 5. Neurodynamic parameters were estimated using EM to be $a=0.68$ and $d=0.83$. For comparison, ZNN initialization (see §2e)

produced parameter estimates $a=0.45$ and $d=1.13$. That is, the assumption of ZNN leads to an underestimate of the neuronal time constant and an overestimate of the input efficacy.

The neuronal time-series, as estimated using BDS Kalman smoothing (bottom panel of figure 5), had a correlation with the generated time-series of $r=0.775$. The ZNN estimate neuronal time-series had a correlation of $r=0.724$. The Wiener deconvolved time-series (second panel in figure 5) gave $r=0.52$ (again). We also show, in the third panel of figure 5, deconvolutions from BDS Kalman filtering.

Figure 5 shows that Wiener estimation recovers the intrinsic dynamics but misses the evoked responses, whereas BDS Kalman filtering recovers the evoked responses but misses the intrinsic dynamics. Deconvolution using BDS Kalman smoothing recovers both.

The smoother can capture the intrinsic dynamics because it uses more information than the filter. It updates the filter estimates of neuronal activity, at a given time point, using information in advance of that time point (i.e. from the future). These updates are implemented using the ‘backward recursions’ described in Appendix B(a). Heuristically, the reason for the improvement is that the best estimates of neuronal activity are obtained using observed fMRI activity about 5 s or so in the future, i.e. at the peak of the haemodynamic response.

(b) *Single word processing fMRI*

We now turn to the analysis of an fMRI dataset recorded during a passive listening task, using epochs of single words presented at different rates. The experimental inputs in figure 6 describe the paradigm in more detail. The driving inputs in the top panel indicate presentation of words in the subject’s headphones and modulatory inputs in the lower panels indicate epochs with different presentation rates.

We focus on a single time-series in primary auditory cortex shown in figure 7. This comprises 120 time points with a sampling period $\Delta t=1.7$ s. Further details of data collection are given in Friston *et al.* (2003). As the experimental inputs are specified at a higher temporal resolution than the fMRI acquisition, we upsampled the fMRI data by a factor of 4 prior to analysis. The input variables were convolved with a ‘canonical’ haemodynamic response function (see top panel of figure 1) to form regressors in a GLM. Figure 7 shows the resulting GLM model fit.

The same input variables and haemodynamic basis function were then used to define a BDS model. The observation and state noise variances were set to $\sigma_e^2 = \sigma_w^2 = 0.1$, and the haemodynamic regression coefficient was set to $\beta=1$. Figure 8 shows the resulting model fit which is clearly superior to the GLM model fit in figure 7 (but see §4). Figure 9 shows neuronal activity as estimated using BDS Kalman smoothing. This includes both event-related responses (the ‘spikes’) and intrinsic activity (slow fluctuations).

Neurodynamic parameters were estimated using EM as $d=0.80$, $a=0.92$, $\mathbf{b}(1)=-0.44$, $\mathbf{b}(2)=-0.08$ and $\mathbf{b}(3)=-0.02$. Thus, epochs with faster stimulus presentations were estimated to have an increasingly inhibitory effect on event-related neurodynamics.

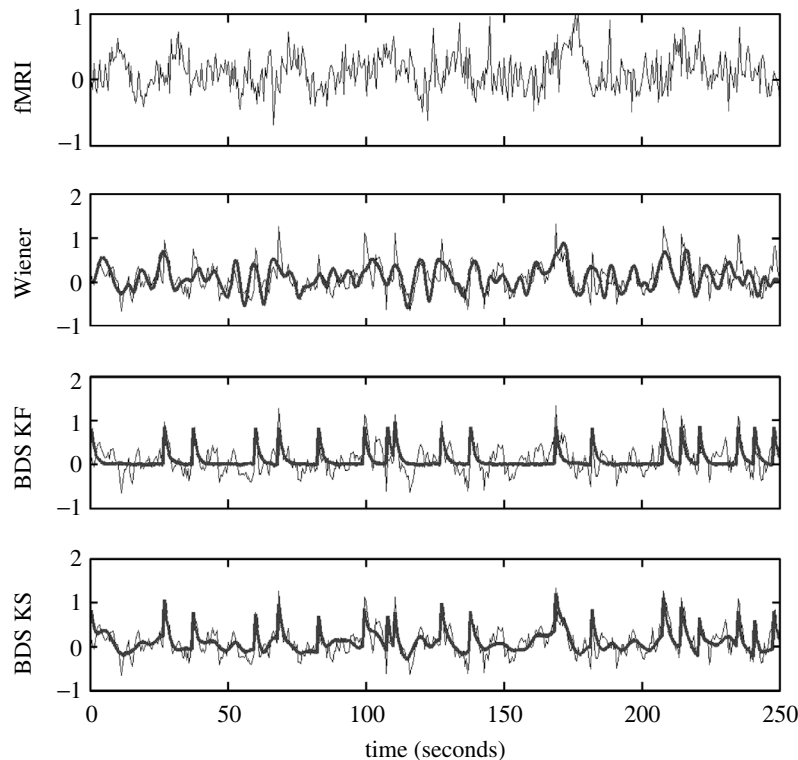


Figure 5. Deconvolving high neuronal noise data. Top (first) panel: fMRI data; second panel: Wiener deconvolution (thick line); third panel: BDS Kalman filtering (thick line); bottom (fourth) panel: BDS Kalman smoothing (thick line). In the bottom three panels, the original simulated neuronal activity is shown as a thin line. Wiener estimation recovers the intrinsic dynamics but misses the evoked responses (the ‘spikes’), whereas BDS Kalman filtering recovers the evoked responses but misses the intrinsic dynamics (slow fluctuations). Deconvolution using BDS Kalman smoothing recovers both.

This pattern was robust across a wide range of settings of the noise variance parameters. This finding is consistent with neuronal repetition suppression effects and is in agreement with DCM of these data (Friston *et al.* 2003).

4. DISCUSSION

We have proposed a new algorithm, based on a BDS, for model-based deconvolution of fMRI time-series. The importance of the work is that haemodynamic time-series can be deconvolved, in an informed way, to disclose the underlying neuronal activity. Being able to estimate neuronal responses in a particular brain region is fundamental for many models of functional integration and connectivity in the brain. Of course, these estimates can only describe those components of neuronal activity to which fMRI is sensitive. They should, nevertheless, be useful for making fMRI-based assessments of connectivity (Friston *et al.* 1997; Gitelman *et al.* 2003).

BDS is fitted to data using a novel EM algorithm where deconvolution is instantiated in an E-step and model parameters are updated in an M-step. Simulations showed EM to be faster than maximum-likelihood optimization based on simplex or pseudo-Newton methods. Deconvolution can be based either on a full E-step, using Kalman smoothing, or a partial E-step based on Kalman filtering. Kalman smoothing uses the full data record whereas Kalman filtering only uses information from the past.

Simulations showed that our model-based deconvolution is more accurate than blind deconvolution

methods (Wiener filtering). This is because BDS uses information about the paradigm. We also observed the following trends. Wiener estimation recovers the intrinsic dynamics but misses the evoked responses, whereas BDS Kalman filtering recovers the evoked responses but misses the intrinsic dynamics. Deconvolution using BDS Kalman smoothing recovers both.

Simulations also suggest that if dynamics are indeed of a stochastic nature, as is assumed in BDS, then if we mistakenly assume deterministic dynamics, estimation of neuronal efficacies and time constants will become inaccurate. This has implications for models that assume deterministic dynamics, such as DCM (Friston *et al.* 2003).

Our applications of BDS provide good examples of Kalman smoothing providing better deconvolutions than Kalman filtering. The reason is that smoothing also uses future observations and the best estimates of neuronal activity are obtained using observed fMRI activity about 5 s or so into the future, that is, at the peak of the haemodynamic response. This property should hold for any state-space model of fMRI (see, for example, Riera *et al.* 2004).

A comparison of GLM and BDS model fits in figures 7 and 8 clearly shows that BDS is superior. Although this is somewhat encouraging, this observation should be tempered with a note of caution. This is because the BDS model is more complex and no penalty was paid for this during model fitting. The BDS model may therefore be overfitted. In particular, as fMRI time-series are known to contain aliased cardiac and respiratory artefacts, the finer details that

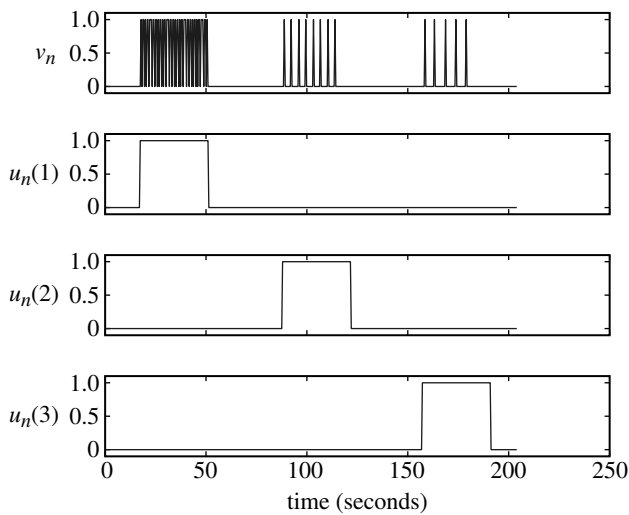


Figure 6. Word fMRI inputs. Driving input v_n (top panel) and modulatory inputs u_n (lower panels). The delta functions in the top panel indicate presentation of words in the subject's headphones and the modulatory inputs in the lower panels indicate epochs with different presentation rates.

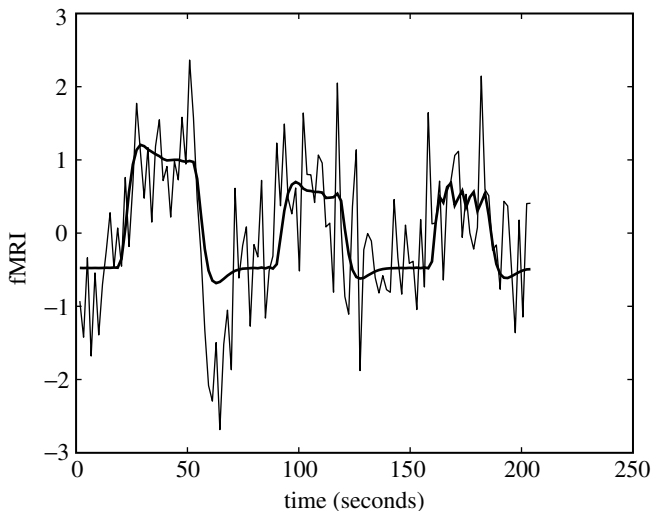


Figure 7. GLM fit to word fMRI data. fMRI time-series from primary auditory cortex (thin line) and GLM model fit (thick line).

BDS picks up may be of artefactual rather than neuronal origin. To eliminate this possibility one would need to estimate model parameters (especially the state noise variance) using Bayesian or cross-validation (Yamashita *et al.* 2004) methods. Fortunately, Bayesian methods have already been developed for linear dynamical systems (Ghahramani & Beal 2001) that should not require too many changes to accommodate the bilinear term.

Recently, Riera *et al.* (2004) have proposed a state-space model for fMRI time-series which allows for nonlinear state-output relations. They showed how model parameters could be estimated using a maximum-likelihood procedure based on extended Kalman filtering. Moreover, they showed how the parameters could be related to a stochastic differential equation implementation of the Balloon model. Thus, unlike BDS which assumes linear haemodynamics, their model can describe nonlinear properties such as

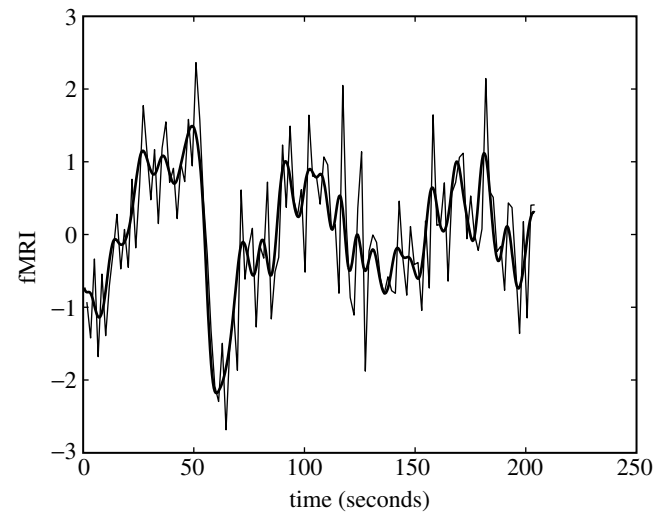


Figure 8. BDS fit to word fMRI data. fMRI time-series from primary auditory cortex (thin line) and BDS model fit (thick line).

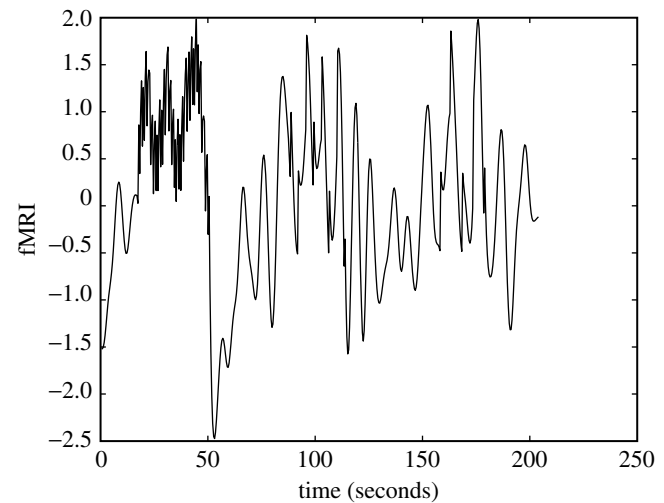


Figure 9. BDS deconvolution of word fMRI data estimation of neuronal activity using BDS Kalman smoothing. Event-related responses can be seen as spikes superimposed on intrinsic dynamics.

haemodynamic refractoriness. This is important, as the neuronal refractoriness inferred using BDS in §3b, for example, may actually be of haemodynamic rather than neuronal origin. In fact, a DCM analysis of these data (Friston *et al.* 2003), which allows for both neuronal and haemodynamic refractoriness, concluded that both effects were present.

The model proposed by Riera *et al.* (2004) also differs from BDS in the input-state relations. In BDS, this is governed by driving inputs that excite linear neurodynamics, which can be changed according to modulatory inputs. This means that neurodynamics and changes in them can be directly attributed to changes in experimental context. This modelling approach is therefore appropriate for designed experiments where the inputs are known. In the Riera model, the inputs are not assumed to be known. Instead, a more flexible radial basis function approach is used to make inferences about the onsets of linear neurodynamical processes.

In previous work we have proposed an fMRI deconvolution model based on the GLM

(Gitelman *et al.* 2003). This uses a forward model in which neuronal activity, represented using a temporal basis set and corresponding coefficients, is convolved with a known haemodynamic kernel to produce an fMRI time-series. Observation noise is then added. Deconvolution is achieved by estimating the coefficients of the temporal basis functions. In Gitelman *et al.* (2003), we used full-rank Fourier basis sets and overcomplete basis sets that used information about experimental design. Coefficients were estimated using priors and parametric empirical bayes. A problem with the method, however, is that for \mathcal{J} coefficients (where \mathcal{J} is typically the length of the time-series or longer) one must store and invert $\mathcal{J} \times \mathcal{J}$ covariance matrices. A computational benefit of the BDS approach is that, by making use of factorizations that derive from Markov properties of the state-space model, one need only manipulate $L \times L$ covariance matrices, where L is the temporal embedding dimension ($L \ll \mathcal{J}$). We have also argued, in §3b, that BDS is more parsimonious and biologically informed than the GLM.

The method we have proposed has been applied to deconvolve data at a single voxel. This is useful in providing a ‘source’ neuronal time-series, for example, for the analysis of psycho-physiological interaction (Friston *et al.* 1997; Gitelman *et al.* 2003). More generally, however, one is interested in making inferences about changes in connectivity in neural networks extended over multiple regions. This requires simultaneous deconvolution at multiple voxels and, ideally, a full model-based spatio-temporal deconvolution. The application of state-space models to this more difficult problem is an exciting area of current research (see, for example, Galka *et al.* in press).

Will Penny and Karl Friston are funded by the Wellcome Trust.

APPENDIX A: EMBEDDING

It is convenient to re-write the neurodynamic model as

$$s_n = \tilde{\mathbf{b}}^T \tilde{\mathbf{u}}_n s_{n-1} + \mathbf{d}^T \mathbf{v}_n + w_n, \quad (\text{A } 1)$$

and $\tilde{\mathbf{u}}_n$ is a vector of modulatory inputs augmented with a 1 as the first entry (note some inputs may be driving and modulatory, in which case they appear in both $\tilde{\mathbf{u}}_n$ and \mathbf{v}_n) and $\tilde{\mathbf{b}}^T = [a, \mathbf{b}]^T$ absorbs a .

The state-space equations can then be written in terms of the embedded neuronal activity, \mathbf{x}_n , as

$$\left. \begin{aligned} \mathbf{x}_n &= \mathbf{A}_n \mathbf{x}_{n-1} + \mathbf{D} \mathbf{v}_n + \tilde{\mathbf{w}}_n \\ y_n &= \mathbf{c}^T \mathbf{x}_n + e_n, \end{aligned} \right\} \quad (\text{A } 2)$$

where $\mathbf{c}^T = \beta^T \Phi$. The embedded state transition matrix is

$$\mathbf{A}_n = \begin{bmatrix} \tilde{\mathbf{b}}^T \mathbf{F}_n \\ \Psi_L \end{bmatrix}, \quad (\text{A } 3)$$

where

$$\mathbf{F}_n = \tilde{\mathbf{u}}_n \mathbf{z}_L^T \quad (\text{A } 4)$$

is a $(B+1) \times L$ matrix (where B is the number of modulatory inputs) and Ψ_L is the $(L-1) \times L$ delay matrix that fills the lower $L-1$ rows of \mathbf{A}_n . This ensures

that the embedded time-series are shifted one time-step each time \mathbf{A}_n is applied. The input matrix is given by

$$\mathbf{D} = \begin{bmatrix} \mathbf{d}^T \\ \mathbf{0}_{L-1, M} \end{bmatrix}, \quad (\text{A } 5)$$

where M is the number of driving inputs. The state noise in equation (A 2) is given by $\tilde{\mathbf{w}}_n = w_n \mathbf{z}_L^T$. The covariance of $\tilde{\mathbf{w}}_n$ is \mathbf{Q} and the only non-zero entry is $\mathbf{Q}(1, 1) = \sigma_w^2$.

If we have a BDS with a single driving input, no modulatory inputs and $L=4$, $a=0.92$ and $d=0.80$, then the embedded neurodynamic model is

$$\mathbf{x}_n = \mathbf{A}_n \mathbf{x}_{n-1} + \mathbf{D} \mathbf{v}_n + \mathbf{w}_n \quad (\text{A } 6)$$

$$\begin{bmatrix} s_n \\ s_{n-1} \\ s_{n-2} \\ s_{n-3} \end{bmatrix} = \begin{bmatrix} 0.92 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} s_{n-1} \\ s_{n-2} \\ s_{n-3} \\ s_{n-4} \end{bmatrix} + \begin{bmatrix} 0.80 \\ 0 \\ 0 \\ 0 \end{bmatrix} \mathbf{v}_n + \begin{bmatrix} w_n \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Modulatory inputs would change the first entry in \mathbf{A}_n .

APPENDIX B: EM ALGORITHM

In probabilistic models with hidden variables, maximum-likelihood learning can be implemented using an EM algorithm (Dempster *et al.* 1977). This requires us to maximize the auxiliary function

$$F = \int p(\mathbf{X} | \mathbf{y}) \log p(\mathbf{X}, \mathbf{y}) d\mathbf{X}, \quad (\text{B } 1)$$

where $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n, \dots, \mathbf{x}_N\}$ are the hidden variables and $\mathbf{y} = \{y_1, \dots, y_n, \dots, y_N\}$ are the observed variables (Ghahramani & Hinton 1996). In BDS, the observed variables are the fMRI time-series and the hidden variables are the neuronal activities. Using the Markov property, we can write

$$p(\mathbf{X}, \mathbf{y}) = p(\mathbf{x}_1) \prod_{n=2}^N p(\mathbf{x}_n | \mathbf{x}_{n-1}) \prod_{n=1}^N p(y_n | \mathbf{x}_n). \quad (\text{B } 2)$$

The initial, transition and output probabilities are given by

$$p(\mathbf{x}_1) = N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1), \quad (\text{B } 3)$$

$$p(\mathbf{x}_n | \mathbf{x}_{n-1}) = N(\mathbf{A}_n \mathbf{x}_{n-1} + \mathbf{D} \mathbf{v}_n, \mathbf{Q}), \quad (\text{B } 4)$$

$$p(y_n | \mathbf{x}_n) = N(\mathbf{c}^T \mathbf{x}_n, \sigma_e^2), \quad (\text{B } 5)$$

which define the observation model, state transition model and initial state distribution. In the above expression, the quantity \mathbf{c} is as defined in equation (A 2). Therefore, the joint log-likelihood is a sum of

quadratic terms

$$\begin{aligned}
 L_{\mathcal{Y}} &= \log p(\mathbf{X}, \mathbf{y}) \\
 &= - \sum_{n=1}^N \frac{1}{2\sigma_e^2} \left[(\mathbf{y}_n - \mathbf{c}^T \mathbf{x}_n)^T (\mathbf{y}_n - \mathbf{c}^T \mathbf{x}_n) \right] - \frac{N}{2} \log |\sigma_e^2| \\
 &\quad - \sum_{n=2}^N \frac{1}{2} \left[(\mathbf{x}_n - \mathbf{A}_n \mathbf{x}_{n-1} - \mathbf{D} \mathbf{v}_n)^T \right. \\
 &\quad \left. \times \mathbf{Q}^{-1} (\mathbf{x}_n - \mathbf{A}_n \mathbf{x}_{n-1} - \mathbf{D} \mathbf{v}_n) \right] \\
 &\quad - \frac{1}{2} \left[(\mathbf{x}_1 - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}_1^{-1} (\mathbf{x}_1 - \boldsymbol{\mu}_1) \right] - \frac{1}{2} \log |\boldsymbol{\Sigma}_1| \\
 &\quad - \frac{N-1}{2} \log |\mathbf{Q}| - \frac{NL}{2} \log 2\pi.
 \end{aligned}$$

In the above equation, the quantity \mathbf{Q} refers to an arbitrary covariance matrix. In appendix B(b) we show how $L_{\mathcal{Y}}$, as a function of the state variables \mathbf{x}_n , simplifies for the \mathbf{Q} defined for the BDS model in appendix A. The expectation over the terms in the above equation can be maximized as shown in the following sections.

(a) E-step

The objective of the E-step is to compute the probability of the hidden variables given the data. Because the initial, transition and output probability distributions are Gaussian, this can be achieved by updating the conditional mean and conditional covariance. Following Ghahramani & Hinton (1996), we write the expected value of \mathbf{x}_n conditioned on all data up to time n as $\mathbf{x}_n^n \equiv E[\mathbf{x}_n | \mathbf{y}_1^n]$. Similarly, the corresponding covariance is given by $\boldsymbol{\Sigma}_n^n \equiv \text{Var}[\mathbf{x}_n | \mathbf{y}_1^n]$.

(i) Kalman filtering

The objective of Kalman filtering is to compute the probability of the hidden variables given all observed variables up to that time point, that is, to compute $p(\mathbf{x}_n | \mathbf{y}_1^n)$. Kalman filtering implements the recursive computation of \mathbf{x}_n^n and $\boldsymbol{\Sigma}_n^n$ from \mathbf{x}_{n-1}^{n-1} and $\boldsymbol{\Sigma}_{n-1}^{n-1}$ in two steps. Firstly, in the time update step

$$p(\mathbf{x}_n | \mathbf{y}_1^{n-1}) = \int p(\mathbf{x}_n | \mathbf{x}_{n-1}) p(\mathbf{x}_{n-1} | \mathbf{y}_1^{n-1}) d\mathbf{x}_{n-1}. \tag{B 7}$$

This is implemented using

$$\left. \begin{aligned}
 \mathbf{x}_n^{n-1} &= \mathbf{A}_n \mathbf{x}_{n-1}^{n-1} + \mathbf{D} \mathbf{v}_n \\
 \boldsymbol{\Sigma}_n^{n-1} &= \mathbf{A}_n \boldsymbol{\Sigma}_{n-1}^{n-1} \mathbf{A}_n^T + \mathbf{Q}.
 \end{aligned} \right\} \tag{B 8}$$

Secondly, in the measurement update step

$$p(\mathbf{x}_n | \mathbf{y}_1^n) = \frac{p(\mathbf{y}_n | \mathbf{x}_n) p(\mathbf{x}_n | \mathbf{y}_1^{n-1})}{\int p(\mathbf{y}_n | \mathbf{x}_n) p(\mathbf{x}_n | \mathbf{y}_1^{n-1}) d\mathbf{x}_n}. \tag{B 9}$$

This is simply Bayes rule where $p(\mathbf{x}_n | \mathbf{y}_1^{n-1})$ (from equation (B 7)) describes our belief in \mathbf{x}_n before observing y_n . The measurement update is implemented using

$$\left. \begin{aligned}
 \mathbf{K}_n &= \boldsymbol{\Sigma}_n^{n-1} \mathbf{c} (\mathbf{c}^T \boldsymbol{\Sigma}_n^{n-1} \mathbf{c} + \sigma_e^2)^{-1} \\
 \mathbf{x}_n^n &= \mathbf{x}_n^{n-1} + \mathbf{K}_n (\mathbf{y}_n - \mathbf{c}^T \mathbf{x}_n^{n-1}) \\
 \boldsymbol{\Sigma}_n^n &= \boldsymbol{\Sigma}_n^{n-1} - \mathbf{K}_n \mathbf{c}^T \boldsymbol{\Sigma}_n^{n-1},
 \end{aligned} \right\} \tag{B 10}$$

where \mathbf{K}_n , known as the Kalman gain matrix, operates as an adaptive step size parameter for each hidden variable.

In the above expressions, the quantity \mathbf{c} is as defined in equation (A 2). The procedure is initialized using $\mathbf{x}_1^0 = \boldsymbol{\mu}_1$ and $\boldsymbol{\Sigma}_1^0 = \boldsymbol{\Sigma}_1$. These updates are exactly as described in Ghahramani & Hinton (1996), except that (i) \mathbf{A}_n is used instead of \mathbf{A} because our dynamics are input dependent and (ii) we use \mathbf{c} instead of \mathbf{C} because our observations are univariate.

(ii) Kalman smoothing

The objective of Kalman smoothing is to compute the probability of the hidden variables given all observed variables, that is, to compute $p(\mathbf{x}_n | \mathbf{y}_1^N)$. They are implemented using a set of ‘backward recursions’ which compute \mathbf{x}_{n-1}^N and $\boldsymbol{\Sigma}_{n-1}^N$ from the forward estimates \mathbf{x}_{n-1}^{n-1} , $\boldsymbol{\Sigma}_{n-1}^{n-1}$. Because these formulae are also almost identical to those described in Ghahramani & Hinton (1996), we do not reproduce them here. The only difference is that \mathbf{A}_n is used instead of \mathbf{A} .

(iii) Expectations

The M-step requires a number of expectations that can be derived from the E-step

$$\left. \begin{aligned}
 \mathbf{m}_n &= \int p(\mathbf{x}_n | \mathbf{y}) \mathbf{x}_n d\mathbf{x}_n \\
 \mathbf{P}_n &= \int p(\mathbf{x}_n | \mathbf{y}) \mathbf{x}_n \mathbf{x}_n^T d\mathbf{x}_n \\
 \mathbf{P}_{n,n-1} &= \int p(\mathbf{x}_n, \mathbf{x}_{n-1} | \mathbf{y}) \mathbf{x}_n \mathbf{x}_{n-1}^T d\mathbf{x}_n.
 \end{aligned} \right\} \tag{B 11}$$

These can be computed as shown in Ghahramani & Hinton (1996), with the minor modification that the updates for \mathbf{P}_n and $\mathbf{P}_{n,n-1}$ depend on \mathbf{A}_n instead of \mathbf{A} .

(b) M-step for neurodynamics

Because the only non-zero element in \mathbf{Q} (see Appendix A) is the first entry, the joint log-likelihood can be written as a function of the neurodynamic parameters as follows:

$$\begin{aligned}
 L_{\mathcal{Y}} &= -\frac{1}{2} \sum_{n=2}^N (\mathbf{x}_n(1) - \mathbf{x}_{n-1} \mathbf{F}_n^T \tilde{\mathbf{b}} - \mathbf{v}_n^T \mathbf{d})^2 \\
 &= -\frac{1}{2} \sum_{n=2}^N (\mathbf{x}_n(1) - [\mathbf{x}_{n-1}^T \mathbf{F}_n^T, \mathbf{v}_n^T] \boldsymbol{\theta})^2,
 \end{aligned} \tag{B 12}$$

where $\mathbf{x}_n(1)$ is the first entry in \mathbf{x}_n and

$$\boldsymbol{\theta} = \begin{bmatrix} \tilde{\mathbf{b}} \\ \mathbf{d} \end{bmatrix}. \tag{B 13}$$

Taking expectations and derivatives leads to the update

$$\hat{\boldsymbol{\theta}} = \left(\sum_n \tilde{\mathbf{S}}_n \right)^{-1} \left(\sum_n \tilde{\mathbf{v}}_n \right), \tag{B 14}$$

where

$$\tilde{\mathbf{S}}_n = \begin{bmatrix} \mathbf{F}_n \mathbf{P}_{n-1} \mathbf{F}_n^T & \mathbf{F}_n \mathbf{m}_{n-1} \mathbf{v}_n^T \\ \mathbf{v}_n \mathbf{m}_{n-1}^T \mathbf{F}_n^T & \mathbf{v}_n \mathbf{v}_n^T \end{bmatrix} \tag{B 15}$$

and

$$\tilde{\mathbf{v}}_n = [\mathbf{P}_{n,n-1}(1) \mathbf{F}_n^T, \mathbf{m}_n(1) \mathbf{v}_n^T]. \tag{B 16}$$

The quantities \mathbf{m}_n , \mathbf{m}_{n-1} , \mathbf{P}_{n-1} and $\mathbf{P}_{n,n-1}$ are computed after Kalman smoothing, \mathbf{v}_n are the driving inputs and \mathbf{F}_n is a matrix derived from the modulatory inputs defined in equation (A 4).

(c) M-step for haemodynamics

The output kernel coefficients can be updated using

$$\hat{\boldsymbol{\beta}} = \left(\sum_n \tilde{\mathbf{P}}_n(1) \right)^{-1} \left(\sum_n \boldsymbol{\Phi} \mathbf{m}_n y_n \right), \quad (\text{B } 17)$$

where

$$\tilde{\mathbf{P}}_n = \boldsymbol{\Phi} \mathbf{P}_n \boldsymbol{\Phi}^T. \quad (\text{B } 18)$$

The quantities \mathbf{m}_n and \mathbf{P}_n are computed after Kalman smoothing, $\boldsymbol{\Phi}$ is the matrix of haemodynamic basis functions and y_n is the observed fMRI time-series.

APPENDIX C: LIKELIHOOD

The likelihood is given by

$$L = \log p(\mathbf{y}_1^N) = \log p(\mathbf{y}_1) + \sum_{n=2}^N \log p(y_n | \mathbf{y}_1^{n-1}) \quad (\text{C } 1)$$

where

$$p(y_n | \mathbf{y}_1^{n-1}) = N(\mathbf{c}^T \mathbf{x}_n^{n-1}, \mathbf{c}^T \boldsymbol{\Sigma}_n^{n-1} \mathbf{c} + \sigma_e^2). \quad (\text{C } 2)$$

The quantities \mathbf{x}_n^{n-1} and $\boldsymbol{\Sigma}_n^{n-1}$ are obtained from Kalman filtering (see Appendix B(a)).

REFERENCES

- Baillet, S., Mosher, J. C. & Leahy, R. M. 2001 Electromagnetic brain mapping. *IEEE Signal Processing Mag.* **18**, 14–30.
- Breakspear, M. & Terry, J. R. 2002 Nonlinear interdependence in neural systems: motivation, theory and relevance. *Int. J. Neurosci.* **112**, 1263–1284.
- Buchel, C. & Friston, K. J. 1997 Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modelling and fMRI. *Cereb. Cortex* **7**, 768–778.
- Buxton, R. B., Wong, E. C. & Frank, L. R. 1998 Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magn. Reson. Med.* **39**, 855–864.
- David, O. & Friston, K. J. 2003 A neural mass model for MEG/EEG: coupling and neuronal dynamics. *NeuroImage* **20**, 1743–1755.
- Dayan, P. & Abbott, L. F. 2001 *Theoretical neuroscience: computational and mathematical modeling of neural systems*. Cambridge, MA: MIT Press.
- Dempster, A. P., Laird, N. M. & Rubin, D. B. 1977 Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. B* **39**, 1–38.
- Frackowiak, R. S. J., Friston, K. J., Frith, C., Dolan, R., Price, C. J., Zeki, S., Ashburner, J. & Penny, W. D. (ed.) 2003 *Human brain function*, 2nd edn. London: Academic Press.
- Friston, K. J. 2002 Bayesian estimation of dynamical systems: an application to fMRI. *NeuroImage* **16**, 513–530.
- Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. B., Frith, C. & Frackowiak, R. S. J. 1995 Statistical parametric maps in functional imaging: a general linear approach. *Hum. Brain Mapp.* **2**, 189–210.
- Friston, K. J., Buchel, C., Fink, G. R., Morris, J., Rolls, E. & Dolan, R. 1997 Psychophysiological and modulatory interactions in neuroimaging. *NeuroImage* **6**, 218–229.
- Friston, K. J., Harrison, L. & Penny, W. D. 2003 Dynamic causal modelling. *NeuroImage* **19**, 1273–1302.
- Galka, A., Yamashita, O., Ozaki, T., Biscay, R. & Valdes-Sosa, P. In press A solution to the dynamical inverse problem of EEG generation using spatiotemporal Kalman filtering. *NeuroImage* **23**, 435–453.
- Ghahramani, Z. & Beal, M. J. 2001 Propagation algorithms for variational Bayesian learning. In *NIPS 13* (ed. T. Leen, T. G. Dietterich & V. Tresp), pp. 507–513. Cambridge, MA: MIT Press.
- Ghahramani, Z. & Hinton, G. E. 1996 *Parameter estimation for linear dynamical systems*. Technical Report CRG-TR-96-2.: Department of Computer Science, University of Toronto. (Also available from <http://www.gatsby.ucl.ac.uk/~zoubin/papers.html>).
- Gitelman, D. R., Penny, W. D., Ashburner, J. & Friston, K. J. 2003 Modeling regional and psychophysiological interactions in fMRI: the importance of hemodynamic deconvolution. *NeuroImage* **19**, 200–207.
- Glover, G. H. 1999 Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage* **9**, 416–429.
- Haykin, S. 1996 *Adaptive filter theory*, 3rd edn. Englewood Cliffs, NJ: Prentice-Hall.
- Henson, R. N. A. 2003 *Human brain function*, ch. 40, 2nd edn. London: Academic Press.
- Henson, R. N. A., Rugg, M. D. & Friston, K. J. 2001 The choice of basis functions in event-related fMRI. *NeuroImage* **13**(Suppl. 1), 127.
- Ingber, L. 1995 Statistical mechanics of multiple scales of neocortical interactions. In *Neocortical dynamics and human EEG rhythms* (ed. P. Nunez). Oxford: Oxford University Press.
- Logothetis, N. K., Pauls, J., Augath, M., Trinath, T. & Oeltermann, A. 2001 Neurophysiological investigation of the basis of the fMRI signal. *Nature* **412**, 150–157.
- McKeown, M. J., Makeig, S., Brown, G. G., Jung, T. P., Kindermann, S. S., Bell, A. J. & Sejnowski, T. J. 1998 Analysis of fMRI data by blind separation into independent spatial components. *Hum. Brain Mapp.* **6**, 160–188.
- Makeig, S., Westerfield, M., Jung, T.-P., Enghoff, S., Townsend, J., Courchesne, E. & Sejnowski, T. J. 2002 Dynamic brain sources of visual evoked responses. *Science* **295**, 690–694.
- Penny, W. D., Stephan, K. E., Mechelli, A. & Friston, K. J. 2004 Comparing dynamic causal models. *NeuroImage* **22**, 1157–1172.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. V. P. 1992 *Numerical recipes in C*. Cambridge: Cambridge University Press.
- Rao, C. R. & Toutenberg, H. 1995 *Linear models: least squares and alternatives*. Heidelberg: Springer.
- Riera, J., Watanabe, J., Iwata, K., Miura, N., Aubert, E., Ozaki, T. & Kawashima, R. 2004 A state-space model of the hemodynamic approach: nonlinear filtering of BOLD signals. *NeuroImage* **21**, 547–567.
- Roweis, S. & Ghahramani, Z. 1999 A unifying review of linear Gaussian models. *Neural Comput.* **11**, 305–346.
- Stephan, K. E., Harrison, L., Penny, W. D. & Friston, K. J. 2004 Biophysical models of fMRI responses. *Curr. Opin. Neurobiol.* **14**, 629–635.
- 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.
- Weigend, A. S. & Gershenfeld, N. A. 1994 *Time series prediction: forecasting the future and understanding the past*. Reading, MA: Addison-Wesley.
- Yamashita, O., Galka, A., Ozaki, T., Biscay, R. & Valdes-Sosa, P. 2004 Recursive penalised least squares solution for dynamical inverse problems of EEG generation. *Hum. Brain Mapp.* **21**, 221–235.