5 *Bayesian Treatments of Neuroimaging Data*

Will Penny and Karl Friston

5.1 Introduction

In this chapter we discuss the application of Bayesian methods to neuroimaging data. This includes data from positron emission tomography (PET), functional magnetic resonance imaging (fMRI), electroencephalography (EEG), and magnetoencephalography (MEG). We concentrate on fMRI but the concepts, methodologies, and modeling approaches apply to all modalities.

A general issue in the analysis of brain imaging data is the relationship between the neurobiological hypothesis one posits and the statistical models adopted to test that hypothesis. One key distinction is between functional specialization and integration.

Briefly, fMRI was originally used to provide functional maps showing which regions are specialized for specific functions, a classic example being the study by Zeki et al. [27] who identified V4 and V5 as being specialized for the processing of color and motion, respectively. More recently, these analyses have been augmented by functional integration studies, which describe how functionally specialized areas interact and how these interactions depend on changes of context. A recent example is the study by Buchel et al. [1] who found that the success with which a subject learned an object-location association task was correlated with the coupling between regions in the dorsal and ventral visual streams.

Functionally specialized brain responses are typically characterized using the general linear model (GLM). An fMRI data set comprises a time series of volumetric data. In "mass-univariate" approaches GLMs are fitted to fMRI time series at each voxel (volume element), resulting in a set of voxel-specific parameters. These parameters are then used to form posterior probability maps (PPMs) that characterize regionally specific responses to experimental manipulation. Figure 5.5, for example, shows a PPM highlighting regions that are sensitive to visual motion stimuli.

Analyses of functional integration are implemented using multivariate approaches that examine the changes in multiple brain areas induced by experimental manipulation. Although there a number of methods for doing this we

focus on a recent approach called dynamic causal modeling (DCM).

This chapter is structured as follows. Section 5.2 introduces an fMRI data set that is analyzed in later sections. Section 5.3 describes the GLM and section 5.4 describes Bayesian estimation procedures for GLMs and extensions for nonlinear models. Section 5.5 describes PPM for making inferences about functional specialization, and section 5.6 describes DCM for making inferences about functional integration.

5.2 Attention to Visual Motion

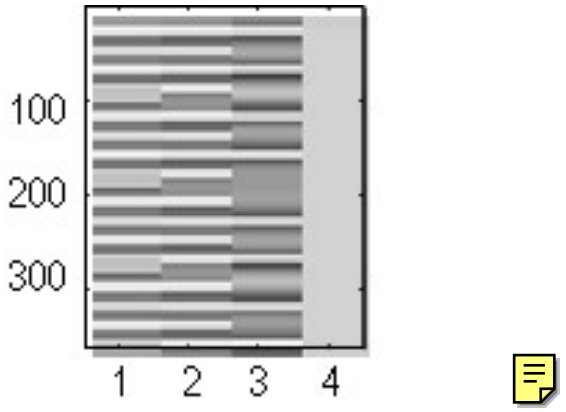

Figure 5.1 Design matrix for study of attention to visual motion. The four columns correspond to "Photic" stimulation, "Motion", "Attention", and a constant. There are as many rows in the design matrix as there are time points in the imaging time series. For this data set there are 360 images. The first three columns in the design matrix were formed by representing the experimental condition using a boxcar regressor and then convolving with a canonical hemodynamic response function (see fgure 5.2).

This section describes a data set that will be analyzed using PPMs and DCMs. Subjects were studied with fMRI under identical stimulus conditions (visual motion subtended by radially moving dots) while manipulating the attentional component of the task (detection of velocity changes).

The data were acquired from normal subjects at 2 Tesla using a Magnetom VISION (Siemens, Erlangen, Germany) whole-body MRI system, equipped with a head volume coil. Contiguous multislice T2*-weighted fMRI images were obtained with a gradient echo-planar sequence (TE = 40 ms, TR = 3.22 seconds, matrix size = 64 x 64 x 32, voxel size 3 x 3 x 3 mm).

Each subject had four consecutive 100-scan sessions comprising a series of ten scan blocks under five different conditions: D F A F N F A F N S. The first condition (D) was a dummy condition to allow for magnetic saturation effects. F (Fixation) corresponds to a low-level baseline where the subjects viewed a fixation point at the center of a screen.

In condition A (Attention) subjects viewed 250 dots moving radially from the center at 4.7 degrees per second and were asked to detect changes in radial velocity (which did not actually occur). This attentional manipulaton was validated post hoc using psychophysics and the motion aftereffect. In condition N (No attention) the subjects were asked simply to view the moving dots. In condition S (Stationary) subjects viewed stationary dots. The order of A and N was swapped for the last two sessions. In all conditions subjects fixated the center of the screen.

In a prescanning session the subjects were given five trials with five speed changes (reducing to 1%). During scanning there were no speed changes and no overt response was required in any condition. In this chapter we analyze data from the first subject.

For the purpose of the analyses in this chapter the above experimental conditions were formulated using the following factors or causes. "Photic" stimulation comprised the $\{A, B, S\}$ conditions, "motion" comprised the $\{N, A\}$ conditions and "Attention" comprised the A condition. These three variables are encoded into the design matrix in figure 5.1. The relative contribution of each of these variables can be assessed using standard least squares or Bayesian estimation. Classical inferences about these contributions are made using T or F statistics, depending upon whether one is looking at a particular linear combination (e.g. a subtraction), or all of them together. Bayesian inferences are based on the posterior or conditional probability that the contribution exceeded some threshold, usually zero.

5.3 The General Linear Model

The general linear model is an equation that expresses an observed response variable y in terms of a linear combination of explanatory variables X:

$$y = X\beta + e, \tag{5.1}$$

where y is a $T \times 1$ vector comprising responses at, e.g. T time points, X is a $T \times K$ design matrix, β is a vector of regression coefficients, and e is a $T \times 1$ error vector.

The general linear model is variously known as "analysis of covariance" or "multiple regression analysis" and subsumes simpler variants, like the "T test" for a difference in means, to more elaborate linear convolution models such as finite impulse response (FIR) models. The matrix X that contains the explanatory variables (e.g. designed effects or confounds) is called the design matrix.

Each column of the design matrix corresponds to some effect one has built into the experiment or that may confound the results. These are referred to as explanatory variables, covariates, or regressors. The design matrix can contain both covariates and indicator variables. Each column of X has an associated unknown parameter. Some of these parameters will be of interest, e.g. the effect of a particular sensorimotor or cognitive condition or the regression coefficient of hemodynamic responses on reaction time. The remaining parameters

will be of no interest and pertain to confounding effects, e.g. the effect of being a particular subject.

A standard method for neuroimaging analysis is the mass-univariate approach [12] where GLMs are fitted to each voxel. This allows one to test for the same effects at each point in the brain. Due primarily to the presence of aliased biorhythms and unmodeled neuronal activity in fMRI, the errors in the GLM will be temporally autocorrelated. The general linear model can accommodate this as shown in section 5.4.

5.3.1 Contrasts

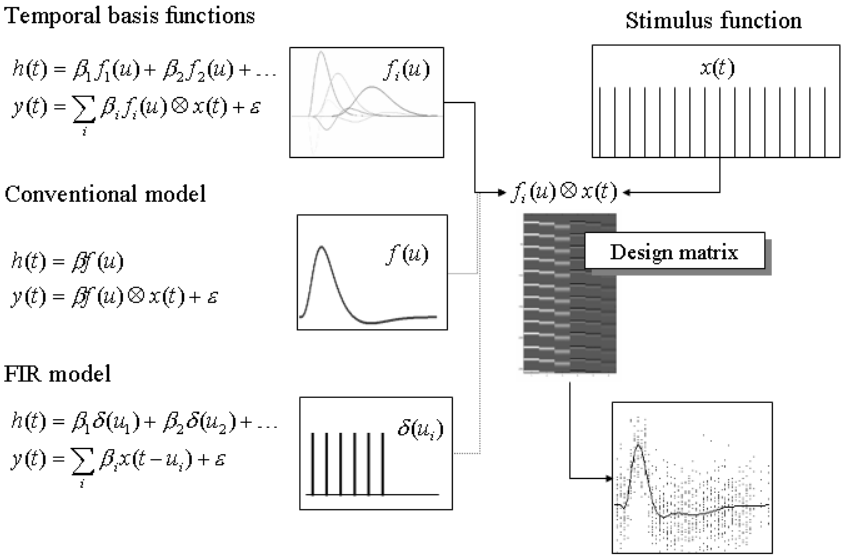

Figure 5.2 Temporal basis functions offer useful constraints on the form of the estimated response that retain (i) the flexibility of finite impulse response (FIR) models and (ii) the efficiency of single regressor models. The specification of these models involves setting up stimulus functions x(t) that model expected neuronal changes, e.g., boxcars of epoch-related responses or spikes (delta functions) at the onset of specific events or trials. These stimulus functions are then convolved with a set of basis functions of peristimulus time u, that model the HRF, in some linear combination. The ensuing regressors are assembled into the design matrix. The basis functions can be as simple as a single canonical HRF (middle), through to a series of delayed delta functions (bottom). The latter case corresponds to a FIR model and the coefficients constitute estimates of the impulse response function at a finite number of discrete sampling times. Selective averaging in event-related fMRI [4] is mathematically equivalent to this limiting case.

To assess effects of interest that are spanned by one or more columns in the design matrix one uses a contrast (i.e., a linear combination of parameter es-

timates). An example of a contrast weight vector would be [-1 1 0 0.....] to compare the difference in responses evoked by two conditions, modeled by the first two condition-specific regressors in the design matrix. Sometimes several contrasts of parameter estimates are jointly interesting. For example, when using polynomial or basis function expansions of some experimental factor. In these instances, a matrix of contrast weights is used that can be thought of as a collection of effects that one wants to test together. Such a contrast may look like

$$c^T = \begin{bmatrix} -1 & 0 & 0 & 0 & . & . & . \\ 0 & 1 & 0 & 0 & . & . & . \end{bmatrix}, \tag{5.2}$$

which would test for the significance of the first or second parameter estimates. The fact that the first weight is -1 as opposed to 1 has no effect on the test because F statistics are based on sums of squares [7]. .

5.3.2 Temporal Basis Functions

Functional MRI using blood oxygen level dependent (BOLD) contrast provides an index of neuronal activity indirectly via changes in blood oxygenation levels. This relationship can be characterized using temporal basis functions as shown in figure 5.2.

For a given impulse of neuronal activity the fMRI signal peaks some 4-6 seconds later, then after 10 seconds or so drops below zero and returns to baseline after 20 to 30 seconds. This response varies from subject to subject and from voxel to voxel and this variation can be captured using temporal basis functions. In [13] the form of the hemodynamic impulse response function (HRF) was estimated using a least squares deconvolution and a time invariant model, where evoked neuronal responses are convolved with the HRF to give the measured hemodynamic response. This simple linear framework is the cornerstone for making statistical inferences about activations in fMRI with the GLM. An impulse response function is the response to a single impulse, measured at a series of times after the input. It characterizes the input-output behavior of the system (i.e. voxel) and places important constraints on the sorts of inputs that will excite a response. The HRFs, estimated in [13] resembled a Poisson or gamma function, peaking at about 5 seconds.

The basic idea behind temporal basis functions is that the hemodynamic response induced by any given trial type can be expressed as the linear combination of several basis functions of peristimulus time. As shown in figure 5.2, the convolution model for fMRI responses takes a stimulus function encoding the supposed neuronal responses and convolves it with an HRF to give a regressor that enters into the design matrix. When using basis functions, the stimulus function is convolved with all the basis functions to give a series of regressors. The associated parameter estimates are the coefficients or weights that determine the mixture of basis functions that best models the HRF for the trial type and voxel in question. We find the most useful basis set to be a canonical HRF and its derivatives with respect to the key parameters that determine

its form (e.g. latency and dispersion) [16]. This is known as an informed basis set. The nice thing about this approach is that it can partition differences among evoked responses into differences in magnitude, latency, or dispersion, that can be tested for using specific contrasts [9].

Temporal basis functions are important because they enable a graceful transition between conventional multilinear regression models with one stimulus function per condition and FIR models with a parameter for each time point following the onset of a condition or trial type. Figure 5.2 illustrates this graphically (see figure legend).

In summary, temporal basis functions offer useful constraints on the form of the estimated response that retain (i) the flexibility of FIR models and (ii) the efficiency of single regressor models. The advantage of using several temporal basis functions (as opposed to an assumed form for the HRF) is that one can model voxel-specific forms for hemodynamic responses and formal differences (e.g. onset latencies) among responses to different sorts of events. The advantages of using an informed basis set over FIR models are that (i) the parameters are estimated more efficiently and (ii) stimuli can be presented at any point in the interstimulus interval. The latter is important because time-locking stimulus presentation and data acquisition give a biased sampling over peristimulus time and can lead to differential sensitivities, in multislice acquisition, over the brain.

5.4 Parameter Estimation

This section describes how the parameters of GLMs can be estimated using maximum likelihood and Bayesian approaches. We also show how the Bayesian approach can be applied to nonlinear models.

5.4.1 Maximum Likelihood

In what follows, $\mathsf{N}(x; m, \Sigma)$ specifies a multivariate Gaussian distribution over x with mean m and covariance Σ. The likelihood specified by the GLM described in the previous section is given by

$$p(y|\beta) = \mathsf{N}(y; X\beta, C_e), \tag{5.3}$$

where C_e is an error covariance matrix. The maximum-likelihood (ML) estimator is given by [25]:

$$\hat{\beta}_{ML} = \left(X^T C_e^{-1} X\right)^{-1} X^T C_e^{-1} y \tag{5.4}$$

This estimator is suitable for fMRI where the temporal autocorrelation in the errors is specified by the matrix C_e. For the moment we consider that C_e is known, although in section 5.4.3 we show how it can be estimated. For other modalities $C_e = \sigma^2 I$ often suffices. The ML estimate then reduces to the ordinary least squares (OLS) estimator

$$\hat{\beta}_{OLS} = \left(X^T X\right)^{-1} X^T y \tag{5.5}$$

5.4.2 Bayes Rule for GLMs

The Bayesian framework allows us to incorporate our prior beliefs about parameter values. This allows one to combine information from multiple modalities or to implement soft neurobiological constraints.

There are many examples of this. In fMRI, the search for activations can be restricted to regions thought to be gray matter as identified from structural MRI [17]. In EEG source localization, source orientations can be softly constrained to be perpendicular to the structural MRI estimates of the cortical surface [24]. More challengingly, estimation of neuronal activity can be based on simultaneously acquired EEG and fMRI [19].

In the context of dynamic causal models, described in section 5.7, parameter estimation can be constrained by neurophysiological data. Hemodynamic transit times, for example, can be constrained to realistic values and parameters governing neuronal dynamics can be constrained to give rise to stable dynamical systems.

If our beliefs can be specified using the Gaussian distribution

$$p(\beta) = \mathsf{N}(\beta; \mu_p, C_p), \tag{5.6}$$

where μ_p is the prior mean and C_p is the prior covariance, then the posterior distribution is [18]

$$p(\beta|Y) = \mathsf{N}(\beta; \mu, C), \tag{5.7}$$

where

$$\begin{aligned} C^{-1} &= X^T C_e^{-1} X + C_p^{-1} \\ \mu &= C(X^T C_e^{-1} y + C_p^{-1} \mu_p). \end{aligned} \tag{5.8}$$

The posterior precision, C^{-1}, is equal to the sum of the prior precision plus the data precision. The posterior mean, μ, is given by the sum of the prior mean plus the data mean, but where each is weighted according to its relative precision. In the abscence of prior information, i.e. $C_p^{-1} = 0$, the above estimate reduces to the ML estimate.

Bayes rule is illustrated for GLMs with two regression coefficients in figure 5.3. The figure shows an example where the prior is centered at zero. This is known as a shrinkage prior because parameter estimates are shrunk toward zero. The figure shows that the shrinkage is greater for β_2, the parameter about which the data are less informative.

5.4.3 Variance Components

It is also possible to compute the probability of the data after integrating out the dependence on model parameters

$$p(y) = \int p(y|\beta)p(\beta)d\beta \tag{5.9}$$

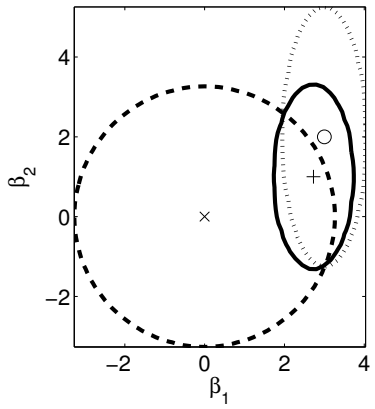

Figure 5.3 Bayesian estimation for GLMs with two parameters. The prior (dashed line) has mean $\mu_p = 0$ (cross) and precision $C_p^{-1} = \mathsf{diag}([1, 1])$. The likelihood (dotted line) has mean $X^T y = [3, 2]^T$ (circle) and precision $(X^T C_e^{-1} X)^{-1} = \mathsf{diag}([10, 1])$. The posterior (solid line) has mean $\mu = [2.73, 1]^T$ (cross) and precision $C^{-1} = \mathsf{diag}([11, 2])$. In this example, the measurements are more informative about β_1 than β_2. This is reflected in the posterior distribution.

This is known as the model evidence.

Bayesian estimation as described in the previous section assumed that we knew the prior covariance, C_p, and error covariance, C_e. This information is, however, rarely available. In [15] a parametric empirical Bayesian (PEB) framework was developed in which these covariances are expressed as

$$\begin{aligned} C_p &= \sum_i \lambda_i Q_i \\ C_e &= \sum_j \lambda_j Q_j, \end{aligned} \tag{5.10}$$

where Q_i and Q_j are known "covariance components" and λ_i, λ_j are hyperparameters that can be set so as to maximize the evidence. For example, Q_j can be set up to describe the temporal autocorrelations that are typically observed in fMRI [10]. This algorithm is used in the estimation of PPMs (see section 5.5).

If the priors are correct, then Bayesian estimation is more accurate than ML estimation [3]. But in general it is not possible to ensure that the "correct" priors are used. We can, however, use the best prior in a family of priors by using the PEB approach described above. It is also possible to compare different families of priors using Bayesian model comparison based on the model evidence [20, 21].

5.4.4 Nonlinear Models

For nonlinear models,

$$y = b(\theta) + e, \tag{5.11}$$

where $b(\theta)$ is a nonlinear function of parameter vector θ, the PEB framework can be applied by locally linearizing the nonlinearity, about a "current" estimate μ_i, using a first-order Taylor series expansion

$$b(\theta) = b(\mu_i) + \frac{\partial b(\mu_i)}{\partial \theta}(\theta - \mu_i) \tag{5.12}$$

Substituting this into equation (5.11) and defining $r \equiv y - b(\mu_i)$, $J \equiv \frac{\partial b(\mu_i)}{\partial \theta}$ and $\Delta\theta \equiv \theta - \mu_i$ gives

$$r = J\Delta\theta + e, \tag{5.13}$$

which now conforms to a GLM (cf. equation (5.1)). The prior, likelihood and posterior are now given by

$$\begin{aligned} p(\Delta\theta) &= \mathsf{N}(\Delta\theta; \mu_p - \mu_i, C_p) \\ p(r|\Delta\theta) &= \mathsf{N}(r; J\Delta\theta, C_e) \\ p(\Delta\theta|r) &= \mathsf{N}(\Delta\theta; \Delta\mu, C_{i+1}). \end{aligned} \tag{5.14}$$

The quantities $\Delta\mu$ and C_{i+1} can be found using the result for the linear case (substitute r for Y and J for X in equation 5.8). If we define our "new" parameter estimate as $\mu_{i+1} = \mu_i + \Delta\mu$, then

$$\begin{aligned} C_{i+1}^{-1} &= J^T C_e^{-1} J + C_p^{-1} \\ \mu_{i+1} &= \mu_i + C_{i+1}(J^T C_e^{-1} r + C_p^{-1}(\mu_p - \mu_i)). \end{aligned} \tag{5.15}$$

This update is applied recursively. It can also be combined with hyperparameter estimates, to characterize C_p and C_e, as described in the previous section and in [8]. This algorithm is used to estimate parameters of dynamic causal models (see section 5.6). In this instance $b(\theta)$ corresponds to the integration of a dynamic system or, equivalently, a convolution operator.

5.5 Posterior Probability Mapping

The dominant paradigm for the analysis of neuroimaging data to date has been statistical parametric mapping [12, 7]. This is a mass-univariate approach in which GLMs are fitted at each voxel using ML or OLS estimators. Contrasts of parameter estimates are then used to make maps of statistical parameters (e.g. T or F values) that quantify effects of interest. These maps are then thresholded using random field theory (RFT)[7] to correct for the multiple statistical tests made over the image volume.

The RFT correction is important as it protects against false-positive inferences. A naive or "uncorrected p-value" of 0.05, used with images containing 100,000 voxels, will lead to SPMs containing on average 5,000 false-positives. This is clearly too liberal. RFT protects instead against family-wise errors (FWEs) where the "family" is the set of voxels in an image. An FWE rate or "corrected p-value" of 0.05 implies on average only a single error in 20 SPMs. This much more conservative criterion has become a standard in the analysis of PET and fMRI data.

Despite its success, statistical parametric mapping has a number of fundamental limitations. In SPM the p-value, ascribed to a particular effect, does not reflect the likelihood that the effect is present but simply the probability of getting the observed data in the effect's absence. If sufficiently small, this p-value can be used to reject the null hypothesis that the effect is negligible. There are several shortcomings of this classical approach.

Firstly, one can never reject the alternate hypothesis (i.e. say that an activation has not occurred) because the probability that an effect is exactly zero is itself zero. This is problematic, for example, in trying to establish double dissociations or indeed functional segregation; one can never say one area responds to color but not motion and another responds to motion but not color.

Secondly, because the probability of an effect being zero is vanishingly small, given enough scans or subjects one can always demonstrate a significant effect at every voxel. This fallacy of classical inference is becoming relevant practically, with the thousands of scans entering into some fixed-effect analyses of fMRI data. The issue here is that a trivially small activation can be declared significant if there are sufficient degrees of freedom to render the variability of the activation's estimate small enough.

A third problem, that is specific to SPM, is the correction or adjustment applied to the p-values to resolve the multiple comparison problem. This has the somewhat nonsensical effect of changing the inference about one part of the brain in a way that is contingent on whether another part is examined. Put simply, the threshold increases with search volume, rendering inference very sensitive to what that inference encompasses. Clearly the probability that any voxel has activated does not change with the search volume and yet the classical p-value does.

All these problems would be eschewed by using the probability that a voxel had activated, or indeed its activation was greater than some threshold. This sort of inference is precluded by classical approaches, which simply give the likelihood of getting the data, given no activation. What one would really like is the probability distribution of the activation given the data. This is the posterior probability used in Bayesian inference.

The posterior distribution requires both the likelihood, afforded by assumptions about the signal and distribution of errors, and the prior probability of activation. These priors can enter as known values or can be estimated from the data. To date, two types of prior have been advocated for fMRI. A global shrinkage prior [14], which for the ith voxel and jth regression coefficient is

given by

$$p(\beta_{ij}) = \mathsf{N}(\beta_{ij}; 0, \sigma_j^2), \tag{5.16}$$

where σ_j^2 quantifies the effect variability over the brain. This prior encodes a belief that every region of the brain is activated in every task but to a greater or lesser degree. More recently gaussian markov random field (GMRF) priors have been proposed [23]. These can be written

$$p(\beta_{ij}) = \mathsf{N}(\beta_{ij}; \sum_{k \epsilon N_i} a_k \beta_{kj}, \sigma_j^2), \tag{5.17}$$

where N_i defines a set of voxels in the neighborhood of voxel i and a_k are spatial weighting coefficients. This encourages parameter estimates to be similar to those at nearby voxels. These priors provide an adaptive spatial regularization that renders the resulting Bayesian inference more sensitive than the equivalent classical inference [23].

5.5.1 Attention to Visual Motion

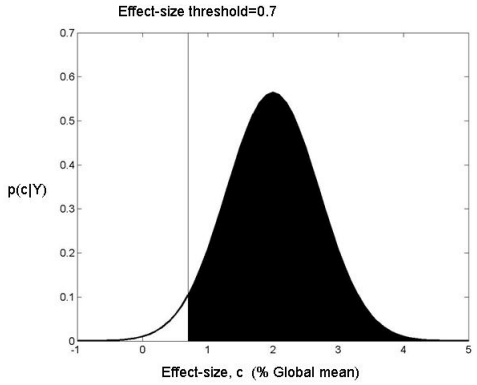

Figure 5.4 Posterior distribution of effect size at a particular voxel. The shaded area gives the posterior probability that the effect size is larger than the chosen threshold of 0.7%. For this voxel, this probability is 0.967. This is larger than the probability threshold 0.95 so this voxel would appear in the PPM shown in figure 5.5.

In this section we compare Bayesian and classical inference using PPMs and SPMs based on real data. These data were analyzed using a conventional SPM procedure [7, 12] and the PEB approach (using global shrinkage priors) described in the previous section. Inference in PPMs is based on the posterior distribution of effect sizes, which is shown for a single voxel in figure 5.4. This collection of posterior distributions can be turned into a PPM by specifying (i)

an effect-size threshold and (ii) a probability threshold [14]. Here we used an effect-size threshold of 0.7% of whole-brain mean, and a probability threshold of 0.95.

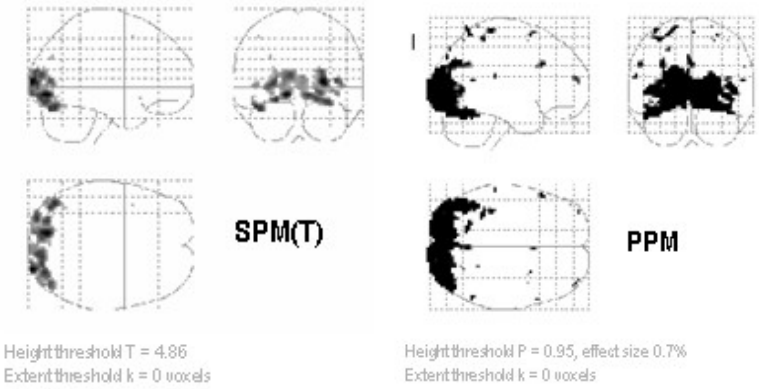

Figure 5.5 Maximum intensity projections (MIPs) of SPM (left) and PPM (right) for the fMRI study of attention to visual motion. The SPM uses a threshold corrected for multiple comparisons at p = 0.05. The PPM uses an effect-size threshold of 0.7% and a probability threshold of 0.95.

The SPM and PPM for these data are presented in figures 5.5 and 5.6. We used a contrast that tested for the effect of visual motion above and beyond that due to photic stimulation with stationary dots, i.e. $c^T = [0100]$ (cf. the design matrix in figure 5.1). These effects are restricted to visual and extrastriate cortex involved in motion processing.

The difference between the PPM and SPM is immediately apparent on inspection of figure 5.6. The critical thing to note is that the SPM identifies a smaller number of voxels than the PPM. Indeed the SPM appears to have missed a critical and bilaterally represented part of the V5 complex (circled cluster on the PPM in the right panel of figure 5.6). The SPM is more conservative because the correction for multiple comparisons in these data is very severe, rendering classical inference relatively insensitive. It is interesting to note that dynamic motion in the visual field has such widespread (if small) effects at a hemodynamic level.

5.6 Dynamic Causal Modeling

Dynamic causal modeling (DCM) is used to make inferences about functional integration and has been formulated for the analysis of fMRI time series in [11]. The term "causal" in DCM arises because the brain is treated as a deterministic dynamical system in which external inputs cause changes in neuronal activity which in turn cause changes in the resulting BOLD signal that is measured with fMRI. This is to be contrasted with a conventional GLM where there is no

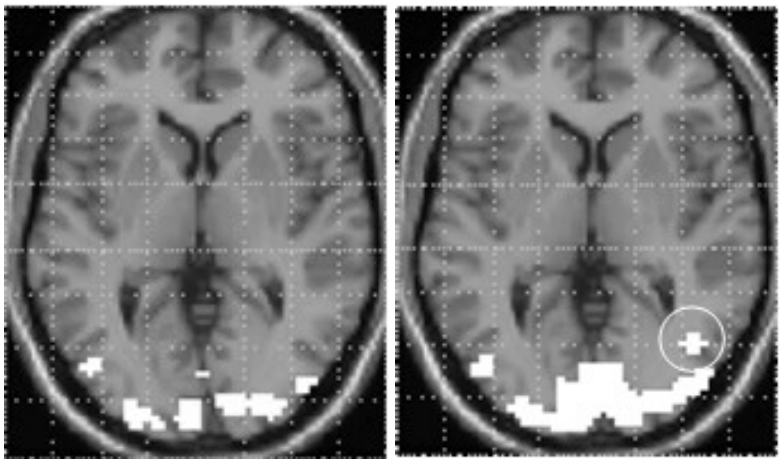

Figure 5.6 Overlays for SPM (left) and PPM (right) for the fMRI study of attention to visual motion, showing axial slices at $z = 3$ mm through extrastriate cortex. The thresholds are the same as those used in the MIPs in figure 5.5

explicit representation of neuronal activity. The second main difference to the GLM is that DCM allows for interactions between regions. Of course, it is this interaction which is central to the study of functional integration.

Current DCMs for fMRI comprise a bilinear model for the neurodynamics and an extended balloon model for the hemodynamics [2, 8]. The neurodynamics are described by the multivariate differential equation

$$\dot{z} = \left(A + \sum_i u_i B_i \right) z + Cu, \tag{5.18}$$

where z is a vector of neuronal activity, $\dot{z}$ denotes the time derivative, and u is a vector of experimental stimuli. This is known as a bilinear model because the dependent variable, $\dot{z}$, is linearly dependent on the product of u_i and z. That u_i and z combine in a multiplicative fashion endows the model with "nonlinear" dynamics that can be understood as a nonstationary linear system that changes according to experimental manipulation u. Importantly, because u is known, parameter estimation is relatively simple.

Connectivity in DCM is characterized by a set of "intrinsic connections", A, that specify which regions are connected and whether these connections are unidirectional or bidirectional. We also define a set of input connections, C, that specify which inputs are connected to which regions, and a set of modulatory or bilinear connections, B, that specify which intrinsic connections can be changed by which inputs. The overall specification of input, intrinsic, and modulatory connectivity comprise our assumptions about model structure. This in turn represents a scientific hypothesis about the structure of the large-scale neuronal network mediating the underlying sensorimotor or cogni-

tive function.

In DCM, neuronal activity gives rise to fMRI activity by a dynamic process described by an extended balloon model for each region or node. This involves a set of hemodynamic state variables, state equations, and hemodynamic parameters, h. In brief, for the ith region, neuronal activity z_i causes an increase in the vasodilator signal s_i that is subject to autoregulatory feedback. Inflow f_i responds in proportion to this signal with concomitant changes in blood volume v_i and deoxyhemoglobin content q_i.

$$\begin{aligned} \dot{s}_i &= z_i - \kappa_i s_i - \gamma_i(f_i - 1) \\ \dot{f}_i &= s_i \\ \tau_i \dot{v}_i &= f_i - v_i^{1/\alpha} \\ \tau_i \dot{q}_i &= f_i \frac{E(f_i, \rho_i)}{\rho_i} - v_i^{1/\alpha} \frac{q_i}{v_i} \end{aligned} \tag{5.19}$$

Outflow is related to volume $f_{out} = v^{1/\alpha}$ through Grubb's exponent α [11]. The oxygen extraction is a function of flow $E(f, \rho) = 1 - (1 - \rho)^{1/f}$ where ρ is resting oxygen extraction fraction. The BOLD signal is then taken to be a static nonlinear function of volume and deoxyhemoglobin content [11].

Together these equations describe a nonlinear hemodynamic process that may be regarded as a biophysically informed generalization of the linear convolution models used in the GLM. This process converts neuronal activity in the ith region z_i to the hemodynamic response. Full details are given in [8, 11].

Fitting a DCM to fMRI time series Y involves the following steps. Firstly, the parameters $\theta = \{A, B, C, h\}$ are initialized by setting them to their prior values. These priors incorporate biophysical and dynamic constraints. The neurodynamic (equation (5.18)) and hemodynamic processes (equation (5.19)) are numerically integrated to obtain a predicted fMRI time series. This integration is efficient because most fMRI experiments employ input vectors that are highly sparse by experimental design. The parameters θ are then updated using the PEB algorithm described in section 5.4.4. The PEB and integration steps are iterated until convergence. Inferences about connections can then be made from the posterior density $p(\theta|Y)$.

5.6.1 Attention to Visual Motion

We now return to the attention to visual motion study so as to make inferences about functional integration. Firstly a number of regions were selected. These were located in primary visual cortex V1, motion-sensitive area V5, superior parietal cortex, (SPC), and (inferior frontal gyrus), IFG. These regions were based on maxima from conventional SPMs testing for the effects of photic stimulation, motion, and attention. The DCM was set up as shown in figure 5.7. Regional time courses were taken as the first eigenvariate of 8 mm spherical volumes at each location. This eigenvariate approach was used so as to obtain a single representative time series from the cluster of active voxels in each region and was implemented using a singular value decomposition (SVD)

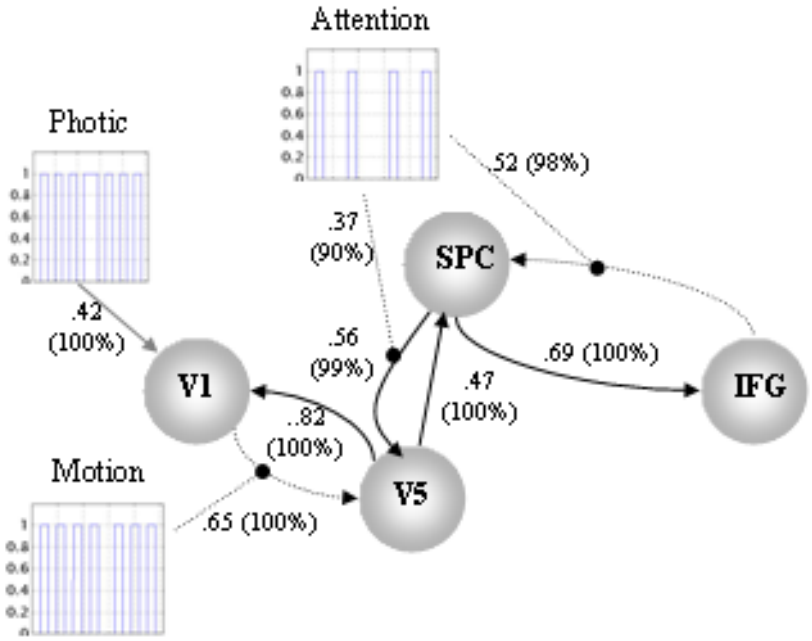

Figure 5.7 DCM for the fMRI study of attention to visual motion. The most interesting aspects of this connectivity involve the role of motion and attention in exerting bilinear effects. Critically, the influence of motion is to enable connections from V1 to the motion-sensitive area V5. The influence of attention is to enable backward connections from the inferior frontal gyrus (IFG) to the superior parietal cortex (SPC). Furthermore, attention increases the influence of SPC on V5. Dotted arrows connecting regions represent significant bilinear effects in the absence of a significant intrinsic coupling. Numbers in brackets represent the posterior probability, expressed as a percentage, that the effect size is larger than 0.17. This cutoff corresponds to a time constant of 4 seconds or less - in DCM stronger effects have faster time constants.

The inputs, in this example, comprise one sensory perturbation and two contextual inputs. The sensory input was simply the presence of photic stimulation and the first contextual one was presence of motion in the visual field. The second contextual input, encoding attentional set, was unity during attention to speed changes and zero otherwise. The outputs corresponded to the four regional eigenvariates. The intrinsic connections were constrained to conform to a hierarchical pattern in which each area was reciprocally connected to its supraordinate area. Photic stimulation entered at, and only at, V1. The effect of motion in the visual field was modeled as a bilinear modulation of the V1 to V5 connectivity and attention was allowed to modulate the backward connec-

tions from IFG and SPC.

The results of the DCM are shown in figure 5.7. Of primary interest here is the modulatory effect of attention that is expressed in terms of the bilinear coupling parameters for this third input. As hoped, we can be highly confident that attention modulates the backward connections from IFG to SPC and from SPC to V5. Indeed, the influences of IFG on SPC are negligible in the absence of attention (dotted connection in figure 5.7). It is important to note that the only way that attentional manipulation could affect brain responses was through this bilinear effect. Attention-related responses are seen throughout the system. This attentional modulation is accounted for by changing just two connections. This change is, presumably, instantiated by instructional set at the beginning of each epoch.

This analysis also illustrates how functional segregation is modeled in DCM. Here one can regard V1 as "segregating" motion from other visual information and distributing it to the motion-sensitive area V5. This segregation is modeled as a bilinear "enabling" of V1 to V5 connections when, and only when, motion is present. Note that in the absence of motion the intrinsic V1 to V5 connection was trivially small (in fact the posterior mean was -0.04). The key advantage of entering motion through a bilinear effect, as opposed to a direct effect on V5, is that we can finesse the inference that V5 shows motion-selective responses with the assertion that these responses are mediated by afferents from V1. The two bilinear effects above represent two important aspects of functional integration that DCM was designed to characterize.

Finally, we note that the inferences we have made are dependent on the particular model we have chosen, that is, the pattern of intrinsic and modulatory connectivity. As alluded to in section 5.4.3, it is possible to compare models using the Bayesian model comparison base on the model evidence. This has been implemented for DCMs applied to these data and resulted in two main findings [21]. Firstly, models with reciprocal connections between hierarchical areas are favored over models with purely feedforward connections or models with a full connectivity. Secondly, attentional effects are best explained by modulation of forward rather than backward connections.

5.7 Discussion

This chapter has described the application of Bayesian methods to neuroimaging data. A key issue we have addressed is the relationship between the neurobiological hypothesis one posits and the statistical models adopted to test that hypothesis.

Historically, the focus of neuroimaging analysis has been the identification of regions that are specialized for certain functions. More recently these approaches have been augmented with multivariate approaches that describe imaging data in terms of distributed network models [22].

Ideally one would like to take neural network models defined by computational neuroscientists and use imaging data to estimate unknown parameters.

While this is some way off steps are nevertheless being made to close the gap. A key step in this direction is the development of the DCM framework. This is currently being extended to explain event-related potentials derived from EEG/MEG data. The underlying neuronal models contain excitatory and inhibitory parameters that govern the dynamics of neuronal masses [6]. These parameters are estimated using the PEB framework described in this chapter.

This illustrates a key point that guides the development of our methodologies. As our models become more realistic and therefore complex they need to be constrained in some way and a simple and principled way of doing this is to use priors in a Bayesian context.

Due to the concise nature of this review we have been unable to cover a number of important topics in the analysis of imaging data.

The first is the use of spatial "preprocessing" methods to remove the effects of motion, and to spatially normalize data to transform it into a standard anatomical space. This is necessary so that regionally specific effects can be reported in a frame of reference that can be related to other studies. It is also necessary for analyses of group data. These methods adopt Bayesian paradigms to constrain the estimated transformations. Further details can be found in [7].

The second area is multimodal integration. We have also applied the PEB framework described in this chapter to the problem of source localization in EEG/MEG [24]. See also [26] for a recent approach using variational Bayes (VB). We, and many others (see e.g. [5]), are currently developing models that will integrate information from EEG, to find out primarily when activations occur, with information from fMRI/MRI, to find out where activations occur. This is an exciting area that would significantly strengthen the bridge between modalities in imaging neuroscience and our understanding of the neurobiology underlying cognitive processing.

References

[1] Buchel C, Coull JT, Friston KJ (1999) The predictive value of changes in effective connectivity for human learning. *Science*, 283:1538–1541.

[2] Buxton RB, Wong EC, Frank LR (1998) Dynamics of blood flow and oxygenation changes during brain activation: The Balloon Model. *Magnetic Resonance in Medicine*, 39:855–864.

[3] Copas JB (1983) Regression prediction and shrinkage. *Journal of the Royal Statistical Society Series B*, 45:311–354.

[4] Dale A, Buckner R (1997) Selective averaging of rapidly presented individual trials using fMRI. *Human Brain Mapping*, 5:329–340.

[5] Dale AM, Sereno MI (1993) Improved localisation of cortical activity by combining Eeg and Meg with Mri cortical surface reconstruction. 5:162–176.

[6] David O, Friston KJ (2003) A neural mass model for MEG/EEG: coupling and neuronal dynamics. *NeuroImage*, 20(3):1743–1755.

[7] Frackowiak RSJ, Friston KJ, Frith C, Dolan R, Price CJ, Zeki S, Ashburner J, Penny WD (2003) *Human Brain Function*, 2nd edition. New York: Academic Press.

[8] Friston KJ (2002) Bayesian estimation of dynamical systems: an application to fMRI. *NeuroImage*, 16:513–530.

[9] Friston KJ, Fletcher P, Josephs O, Holmes AP, Rugg MD, Turner R (1998) Event-related fMRI: characterizing differential responses. *NeuroImage*, 7:30–40.

[10] Friston KJ, Glaser DE, Henson RNA, Kiebel SJ, Phillips C, Ashburner J (2002) Classical and Bayesian inference in neuroimaging: applications. *NeuroImage*, 16:484–512.

[11] Friston KJ, Harrison L, Penny WD (2003) Dynamic causal modelling. *NeuroImage*, 19(4):1273–1302.

[12] Friston KJ, Holmes AP, Worsley KJ, Poline JB, Frith C, Frackowiak RSJ (1995) Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping*, 2:189–210.

[13] Friston KJ, Jezzard P, Turner R (1994) Analysis of functional MRI time-series. *Human Brain Mapping*, 1:153–171.

[14] Friston KJ, Penny WD (2003) Posterior probability maps and SPMs. *NeuroImage*, 19(3):1240–1249.

[15] Friston KJ, Penny WD, Phillips C, Kiebel SJ, Hinton G, Ashburner J (2002) Classical and Bayesian inference in neuroimaging: theory. *NeuroImage*, 16:465–483.

[16] Henson RNA, Rugg MD, Friston KJ (2001) The choice of basis functions in event-related fMRI. *NeuroImage*, 13(6):127, June. Supplement 1.

[17] Kiebel SJ, Goebel R, Friston KJ (2000) Anatomically informed basis functions. *NeuroImage*, 11(6):656–667.

[18] Lee PM (1997) *Bayesian Statistics: An Introduction*, 2nd Edition. London: Arnold.

[19] Melie-Garcia L, Trujillo-Barreto N, Martinez-Montes E, Valdes-Sosa P (1999) A symmetrical bayesian model for fMRI and EEG/MEG neuroimage fusion. *International Journal of Biolelectromagnetism*, 3(1).

[20] Penny WD, Flandin G, Trujillo-Bareto N (2005) Bayesian comparison of spatially regularised general linear models. *Human Brain Mapping*.

[21] Penny WD, Stephan KE, Mechelli A, Friston KJ (2004) Comparing dynamic causal models. *NeuroImage*, 22(3):1157–1172.

[22] Penny WD, Stephan KE, Mechelli A, Friston KJ (2004) Modelling functional integration: a comparison of structural equation and dynamic causal and models. *NeuroImage*, 23:264-274.

[23] Penny WD, Trujillo-Barreto N, Friston KJ (2005) Bayesian fMRI time series analysis with spatial priors. *NeuroImage*, 24(2):350–362.

[24] Phillips C, Rugg MD, Friston KJ (2002) Systematic regularization of linear inverse solutions of the EEG source localization problem. *NeuroImage*, 17(1):287–301.

[25] Rao CR, Toutenberg H (1995) *Linear Models: Least Squares and Alternatives*. New York: Springer-Verlag.

[26] Sato M, Yoshioka T, Kajiwara S, Toyoma K, Goda N, Doya K, Kawato (2004) Hierarchical Bayesian estimation for MEG inverse problem. *Neuroimage*, 23:806–826.

[27] Zeki S, Lueck CJ, Friston KJ, Kennard C, Watson JD, Frackowiak RS (1991) A direct demonstration of functional specialization in human visual cortex. *Journal of Neuroscience*, 11(3):641–649.