CHAPTER

2

Statistical parametric mapping

K. Friston

INTRODUCTION

This chapter summarizes the ideas and procedures used in the analysis of brain imaging data. It provides sufficient background to understand the principles of experimental design and data analysis and serves to introduce the main themes covered by subsequent chapters. These chapters have been organized into six parts. The first three parts follow the key stages of analysis: image transformations, modelling, and inference. These parts focus on identifying, and making inferences about, regionally specific effects in the brain. The final three parts address biophysical models of distributed neuronal responses, closing with analyses of functional and effective connectivity.

Characterizing a regionally specific effect rests on estimation and inference. Inferences in neuroimaging may be about differences expressed when comparing one group of subjects to another or, within subjects, changes over a sequence of observations. They may pertain to structural differences (e.g. in voxel-based morphometry) (Ashburner and Friston, 2000) or neurophysiological measures of brain functions (e.g. fMRI or functional magnetic resonance imaging). The principles of data analysis are very similar for all of these applications and constitute the subject of this and subsequent chapters. We will focus on the analysis of fMRI time-series because this covers many of the issues that are encountered in other modalities. Generally, the analyses of structural and PET (positron emission tomography) data are simpler because they do not have to deal with correlated errors from one scan to the next. Conversely, EEG and MEG (electro- and magnetoencephalography) present special problems for model inversion, however, many of the basic principles are shared by fMRI and EEG, because they are both caused by distributed neuronal dynamics. This chapter focuses on the design and analysis of neuroimaging studies. In the next chapter, we will look at conceptual and mathematical models that underpin the operational issues covered here.

Background

Statistical parametric mapping is used to identify regionally specific effects in neuroimaging data and is a prevalent approach to characterizing functional anatomy, specialization and disease-related changes. The complementary perspective, namely functional integration, requires a different set of approaches that examine the relationship among changes in one brain region relative to changes in others. Statistical parametric mapping is a voxel-based approach, employing topological inference, to make some comment about regionally specific responses to experimental factors. In order to assign an observed response to a particular brain structure, or cortical area, the data are usually mapped into an anatomical space. Before considering statistical modelling, we deal briefly with how images are realigned and normalized into some standard anatomical space. The general ideas behind statistical parametric mapping are then described and illustrated with attention to the different sorts of inferences that can be made with different experimental designs.

EEG, MEG and fMRI data lend themselves to a signal processing perspective. This can be exploited to ensure that both the design and analysis are as efficient as possible. Linear time invariant models provide the bridge between inferential models employed by statistical mapping and conventional signal processing approaches. We will touch on these and develop them further in the next chapter. Temporal autocorrelations in noise processes represent another important issue, especially in fMRI, and approaches to maximizing efficiency in the context of serially correlated errors will be discussed. We will also consider event and epoch-related designs in terms of efficiency. The chapter closes by looking at the distinction

Statistical Parametric Mapping, by Karl Friston et al.
ISBN–13: 978-0-12-372560-8 ISBN–10: 0-12-372560-7

between fixed and random-effect analyses and how this relates to inferences about the subjects studied or the population from which these subjects came.

In summary, this chapter reviews the three main stages of data analysis: spatial or image transforms, modelling and inference; these are the areas covered in the first three parts of this book and are summarized schematically in Plate 1 (see colour plate section). We then look at experimental design in light of the models covered in earlier parts. The next chapter deals with different models of distributed responses and previews the material covered in the final three parts of this book.

SPATIAL TRANSFORMS AND COMPUTATIONAL ANATOMY

A central theme in this book is the inversion of forward or generative models of how data are caused. We will see this in many different contexts, from the inversion of linear models of fMRI time-series to the inversion of dynamic causal models of distributed EEG responses. Image reconstruction, in imaging modalities like PET and fMRI, can be regarded as inverting a forward model of how signals, deployed in anatomical space, conspire to produce measured signals. In other modalities, like EEG and MEG, this inversion, or source reconstruction, can be a substantial problem in its own right. In most instances, it is expedient to decompose the inversion of forward spatiotemporal models into spatial and temporal parts. Operationally, this corresponds to reconstructing the spatial signal at each time point and then inverting a temporal model of the time-series at each spatial source (although we will consider full spatiotemporal models in Chapters 25 and 26). This view of source or image reconstruction as model inversion can be extended to cover the inversion of anatomical models describing anatomical variation within and between subjects. The inversion of these models corresponds to registration and normalization respectively. The aim of these anatomical inversions or transformations is to remove or characterize anatomical differences. Chapters 4 to 6 deal with the inversion of anatomical models for imaging modalities. Figure 2.1 shows an example of a generative model for structural images that is presented in Chapter 6. Chapters 28 and 29 deal with the corresponding inversion for EEG and MEG data.

This inversion corresponds to a series of spatial transformations that try to reduce unwanted variance components in the voxel time-series. These components are induced by movement or shape differences among a series of scans. Voxel-based analyses assume that data from a particular voxel derive from the same part of the brain. Violations of this assumption will introduce artefactual changes in the time-series that may obscure changes, or differences, of interest. Even single-subject analyses usually proceed in a standard anatomical space, simply to enable reporting of regionally-specific effects in a frame of reference that can be related to other studies. The first step is to realign the data to undo the effects of subject movement during the scanning session (see Chapter 4). After realignment, the data are then transformed using linear or non-linear warps into a standard anatomical space (see Chapters 5 and 6). Finally, the data are usually spatially smoothed before inverting the temporal part of the model.

Realignment

Changes in signal intensity over time, from any one voxel, can arise from head motion and this represents a serious confound, particularly in fMRI studies. Despite restraints on head movement, cooperative subjects still show displacements of up several millimetres. Realignment involves estimating the six parameters of an affine 'rigid-body' transformation that minimizes the differences between each successive scan and a reference scan (usually the first or the average of all scans in the time series). The transformation is then applied by resampling the data using an interpolation scheme. Estimation of the affine transformation is usually effected with a first-order approximation of the Taylor expansion of the effect of movement on signal intensity using the spatial derivatives of the images (see below). This allows for a simple iterative least square solution that corresponds to a Gauss-Newton search (Friston *et al*., 1995a). For most imaging modalities this procedure is sufficient to realign scans to, in some instances, a hundred microns or so (Friston *et al*., 1996a). However, in fMRI, even after perfect realignment, movement-related signals can still persist. This calls for a further step in which the data are *adjusted* for residual movement-related effects.

Adjusting for movement-related effects

In extreme cases, as much as 90 per cent of the variance in fMRI time-series can be accounted for by the effects of movement *after* realignment (Friston *et al*., 1996a). Causes of these movement-related components are due to movement effects that cannot be modelled using a linear model. These non-linear effects include: subject movement between slice acquisition, interpolation artefacts (Grootoonk *et al*., 2000), non-linear distortion due to magnetic field inhomogeneities (Andersson *et al*., 2001) and spin-excitation history effects (Friston *et al*., 1996a). The

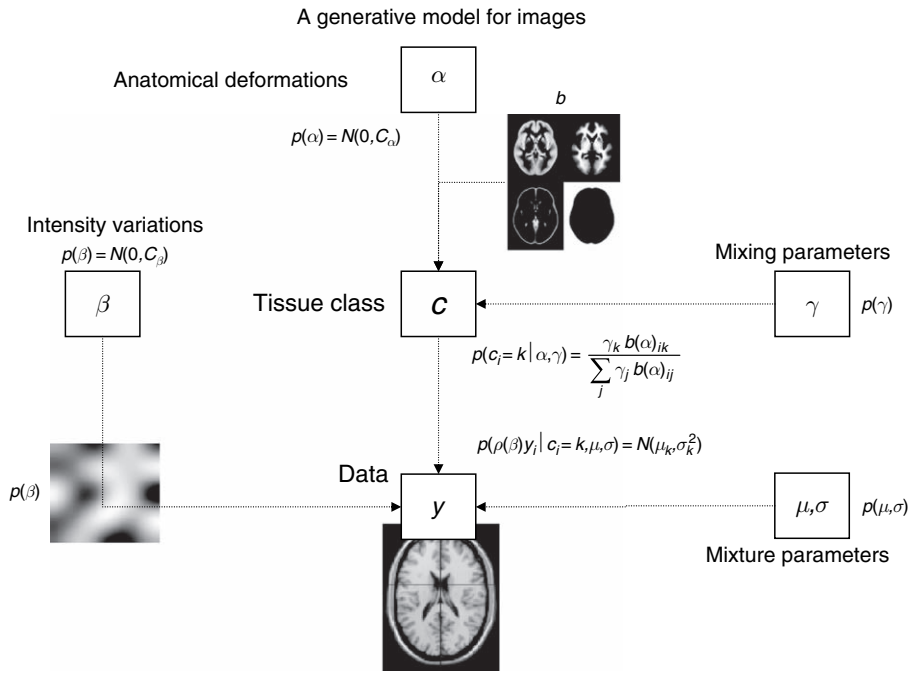

FIGURE 2.1 A graphical model describing the generation of an image. The boxes or 'nodes' represent quantities required to generate an image and the lines or 'edges' encode conditional dependencies among these quantities. This graphical description is a useful way to describe a generative model and makes all the conditional dependencies explicit. In this example, one starts by sampling some warping parameters α from their prior density $p(\alpha)$. These are used to resample (i.e. warp) a series of tissue-class maps to give $b(\alpha)_{ik}$ for each voxel and tissue class. The warping parameters model subject-specific anatomical deformations. Mixing parameters γ are then selected from their prior density $p(\gamma)$; these control the relative proportions of different tissue-classes over the brain. The mixing parameters scale the tissue-class maps to provide a density from which a voxel-specific tissue-class c_i is sampled. This specifies a mixture of Gaussians from which the voxel intensity is sampled. This mixture is specified in terms of the expectations μ and variances σ of their constituent Gaussians that are sampled from the prior density $p(\mu, \sigma)$. The final stage of image construction is to scale the voxel values with some slowly varying intensity field whose parameters β are sampled from their prior $p(\beta)$. The resulting image embodies random effects expressed at the level of anatomical deformation, amount of different tissue types, the expression of those tissues in the measurement, and image-specific inhomogeneities. Inversion of this generative model implicitly corrects for intensity variations, classifies each voxel probabilistically (i.e. segments) and spatially normalizes the image. Critically, this inversion accounts properly for all the conditional dependencies among the model's parameters and provides the most likely estimates given the data (see Chapter 6 for details of this model and its inversion).

latter can be pronounced if the repetition time approaches T1 making the current signal a function of movement history. These effects can render the movement-related signal a non-linear function of displacement in the n-th and previous scans:

$$y_n = f(x_n, x_{n-1}, \ldots)$$

By assuming a sensible form for this function, one can include these effects in the temporal model, so that they are explained away when making inferences about activations. This relies on accurate displacement estimates from the realignment and assumes activations are not correlated with the movements (any component that is correlated will be explained away).

The form for $f(x_n, x_{n-1}, \ldots)$, proposed in Friston *et al.* (1996a), was a non-linear autoregression model that used polynomial expansions to second order. This model was motivated by spin-excitation history effects and allowed displacement in previous scans to explain movement-related signal in the current scan. However, it is also a reasonable model for other sources of movement-related confounds. Generally, for repetition times (TR) of several seconds, interpolation artefacts supersede (Grootoonk *et al.*, 2000) and first-order terms, comprising an expansion of the current displacement in terms of periodic basis functions, are sufficient.

This section has considered *spatial* realignment. In multislice acquisition, different slices are acquired at different times. This raises the possibility of *temporal* realignment to ensure that the data from any given volume were sampled at the same time. This is usually performed using interpolation over time and only when the TR is sufficiently small to permit interpolation. Generally, timing effects of this sort are not considered problematic because they manifest as artefactual latency differences in evoked responses from region to region. Given that biological latency differences are in the order of a few seconds, inferences about

these differences are only made when comparing different trial types at the *same* voxel. Provided the effects of latency differences are modelled (see Chapter 14) temporal realignment is unnecessary in most applications.

Spatial normalization

In realignment, the generative model for within-subject movements is a rigid-body displacement of the first image. The generative model for spatial normalization is a canonical image or template that is distorted to produce a subject-specific image. Spatial normalization inverts this model by undoing the warp using a template-matching procedure. We focus on this simple model here, but note that more comprehensive models can be adopted (see Figure 2.1 and Chapter 6).

After realigning the data, a mean image of the series, or some other co-registered (e.g. a T1-weighted) image, is used to estimate some warping parameters that map it onto a template that already conforms to some standard anatomical space (e.g. Talairach and Tournoux, 1988). This estimation can use a variety of models for the mapping, including: a twelve-parameter affine transformation, where the parameters constitute a spatial transformation matrix; low-frequency basis functions, usually a discrete cosine set or polynomials, where the parameters are the coefficients of the basis functions employed; or a vector field specifying the mapping for each control point (e.g. voxel). In the latter case, the parameters are vast in number and constitute a vector field that is bigger than the image itself. Estimation of the parameters of all these models can be accommodated in a Bayesian framework, in which one is trying to find the warping parameters θ that have the maximum posterior probability $p(\theta|y)$ given the data y, where $p(\theta|y)p(y) = p(y|\theta)p(\theta)$. Put simply, one wants to find the deformation that is most likely given the data. This deformation can be found by maximizing the probability of getting the data, given the current parameters, times the probability of those parameters. In practice, the deformation is updated iteratively using a Gauss-Newton scheme to maximize $p(\theta|y)$. This involves jointly minimizing the likelihood and prior potentials $H(y|\theta) = \ln p(y|\theta)$ and $H(\theta) = \ln p(\theta)$. The likelihood potential is generally taken to be the sum of squared differences between the template and deformed image and reflects the probability of actually getting that image if the transformation was correct. The prior potential can be used to incorporate prior information or constraints on the warp. Priors can be determined empirically or motivated by constraints on the mappings. Priors play a more essential role as the number of parameters specifying the mapping increases and are central to high-dimensional warping schemes (Ashburner *et al.*, 1997 and see Chapter 5).

In practice, most people use an affine or spatial basis function warps and iterative least squares to minimize the posterior potential. A nice extension of this approach is that the likelihood potential can be refined and taken as the difference between the index image and a mixture of templates (e.g. depicting grey, white and skull tissue partitions). This models intensity differences that are unrelated to registration differences and allows different modalities to be co-registered (see Friston *et al.*, 1995a; Figure 2.2).

A special consideration is the spatial normalization of brains that have gross anatomical pathology. This pathology can be of two sorts: quantitative changes in the amount of a particular tissue compartment (e.g. cortical atrophy), or qualitative changes in anatomy involving the insertion or deletion of normal tissue compartments (e.g. ischaemic tissue in stroke or cortical dysplasia). The former case is, generally, not problematic in the sense that changes in the amount of cortical tissue will not affect its optimum spatial location in reference to some template (and, even if it does, a disease-specific template is easily constructed). The second sort of pathology can introduce bias in the normalization (because the generative model does not have a lesion) unless special precautions are taken. These usually involve imposing constraints on the warping to ensure that the pathology does not bias the deformation of undamaged tissue. This involves hard constraints implicit in using a small number of basis functions or soft constraints implemented by increasing the role of priors in Bayesian estimation. This can involve decreasing the precision of the data in the region of pathology so that more importance is afforded to the priors (cf. lesion masking). An alternative strategy is to use another modality that is less sensitive to the pathology as the basis of the spatial normalization procedure.

Registration of functional and anatomical data

It is sometimes useful to co-register functional and anatomical images. However, with echo-planar imaging, geometric distortions of T2* images, relative to anatomical T1-weighted data, can be a serious problem because of the very low frequency per point in the phase encoding direction. Typically, for echo-planar fMRI, magnetic field inhomogeneity, sufficient to cause de-phasing of 2π through the slice, corresponds to an in-plane distortion of a voxel. Un-warping schemes have been proposed to correct for the distortion effects (Jezzard and Balaban, 1995). However, this distortion is not an issue if one spatially normalizes the functional data.

Spatial smoothing

The motivations for smoothing the data are fourfold. By the matched filter theorem, the optimum

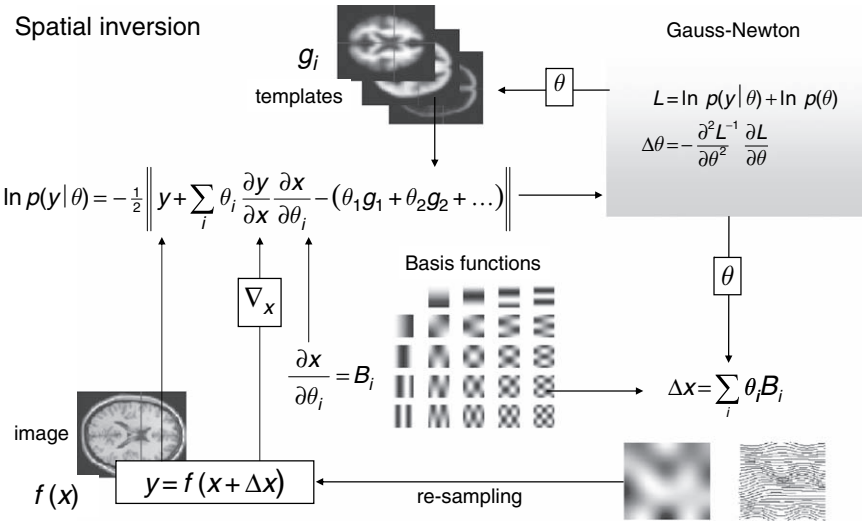

FIGURE 2.2 Schematic illustrating a Gauss-Newton scheme for maximizing the posterior probability $p(\theta|y)$ of the parameters required to spatially normalize an image. This scheme is iterative. At each step, the conditional estimate of the parameters obtains by jointly minimizing the likelihood and the prior potentials. The former is the difference between a resampled (i.e. warped) version $y = f(x + \Delta x)$ of the image $f(x)$ and the best linear combination of some templates $g(x)_1, \ldots$. These parameters are used to mix the templates and resample the image to reduce progressively both the spatial and intensity differences. After convergence, the resampled image can be considered normalized.

smoothing kernel corresponds to the size of the effect that one anticipates. The spatial scale of haemodynamic responses is, according to high-resolution optical imaging experiments, about 2–5 mm. Despite the potentially high resolution afforded by fMRI, an equivalent smoothing is suggested for most applications. By the central limit theorem, smoothing the data will render the errors more normal in their distribution and ensure the validity of inferences based on parametric tests. When making inferences about regional effects using random field theory (see below) the assumption is that the error terms are a reasonable lattice representation of an underlying continuous scalar field. This necessitates smoothness to be substantially greater than voxel size. If the voxels are large, then they can be reduced by sub-sampling the data and smoothing (with the original point spread function) with little loss of intrinsic resolution. In the context of inter-subject averaging it is often necessary to smooth more (e.g. 8 mm in fMRI or 16 mm in PET) to project the data onto a spatial scale where homologies in functional anatomy are expressed among subjects.

Summary

Spatial registration and normalization can proceed at a number of spatial scales depending on how one parameterizes variations in anatomy. We have focused on the role of normalization to remove unwanted differences to enable subsequent analysis of the data. However, it is important to note that the products of spatial normalization are twofold: a spatially normalized image and a deformation field (Plate 2). This deformation field contains important information about anatomy, in relation to the template used in the normalization procedure. The analysis of this information forms a key part of computational neuroanatomy. The tensor fields can be analysed directly (deformation-based morphometry – Ashburner *et al.*, 1998; Chung *et al.*, 2001) or used to create maps of specific anatomical attributes (e.g. compression, shears etc.). These maps can then be analysed on a voxel by voxel basis (tensor-based morphometry). Finally, the normalized structural images can themselves be subject to statistical analysis after some suitable segmentation procedure. This is known as *voxel-based morphometry*. Voxel-based morphometry is the most commonly used voxel-based neuroanatomical procedure and can easily be extended to incorporate tensor-based approaches (see Chapters 6 and 7).

STATISTICAL PARAMETRIC MAPPING AND THE GENERAL LINEAR MODEL

Functional mapping studies are usually analysed with some form of statistical parametric mapping. Statistical parametric mapping entails the construction of continuous statistical processes to test hypotheses about regionally specific effects (Friston *et al.*, 1991). Statistical parametric maps (SPMs) are images or fields with

values that are, under the null hypothesis, distributed according to a known probability density function, usually the Student's *t* or *F*-distributions. These are known colloquially as *t*- or *F*-maps. The success of statistical parametric mapping is due largely to the simplicity of the idea. Namely, one analyses each and every voxel using any standard (univariate) statistical test. The resulting statistical parameters are assembled into an image – the SPM. SPMs are interpreted as continuous statistical processes by referring to the probabilistic behaviour of random fields (Adler, 1981; Worsley *et al*., 1992, 1996; Friston *et al*., 1994). Random fields model both the univariate probabilistic characteristics of an SPM and any non-stationary spatial covariance structure. 'Unlikely' topological features of the SPM, like peaks, are interpreted as regionally specific effects, attributable to the experimental manipulation.

Over the years, statistical parametric mapping has come to refer to the conjoint use of *the general linear model* (GLM) and *random field theory* (RFT) theory to analyse and make classical inferences about topological features of the statistical parametric maps (SPM). The GLM is used to estimate some parameters that explain continuous data in exactly the same way as in conventional analyses of discrete data (see Part 2). RFT is used to resolve the multiple comparison problem that ensues when making inferences over the volume analysed (see Part 3). RFT provides a method for adjusting *p*-values for the search volume and plays the same role for continuous data (i.e. images) as the Bonferroni correction for a number of discontinuous or discrete statistical tests.

The approach was called SPM for three reasons:

1 To acknowledge *s*ignificance *p*robability *m*apping, where interpolated pseudo-maps of *p*-values are used to summarize the analysis of multichannel event-related potential (ERP) studies.
2 For consistency with the nomenclature of parametric maps of physiological or physical parameters (e.g. parametric maps of regional cerebral blood flow (rCBF) or volume).
3 In reference to the *parametric* statistics that populate the maps.

Despite its simplicity, there are some fairly subtle motivations for the approach that deserve mention. Usually, given a response or dependent variable, comprising many thousands of voxels, one would use *multivariate* analyses as opposed to the *mass-univariate* approach that SPM represents. The problems with multivariate approaches are that:

1 they do not support inferences about regionally specific effects (i.e. topological features with a unique localizing attribute)
2 they require more observations than the dimension of the response variable (i.e. need more scans than voxels)
3 even in the context of dimension reduction, they are less sensitive to focal effects than mass-univariate approaches.

A heuristic, for their relative lack of power, is that multivariate approaches estimate the model's error covariances using lots of parameters (e.g. the covariance between the errors at all pairs of voxels). Conversely, SPM characterizes spatial covariance with a smoothness parameter, for each voxel. In general, the more parameters (and hyperparameters) an estimation procedure has to deal with, the more variable the estimate of any one parameter becomes. This renders inferences about any single estimate less efficient.

Multivariate approaches consider voxels as different levels of an experimental or treatment factor and use classical analysis of variance, not at each voxel but by considering the data sequences from all voxels together, as replications over voxels. The problem here is that regional changes in error variance, and spatial correlations in the data, induce profound non-sphericity[1] in the error terms. This non-sphericity would again require large numbers of parameters to be estimated for each voxel using conventional techniques. In SPM, the non-sphericity is parameterized in a parsimonious way with just two parameters for each voxel. These are the error variance and smoothness estimators (see Part 3). This minimal parameterization lends SPM a sensitivity that surpasses multivariate approaches. SPM can do this because RFT implicitly harnesses constraints on the non-sphericity implied by the continuous (i.e. analytic) nature of the data. This is something that conventional multivariate and equivalent univariate approaches cannot accommodate, to their cost.

Some analyses use statistical maps based on non-parametric tests that eschew distributional assumptions about the data (see Chapter 21). These approaches are generally less powerful (i.e. less sensitive) than parametric approaches (see Aguirre *et al*., 1998). However, they have an important role in evaluating the assumptions

[1] Sphericity refers to the assumption of identically and independently distributed error terms (IID). Under IID assumptions the probability density function of the errors, from all observations, has spherical iso-contours, hence *sphericity*. Deviations from either of the IID criteria constitute non-sphericity. If the error terms are not identically distributed then different observations have different error variances. Correlations among errors reflect dependencies among the error terms (e.g. serial correlation in fMRI time series) and constitute the second component of non-sphericity. In neuroimaging both spatial and temporal non-sphericity can be quite profound.

behind parametric approaches and may supervene in terms of sensitivity when these assumptions are violated (e.g. when degrees of freedom are very small and voxel sizes are large in relation to smoothness).

In Part 4 we consider Bayesian alternatives to classical inference with SPMs. This rests on conditional inferences about an effect, given the data, as opposed to classical inferences about the data, given the effect is zero. Bayesian inferences on continuous fields or images use posterior probability maps (PPMs). Although less commonly used than SPMs, PPMs are potentially useful, not least because they do not have to contend with the multiple comparisons problem induced by classical inference. In contradistinction to SPM, this means that inferences about a given regional response do not depend on inferences about responses elsewhere. Next we consider parameter estimation in the context of the GLM. This is followed by an introduction to the role of RFT when making classical inferences about continuous data.

The general linear model

Statistical analysis of imaging data corresponds to inverting generative models of the data to partition observed responses into components of interest, confounds and error. Inferences are then pursued using statistics that compare interesting effects and the error. This classical inference can be regarded as a direct comparison of the variance due to an interesting experimental manipulation with the error variance (compare with the *F*-statistic and other likelihood ratios). Alternatively, one can view the statistic as an estimate of the response, or difference of interest, divided by an estimate of its standard deviation. This is a useful way to think about the *t*-statistic.

A brief review of the literature may give the impression that there are numerous ways to analyse PET and fMRI time-series with a diversity of statistical and conceptual approaches. This is not the case. With very few exceptions, every analysis is a variant of the general linear model. This includes simple *t*-tests on scans assigned to one condition or another, correlation coefficients between observed responses and boxcar stimulus functions in fMRI, inferences made using multiple linear regression, evoked responses estimated using linear time invariant models, and selective averaging to estimate event-related responses in fMRI. Mathematically, they are all formally identical and can be implemented with the same equations and algorithms. The only thing that distinguishes among them is the design matrix encoding the temporal model or experimental design. The use of the correlation coefficient deserves special mention because of its popularity in fMRI (Bandettini *et al*., 1993). The significance of a correlation is identical to the significance of the equivalent *t*-statistic testing for a regression of the data on a stimulus function. The correlation coefficient approach is useful but the inference is effectively based on a limiting case of multiple linear regression that obtains when there is only one regressor. In fMRI, many regressors usually enter a statistical model. Therefore, the *t*-statistic provides a more versatile and generic way of assessing the significance of regional effects and is usually preferred over the correlation coefficient.

The general linear model is an equation $Y = X\beta + \varepsilon$ that expresses the observed response variable in terms of a linear combination of explanatory variables X plus a well behaved error term (Figure 2.3 and Friston *et al*., 1995b). 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 that contains the explanatory variables (e.g. designed effects or confounds) is called the *design matrix*. Each column of the design matrix corresponds to an effect one has built into the experiment or that may confound the results. These are referred to as explanatory variables, covariates or regressors. The example in Plate 1 relates to an fMRI study of visual stimulation under four conditions. The effects on the response variable are modelled in terms of functions of the presence of these conditions (i.e. boxcars smoothed with a haemodynamic response function) and constitute the first four columns of the design matrix. There then follows a series of terms that are designed to remove or model low-frequency variations in signal due to artefacts such as aliased biorhythms and other drift terms. The final column is whole brain activity. The relative contribution of each of these columns is assessed using standard maximum likelihood and 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. The operational equations are depicted schematically in Figure 2.3. In this scheme, the general linear model has been extended (Worsley and Friston, 1995) to incorporate intrinsic non-sphericity, or correlations among the error terms, and to allow for some temporal filtering of the data with the matrix S. This generalization brings with it the notion of *effective degrees of freedom*, which are less than the conventional degrees of freedom under IID assumptions (see footnote). They are smaller because the temporal correlations reduce the effective number of independent observations. The statistics are constructed using the approximation of Satterthwaite. This is the same approximation used in classical non-sphericity corrections such as the Geisser-Greenhouse correction. However, in the

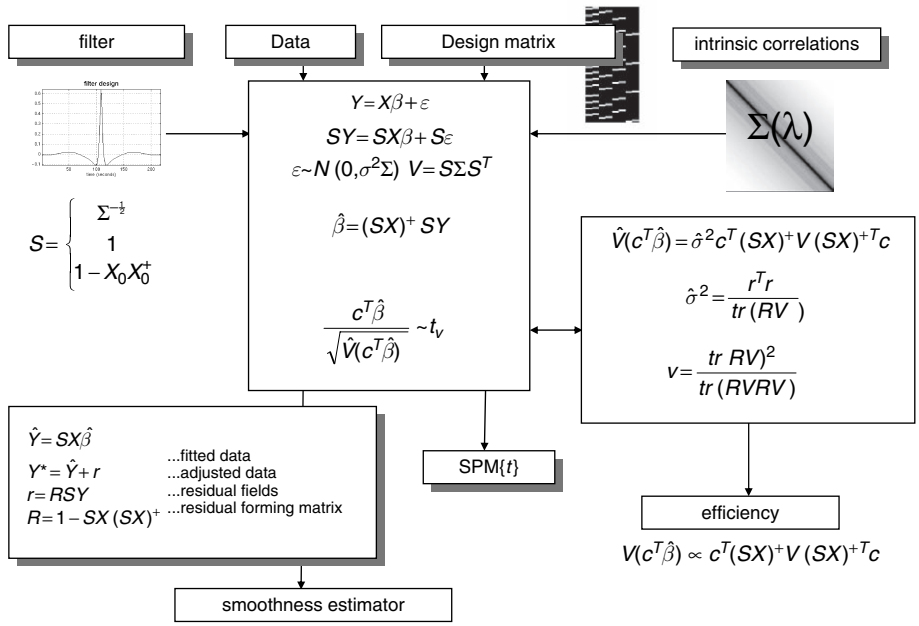

FIGURE 2.3 The general linear model. The general linear model is an equation expressing the response variable Y in terms of a linear combination of explanatory variables in a design matrix X and an error term with assumed or known autocorrelation Σ. The data can be filtered with a convolution or residual forming matrix (or a combination) S, leading to a generalized linear model that includes [intrinsic] serial correlations and applied [extrinsic] filtering. Different choices of S correspond to different estimation schemes. The parameter estimates obtain in a least squares sense using the pseudo-inverse (denoted by $+$) of the filtered design matrix. Generally, an effect of interest is specified by a vector of contrast weights c that give a weighted sum or compound of parameter estimates referred to as a *contrast*. The t-statistic is simply this contrast divided by its standard error (i.e. square root of its estimated variance). The ensuing t-statistic is distributed with v degrees of freedom. The equations for estimating the variance of the contrast and the degrees of freedom are provided in the right-hand panel. Efficiency is simply the inverse of the variance of the contrast. These expressions are useful when assessing the relative efficiency of different designs encoded in X. The parameter estimates can be examined directly or used to compute the fitted responses (see lower left panel). Adjusted data refer to data from which fitted components (e.g. confounds) have been removed. The residuals r, obtain from applying the residual-forming matrix R to the data. These residual fields are used to estimate the smoothness of the component fields of the SPM and are needed by random field theory (see Figure 2.4).

Worsley and Friston (1995) scheme, this approximation is used to construct the statistics and appropriate degrees of freedom, not simply to provide a *post hoc* correction to the degrees of freedom.

There is a special and important case of temporal filtering. This is when the filtering de-correlates (i.e. whitens) the error terms by using $S = \Sigma^{-1/2}$. This is the filtering scheme used in current implementations of the SPM software and renders the ordinary least squares (OLS) parameter estimates maximum likelihood (ML) estimators. These are optimal in the sense that they are the minimum variance estimators of all unbiased estimators. The estimation of $S = \Sigma^{-1/2}$ uses expectation maximization (EM) to provide restricted maximum likelihood (ReML) estimates of $\Sigma = \Sigma(\lambda)$ in terms of hyperparameters λ corresponding to variance components (see Chapter 11 and Chapter 24 for an explanation of EM). In this case, the effective degrees of freedom revert to the maximum that would be attained in the absence of temporal correlations or non-sphericity.

Contrasts

The equations summarized in Figure 2.3 can be used to implement a vast range of statistical analyses. The issue is therefore not the mathematics but the formulation of a design matrix appropriate to the study design and inferences that are sought. The design matrix can contain both covariates and indicator variables. Each column has an associated unknown or free parameter β. Some of these parameters will be of interest (e.g. the effect of a particular sensorimotor or cognitive condition or the regression coefficient of haemodynamic 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 or the regression slope of voxel activity on global activity). Inferences about the parameter estimates are made using their estimated variance. This allows one to test the null hypothesis, that all the estimates are zero, using the F-statistic to give an SPM{F} or that some particular linear combination (e.g. a subtraction) of the estimates is zero using an SPM{t}. The

t-statistic obtains by dividing a contrast or compound (specified by contrast weights) of the ensuing parameter estimates by the standard error of that compound. The latter is estimated using the variance of the residuals about the least-squares fit. An example of a contrast weight *vector* would be [−1 1 0. . . .] to compare the difference in responses evoked by two conditions, modelled 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 (Büchel *et al*., 1996) or basis function expansions of some experimental factor. In these instances, the SPM{*F*} is used and is specified with a *matrix* of contrast weights that can be thought of as a collection of '*t*-contrasts' (see Chapter 9 for a fuller explanation). An *F*-contrast may look like:

$$\begin{bmatrix} -1 & 0 & 0 & 0 & \dots \\ 0 & 1 & 0 & 0 & \dots \end{bmatrix}$$

This would test for the significance of the first *or* second parameter estimates. The fact that the first weight is negative has no effect on the test because the *F*-statistic is based on sums of squares.

In most analyses, the design matrix contains indicator variables or parametric variables encoding the experimental manipulations. These are formally identical to classical analysis of covariance (i.e. ANCOVA) models. An important instance of the GLM, from the perspective of fMRI, is the linear time-invariant (LTI) model. Mathematically, this is no different from any other GLM. However, it explicitly treats the data-sequence as an ordered time-series and enables a signal processing perspective that can be very useful (see next chapter and Chapter 14).

TOPOLOGICAL INFERENCE AND THE THEORY OF RANDOM FIELDS

Classical inferences using SPMs can be of two sorts, depending on whether one knows where to look in advance. With an anatomically constrained hypothesis, about effects in a particular brain region, the uncorrected *p*-value associated with the height or extent of that region in the SPM can be used to test the hypothesis. With an anatomically open hypothesis (i.e. a null hypothesis that there is no effect anywhere in a specified volume of the brain), a correction for multiple dependent comparisons is necessary. The theory of random fields provides a way of adjusting the *p*-value that takes into account the fact that neighbouring voxels are not independent by virtue of continuity in the original data. Provided the data are smooth the RFT adjustment is less severe (i.e. is more sensitive) than a Bonferroni correction for the number of voxels. As noted above, RFT deals with the multiple comparisons problem in the context of continuous, statistical fields, in a way that is analogous to the Bonferroni procedure for families of discrete statistical tests. There are many ways to appreciate the difference between RFT and Bonferroni corrections. Perhaps the most intuitive is to consider the fundamental difference between an SPM and a collection of discrete *t*-values. When declaring a peak or cluster of the SPM to be significant, we refer collectively to all the voxels associated with that feature. The false positive rate is expressed in terms of peaks or clusters, under the null hypothesis of no activation. This is not the expected false positive rate of voxels. One false positive peak may be associated with hundreds of voxels, if the SPM is very smooth. Bonferroni correction controls the expected number of false positive *voxels*, whereas RFT controls the expected number of false positive *peaks*. Because the number of peaks is always less than the number of voxels, RFT can use a lower threshold, rendering it much more sensitive. In fact, the number of false positive voxels is somewhat irrelevant because it is a function of smoothness. The RFT correction discounts voxel size by expressing the search volume in terms of smoothness or resolution elements (*Resels*) (Figure 2.4). This intuitive perspective is expressed formally in terms of differential topology using the *Euler characteristic* (Worsley *et al*., 1992). At high thresholds the Euler characteristic corresponds to the number of peaks above threshold.

There are only two assumptions underlying the use of the RFT correction:

1 the error fields (but not necessarily the data) are a reasonable lattice approximation to an underlying random field with a multivariate Gaussian distribution
2 these fields are continuous, with a differentiable and invertible autocorrelation function.

A common misconception is that the autocorrelation function has to be Gaussian. It does not. The only way in which these assumptions can be violated is if:

1 the data are not smooth, violating the reasonable lattice assumption or
2 the statistical model is mis-specified so that the errors are not normally distributed.

Early formulations of the RFT correction were based on the assumption that the spatial correlation structure was wide-sense stationary. This assumption can now be relaxed due to a revision of the way in which the smoothness estimator enters the correction procedure (Kiebel *et al*., 1999). In other words, the corrections retain their validity, even if the smoothness varies from voxel to voxel.

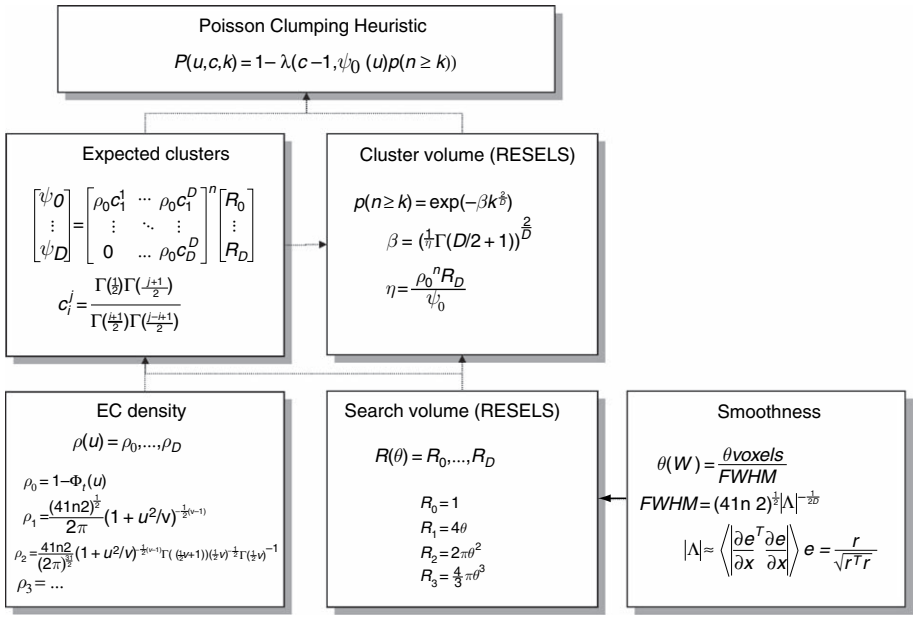

FIGURE 2.4 Schematic illustrating the use of random field theory (RFT) in making inferences about SPMs. If one knew precisely where to look, then inference can be based on the value of the statistic at the specified location in the SPM. However, generally, one does not have a precise anatomical *prior*, and an adjustment for multiple dependent comparisons has to be made to the *p*-values. These corrections use distributional approximations from RFT. This schematic deals with a general case of n SPM{t} whose voxels all survive a common threshold u (i.e. a conjunction of n component SPMs). The central probability, upon which all peak, cluster or set-level inferences are made, is the probability $P(u, c, k)$ of getting c or more clusters with k or more resels (resolution elements) above this threshold. By assuming that clusters behave like a multidimensional Poisson point-process (i.e. the Poisson clumping heuristic) $P(u, c, k)$ is determined simply; the distribution of c is Poisson with an expectation that corresponds to the product of the expected number of clusters, of any size, and the probability that any cluster will be bigger than k resels. The latter probability depends on the expected number of resels per cluster η. This is simply the expected suprathreshold volume, divided by the expected number of clusters. The expected number of clusters ψ_0 is estimated with the Euler characteristic (EC) (effectively the number of blobs minus the number of holes). This depends on the EC density for the statistic in question (with degrees of freedom v) and the resel counts. The EC density is the expected EC per unit of D-dimensional volume of the SPM where the volume of the search is given by the resel counts. Resel counts are a volume measure that has been normalized by the smoothness of the SPMs component fields, expressed in terms of the full width at half maximum (FWHM). This is estimated from the determinant of the variance-covariance matrix of the first spatial derivatives of e, the normalized residual fields r (from Figure 2.3). In this example equations for a sphere of radius θ are given. ϕ denotes the cumulative density function for the statistic in question. (See Appendix 6 for technical details.)

Anatomically closed hypotheses

When making inferences about regional effects (e.g. activations) in SPMs, one often has some idea about where the activation should be. In this instance, a correction for the entire search volume is inappropriate. However, a problem remains in the sense that one would like to consider activations that are 'near' the predicted location, even if they are not exactly coincident. There are two approaches one can adopt: pre-specify a small search volume and make the appropriate RFT correction (Worsley *et al*., 1996); or use the uncorrected *p*-value based on spatial extent of the nearest cluster (Friston, 1997). This probability is based on getting the observed number of voxels, or more, in a given cluster (conditional on that cluster existing). Both these procedures are based on distributional approximations from RFT.

Anatomically open hypotheses and levels of inference

To make inferences about regionally specific effects, the SPM is thresholded using some height and spatial extent thresholds that are specified by the user. Corrected *p*-values can then be derived that pertain to:

1 The number of activated regions (i.e. number of clusters above the height and volume threshold). These are *set-level inferences*.
2 The number of activated voxels (i.e. volume) comprising a particular region. These are *cluster-level inferences*.
3 The *p*-value for each peak within that cluster, i.e. *peak-level inferences*.

These *p*-values are corrected for the multiple dependent comparisons and are based on the probability of obtaining c, or more, clusters with k, or more, voxels,

above a threshold u in an SPM of known or estimated smoothness. This probability has a reasonably simple form (see Figure 2.4 for details).

Set-level refers to the inference that the number of clusters comprising an observed activation profile is highly unlikely to have occurred by chance and is a statement about the activation profile, as characterized by its constituent regions. Cluster-level inferences are a special case of set-level inferences that obtain when the number of clusters $c = 1$. Similarly, peak-level inferences are special cases of cluster-level inferences that result when the cluster can be small (i.e. $k = 0$). Using a theoretical power analysis (see Friston *et al*., 1996b and Chapter 19) of distributed activations, one observes that set-level inferences are generally more powerful than cluster-level inferences and that cluster-level inferences are generally more powerful than peak-level inferences. The price paid for this increased sensitivity is reduced localizing power. Peak-level tests permit individual maxima to be identified as significant, whereas cluster and set-level inferences only allow clusters or sets of clusters to be declared significant. It should be remembered that these conclusions, about the relative power of different inference levels, are based on distributed activations. Focal activation may well be detected with greater sensitivity using tests based on peak height. Typically, people use peak-level inferences and a spatial extent threshold of zero. This reflects the fact that characterizations of functional anatomy are generally more useful when specified with a high degree of anatomical precision.

EXPERIMENTAL AND MODEL DESIGN

This section considers the different sorts of designs that can be employed in neuroimaging studies. Experimental designs can be classified as *single factor* or *multifactor* designs; within this classification the levels of each factor can be *categorical* or *parametric*. We will start by discussing categorical and parametric designs and then deal with multifactor designs. We then move on to some more technical issues that attend the analysis of fMRI experiments. These are considered in terms of model design, using a signal processing perspective.

Categorical designs, cognitive subtraction and conjunctions

The tenet of cognitive subtraction is that the difference between two tasks can be formulated as a separable cognitive or sensorimotor component and that regionally specific differences in haemodynamic responses, evoked by the two tasks, identify the corresponding functionally selective area. Early applications of subtraction range from the functional anatomy of word processing (Petersen *et al*., 1989) to functional specialization in extrastriate cortex (Lueck *et al*., 1989). The latter studies involved presenting visual stimuli with and without some sensory attribute (e.g. colour, motion etc.). The areas highlighted by subtraction were identified with homologous areas in monkeys that showed selective electrophysiological responses to equivalent visual stimuli.

Cognitive conjunctions (Price and Friston, 1997) can be thought of as an extension of the subtraction technique, in the sense that they combine a series of subtractions. In subtraction, one tests a single hypothesis pertaining to the activation in one task relative to another. In conjunction analyses, several hypotheses are tested, asking whether the activations, in a series of task pairs, are collectively significant (cf. an F-test). Consider the problem of identifying regionally specific activations due to a particular cognitive component (e.g. object recognition). If one can identify a series of task pairs whose differences have only that component in common, then the region which activates, in all the corresponding subtractions, can be associated with the common component. Conjunction analyses allow one to demonstrate the context-invariant nature of regional responses. One important application of conjunction analyses is in multisubject fMRI studies, where generic effects are identified as those that are jointly significant in all the subjects studied (see below).

Parametric designs

The premise behind parametric designs is that regional physiology will vary systematically with the degree of cognitive or sensorimotor processing or deficits thereof. Examples of this approach include the PET experiments of Grafton *et al*. (1992) that demonstrated significant correlations between haemodynamic responses and the performance of a visually guided motor tracking task. On the sensory side, Price *et al*. (1992) demonstrated a remarkable linear relationship between perfusion in periauditory regions and frequency of aural word presentation. This correlation was not observed in Wernicke's area, where perfusion appeared to correlate, not with the discriminative attributes of the stimulus, but with the presence or absence of semantic content. These relationships or *neurometric functions* may be linear or non-linear. Using polynomial regression, in the context of the GLM, one can identify non-linear relationships between stimulus parameters (e.g. stimulus duration or presentation rate) and evoked responses. To do this one usually uses an SPM{F} (see Büchel *et al*., 1996).

The example provided in Figure 2.5 illustrates both categorical and parametric aspects of design and analysis. These data were obtained from an fMRI study of visual motion processing using radially moving dots. The stimuli were presented over a range of speeds using *isoluminant* and *isochromatic* stimuli. To identify areas involved in visual motion, a stationary dots condition was subtracted from the moving dots conditions (see the contrast weights in the upper right). To ensure significant motion-sensitive responses, using colour and luminance cues, a conjunction of the equivalent subtractions was assessed under both viewing contexts. Areas V5 and V3a are seen in the ensuing SPM{t}. The t-values in this SPM are simply the minimum of the t-values for each subtraction. Thresholding this SPM ensures that all voxels survive the threshold u in each subtraction separately. This *conjunction* SPM has an equivalent interpretation; it represents the intersection of the excursion sets, defined by the threshold u, of each *component* SPM. This intersection is the essence of a conjunction. The expressions in Figure 2.4 pertain to the general case of the minimum of n t-values. The special case where $n = 1$ corresponds to a conventional SPM{t}.

The responses in left V5 are shown in the lower panel of Figure 2.5 and speak to a compelling inverted 'U' relationship between speed and evoked response that peaks

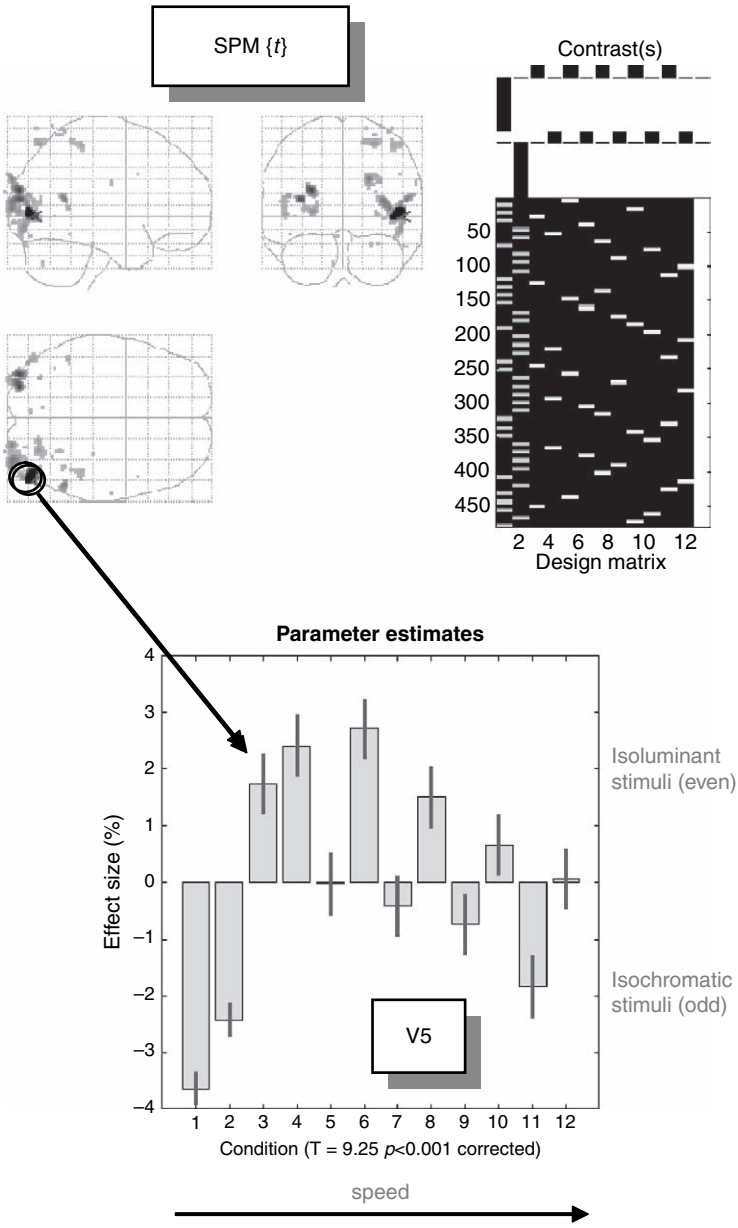

FIGURE 2.5 Top right: design matrix: this is an image representation of the design matrix. Contrasts: these are the vectors of contrast weights defining the linear compounds of parameters tested. The contrast weights are displayed over the column of the design matrix that corresponds to the effects in question. The design matrix here includes condition-specific effects (boxcar-functions convolved with a haemodynamic response function). Odd columns correspond to stimuli shown under isochromatic conditions and even columns model responses to isoluminant stimuli. The first two columns are for stationary stimuli and the remaining columns are for conditions of increasing speed. The final column is a constant term. Top left: SPM{t}: this is a maximum intensity projection conforming to the standard anatomical space of Talairach and Tournoux (1988). The values here are the minimum t-values from both contrasts, thresholded at $p = 0.001$ uncorrected. The most significant conjunction is seen in left V5. Lower panel: plot of the condition-specific parameter estimates for this voxel. The t-value was 9.25 (p <0.001 corrected – see Figure 2.4).

at around eight degrees per second. It is this sort of relationship that parametric designs try to characterize. Interestingly, the form of these speed-dependent responses was similar using both stimulus types, although luminance cues are seen to elicit a greater response. From the point of view of a factorial design there is a *main effect* of cue (isoluminant versus isochromatic), a main effect of speed, but no speed by cue *interaction*.

Clinical neuroscience studies can use parametric designs by looking for the neuronal correlates of clinical (e.g. symptom) ratings over subjects. In many cases, multiple clinical scores are available for each subject and the statistical design can usually be seen as a multilinear regression. In situations where the clinical scores are correlated, principal component analysis or factor analysis is sometimes applied to generate a new, and smaller, set of explanatory variables that are orthogonal to each other. This has proved particularly useful in psychiatric studies where syndromes can be expressed over a number of different dimensions (e.g. the degree of psychomotor poverty, disorganization and reality distortion in schizophrenia; see Liddle *et al*., 1992). In this way, regionally specific correlates of various symptoms may point to their distinct pathogenesis in a way that transcends the syndrome itself. For example, psychomotor poverty may be associated with left dorso-lateral prefrontal dysfunction, irrespective of whether the patient is suffering from schizophrenia or depression.

Factorial designs

Factorial designs are more prevalent than single-factor designs because they enable inferences about interactions. At its simplest, an interaction represents a change in a change. Interactions are associated with factorial designs where two or more factors are combined in the same experiment. The effect of one factor, on the effect of the other, is assessed by the interaction. Factorial designs have a wide range of applications. An early application, in neuroimaging, examined adaptation and plasticity during motor performance by assessing time by condition interactions (Friston *et al*., 1992a). Psychopharmacological activation studies are further examples of factorial designs (Friston *et al*., 1992b). In these studies, cognitively evoked responses are assessed before and after being given a drug. The interaction term reflects the pharmacological modulation of task-dependent activations. Factorial designs have an important role in the context of cognitive subtraction and additive factors logic by virtue of being able to test for interactions, or context-sensitive activations, i.e. to demonstrate the fallacy of pure-insertion (see Friston *et al*., 1996c). These interaction effects can sometimes be interpreted as the integration of the two or more [cognitive] processes or the modulation of one [perceptual] process by another. Figure 2.6 shows an example which takes an unusual perspective on the modulation of event-related responses as the interaction between stimulus presentation and experiential context.

From the point of view of clinical studies, interactions are central. The effect of a disease process on sensorimotor or cognitive activation is simply an interaction and involves replicating a subtraction experiment in subjects with and without the pathology. Factorial designs can also embody parametric factors. If one of the factors has a number of parametric levels, the interaction can be expressed as a difference in regression slope of regional activity on the parameter, under both levels of the other [categorical] factor. An important example of factorial designs, that mix categorical and parameter factors, are those looking for *psychophysiological interactions*. Here the parametric factor is brain activity measured in a particular brain region. These designs have proven useful in looking at the interaction between bottom-up and top-down influences within processing hierarchies in the brain (Friston *et al*., 1997). This issue will be addressed below and in Part 6, from the point of view of effective connectivity.

Designing fMRI studies

In this section, we consider fMRI time-series from a signal processing perspective with particular focus on optimal experimental design and efficiency. fMRI time-series can be viewed as a linear admixture of signal and noise. Signal corresponds to neuronally mediated haemodynamic changes that can be modelled as a convolution of some underlying neuronal process, responding to changes in experimental factors, by a haemodynamic response function. Noise has many contributions that render it rather complicated in relation to some neurophysiological measurements. These include neuronal and non-neuronal sources. Neuronal noise refers to neurogenic signal not modelled by the explanatory variables and has the same frequency structure as the signal itself. Non-neuronal components have both white (e.g. Johnson noise) and coloured components (e.g. pulsatile motion of the brain caused by cardiac cycles and local modulation of the static magnetic field by respiratory movement). These effects are typically low-frequency or wide-band (e.g. aliased cardiac-locked pulsatile motion). The superposition of all these components induces temporal correlations among the error terms (denoted by Σ in Figure 2.3) that can affect sensitivity to experimental effects. Sensitivity depends upon the relative amounts of signal and noise and the efficiency of the experimental design. Efficiency is simply a measure of how reliable the parameter

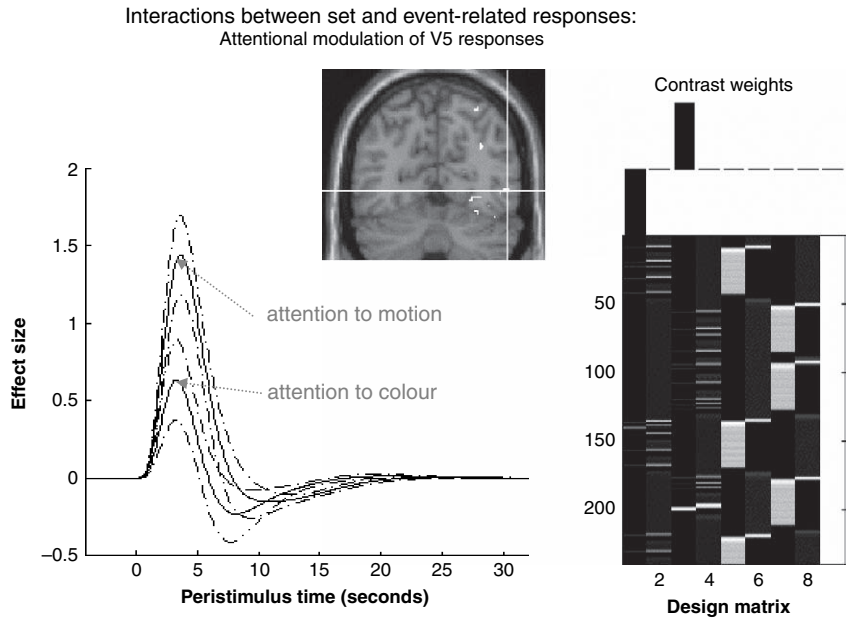

FIGURE 2.6 Results showing attentional modulation of visually-evoked responses. Subjects viewed stationary monochromatic stimuli that occasionally changed colour and moved at the same time. These compound events were presented under two levels of attentional set (attention to colour and attention to motion). The event-related responses are modelled, in an attention-specific fashion by the first four regressors (stick-functions convolved with a haemodynamic response function and its derivative) in the design matrix on the right. The main effects of attention are modelled as similarly convolved boxcars. The interaction between attentional set and visually evoked responses is simply the difference in evoked responses under both levels of attention and is tested for with the appropriate contrast weights (upper right). Only the first 256 rows of the design matrix are shown. The most significant modulation of evoked responses, under attention to motion, was seen in left V5 (insert). The fitted responses and their standard errors are shown on the left as functions of peristimulus time.

estimates are and can be defined as the inverse of the variance of a contrast of parameter estimates (see Figure 2.3). There are two important considerations that arise from this perspective on fMRI time-series: the first pertains to optimal experimental design and the second to optimum de-convolution of the time-series to obtain the most efficient parameter estimates.

The haemodynamic response function and optimum design

As noted above, an LTI model of neuronally mediated signals in fMRI suggests that only those experimentally induced signals that survive convolution with the haemodynamic response function (HRF) can be estimated with any efficiency. By convolution theorem, the frequency structure of experimental variance should therefore be designed to match the transfer function of the HRF. The corresponding frequency profile of this transfer function is shown in Figure 2.7 (solid line). It can be seen that frequencies around 0.03 Hz are optimal, corresponding to periodic designs with 32-second periods (i.e. 16-second epochs). Generally, the first objective of experimental design is to comply with the natural constraints imposed by the HRF and ensure that experimental variance occupies these intermediate frequencies.

Serial correlations and filtering

This is quite a complicated but important area. Conventional signal processing approaches dictate that whitening the data engenders the most efficient parameter estimation. This corresponds to filtering with a convolution matrix $S = K^{-1}$ that is the inverse of the intrinsic convolution matrix, i.e. $KK^T = \Sigma$ (see Figure 2.3). This whitening strategy renders the least-square estimator in Figure 2.3 equivalent to the ML or Gauss-Markov estimator. However, one generally does not know the form of the intrinsic correlations, which means they have to be estimated. This estimation usually proceeds using a restricted maximum likelihood (ReML) estimate of the serial correlations, among the residuals, that properly accommodates the effects of the residual-forming matrix and associated loss of degrees of freedom. However, using this estimate of the intrinsic non-sphericity to form a Gauss-Markov estimator at each voxel is not easy. First, the estimate of non-sphericity can itself be imprecise leading to bias in the standard error (Friston *et al.*, 2000). Second, ReML estimation requires a computationally

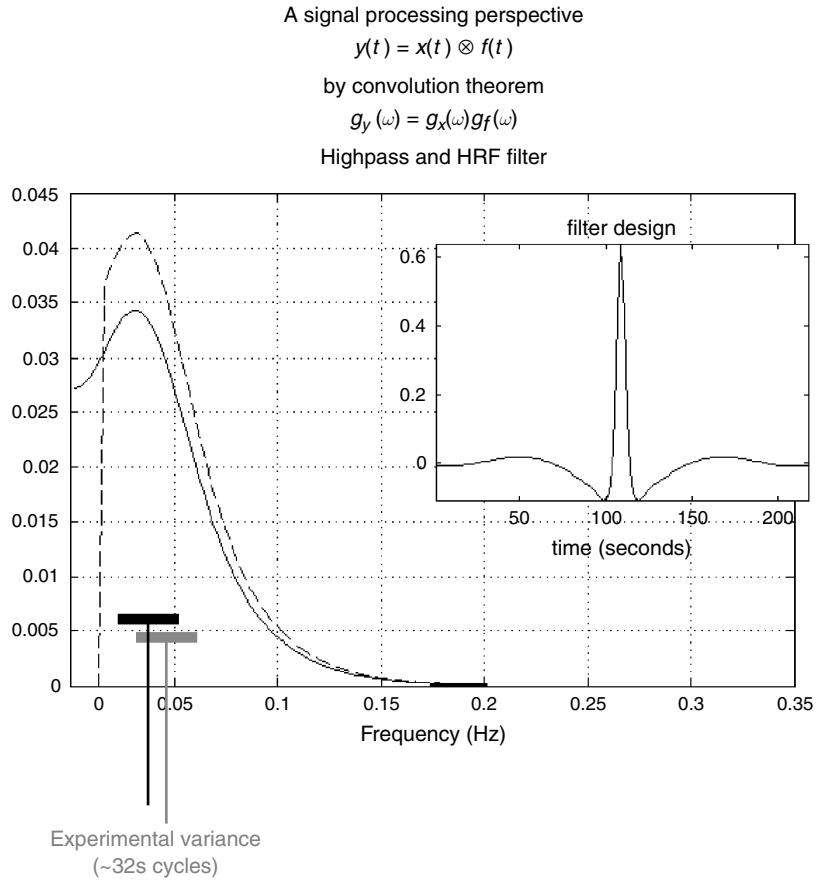

FIGURE 2.7 Transfer function of a canonical haemodynamic response function (HRF), with (broken line) and without (solid line) the application of a highpass filter. This transfer function corresponds to the spectral density of a white-noise process after convolution with the HRF and places constraints on the frequencies that survive haemodynamic convolution. This follows from convolution theorem (summarized in the equations). The insert is the filter expressed in time, corresponding to the spectral density that obtains after convolution with the HRF and highpass filtering.

prohibitive iterative procedure at every voxel. There are a number of approaches to these problems that aim to increase the efficiency of the estimation and reduce the computational burden. The approach adopted in current versions of our software is to use ReML estimates based on all voxels that respond to experimental manipulation. This affords very efficient hyperparameter estimates[2] and, furthermore, allows one to use the same matrices at each voxel when computing the parameter estimates.

Although we usually make $S = \Sigma^{-1/2} = K^{-1}$, using a first-pass ReML estimate of the serial correlations, we will deal with the simpler and more general case where S can take any form. In this case, the parameter estimates are *generalized* least square (GLS) estimators. The GLS estimator is unbiased and, luckily, is identical to the Gauss-Markov estimator if the regressors in the design matrix are periodic.[3] After GLS estimation, the ReML estimate of $V = S\Sigma S^T$ enters into the expressions for the standard error and degrees of freedom provided in Figure 2.3.

[2] The efficiency scales with the number of voxels.

[3] More exactly, the GLS and ML estimators are the same if the design matrix is spanned by the eigenvectors of the Toeplitz autocorrelation matrix Σ.

fMRI noise has been variously characterized as a $1/f$ process (Zarahn *et al*., 1997) or an autoregressive process (Bullmore *et al*., 1996) with white noise (Purdon and Weisskoff, 1998). Irrespective of the exact form these serial correlations take, treating low-frequency drifts as fixed effects can finesse the hyperparameterization of serial correlations. Removing low frequencies from the time-series allows the model to fit serial correlations over a more restricted frequency range or shorter time spans. Drift removal can be implemented by including drift terms in the design matrix or by including the implicit residual forming matrix in S to make it a highpass filter. An example of a highpass filter with a highpass cut-off of 1/64 Hz is shown in the inset of Figure 2.7. This filter's transfer function (the broken line in the main panel) illustrates the frequency structure of neurogenic signals after highpass filtering.

Spatially coherent confounds and global normalization

Implicit in the use of highpass filtering is the removal of low-frequency components that can be regarded as confounds. Other important confounds are signal

components that are artefactual or have no regional specificity. These are referred to as global confounds and have a number of causes. These can be divided into physiological (e.g. global perfusion changes in PET) and non-physiological (e.g. transmitter power calibration or receiver gain in fMRI). The latter generally scale the signal before the MRI sampling process. Other non-physiological effects may have a non-scaling effect (e.g. Nyquist ghosting, movement-related effects etc.). In PET, it is generally accepted that regional changes in rCBF, evoked neuronally, mix additively with global changes to give the measured signal. This calls for a global normalization procedure where the global estimator enters into the statistical model as a confound. In fMRI, instrumentation effects that scale the data motivate a global normalization by proportional scaling, using the whole brain mean, before the data enter into the statistical model.

It is important to differentiate between global confounds and their estimators. By definition, the global mean over intracranial voxels will subsume all regionally specific effects. This means that the global estimator may be partially collinear with effects of interest, especially if evoked responses are substantial and widespread. In these situations, global normalization may induce apparent deactivations in regions not expressing a physiological response. These are not artefacts in the sense that they are real, relative to global changes, but they have less face validity in terms of the underlying neurophysiology. In instances where regionally specific effects bias the global estimator, some investigators prefer to omit global normalization. Provided drift terms are removed from the time-series, this is generally acceptable because most global effects have slow time constants. However, the issue of normalization-induced deactivations is better circumnavigated with experimental designs that use well-controlled conditions, which elicit differential responses in restricted brain systems.

Non-linear system identification approaches

So far, we have only considered linear models and first-order HRFs. Another signal processing perspective is provided by non-linear system identification (Vazquez and Noll, 1998). This section considers non-linear models as a prelude to the next subsection on event-related fMRI, where non-linear interactions among evoked responses provide constraints for experimental design and analysis. We have described an approach to characterizing evoked haemodynamic responses in fMRI based on non-linear system identification, in particular the use of Volterra series (Friston *et al.*, 1998). This approach enables one to estimate Volterra kernels that describe the relationship between stimulus presentation and the haemodynamic responses that ensue. Volterra series are essentially high-order extensions of linear convolution models. These kernels therefore represent a non-linear characterization of the HRF that can model the responses to stimuli in different contexts and interactions among stimuli. In fMRI, the kernel coefficients can be estimated by using a second order approximation to the Volterra series to formulate the problem in terms of a general linear model and expanding the kernels in terms of temporal basis functions (see Chapter 27). This allows the use of the standard techniques described above to estimate the kernels and to make inferences about their significance on a voxel-specific basis using SPMs.

One important manifestation of non-linear effects, captured by second order kernels, is a modulation of stimulus-specific responses by preceding stimuli that are proximate in time. This means that responses at high stimulus presentation rates saturate and, in some instances, show an inverted U behaviour. This behaviour appears to be specific to blood oxygenation-level-dependent (BOLD) effects (as distinct from evoked changes in cerebral blood flow) and may represent a *haemodynamic refractoriness*. This effect has important implications for event-related fMRI, where one may want to present trials in quick succession.

The results of a typical non-linear analysis are given in Figure 2.8. The results in the right panel represent the average response, integrated over a 32-second train of stimuli as a function of stimulus onset asynchrony (SOA) within that train. These responses are based on the kernel estimates (left hand panels) using data from a voxel in the left posterior temporal region of a subject obtained during the presentation of single words at different rates. The solid line represents the estimated response and shows a clear maximum at just less than one second. The dots are responses based on empirical data from the same experiment. The broken line shows the expected response in the absence of non-linear effects (i.e. that predicted by setting the second order kernel to zero). It is clear that non-linearities become important at around two seconds leading to an actual diminution of the integrated response at sub-second SOAs. The implication of this sort of result is that SOAs should not really fall much below one second and at short SOAs the assumptions of linearity are violated. It should be noted that these data pertain to single word processing in auditory association cortex. More linear behaviours may be expressed in primary sensory cortex where the feasibility of using minimum SOAs, as low as 500 ms, has been demonstrated (Burock *et al.*, 1998). This lower bound on SOA is important because some effects are detected more efficiently with high presentation rates. We now consider this from the point of view of event-related designs.

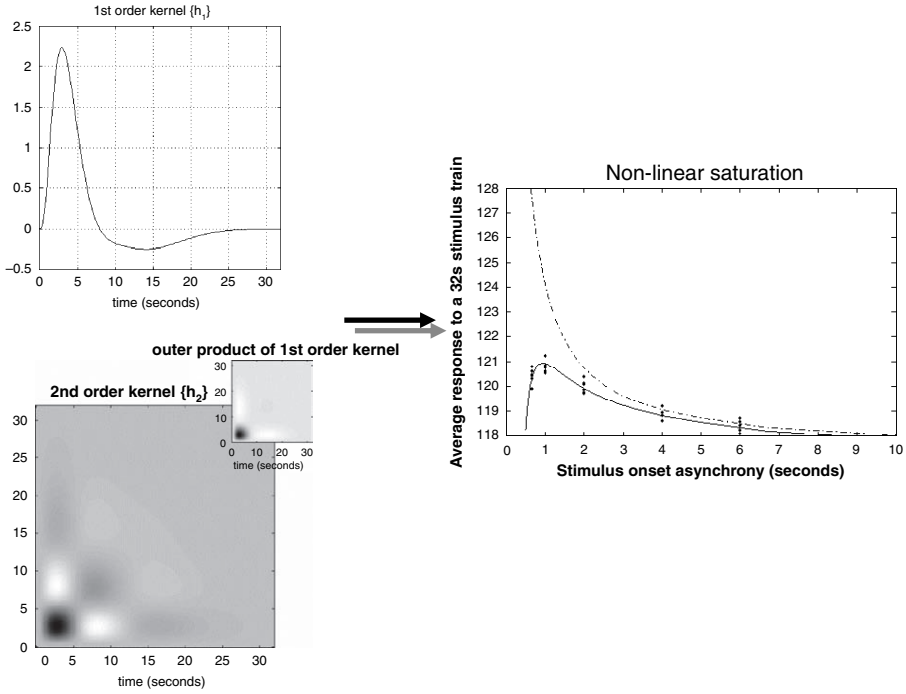

FIGURE 2.8 Left panels: Volterra kernels from a voxel in the left superior temporal gyrus at −56, −28, 12 mm. These kernel estimates were based on a single-subject study of aural word presentation at different rates (from zero to ninety words per minute) using a second order approximation to a Volterra series expansion modelling the observed haemodynamic response to stimulus input (a delta function for each word). These kernels can be thought of as a characterization of the second order haemodynamic response function. The first order kernel (h_1 – upper panel) represents the (first order) component usually presented in linear analyses. The second order kernel (h_2 – lower panel) is presented in image format. The colour scale is arbitrary; white is positive and black is negative. The insert on the right represents $h_1 h_1^T$, the second order kernel that would be predicted by a simple model that involved linear convolution with h_1 followed by some static non-linearity. Right panel: integrated responses over a 32-second stimulus train as a function of SOA. Solid line: estimates based on the non-linear convolution model parameterized by the kernels on the left. Broken line: the responses expected in the absence of second order effects (i.e. in a truly linear system). Dots: empirical averages based on the presentation of actual stimulus trains.

Event and epoch-related designs

A crucial distinction in experimental design for fMRI is that between epoch and event-related designs. In single photon emission computerized tomography (SPECT) and positron emission tomography (PET) only epoch-related responses can be assessed because of the relatively long half-life of the radiotracers used. However, in fMRI there is an opportunity to measure event-related responses, not unlike the paradigm used in electroencephalography (EEG) and magnetoencephalography (MEG). An important issue, in event-related fMRI, is the choice of inter-stimulus interval or more precisely SOA. The SOA, or the distribution of SOAs, is a critical factor and is chosen, subject to psychological or psychophysical constraints, to maximize the efficiency of response estimation. The constraints on the SOA clearly depend upon the nature of the experiment but are generally satisfied when the SOA is small and derives from a random distribution. Rapid presentation rates allow for the maintenance of a particular cognitive or attentional set, decrease the latitude that the subject has for engaging alternative strategies, or incidental processing, and allows the integration of event-related paradigms using fMRI and electrophysiology. Random SOAs ensure that preparatory or anticipatory factors do not confound event-related responses and ensure a uniform context in which events are presented. These constraints speak of the well-documented advantages of event-related fMRI over conventional blocked designs (Buckner *et al*., 1996; Clark *et al*., 1998).

In order to compare the efficiency of different designs, it is useful to have a common framework that encompasses all of them. The efficiency can then be examined in relation to the parameters of the designs. Designs can be *stochastic* or *deterministic* depending on whether there is a random element to their specification. In stochastic designs (Heid *et al*., 1997) one needs to specify the probabilities of an event occurring at all times those events could occur. In deterministic designs, the occurrence probability is unity and the design is completely specified by the times of stimulus presentation or trials. The distinction between stochastic and deterministic designs pertains to how a particular realization or stimulus sequence is created. The efficiency afforded

by a particular event sequence is a function of the event sequence itself, and not of the process generating the sequence (i.e. deterministic or stochastic). However, within stochastic designs, the design matrix X, and associated efficiency, are random variables and the *expected* or average efficiency, over realizations of X is easily computed.

In the framework considered here (Friston *et al*., 1999a), the occurrence probability p of any event occurring is specified at each time that it could occur (i.e. every SOA or stimulus onset asynchrony). Here p is a vector with an element for every SOA. This formulation engenders the distinction between *stationary* stochastic designs, where the occurrence probabilities are constant and *non-stationary* stochastic designs, where they change over time. For deterministic designs, the elements of p are 0 or 1, the presence of a 1 denoting the occurrence of an event. An example of p might be the boxcars used in conventional block designs. Stochastic designs correspond to a vector of identical values and are therefore stationary in nature. Stochastic designs with temporal modulation of occurrence probability have time-dependent probabilities varying between 0 and 1. With these probabilities the expected design matrices and expected efficiencies can be computed. A useful thing about this formulation is that by setting the mean of the probabilities p to a constant, one can compare different deterministic and stochastic designs given the same number of events. Some common examples are given in Figure 2.9 (right panel) for an SOA of one second and 32 expected events or trials over a 64 second period (except for the first deterministic example with four events and an SOA of 16 seconds). It can be seen that the least efficient is the sparse deterministic design (despite the fact that the SOA is roughly optimal for this class), whereas the most efficient is a block design. A slow modulation of occurrence probabilities gives high efficiency while retaining the advantages of stochastic designs and may represent a useful compromise between the high efficiency of block designs and the psychological benefits and latitude afforded by stochastic designs. However, it is important not to generalize these conclusions too far. An efficient design for one effect may

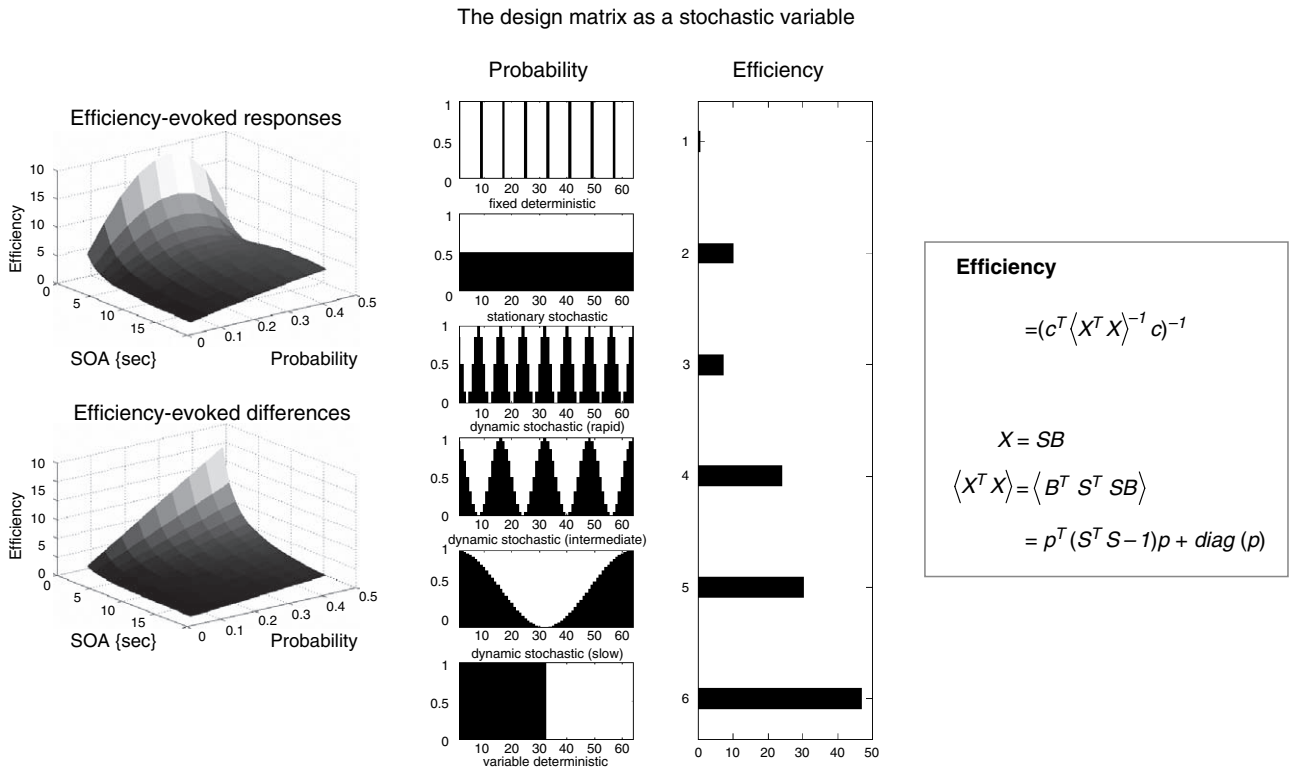

FIGURE 2.9 Efficiency as a function of occurrence probability p for a model X formed by post-multiplying S (a matrix containing n columns, modelling n possible event-related responses every SOA) by B. B is a random binary vector that determines whether the n-th response contributes to the design matrix $X = SB$ where $\langle B \rangle = p$. Right panels: a comparison of some common designs. A graphical representation of the occurrence probability as a function of time (seconds) is shown on the left and the corresponding efficiency is shown on the right. These results assume a minimum SOA of one second, a time-series of 64 seconds and a single trial-type. The expected number of events was 32 in all cases (apart from the first). Left panels: efficiency in a stationary stochastic design with two event types both presented with increasing probability every SOA. The upper graph is for a contrast testing for the response evoked by one trial type and the lower graph is for a contrast testing for differential responses.

not be the optimum for another, even within the same experiment. This can be illustrated by comparing the efficiency with which evoked responses are detected and the efficiency of detecting the difference in evoked responses elicited by two sorts of trials.

Consider a stationary stochastic design with two trial types. Because the design is stationary, the vector of occurrence probabilities, for each trial type, is specified by a single probability. Let us assume that the two trial types occur with the same probability p. By varying p and SOA one can find the most efficient design depending upon whether one is looking for evoked responses *per se* or differences among evoked responses. These two situations are depicted in the left panels of Figure 2.9. It is immediately apparent that, for both sorts of effects, very small SOAs are optimal. However, the optimal occurrence probabilities are not the same. More infrequent events (corresponding to a smaller $p = 1/3$) are required to estimate the responses themselves efficiently. This is equivalent to treating the baseline or control condition as any other condition (i.e. by including null events, with equal probability, as further event types). Conversely, if we are only interested in making inferences about the differences, one of the events plays the role of a null event and the most efficient design ensues when one or the other event occurs (i.e. $p = 1/2$). In short, the most efficient designs obtain when the events subtending the differences of interest occur with equal probability.

Another example of how the efficiency is sensitive to the effect of interest is apparent when we consider different parameterizations of the HRF. This issue is sometimes addressed through distinguishing between the efficiency of response *detection* and response *estimation*. However, the principles are identical and the distinction reduces to how many parameters one uses to model the HRF for each trial type (one basis function is used for detection and a number are required to estimate the shape of the HRF). Here the contrasts may be the same but the shape of the regressors will change depending on the temporal basis set employed. The conclusions above were based on a single canonical HRF. Had we used a more refined parameterization of the HRF, say using three-basis functions, the most efficient design to estimate one basis function coefficient would not be the most efficient for another. This is most easily seen from the signal processing perspective where basis functions with high-frequency structure (e.g. temporal derivatives) require the experimental variance to contain high-frequency components. For these basis functions a randomized stochastic design may be more efficient than a deterministic block design, simply because the former embodies higher frequencies. In the limiting case of finite impulse response (FIR) estimation, the regressors become a series of stick functions all of which have high frequencies. This parameterization of the HRF calls for high frequencies in the experimental variance. However, the use of FIR models is contraindicated by model selection procedures (see Chapter 14) that suggest only two or three HRF parameters can be estimated with any efficiency. Results that are reported in terms of FIRs should be treated with caution because the inferences about evoked responses are seldom based on the FIR parameter estimates. This is precisely because they are estimated inefficiently and contain little useful information.

INFERENCE IN HIERARCHICAL MODELS

In this section, we consider some issues that are generic to brain mapping studies that have repeated measures or replications over subjects. The critical issue is whether we want to make an inference about the effect in relation to the *within-subject variability* or with respect to the *between-subject variability*. For a given group of subjects, there is a fundamental distinction between saying that the response is significant relative to the precision[4] with which that response is measured and saying that it is significant in relation to the inter-subject variability. This distinction relates directly to the difference between *fixed*- and *random*-effect analyses. The following example tries to make this clear. Consider what would happen if we scanned six subjects during the performance of a task and baseline. We then construct a statistical model where task-specific effects were modelled separately for each subject. Unknown to us, only one of the subjects activated a particular brain region. When we examine the contrast of parameter estimates, assessing the mean activation over all subjects, we see that it is greater than zero by virtue of this subject's activation. Furthermore, because that model fits the data extremely well (modelling no activation in five subjects and a substantial activation in the sixth), the error variance, on a scan-to-scan basis, is small and the t-statistic is very significant. Can we then say that the group shows an activation? On the one hand, we can say, quite properly, that the mean group response embodies an activation but, clearly, this does not constitute an inference that the group's response is significant (i.e. that this sample of subjects shows a consistent activation). The problem here is that we are using the *scan-to-scan* error variance and this is not necessarily appropriate for an inference about group responses. To make the inference that the group showed a significant activation, one would have to assess the variability in

[4] Precision is the inverse of the variance.

activation effects from *subject to subject* (using the contrast of parameter estimates for each subject). This variability now constitutes the proper error variance. In this example, the variance of these six measurements would be large relative to their mean and the corresponding *t*-statistic would not be significant.

The distinction between the two approaches above relates to how one computes the appropriate error variance. The first represents a fixed-effects analysis and the second a random-effects analysis (or more exactly a mixed-effects analysis). In the former, the error variance is estimated on a scan-to-scan basis, assuming that each scan represents an independent observation (ignoring serial correlations). Here the degrees of freedom are essentially the number of scans (minus the rank of the design matrix). Conversely, in random-effects analyses, the appropriate error variance is based on the activation from subject to subject where the effect *per se* constitutes an independent observation and the degrees of freedom fall dramatically to the number of subjects. The term 'random effect' indicates that we have accommodated the randomness of different responses from subject to subject. Both analyses are perfectly valid but only in relation to the inferences that are being made: inferences based on fixed-effects analyses are about the particular subject(s) studied. Random-effects analyses are usually more conservative but allow the inference to be generalized to the population from which the subjects were selected.

Random-effects analyses

The implementation of random-effects analyses in SPM is fairly straightforward and involves taking the contrasts of parameters estimated from a *first-level* (within-subject) analysis and entering them into a *second-level* (between-subject) analysis. This ensures that there is only one observation (i.e. contrast) per subject in the second-level analysis and that the error variance is computed using the subject-to-subject variability of estimates from the first level. This is also known as a summary statistic approach and, in the context of fully balanced designs is formally identical to mixed-effects analysis. The nature of the inference made is determined by the contrasts that enter the second level (Figure 2.10). The second-level design matrix simply tests the null hypothesis that the contrasts are zero (and is usually a column of ones, implementing a single-sample *t*-test).

The reason this multistage procedure emulates a full mixed-effects analysis, using a hierarchical observation model (see Chapters 11 and 12), rests upon the fact that the design matrices for each subject are the same (or sufficiently similar). In this special case, the estimator of the

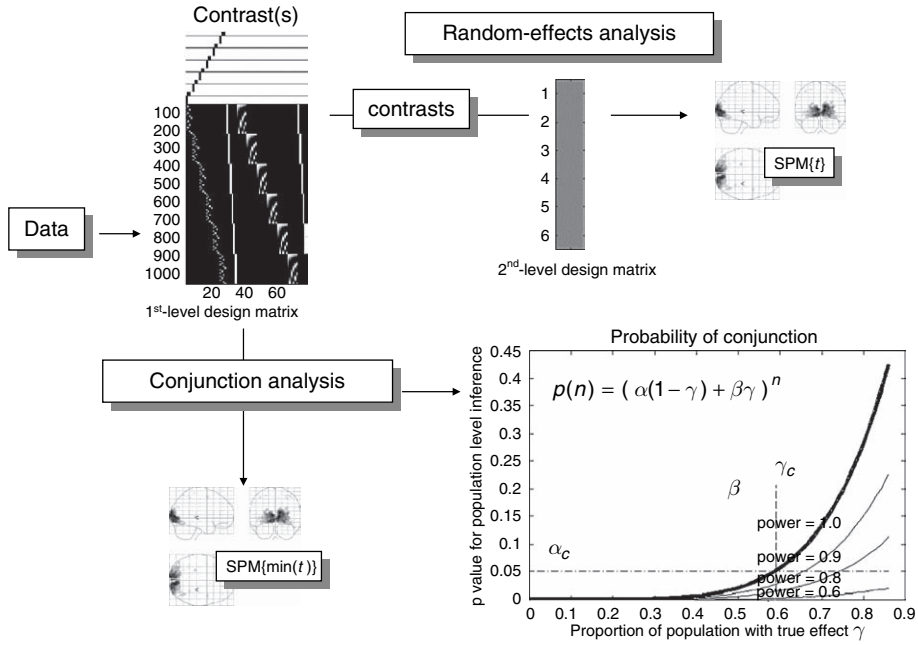

FIGURE 2.10 Schematic illustrating the implementation of random-effect and conjunction analyses for population inference. The lower right graph shows the probability $p(n) = (\alpha(1-\gamma)+\beta\gamma)^n$ of obtaining a conjunction over n subjects, conditional on a certain proportion γ of the population expressing the effect, for a test with specificity of $\alpha = 0.05$, at several sensitivities ($\beta = 1$, 0.9, 0.8 and 0.6). The broken lines denote the critical specificity for population inference α_c and the associated proportion of the population γ_c (see Friston *et al*., 1999b for details).

variance at the second level contains the right mixture of within- and between-subject error. It is important to appreciate this because the efficiency of the design at the first level percolates to higher levels. It is therefore important to use efficient strategies at all levels in a hierarchical design.

Conjunction analyses and population inferences

In some instances, a fixed-effects analysis is more appropriate, particularly when reporting single-case studies. With a series of single cases, it is natural to ask what are common features of functional anatomy (e.g. the location of V5) and what aspects are subject specific (e.g. the location of ocular dominance columns)? One way to address commonalties is to use a conjunction analysis over subjects. It is important to understand the nature of the inference provided by conjunction analyses, because there has been some confusion (see Nichols *et al*., 2005; Friston *et al*., 2005). Imagine that, in sixteen subjects the activation in V5, elicited by a motion stimulus, was greater than zero. The probability of this occurring by chance, in the same area, is extremely small and is the *p*-value returned by a conjunction analysis using a threshold of $p = 0.5$ (i.e. $t = 0$) for each subject. This constitutes evidence that V5 is engaged by motion. However, it is not an assertion that each subject activated significantly (we only require the *t*-value to be greater than zero for each subject). In other words, a significant conjunction is not a conjunction of significance.

The motivations for conjunction analyses, in the context of multisubject studies, are twofold. They provide an inference, in a fixed-effects context, testing the null hypotheses of no activation in any of the subjects, which can be much more sensitive than testing for the average activation. Second, they can be used to make inferences about the population in terms of confidence intervals on the proportion of subjects showing an effect (see Friston *et al*., 1999b).

CONCLUSION

In this chapter, we have reviewed the main components of image analysis and have introduced the tenets of statistical parametric mapping. We have also considered the design of experiments and their statistical models, with a special focus on fMRI. This chapter has covered the key operational issues in identifying regionally specific effects in neuroimaging. In the next chapter, we look at models for neuroimaging from a broader perspective and address the functional integration of distributed responses in the brain.

REFERENCES

Adler RJ (1981) *The geometry of random fields*. Wiley, New York

Aguirre GK, Zarahn E, D'Esposito M (1998) A critique of the use of the Kolmogorov-Smirnov (KS) statistic for the analysis of BOLD fMRI data. *Mag Res Med* **39**: 500–05

Andersson JL, Hutton C, Ashburner J *et al*. (2001) Modeling geometric deformations in EPI time series. *NeuroImage* **13**: 903–19

Ashburner J, Friston KJ. (2000) Voxel-based morphometry – the methods. *NeuroImage* **11**: 805–21

Ashburner J, Neelin P, Collins DL *et al*. (1997) Incorporating prior knowledge into image registration. *NeuroImage* **6**: 344–52

Ashburner J, Hutton C, Frackowiak R *et al*. (1998) Identifying global anatomical differences: deformation-based morphometry. *Hum Brain Mapp* **6**: 348–57

Bandettini PA, Jesmanowicz A, Wong EC *et al*. (1993) Processing strategies for time course data sets in functional MRI of the human brain. *Mag Res Med* **30**: 161–73

Büchel C, Wise RJS, Mummery CJ *et al*. (1996) Non-linear regression in parametric activation studies. *NeuroImage* **4**: 60–66

Buckner R, Bandettini P, O'Craven K *et al*. (1996) Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging. *Proc Natl Acad Sci USA* **93**: 14878–83

Bullmore ET, Brammer MJ, Williams SCR *et al*. (1996) Statistical methods of estimation and inference for functional MR images. *Mag Res Med* **35**: 261–77

Burock MA, Buckner RL, Woldorff MG *et al*. (1998) Randomized event-related experimental designs allow for extremely rapid presentation rates using functional MRI. *NeuroReport* **9**: 3735–39

Chung MK, Worsley KJ, Paus T *et al*. (2001) A unified statistical approach to deformation-based morphometry. *NeuroImage* **14**: 595–606

Clark VP, Maisog JM, Haxby JV (1998) fMRI study of face perception and memory using random stimulus sequences. *J Neurophysiol* **76**: 3257–65

Friston KJ (1997) Testing for anatomical specified regional effects. *Hum Brain Mapp* **5**: 133–36

Friston KJ, Frith CD, Liddle PF *et al*. (1991) Comparing functional (PET) images: the assessment of significant change. *J Cereb Blood Flow Metab* **11**: 690–99

Friston KJ, Frith C, Passingham RE *et al*. (1992a) Motor practice and neurophysiological adaptation in the cerebellum: a positron tomography study. *Proc Roy Soc Lon Series B***248**: 223–28

Friston KJ, Grasby P, Bench C *et al*. (1992b) Measuring the neuromodulatory effects of drugs in man with positron tomography. *Neurosci Lett* **141**: 106–10

Friston KJ, Worsley KJ, Frackowiak RSJ *et al*. (1994) Assessing the significance of focal activations using their spatial extent. *Hum Brain Mapp* **1**: 214–20

Friston KJ, Ashburner J, Frith CD *et al*. (1995a) Spatial registration and normalization of images. *Hum Brain Mapp* **2**: 165–89

Friston KJ, Holmes AP, Worsley KJ *et al*. (1995b) Statistical parametric maps in functional imaging: a general linear approach. *Hum Brain Mapp* **2**: 189–210

Friston KJ, Williams S, Howard R *et al*. (1996a) Movement related effects in fMRI time series. *Mag Res Med* **35**: 346–55

Friston KJ, Holmes A, Poline J-B *et al*. (1996b) Detecting activations in PET and fMRI: levels of inference and power. *NeuroImage* **4**: 223–35

Friston KJ, Price CJ, Fletcher P *et al*. (1996c) The trouble with cognitive subtraction. *NeuroImage* **4**: 97–104

Friston KJ, Büchel C, Fink GR *et al*. (1997) Psychophysiological and modulatory interactions in neuroimaging. *NeuroImage* **6**: 218–29

Friston KJ, Josephs O, Rees G *et al*. (1998) Non-linear event-related responses in fMRI. *Mag Res Med* **39**: 41–52

Friston KJ, Zarahn E, Josephs O *et al*. (1999a) Stochastic designs in event-related fMRI. *NeuroImage* **10**: 607–19

Friston KJ, Holmes AP, Price CJ *et al*. (1999b) Multisubject fMRI studies and conjunction analyses. *NeuroImage* **10**: 385–96

Friston KJ, Josephs O, Zarahn E *et al*. (2000) To smooth or not to smooth? Bias and efficiency in fMRI time-series analysis. *NeuroImage* **12**: 196–208

Friston KJ, Penny WD and Glaser DE (2005) Conjunction revisited. *NeuroImage* **25**(3): 661–7

Grafton S, Mazziotta J, Presty S *et al*. (1992) Functional anatomy of human procedural learning determined with regional cerebral blood flow and PET. *J Neurosci* **12**: 2542–48

Grootoonk S, Hutton C, Ashburner J *et al*. (2000) Characterization and correction of interpolation effects in the realignment of fMRI time series. *NeuroImage* **11**: 49–57

Heid O, Gönner F, Schroth G (1997) Stochastic functional MRI. *NeuroImage* **5**: S476

Jezzard P, Balaban RS (1995) Correction for geometric distortion in echo-planar images from B0 field variations. *Mag Res Med* **34**: 65–73

Kiebel SJ, Poline JB, Friston KJ *et al*. (1999) Robust smoothness estimation in statistical parametric maps using standardized residuals from the general linear model. *NeuroImage* **10**: 756–66

Liddle PF, Friston KJ, Frith CD *et al*. (1992) Cerebral blood-flow and mental processes in schizophrenia. *Roy Soc Med* **85**: 224–27

Lueck CJ, Zeki S, Friston KJ *et al*. (1989) The color centre in the cerebral cortex of man. *Nature* **340**: 386–89

Nichols T, Brett M, Andersson J, Wager T, Poline JB (2004) Valid conjunction interference with the minimum statistic. *NeuroImage* **25**(3): 653–60

Petersen SE, Fox PT, Posner MI *et al*. (1989) Positron emission tomographic studies of the processing of single words. *J Cog Neurosci* **1**: 153–70

Price CJ, Friston KJ (1997) Cognitive conjunction: a new approach to brain activation experiments. *NeuroImage* **5**: 261–70

Price CJ, Wise RJS, Ramsay S *et al*. (1992) Regional response differences within the human auditory cortex when listening to words. *Neurosci Lett* **146**: 179–82

Purdon PL, Weisskoff RM (1998) Effect of temporal autocorrelation due to physiological noise and stimulus paradigm on voxel-level false-positive rates in fMRI. *Hum Brain Mapp* **6**: 239–495

Talairach P, Tournoux J (1988) A stereotactic coplanar atlas of the human brain. Thieme, Stuttgart

Vazquez AL, Noll CD (1998) Non-linear aspects of the BOLD response in functional MRI. *NeuroImage* **7**: 108–18

Worsley KJ, Friston KJ (1995) Analysis of fMRI time-series revisited – again. *NeuroImage* **2**: 173–81

Worsley KJ, Evans AC, Marrett S *et al*. (1992) A three-dimensional statistical analysis for rCBF activation studies in human brain. *J Cereb Blood Flow Metab* **12**: 900–18

Worsley KJ, Marrett S, Neelin P *et al*. (1996) A unified statistical approach or determining significant signals in images of cerebral activation. *Hum Brain Mapp* **4**: 58–73

Zarahn E, Aguirre GK, and D'Esposito M (1997) Empirical analyses of BOLD fMRI statistics: I Spatially unsmoothed data collected under null-hypothesis conditions. *NeuroImage* **5**: 179–97