Chapter #1

Statistical Parametric Mapping

Karl J Friston
Wellcome Dept. of Imaging Neuroscience

Abstract:

Key words:

1. INTRODUCTION

This chapter is about making regionally specific inferences in neuroimaging. These inferences may be about differences expressed when comparing one group of subjects to another or, within subjects, over a sequence of observations. They may pertain to structural differences (*e.g.* in voxel-based morphometry - Ashburner and Friston 2000) or neurophysiological indices of brain functions (*e.g.* fMRI). The principles of data analysis are very similar for all of these applications and constitute the subject of this chapter. We will focus on the analysis of fMRI time-series because this covers most of the issues that are likely to be encountered in other modalities. Generally,

the analysis of structural images and PET scans is simpler because they do not have to deal with correlated errors, from one scan to the next.

A general issue, in data analysis, is the relationship between the neurobiological hypothesis one posits and the statistical models adopted to test that hypothesis. This chapter begins by reviewing the distinction between functional *specialization* and *integration* and how these principles serve as the motivation for most analyses of neuroimaging data. We will address the design and analysis of neuroimaging studies from both these perspectives but note that both have to be integrated for a full understanding of brain mapping results.

Statistical parametric mapping is generally used to identify functionally specialized brain regions and is the most prevalent approach to characterizing functional anatomy and disease-related changes. The alternative perspective, namely that provided by functional integration, requires a different set of [multivariate] approaches that examine the relationship between changes in activity in one brain area and another. Statistical parametric mapping is a voxel-based approach, employing classical 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 must conform to a known anatomical space. Before considering statistical modeling, this chapter deals briefly with how a time-series of images are realigned and mapped into some standard anatomical space (*e.g.* a stereotactic 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. fMRI is special, in the sense that the 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. Temporal autocorrelations in noise processes represent another important issue, specific to fMRI, and approaches to maximizing efficiency in the context of serially correlated error terms will be discussed. Nonlinear models of evoked hemodynamics will be considered here because they can be used to indicate when the assumptions behind linear models are violated. fMRI can capture data very fast (in relation to other imaging techniques), engendering the opportunity to measure event-related responses. The distinction between event and epoch-related designs will be discussed from the point of view of efficiency and the constraints provided by nonlinear characterizations. Before considering multivariate analyses we will close the discussion of inferences, about regionally specific effects, by looking at the distinction between fixed and random-effect analyses and how this relates to inferences about the subjects studied or the population from which these subjects came. The final section will deal with functional integration using models of effective connectivity and other multivariate approaches.

2. FUNCTIONAL SPECIALIZATION AND INTEGRATION

The brain appears to adhere to two fundamental principles of functional organization, *functional integration* and *functional specialization,* where the integration within and among specialized areas is mediated by effective connectivity. The distinction relates to that between *localisationism* and *[dis]connectionism* that dominated thinking about cortical function in the nineteenth century. Since the early anatomic theories of Gall, the identification of a particular brain region with a specific function has become a central theme in neuroscience. However functional localization *per se* was not easy to demonstrate: For example, a meeting that took place on August 4th 1881 addressed the difficulties of attributing function to a cortical area, given the dependence of cerebral activity on underlying connections (Phillips et al 1984). This meeting was entitled "Localization of function in the cortex cerebri". Goltz (1881), although accepting the results of electrical stimulation in dog and monkey cortex, considered that the excitation method was inconclusive, in that movements elicited might have originated in related pathways, or current could have spread to distant centers. In short, the excitation method could not be used to infer functional localization because localisationism discounted interactions, or functional integration among different brain areas. It was proposed that lesion studies could supplement excitation experiments. Ironically, it was observations on patients with brain lesions some years later (see Absher and Benson 1993) that led to the concept of

disconnection syndromes and the refutation of localisationism as a complete or sufficient explanation of cortical organization. Functional localization implies that a function can be localized in a cortical area, whereas specialization suggests that a cortical area is specialized for some aspects of perceptual or motor processing, and that this specialization is anatomically segregated within the cortex. The cortical infrastructure supporting a single function may then involve many specialized areas whose union is mediated by the functional integration among them. In this view functional specialization is only meaningful in the context of functional integration and *vice versa*.

2.1 Functional specialization and segregation

The functional role played by any component (*e.g.* cortical area, subarea or neuronal population) of the brain is largely defined by its connections. Certain patterns of cortical projections are so common that they could amount to rules of cortical connectivity. "These rules revolve around one, apparently, overriding strategy that the cerebral cortex uses - that of functional segregation" (Zeki 1990). Functional segregation demands that cells with common functional properties be grouped together. This architectural constraint necessitates both convergence and divergence of cortical connections. Extrinsic connections among cortical regions are not continuous but occur in patches or clusters. This patchiness has, in some instances, a clear relationship to functional segregation. For example, V2 has a distinctive cytochrome oxidase architecture, consisting of thick stripes, thin stripes and inter-stripes. When recordings are made in

V2, directionally selective (but not wavelength or color selective) cells are found exclusively in the thick stripes. Retrograde (*i.e.* backward) labeling of cells in V5 is limited to these thick stripes. All the available physiological evidence suggests that V5 is a functionally homogeneous area that is specialized for visual motion. Evidence of this nature supports the notion that patchy connectivity is the anatomical infrastructure that mediates functional segregation and specialization. If it is the case that neurons in a given cortical area share a common responsiveness (by virtue of their extrinsic connectivity) to some sensorimotor or cognitive attribute, then this functional segregation is also an anatomical one. Challenging a subject with the appropriate sensorimotor attribute or cognitive process should lead to activity changes in, and only in, the area of interest. This is the anatomical and physiological model upon which the search for regionally specific effects is based .

The analysis of functional neuroimaging data involves many steps that can be broadly divided into; (i) spatial processing, (ii) estimating the parameters of a statistical model and (iii) making inferences about those parameter estimates with their associated statistics (see Figure 1). We will deal first with spatial transformations: In order to combine data from different scans from the same subject, or data from different subjects it is necessary that they conform to the same anatomical frame of reference. This is the subject of the next section.

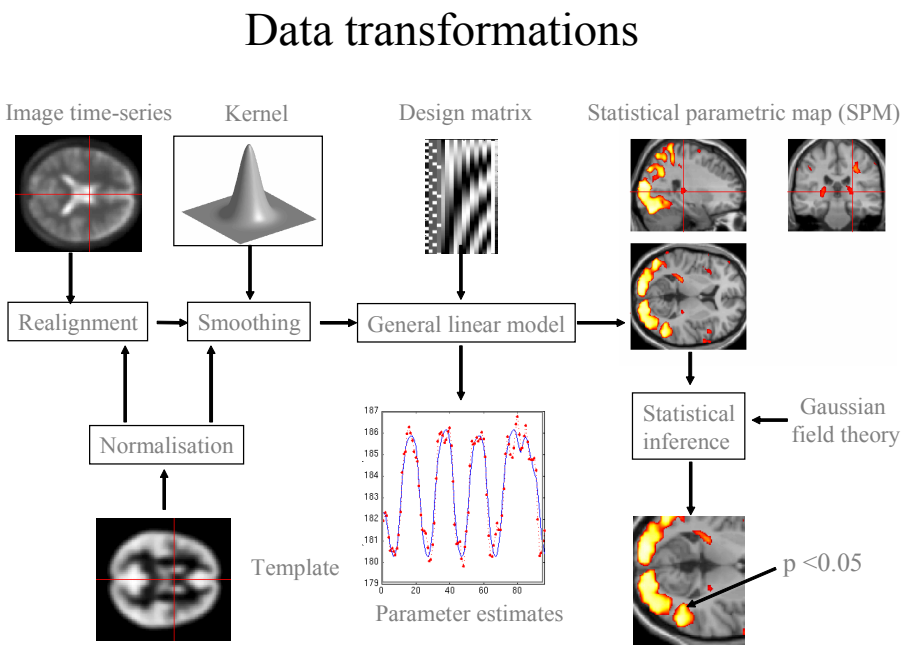


Figure 1. This schematic depicts the transformations that start with an imaging data sequence and end with a statistical parametric map (SPM). SPMs that can be thought of as 'X-rays' of the significance of an effect. Voxel-based analyses require the data to be in the same anatomical space: This is effected by realigning the data (and removing movement-related signal components that persist after realignment). After realignment the images are subject to non-linear warping so that they match a template that already conforms to a standard anatomical space. After smoothing, the general linear model is employed to (i) estimate the parameters of the model and (ii) derive the appropriate univariate test statistic at every voxel (see Figure 3). The test statistics that ensue (usually T or F statistics) constitute the SPM. The final stage is to make statistical inferences on the basis of the SPM and Gaussian random field theory (see Figure 6) and characterize the responses observed using the fitted responses or parameter estimates.

3. SPATIAL REALIGNMENT AND NORMALISATION

The analysis of neuroimaging data generally starts with a series of spatial transformations. These transformations aim to reduce artifactual variance components in the voxel time-series that are

induced by movement or shape differences among a series of scans. Voxel-based analyses assume that the data from a particular voxel all derive from the same part of the brain. Violations of this assumption will introduce artifactual changes in the voxel values that may obscure changes, or differences, of interest.. Even single-subject analyses 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 in order to 'undo' the effects of subject movement during the scanning session. After realignment the data are then transformed using linear or nonlinear warps into a standard anatomical space. Finally, the data are usually spatially smoothed before entering the analysis proper.

3.1 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, co-operative subjects still show displacements of up to a millimeter or so. Realignment involves (i) estimating the 6 parameters of an affine 'rigid-body' transformation that minimizes the [sum of squared] differences between each successive scan and a reference scan (usually the first or the average of all scans in the time series) and (ii) applying the transformation by re-sampling the data using tri-linear, sinc or cubic spline interpolation. 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 squares 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 final step in which the data are *adjusted* for residual movement-related effects.

3.2 Adjusting for movement related effects in fMRI

In extreme cases as much as 90% of the variance, in a 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 modeled using a *linear* affine model. These nonlinear effects include; (i) subject movement between slice acquisition, (ii) interpolation artifacts (Grootoonk *et al* 2000), (iii) nonlinear distortion due to magnetic field inhomogeneities (Andersson *et al* 2001) and (iv) spin-excitation history effects (Friston *et al* 1996a). The latter can be pronounced if the TR (repetition time) approaches T_1 making the current signal a function of movement history. These multiple effects render the movement-related signal (y) a nonlinear function of displacement (x) in the nth and previous scans $y_n = f(x_n, x_{n-1}, \ldots)$. By assuming a sensible form for this function, its parameters can be

estimated using the observed time-series and the estimated movement parameters x from the realignment procedure. The estimated movement-related signal is then simply subtracted from the original data. This adjustment can be carried out as a pre-processing step or embodied in model estimation during the analysis proper. The form for $f(x)$, proposed in Friston *et al* (1996a), was a nonlinear auto-regression 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 the current movement-related signal. However, it is also a reasonable model for many other sources of movement-related confounds. Generally, for TRs of several seconds, interpolation artifacts supersede (Grootoonk *et al* 2000) and first order terms, comprising an expansion of the current displacement in terms of periodic basis functions, appear to be sufficient.

This subsection has considered *spatial* realignment. In multislice acquisition different slices are acquired at slightly 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 sinc interpolation over time and only when (i) the temporal dynamics of evoked responses are important and (ii) the TR is sufficiently small to permit interpolation. Generally timing effects of this sort are not considered problematic because they manifest as artifactual latency differences in evoked responses from region to region. Given that biophysical latency differences may be 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, this renders temporal realignment unnecessary in most instances.

3.3 Spatial Normalization

After realigning the data, a mean image of the series, or some other co-registered (*e.g.* a T_1-weighted) image, is used to estimate the 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: (i) a 12-parameter affine transformation, where the parameters constitute a spatial transformation matrix, (ii) low frequency basis spatial functions (usually a discrete cosine set or polynomials), where the parameters are the coefficients of the basis functions employed and (ii) 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 simple Bayesian framework, in which one is trying to find the deformation parameters θ that have the maximum posterior probability $p(\theta \mid y)$ given the data *y*, where $p(\theta \mid y) = p(y \mid \theta)\, p(\theta)$. 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, assuming the current estimate of the deformation is true, times the probability of that estimate being true. In practice the deformation is updated iteratively

using a Gauss-Newton scheme to maximize $p(\theta \mid y)$. This involves jointly minimizing the likelihood and prior potentials $H(y \mid \theta) = \ln p(y \mid \theta)$ 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 about the likelihood of a given 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).

In practice most people use an affine or spatial basis function warps and use 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 difference between the index image and the best [linear] combination of templates (*e.g.* depicting gray, white, CSF and skull tissue partitions). This models intensity differences that are unrelated to registration differences and allows different modalities to be co-registered (see Figure 2).

A special consideration is the spatial normalization of brains that have gross anatomical pathology. This pathology can be of two sorts (i) quantitative changes in the amount of a particular tissue compartment (*e.g.* cortical atrophy) or (ii) qualitative changes in anatomy involving the insertion or deletion of normal tissue compartments (*e.g.* ischemic 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 substantial 'errors' in the normalization 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. An alternative strategy is to use another modality that is less sensitive to the pathology as the basis of the spatial normalization procedure or to simply remove the damaged region from the estimation by masking it out.

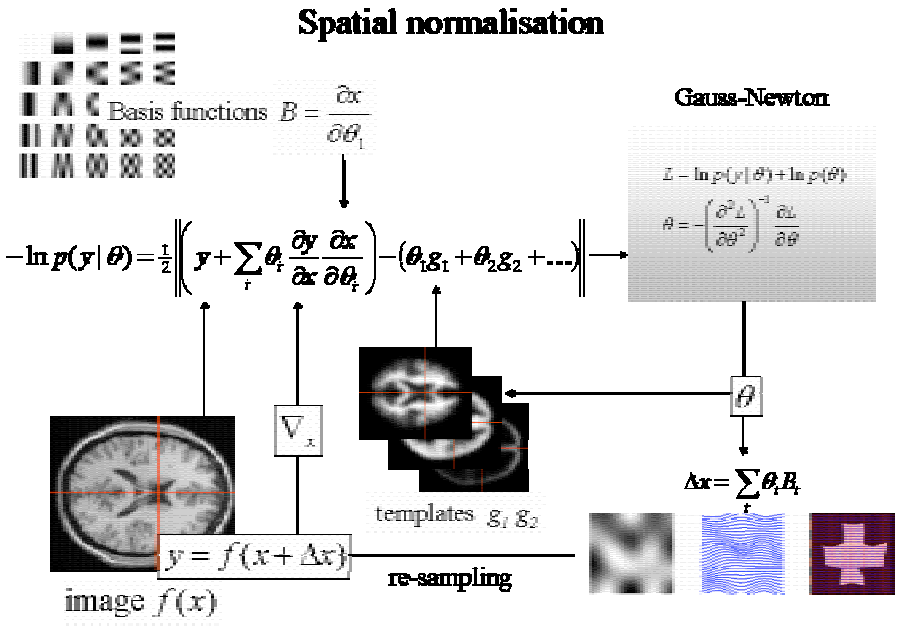


Figure 2. Schematic illustrating a Gauss-Newton scheme for maximizing the posterior probability of the parameters for spatially normalizing an image. This scheme is iterative. At each step the conditional estimate of the parameters is obtained by jointly minimizing the likelihood and the prior potentials. The former is the difference between a resampled (i.e. warped) version y of the image f and the best linear combination of some templates g. These parameters are used to mix the templates and resample the image to progressively reduce both the spatial and intensity differences. After convergence the resampled image can be considered normalized.

3.4 Co-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 T_2* images, relative to anatomical T_1-weighted data, are a particularly 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 dephasing of 2π through the slice, corresponds to an in-plane distortion of a voxel. 'Unwarping' 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.

3.5 Spatial smoothing

The motivations for smoothing the data are fourfold: (i) By the matched filter theorem, the optimum smoothing kernel corresponds to the size of the effect that one anticipates. The spatial scale of hemodynamic responses is, according to high-resolution optical imaging experiments, about 2 to 5mm. Despite the potentially high resolution afforded by fMRI an equivalent smoothing is suggested for

most applications. (ii) 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. (iii) When making inferences about regional effects using Gaussian random field theory (see Section IV) one of the assumptions is that the error terms are a reasonable lattice representation of an underlying and smooth Gaussian 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. (iv) In the context of inter-subject averaging it is often necessary to smooth more (*e.g.* 8 mm in fMRI or 16mm in PET) to project the data onto a spatial scale where homologies in functional anatomy are expressed among subjects.

4. STATISTICAL PARAMETRIC MAPPING

Functional mapping studies are usually analyzed with some form of statistical parametric mapping. Statistical parametric mapping refers to the construction of spatially extended statistical processes to test hypotheses about regionally specific effects (Friston *et al* 1991). Statistical parametric maps (SPMs) are image processes with voxel 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 spatially extended statistical processes by referring to the probabilistic behavior of Gaussian fields (Adler 1981, Worsley *et al* 1992, Friston *et al* 1994a, Worsley *et al* 1996). Gaussian random fields model both the univariate probabilistic characteristics of a SPM and any non-stationary spatial covariance structure. 'Unlikely' excursions of the SPM are interpreted as regionally specific effects, attributable to the sensorimotor or cognitive process that has been manipulated experimentally.

Over the years statistical parametric mapping has come to refer to the conjoint use of *the general linear model* (GLM) and *Gaussian random field* (GRF) theory to analyze and make classical inferences about spatially extended data through statistical parametric maps (SPMs). The GLM is used to estimate some parameters that could explain the data in exactly the same way as in conventional analysis of discrete data. GRF theory is used to resolve the multiple comparison problem that ensues when making inferences over a volume of the brain. GRF theory provides a method for correcting *p* values for the search volume of a SPM and plays the same role for *continuous* data (*i.e.* images) as the Bonferonni correction for the number of discontinuous or *discrete* statistical tests.

The approach was called SPM for three reasons; (i) To acknowledge *S*ignificance *P*robability *M*apping, the use of

interpolated pseudo-maps of *p* values used to summarize the analysis of multi-channel ERP studies. (ii) For consistency with the nomenclature of parametric maps of physiological or physical parameters (*e.g.* regional cerebral blood flow rCBF or volume rCBV parametric maps). (iii) In reference to the *parametric* statistics that comprise 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; (i) they do not support inferences about regionally specific effects, (ii) they require more observations than the dimension of the response variable (*i.e.* number of voxels) and (iii), even in the context of dimension reduction, they are usually less sensitive to focal effects than mass-univariate approaches. A heuristic argument, 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). In general, the more parameters (and hyper-parameters) 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.

An alternative approach would be to consider different voxels as different levels of an experimental or treatment factor and use classical analysis of variance, not at each voxel (*c.f.* SPM), 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 [hyper]parameters to be estimated for each voxel using conventional techniques. In SPM the non-sphericity is parameterized in the most parsimonious way with just two [hyper]parameters for each voxel. These are the error variance and smoothness estimators (see Section IV.B and Figure 2). This minimal parameterization lends SPM a sensitivity that usually surpasses other approaches. SPM can do this because GRF theory implicitly imposes constraints on the non-sphericity implied by the continuous and [spatially] extended nature of the data. This is the only constraint on the behavior of the error terms implied by the use of GRF theory and is something that conventional multivariate and equivalent univariate approaches are unable to accommodate, to their cost.

Some analyses use statistical maps based on non-parametric tests that eschew distributional assumptions about the data. These approaches may, in some instances, be useful but are generally less powerful (*i.e.* less sensitive) than parametric approaches (see Aguirre *et al* 1998). Their original motivation in fMRI was based on the [specious] assumption that the residuals were not normally distributed. Next we consider parameter estimation in the context of the GLM.

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

This is followed by an introduction to the role of GRF theory when making classical inferences about continuous data.

4.1 The general linear model

Statistical analysis of imaging data corresponds to (i) modeling the data to partition observed neurophysiological responses into components of interest, confounds and error and (ii) making inferences about the interesting effects in relation to the error variance. This inference can be regarded as a direct comparison of the variance due to an interesting experimental manipulation with the error variance (*c.f.* 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 analyze PET and fMRI time-series with a diversity of statistical and conceptual approaches. This is not the case. With very a few exceptions, every analysis is a variant of the general linear model. This includes; (i) simple T tests on scans assigned to one condition or another, (ii) correlation coefficients between observed responses and boxcar stimulus functions in fMRI, (iii) inferences made using multiple linear regression, (iv) evoked responses estimated using linear time invariant models and (v) selective averaging to estimate event-related responses in fMRI. Mathematically they are all identical. 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 the 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 into a statistical model. Therefore, the T statistic provides a more versatile and generic way of assessing the significance of regional effects and is preferred over the correlation coefficient.

The general linear model is an equation $Y = X\beta + \varepsilon$ that expresses the observed response variable Y in terms of a linear combination of explanatory variables X plus a well behaved error term (see Figure 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 X that contains the explanatory variables (*e.g.* designed effects or confounds) is called the *design matrix*. Each column of the design matrix corresponds to some effect one has built into the experiment or that may confound the results. These are referred to as explanatory variables, covariates or regressors. The example in Figure 1 relates to a fMRI study of visual stimulation under four conditions. The effects on the response variable are modeled in terms of functions of the presence of these conditions (*i.e.* boxcars smoothed with a hemodynamic 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 artifacts 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 least squares 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 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 specified temporal filtering of the data. This generalization brings with it the notion of *effective degrees of freedom*, which are less than the conventional degrees of freedom under i.i.d. assumptions (see footnote). They are smaller because the temporal correlations reduce the effective number of independent observations. The T and F statistics are constructed using Satterthwaite's approximation. This is the same approximation used in classical non-sphericity corrections such as the Geisser-Greenhouse correction. However, in the Worsley and Friston (1995) scheme, Satherthwaite's 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.

The equations summarized in Figure 3 can be used to implement a vast range of statistical analyses. The issue is therefore not so much the mathematics but the formulation of a design matrix X appropriate to the study design and inferences that are sought. The design matrix can contain both covariates and indicator variables. Each column of X

has an associated unknown parameter. Some of these parameters will be of interest (*e.g.* the effect of particular sensorimotor or cognitive condition or the regression coefficient of hemodynamic responses on reaction time). The remaining parameters will be of no interest and pertain to confounding effects (*e.g.* the effect of being a particular subject 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 a 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 0.....] to compare the difference in responses evoked by two conditions, modeled by the first two condition-specific regressors in the design matrix. Sometimes several contrasts of parameter estimates are jointly interesting. For example, when using polynomial (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' that one wants to test together. A 'F-contrast' may look like,

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

which would test for the significance of the first *or* second parameter estimates. The fact that the first weight is –1 as opposed to 1 has no effect on the test because the F statistic is based on sums of squares.

. In most analysis the design matrix contains indicator variables or parametric variables encoding the experimental manipulations. These are formally identical to classical analysis of [co]variance (*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.

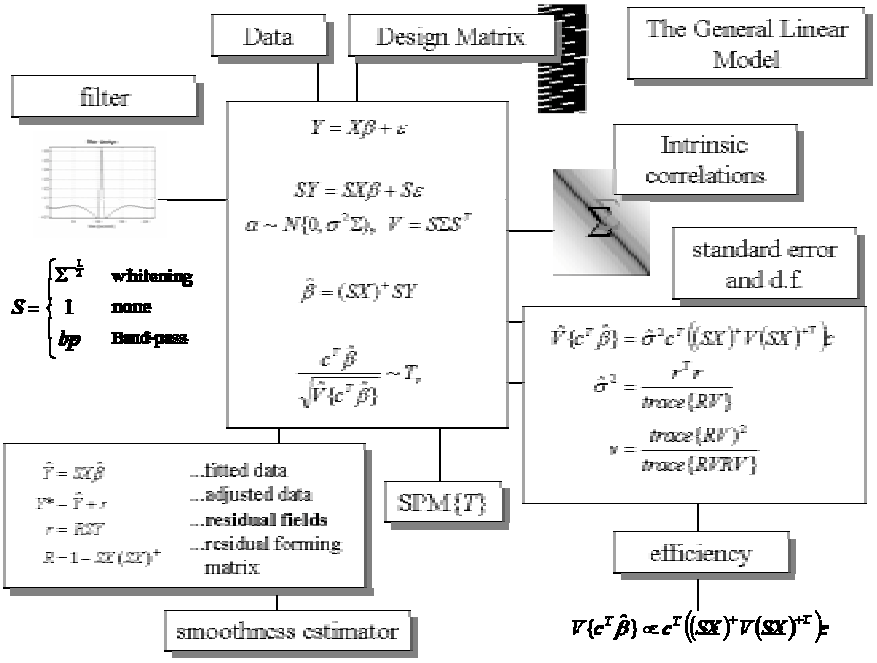


Figure 3. The general linear model. The general linear model is an equation expressing the response variable Y in terms of a linear combination of explanatory variables in a design matrix X and an error term with assumed or known autocorrelation □. In fMRI the data can

be filtered with a convolution matrix S, leading to a generalized linear model that includes [intrinsic] serial correlations and applied [extrinsic] filtering. Different choices of S correspond to different [de]convolution schema as indicated on the upper left. The parameter estimates obtain in a least squares sense using the pseudoinverse (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 the estimated 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 associated with the error variance 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 an experimental design. The parameter estimates can either be examined directly or used to compute the fitted responses (see lower left panel). Adjusted data refers to data from which estimated 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 used in Gaussian random field theory (see Figure 6).

4.1.1 Linear Time Invariant (LTI) systems and temporal basis functions

In Friston *et al* (1994b) the form of the hemodynamic impulse response function (HRF) was estimated using a least squares de-convolution and a time invariant model, where evoked neuronal responses are convolved with the HRF to give the measured hemodynamic response (see also Boynton *et al* 1996). This simple linear framework is the cornerstone for making statistical inferences about activations in fMRI with the GLM. An impulse response function is the response to a single impulse, measured at a series times after the input. It characterizes the input-output behavior of the system (*i.e.*. voxel) and places important constraints on the sorts of inputs that will excite a response. The HRFs, estimated in Friston *et al* (1994b) resembled a Poisson or Gamma function, peaking at about 5 seconds. Our understanding of the biophysical and physiological mechanisms that underpin the HRF has grown considerably in the past

few years (*e.g.* Buxton and Frank 1997). Figure 4 shows some simulations based on the hemodynamic model described in Friston *et al* (2000a). Here, neuronal activity induces some auto-regulated signal that causes transient increases in rCBF. The resulting flow increases dilate the venous balloon increasing its volume (v) and diluting venous blood to decrease deoxyhemoglobin content (q). The BOLD signal is roughly proportional to the concentration of deoxyhemoglobin (q/v) and follows the rCBF response with about a seconds delay.

Knowing the forms that the HRF can take is important for several reasons, not least because it allows for better statistical models of the data. The HRF may vary from voxel to voxel and this has to be accommodated in the GLM. To allow for different HRFs in different brain regions the notion of temporal basis functions, to model evoked responses in fMRI, was introduced (Friston *et al* 1995c) and applied to event-related responses in Josephs *et al* (1997) (see also Lange and Zeger 1997). The basic idea behind temporal basis functions is that the hemodynamic response induced by any given trial type can be expressed as the linear combination of several [basis] functions of peristimulus time. The convolution model for fMRI responses takes a stimulus function encoding the supposed neuronal responses and convolves it with a HRF to give a regressor that enters into the design matrix. When using basis functions the stimulus function is convolved with all the basis functions to give a series of regressors. The associated parameter estimates are the coefficients or weights that determine the mixture of basis functions that best models the HRF for the trial type and voxel in question. We find the most useful basis set

to be a canonical HRF and its derivatives with respect to the key parameters that determine its form (*e.g.* latency and dispersion). The nice thing about this approach is that it can partition differences among evoked responses into differences in magnitude, latency or dispersion, that can be tested for using specific contrasts and the SPM{T} (see Friston *et al* 1998b for details).

Temporal basis functions are important because they enable a graceful transition between conventional multi-linear regression models with one stimulus function per condition and FIR models with a parameter for each time point following the onset of a condition or trial type. Figure 5 illustrates this graphically (see Figure legend). In summary, temporal basis functions offer useful constraints on the form of the estimated response that retain (i) the flexibility of FIR models and (ii) the efficiency of single regressor models. The advantage of using several temporal basis functions (as opposed to an assumed form for the HRF) is that one can model voxel-specific forms for hemodynamic responses and formal differences (*e.g.* onset latencies) among responses to different sorts of events. The advantages of using basis functions over FIR models are that (i) the parameters are estimated more efficiently and (ii) stimuli can be presented at any point in the inter-stimulus interval. The latter is important because time-locking stimulus presentation and data acquisition gives a biased sampling over peristimulus time and can lead to differential sensitivities, in multi-slice acquisition, over the brain.

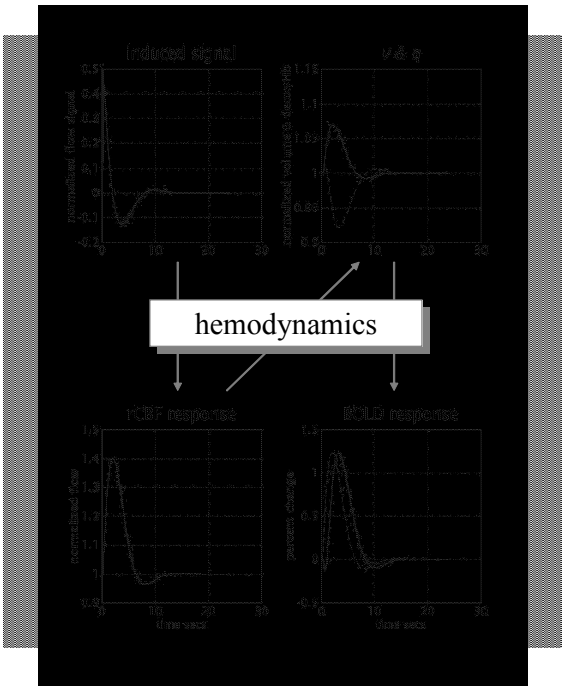


Figure 4. Hemodynamics elicited by an impulse of neuronal activity as predicted by a dynamical biophysical model (see Friston et al 2000a for details). A burst of neuronal activity causes an increase in flow inducing signal that decays with first order kinetic and is down regulated by local flow. This signal increases rCBF with dilates the venous capillaries increasing its volume (v). Concurrently, venous blood is expelled from the venous pool decreasing deoxyhemoglobin content (q). The resulting fall in deoxyhemoglobin concentration leads to a transient increases in BOLD (blood oxygenation level dependent) signal and a subsequent undershoot.

4.2 Statistical inference and the theory of Gaussian fields

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 Gaussian random fields provides a way of correcting the *p*-value that takes into account the fact that neighboring voxels are not independent by virtue of continuity in the original data. Provided the data are sufficiently smooth the GRF correction is less severe (*i.e.* is more sensitive) than a Bonferroni correction for the number of voxels. As noted above GRF theory deals with the multiple comparisons problem in the context of continuous, spatially extended 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 GRF and Bonferroni corrections. Perhaps the most intuitive is to consider the fundamental difference between a SPM and a collection of discrete T values. When declaring a connected volume or region of the SPM to be significant, we refer collectively to all the voxels that comprise that volume. The false positive rate is expressed in terms of connected [excursion] sets of voxels above some threshold, under the null hypothesis of no activation. This is not the expected number of false positive voxels. One false positive volume may contain hundreds of voxels, if the SPM is very smooth. A Bonferroni correction would control the expected number of false positive *voxels*, whereas GRF theory controls the expected number of false positive *regions*. Because a false positive region can contain many voxels the corrected threshold under a GRF correction is much lower, rendering it much more sensitive. In fact the number of voxels in a region is somewhat irrelevant because it is a function of smoothness. The GRF correction discounts voxel size by expressing

the search volume in terms of smoothness or resolution elements (*Resels*). See Figure 6. 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 regions exceeding the threshold.

There are only two assumptions underlying the use of the GRF correction: (i) The error fields (but not necessarily the data) are a reasonable lattice approximation to an underlying random field with a multivariate Gaussian distribution. (ii) These fields are continuous, with a twice-differentiable 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; (i) the data are not smoothed (with or without sub-sampling of the data to preserve resolution), violating the reasonable lattice assumption or (ii) the statistical model is mis-specified so that the errors are not normally distributed. Early formulations of the GRF 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.

4.2.1 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; (i) pre-specify a small search volume and make the appropriate GRF correction (Worsley *et al* 1996) or (ii) used 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 GRF theory.

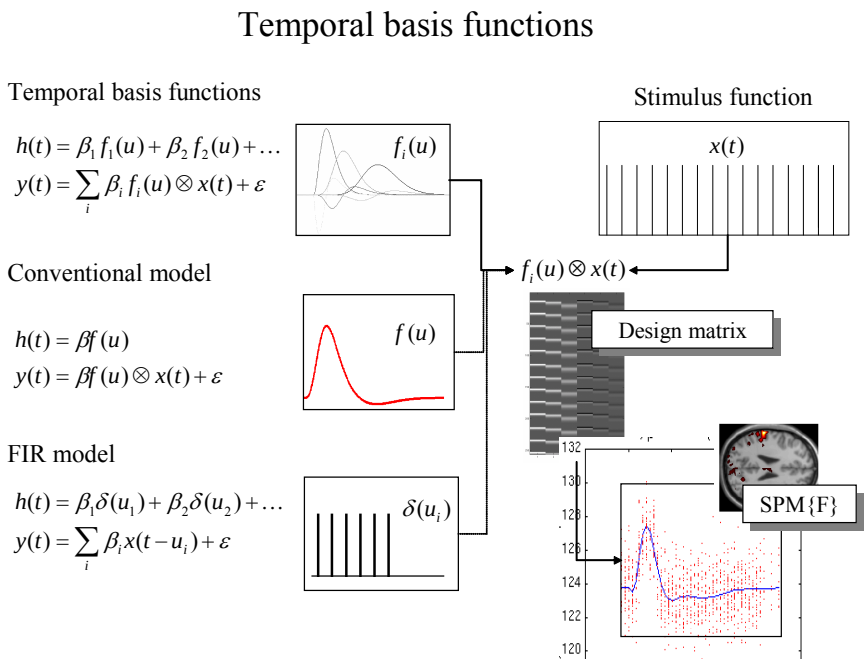


Figure 5. Temporal basis functions offer useful constraints on the form of the estimated response that retain (i) the flexibility of FIR models and (ii) the efficiency of single regressor models. The specification of these constrained FIR models involves setting up stimulus functions x(t) that model expected neuronal changes [e.g. boxcars of epoch-related responses or spikes (delta functions) at the onset of specific events or trials]. These stimulus functions are then convolved with a set of basis functions of peri-stimulus time u, that model the HRF, in some linear combination. The ensuing regressors are assembled into the design matrix. The basis functions can be as simple as a single canonical HRF (middle), through to a series of delayed delta functions (bottom). The latter case corresponds to a FIR model and the coefficients constitute estimates of the impulse response function at a finite number of discrete sampling times, for the event or epoch in question. Selective averaging in event-related fMRI (Dale and Buckner 1997) is mathematically equivalent to this limiting case.

4.2.2 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; (i) the number of activated regions (*i.e.* number of clusters

above the height and volume threshold) - *set level inferences*, (ii) the number of activated voxels (*i.e.* volume) comprising a particular region - *cluster level inferences* and (iii) the *p*-value for each voxel within that cluster - *voxel 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 6 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 voxel-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 (Friston *et al* 1996b) 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 voxel-level inferences. The price paid for this increased sensitivity is reduced localizing power. Voxel-level tests permit individual voxels 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 voxel-

level tests based on peak height. Typically, people use voxel-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.

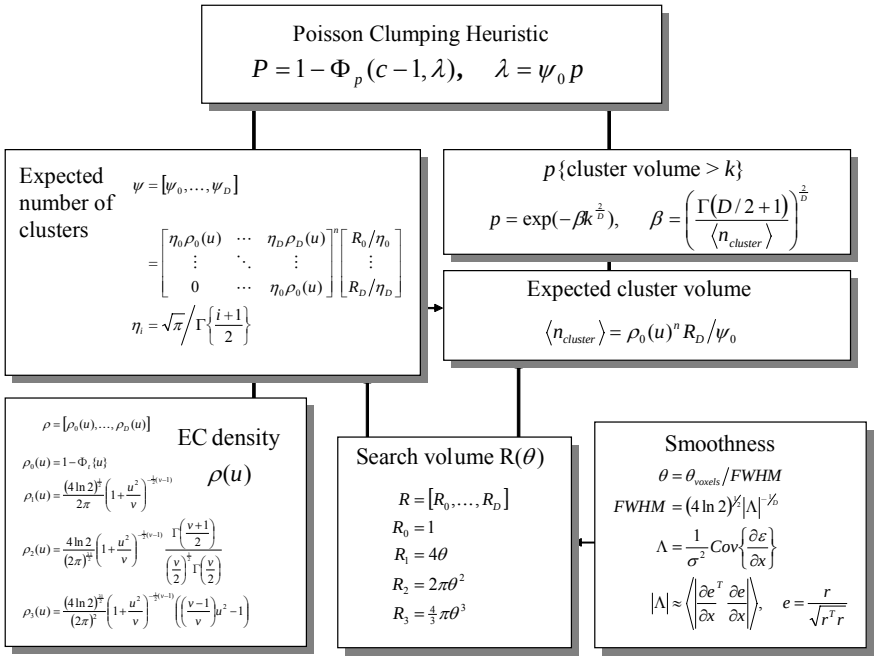


Figure 6. Schematic illustrating the use of Gaussian random field theory in making inferences about activations in SPMs. If one knew where to look exactly, then inference can be based on the value of the statistic at a specified location in the SPM, without correction. However, if one did not have an anatomical constraint a priori, then a correction for multiple dependent comparisons has to be made. These corrections are usually made using distributional approximations from GRF theory. This schematic deals with a general case of n SPM{T}s whose voxels all survive a common threshold u (i.e. a conjunction of n component SPMs). The central probability, upon which all voxel, cluster or set-level inferences are made is the probability P 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 is simply determined. 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 is shown using a form for a single Z-variate field constrained by the expected number of resels per cluster (<.> denotes expectation or average). The expected

number of resels per cluster is simply the expected number of resels in total divided by the expected number of clusters. The expected number of clusters is estimated with the Euler characteristic (EC) (effectively the number of blobs minus the number of holes). This estimate is in turn a function of 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 D dimensional volume of the search space is given by the corresponding element in the vector of resel counts. Resel counts can be thought of as a volume metric 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 3). In this example equations for a sphere of radius are given. denotes the cumulative density function for the sub-scripted statistic in question.

5. EXPERIMENTAL 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 *multifactorial* designs, within this classification the levels of each factor can be *categorical* or *parametric*. We will start be discussing categorical and parametric designs and then deal with multifactorial designs.

5.1 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 hemodynamic responses, evoked by the two tasks, identify the corresponding functionally specialized 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.* color, 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 ones tests a *single* hypothesis pertaining to the activation in one task relative to another. In conjunction analyses *several* hypotheses are tested, asking whether all the activations, in a series of task pairs, are jointly significant. 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 multi-subject fMRI studies, where generic effects are identified as those that are conjointly significant in all the subjects studied (see Section VII).

5.2 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 hemodynamic 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 peri-auditory 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 nonlinear. Using polynomial regression, in the context of the GLM, one can identify nonlinear relationships between stimulus parameters (*e.g.* stimulus duration or presentation rate) and evoked responses. To do this one usually uses a SPM{F} (see Büchel *et al* 1996).

The example provided in Figure 7 illustrates both categorical and parametric aspects of design and analysis. These data were obtained from a 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 on the upper right). To ensure significant motion-sensitive responses, using both color 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{T_{min}} 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 6 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 7 and speak to a compelling inverted 'U' relationship between speed and evoked response that peaks at around 8 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 vs. isochromatic), a main [nonlinear] 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 dorsolateral prefrontal dysfunction irrespective of whether the patient is suffering from schizophrenia or depression.

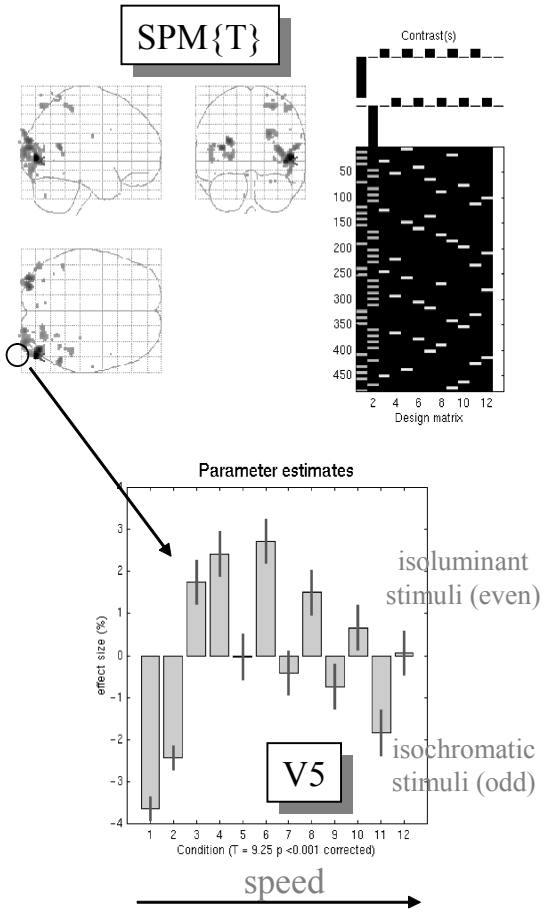


Figure 7. Top right: Design matrix: This is an image representation of the design matrix. Contrasts: These are the vectors of contrast weights defining the linear compounds of parameters tested. The contrast weights are displayed over the column of the design matrix that corresponds to the effects in question. The design matrix here includes condition-specific effects (boxcars convolved with a hemodynamic 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 of the SPM{T} conforming to the standard anatomical space of Talairach and Tournoux (1988). The T 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 6).

5.3 Multifactorial designs

Factorial designs are becoming 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 term. Factorial designs have a wide range of applications. An early application, in neuroimaging, examined physiological 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 (i) the integration of the two or more [cognitive] processes or (ii) the modulation of one [perceptual] process by another being manipulated. See Figure 8 for an example. 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 pathophysiology being studied. 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 in Section VIII from the point of view of effective connectivity.

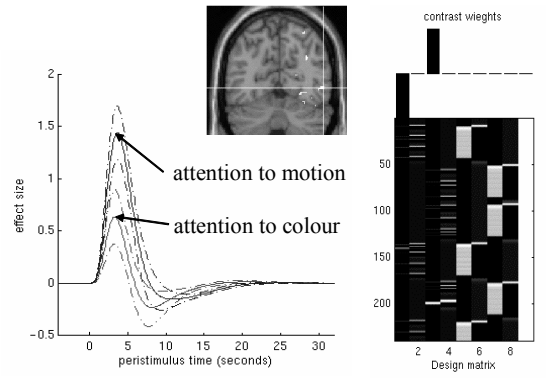


Figure 8. Results showing how to assess an interaction using an event-related design. Subjects viewed stationary monochromatic stimuli that occasionally changed color and moved at the same time. These compound events were presented under two levels of attentional set (attention to color and attention to motion), The event-related responses are modeled, in an attention-specific fashion by the first four regressors (delta functions convolved with a hemodynamic response function and its derivative) in the design matrix on the right. The simple main effects of attention are modeled 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.

6. DESIGNING FMRI STUDIES

In this section we consider fMRI time-series from a signal processing perspective with particular reference to optimal experimental design and efficiency. fMRI time-series can be viewed as a linear admixture of signal and noise. Signal corresponds to neuronally mediated hemodynamic changes that can be modeled as a [non]linear convolution of some underlying neuronal process, responding to changes in experimental factors, by a HRF. Noise has many contributions that render it rather complicated in relation to other neurophysiological measurements. These include neuronal and non-neuronal sources. Neuronal noise refers to neurogenic signal not modeled by the explanatory variables and has the same frequency structure as the signal itself. Non-neuronal components have both white [*e.g.* R.F. (Johnson) noise] and colored components [*e.g.* pulsatile motion of the brain caused by cardiac cycles and local modulation of the static magnetic field B_0 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 3) that can effect sensitivity to experimental effects. Sensitivity depends upon (i) the relative amounts of signal and noise and (ii) the efficiency of the experimental design. Efficiency is simply a measure of how reliable the parameter estimates are and can be defined as the inverse of the variance of a contrast of parameter estimates (see Figure 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.

6.1 The hemodynamic response function and optimum design

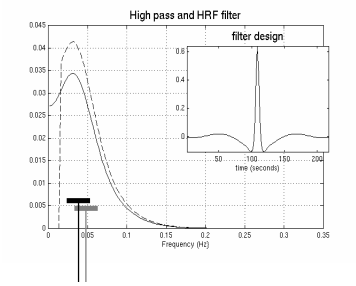
As noted above, a LTI model of neuronally mediated signals in fMRI suggests that, whatever the frequency structure of experimental variables, only those that survive convolution with the hemodynamic response function (HRF) can be estimated with any efficiency. By convolution theorem the experimental 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 9 - solid line). It is clear 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.

A Signal processing perspective

$$y(t) = x(t) \otimes f(t)$$

by convolution theorem

$$g_y(\omega) = g_f(\omega) g_x(\omega)$$



Experimental variance
(~32s cycles)

Figure 9. Modulation transfer function of a canonical hemodynamic response function (HRF), with (broken line) and without (solid line) the addition of a high pass 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 are survive convolution with the HRF. 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 high-pass filtering.

6.2 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 (see Figure 3) that is the inverse of the intrinsic convolution matrix K ($KK^T = \Sigma$). This *whitening* strategy renders the generalized least squares estimator in Figure 3 equivalent

to the Gauss-Markov estimator or minimum variance estimator. However, one generally does not know the form of the intrinsic correlations, which mean it has 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 has several problems. (i) First the estimate of non-sphericity can itself be very inefficient leading to bias in the standard error (Friston *et al* 2000b). (ii) ReML estimation requires an iterative procedure at every voxel and this is computationally prohibitive. (iii) Adopting a different form for the serial correlations at each voxel means the effective degrees of freedom and the null distribution of the statistic will change from voxel to voxel. This violates the assumptions of GRF results for T and F fields (although not very seriously). There are a number of different approaches to these problems that aim to increase the efficiency of the estimation and reduce the computational burden. The approach we have chosen is to forgo the efficiency of the Gauss-Markov estimator and use a generalized least square GLS estimator, after approximately whitening the data with a high-pass filter. The GLS estimator is unbiased and, luckily, is identical to the Gauss-Markov estimator if the regressors in the design matrix are periodic[2]. After GLS estimation Σ is estimated

[2] More exactly, the GLS and ML estimators are the same if X lies within the space spanned by the eigenvectors of Toeplitz autocorrelation matrix Σ.

using ReML and the resulting estimate of $V = S\Sigma S^T$ entered into the expression for the standard error and degrees of freedom provided in Figure 3. To ensure this non-sphericity estimate is robust, we assume it is the same at all voxels. Clearly this is an approximation but can be motivated by the fact we have applied the same high-pass temporal convolution matrix S to all voxels. This ameliorates any voxel to voxel variations in $V = S\Sigma S^T$ (see Figure 3)

The reason that high-pass filtering approximates a whitening filter is that there is a preponderance of low frequencies in the noise. 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, high-pass filtering suppresses low frequency components in the same way that whitening would. An example of a band-pass filter with a high-pass cut-off of 1/64 Hz is shown in inset of Figure 7. This filter's transfer function (the broken line in the main panel) illustrates the frequency structure of neurogenic signals after high-pass filtering.

6.3 Spatially coherent confounds and global normalization

Implicit in the use of high pass filtering is the removal of low frequency components that can be regarded as confounds. Other important confounds are signal components that are artifactual 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, mediated by changes in pCO_2) and non-physiological (*e.g.* transmitter power calibration, B_1 coil profile and 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 intra-cranial voxels will subsume regionally specific effects. This means that the global estimator may be partially collinear with effects of interest, especially if the 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 artifacts in the sense they are real, relative to global changes, but they have little 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.

6.4 Nonlinear system identification approaches

So far we have only considered LTI models and first order HRFs. Another signal processing perspective is provided by nonlinear system identification (Vazquez and Noll 1998). This section considers nonlinear models as a prelude to the next subsection on event-related fMRI, where nonlinear interactions among evoked responses provide constraints for experimental design and analysis. We have described an approach to characterizing evoked hemodynamic responses in fMRI based on nonlinear system identification, in particular the use of *Volterra series* (Friston *et al* 1998a). This approach enables one to estimate Volterra kernels that describe the relationship between stimulus presentation and the hemodynamic responses that ensue. Volterra series are essentially high order extensions of linear convolution models. These kernels therefore represent a nonlinear 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 (i) using a second order approximation to the Volterra series to formulate the problem in terms of a general linear model and (ii) expanding the kernels in terms of temporal basis functions. 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 the nonlinear effects, captured by the 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 behavior. This behavior appears to be specific to BOLD effects (as distinct from evoked changes in cerebral blood flow) and may represent a *hemodynamic refractoriness*. This effect has important implications for event-related fMRI, where one may want to present trials in quick succession (see below).

The results of a typical nonlinear analysis are given in Figure 10. 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 were 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 nonlinear effects (*i.e.* that predicted by setting the second order kernel to zero). It is clear that nonlinearities 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 (i) SOAs should not really fall much below one second and (ii) 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 behaviors may be expressed in primary sensory cortex where the feasibility of using minimum SOAs as low as 500ms 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 and event-related designs.

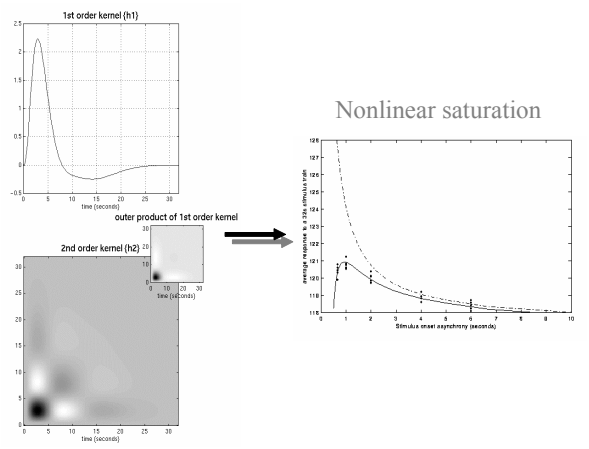


Figure 10. Left panels: Volterra kernels from a voxel in the left superior temporal gyrus at -56, -28, 12mm. These kernel estimates were based on a single subject study of aural word presentation at different rates (from 0 to 90 words per minute) using a second order approximation to a Volterra series expansion modeling the observed hemodynamic response to stimulus input (a delta function for each word). These kernels can be thought of as a characterization of the second order hemodynamic response function. The first order kernel (upper panel) represents the (first order) component usually presented in linear analyses. The second order kernel (lower panel) is presented in image format. The color scale is arbitrary; white is positive and black is negative. The insert on the right represents the second order

kernel that would be predicted by a simple model that involved linear convolution with followed by some static nonlinearly.

Right panel: Integrated responses over a 32-second stimulus train as a function of SOA. Solid line: Estimates based on the nonlinear 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.

6.5 Event and epoch-related designs

A crucial distinction, in experimental design for fMRI, is that between *epoch* and *event*-related designs. In SPECT and 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 that may be important in some cognitive and clinical contexts. 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 in experimental design 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 to 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 some common framework that accommodates them all. 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). 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 11 (right panel) for an SOA of 1 second and 32 expected events or trials over a 64 second period (except for the first deterministic example with 4 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 whilst 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 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 11. 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 trail 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 FIR estimation the regressors become a series of stick functions (see Figure 5) 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 (Henson *et al* in preparation) 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.

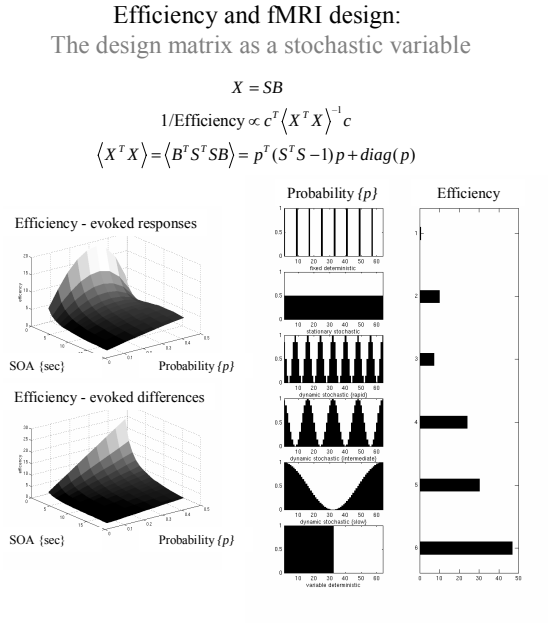


Figure 11. Efficiency as a function of occurrence probabilities p. for a model X formed by post-multiplying S (a matrix containing n columns, modeling n possible event-related responses every SOA) by B. B is a random binary vector that determines whether the nth response is included in X or not, where . Right panels: A comparison of some common designs. A graphical representation of the occurrence probabilities p expressed 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 probability p 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.

7. INFERENCES ABOUT SUBJECTS AND POPULATIONS

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 variability with which that response in 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 constructed 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 the 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. In order to make the inference that the group showed a significant activation one would have to assess the

variability in 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-effect analysis and the second a random-effect 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-effect 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 differential responses by comparing the mean activation to the variability in activations 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.

7.1 Random-effects analyses

The implementation of random-effect analyses in SPM is fairly straightforward and involves taking the contrasts of parameters estimated from a *first-level* (fixed-effect) analysis and entering them into a *second-level* (random-effect) 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. The nature of the inference made is determined by the contrasts entered into the second level (see Figure 12). 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 analyses, using a hierarchical observation model, 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 variance at the second level contains the right mixture of variance induced by observation error at the first level and between-subject error at the second. It is important to appreciate this because the efficiency of the design at the first level percolates up to higher levels. It is therefore important to use efficient strategies at all levels in a hierarchical design.

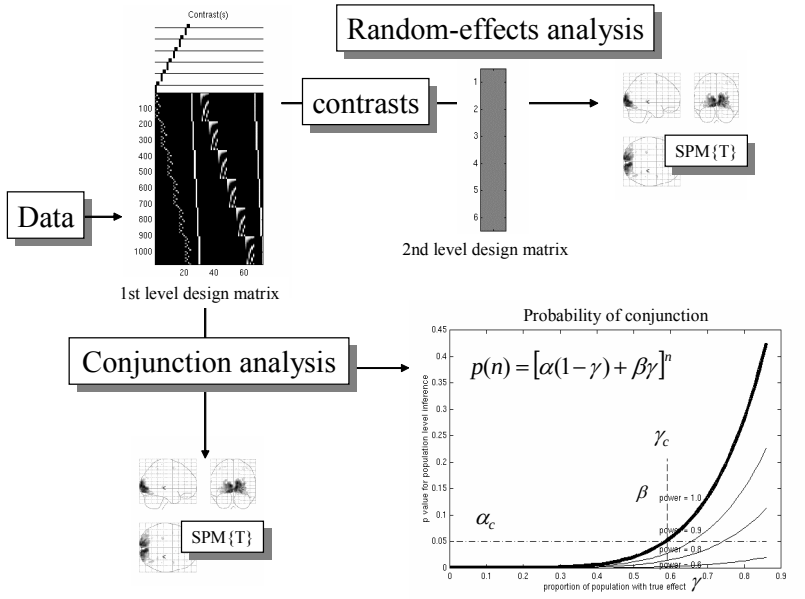


Figure 12. Schematic illustrating the implementation of random-effect and conjunction analyses for population inference. The lower right graph shows the probability p(n) of obtaining a conjunction over n subjects, conditional on a certain proportion □ of the population expressing the effect, for a test with specificity of □ = 0.05, at several sensitivities (□ = 1, 0.9, 0.8 and 0.6). The critical specificity for population inference □c and the associated proportion of the population □c are denoted by the broken lines.

7.2 Conjunction analyses and population inferences

In some instances a fixed effects analysis is more appropriate, particularly to facilitate the reporting of a series of single-case studies. Among these 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 of this sort. Imagine that in 16 subjects the activation in V5, elicited by a motion stimulus, was great 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 $p = 0.5$ (T = 0) for each subject. This result constitutes evidence that V5 is involved in motion processing. However, note that this 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 of activations is not synonymous with a conjunction of significant activations.

The motivations for conjunction analyses, in the context of multi-subject studies are twofold. (i) They provide an inference, in a fixed-effect analysis testing the null hypotheses of no activation in any of the subjects, that can be much more sensitive than testing for the average activation. (ii) They can be extended to make inferences about the population as described next

If, for any given contrast, one can establish a conjunction of effects over n subjects using a test with a specificity of α and sensitivity β, the probability of this occurring by chance can be expressed as a function of γ, the proportion of the population that would have activated (see the equation in Figure 12 - lower right panel). This probability has an upper bound α_c corresponding to a critical proportion γ_c that is realized when (the generally unknown) sensitivity is one. In other words, under the null hypothesis that the proportion of the population evidencing this effect is less than or equal

to γ_c, the probability of getting a conjunction over *n* subjects is equal to, or less than, α_c. In short a conjunction allows one to say, with a specificity of α_c, that more than γ_c of the population show the effect in question. Formally, we can view this analysis as a conservative 100(1 - α_c)% confidence region for the unknown parameter γ. These inferences can be construed as statements about how typical the effect is, without saying that it is necessarily present in every subject.

In practice, a conjunction analysis of a multi-subject study comprises the following steps: (i) A design matrix is constructed where the explanatory variables pertaining to each experimental condition are replicated for each subject. This subject-separable design matrix implicitly models subject by condition interactions (*i.e.* different condition-specific responses among sessions). (ii) Contrasts are then specified that test for the effect of interest in each subjects to give a series of SPM{T} that can be reported as a series of 'single-case' studies in the usual way. (iii) These SPM{T} are combined at a threshold *u* (corresponding to the specificity α in Figure 12) to give a SPM{T_{min}} (*i.e.* conjunction SPM). The corrected *p*-values associated with each voxel are computed as described in Figure 6. These *p*-values provide for inferences about effects that are common to the particular subjects studied. Because we have demonstrated regionally specific conjunctions, one can also proceed to make an inference about the population from which these subjects came using the confidence region approach described above (see Friston *et al* 1999b for a fuller discussion).

8. EFFECTIVE CONNECTIVITY

8.1 Functional and Effective connectivity

Imaging neuroscience has firmly established functional specialization as a principle of brain organization in man. The functional integration of specialized areas has proven more difficult to assess. Functional integration is usually inferred on the basis of correlations among measurements of neuronal activity. Functional connectivity has been defined as *correlations between remote neurophysiological events*. However correlations can arise in a variety of ways: For example in multi-unit electrode recordings they can result from stimulus-locked transients evoked by a common input or reflect stimulus-induced oscillations mediated by synaptic connections (Gerstein and Perkel 1969). Integration within a distributed system is usually better understood in terms of effective connectivity: Effective connectivity refers explicitly to *the influence that one neural system exerts over another*, either at a synaptic (*i.e.* synaptic efficacy) or population level. It has been proposed that "the [electrophysiological] notion of effective connectivity should be understood as the experiment- and time-dependent, simplest possible circuit diagram that would replicate the observed timing relationships between the recorded neurons" (Aertsen and Preißl 1991). This speaks to two important points: (i) Effective connectivity is dynamic,

i.e. activity- and time-dependent and (ii) it depends upon a model of the interactions. The estimation procedures employed in functional neuroimaging can be classified as (i) those based directly on regression (Friston *et al* 1995d) or (ii) structural equation modeling (McIntosh *et al* 1994) (*i.e.* path analysis).

There is a necessary relationship between approaches to characterizing functional integration and multivariate analyses because the latter are necessary to model interactions among brain regions. Multivariate approaches can be divided into those that are inferential in nature and those that are data led or exploratory.. We will first consider multivariate approaches that are universally based on functional connectivity or covariance patterns (and are generally exploratory) and then turn to models of effective connectivity (that usually allow for some form of inference)

8.2 Eigenimage analysis and related approaches

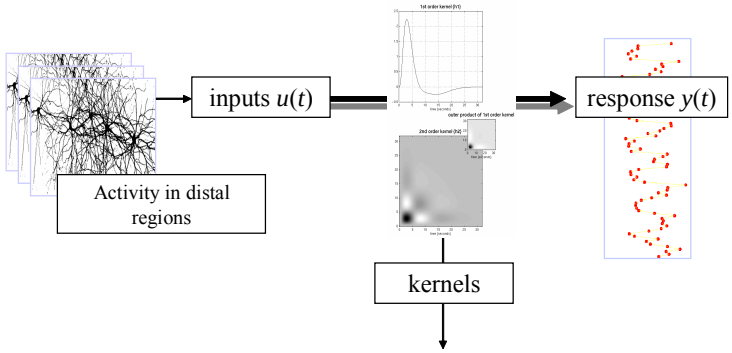
Most analyses of covariances among brain regions are based on the singular value decomposition (SVD) of the between-voxel covariances in a neuroimaging time-series In Friston *et al* (1993) we introduced voxel-based principal component analysis (PCA) of neuroimaging time-series to characterize distributed brain systems implicated in sensorimotor, perceptual or cognitive processes. These distributed systems are identified with principal components or *eigenimages* that correspond to spatial modes of coherent brain activity. This approach represents one of the simplest multivariate characterizations of functional neuroimaging time-series and falls into

the class of exploratory analyses. Principal component or eigenimage analysis generally uses SVD to identify a set of orthogonal spatial modes that capture the greatest amount of variance, expressed over time. As such the ensuing modes embody the most prominent aspects of the variance-covariance structure of a given time-series. Noting that the covariances among brain regions is equivalent to functional connectivity renders eigenimage analysis particularly interesting because it was among the first ways of addressing functional integration (*i.e.* connectivity) with neuroimaging data. Subsequently, eigenimage analysis has been elaborated in a number of ways. Notable among these are canonical variate analysis (CVA) and multidimensional scaling (Friston et al 1996d,e). Canonical variate analysis was introduced in the context of ManCova (multiple analysis of covariance) and uses the generalized eigenvector solution to maximize the variance that can be explained by some explanatory variables relative to error. CVA can be thought of as an extension of eigenimage analysis that refers explicitly to some explanatory variables and allows for statistical inference.

In fMRI eigenimage analysis (Sychra *et al* 1994) is generally used as an exploratory device to characterize coherent brain activity. These variance components may, or may not be, related to experimental design and endogenous coherent dynamics have been observed in the motor system (Biswal *et al* 1995). Despite its exploratory power eigenimage analysis is fundamentally limited for two reasons. Firstly, it offers only a linear decomposition of any set of neurophysiological measurements and secondly the particular set of eigenimages or spatial modes obtained is uniquely determined by

constraints that are biologically implausible. These aspects of PCA confer inherent limitations on the interpretability and usefulness of eigenimage analysis of biological time-series and have motivated the exploration of nonlinear PCA and neural network approaches (*e.g.* Mørch et al 1995).

Two other important approaches deserve mention here. The first is independent component analysis (ICA). ICA uses entropy maximization to find, using iterative schemes, spatial modes or their dynamics that are approximately *independent*. This is a stronger requirement than *orthogonality* in PCA and involves removing high order correlations among the modes (or dynamics). It was initially introduced as *spatial* ICA (McKeown *et al* 1998) in which the independence constraint was applied to the modes (with no constraints on their temporal expression). More recent approaches use, by analogy with magneto- and electrophysiological time-series analysis, *temporal* ICA where the dynamics are enforced to be independent. This requires an initial dimension reduction (usually using conventional eigenimage analysis). Finally, there has been an interest in cluster analysis (Baumgartner et al 1997). Conceptually, this can be related to eigenimage analysis through multidimensional scaling and principal coordinates analysis. In cluster analysis voxels in a multidimensional scaling space are assigned belonging probabilities to a small number of clusters, thereby characterizing the temporal dynamics (in terms of the cluster centroids) and spatial modes (defined by the belonging probability for each cluster). These approaches eschew many of the unnatural constraints imposed by eigenimage analysis and can be a useful exploratory device.



Volterra series a general nonlinear input-state-output characterization

$$y(t) = \kappa_0 + \sum_{i=1}^{\infty} \int_0^t \cdots \int_0^t \kappa_i(\sigma_1, \ldots \sigma_i) u(\sigma_1) \ldots u(\sigma_i) d\sigma_1 \ldots d\sigma_i$$

V5

$$\kappa_i(\sigma_1) = \frac{\partial^i y(t)}{\partial u(\sigma_1)}, \qquad \kappa_i(\sigma_1, \sigma_2) = \frac{\partial^2 y(t)}{\partial u(\sigma_1) \partial u(\sigma_2)}, \qquad \cdots$$

PPC V2

n.b. Volterra kernels are synonymous with effective connectivity

Figure 13. Schematic depicting the causal relationship between the outputs and the recent history of the inputs to a nonlinear dynamical system, in this instance a brain region or voxel. This relationship can be expressed as a Volterra series, which expresses the response or output y(t) as a nonlinear convolution of the inputs u(t), critically without reference to any [hidden] state variables. This series is simply a functional Taylor expansion of y(t) as a function of the inputs over the recent past. is the ith order kernel. Volterra series have been described as a 'power series with memory' and are generally thought of as a high-order or 'nonlinear convolution' of the inputs to provide an output. Volterra kernels are useful in characterizing the effective connectivity or influences that one neuronal system exerts over another because they represent the causal characteristics of the system in question. Neurobiologically they have a simple and compelling interpretation – they are synonymous with effective connectivity. It is evident that the first-order kernel embodies the response evoked by a change in input at . In other words it is a time-dependant measure of driving efficacy. Similarly the second order kernel reflects the modulatory influence of the input at on the evoked response at . And so on for higher orders.

8.3 Characterizing nonlinear coupling among brain areas

Linear models of effective connectivity assume that the multiple inputs to a brain region are linearly separable. This assumption precludes activity-

dependent connections that are expressed in one context and not in another. The resolution of this problem lies in adopting nonlinear models like the Volterra formulation that include interactions among inputs. These interactions can be construed as a context- or activity-dependent modulation of the influence that one region exerts over another, where that context is instantiated by activity in further brain regions exerting modulatory effects. These nonlinearities can be introduced into structural equation modeling using so-called 'moderator' variables that represent the interaction between two regions in causing activity in a third (Büchel *et al* 1997). From the point of view of regression models modulatory effects can be modeled with nonlinear input-output models and in particular the Volterra formulation described above. In this instance the inputs are not stimuli but activities from other regions. Because the kernels are high-order they embody interactions over time and among inputs and can be thought of as explicit measures of effective connectivity (see Figure 13). An important thing about the Volterra formulation is that it has a high face validity and biological plausibility. The only thing it assumes is that the response of a region is some analytic nonlinear function of the inputs over the recent past. This function exists even for complicated dynamical systems with many [unobservable] state variables. Within these models, the influence of one region on another has two components; (i) the direct or *driving* influence of input from the first (*e.g.* hierarchically lower) region, irrespective of the activities elsewhere and (ii) an activity-dependent, *modulatory* component that represents an interaction with inputs from the remaining (*e.g.* hierarchically higher) regions. These are mediated by the first and second order kernels respectively. The example provided in Figure 14 addresses the modulation of visual cortical responses by attentional mechanisms (*e.g.* Treue and Maunsell 1996) and the mediating role of activity-dependent changes in effective connectivity:

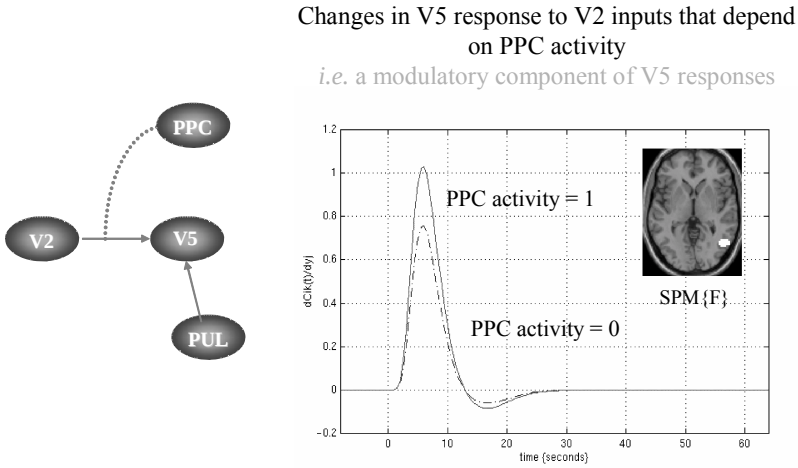


Figure 14. Left: Brain regions and connections comprising the effective connectivity model formulated in terms of a Volterra series (see Figure 13). Right: Characterization of the effects of V2 inputs on V5 and their modulation by posterior parietal cortex (PPC). The broken lines represent estimates of V5 responses when PPC activity is zero, according to a second order Volterra model of effective connectivity with inputs to V5 from V2, PPC and the pulvinar (PUL). The solid curves represent the same response when PPC activity is one standard deviation of its variation over conditions. It is evident that V2 has an activating effect on V5 and that PPC increases the responsiveness of V5 to these inputs. The insert shows all the voxels in V5 that evidenced a modulatory effect ($p < 0.05$ uncorrected). These voxels were identified by thresholding a SPM{F} testing for the contribution of second order kernels involving V2 and PPC (treating all other terms as nuisance variables). The data were obtained with fMRI under identical stimulus conditions (visual motion subtended by radially moving dots) whilst manipulating the attentional component of the task (detection of velocity changes).

The right panel in Figure 14 shows a characterization of this modulatory effect in terms of the increase in V5 responses, to a simulated V2 input, when posterior parietal activity is zero (broken line) and when it is high (solid lines). In this study subjects were studied with fMRI under identical stimulus conditions (visual motion

subtended by radially moving dots) whilst manipulating the attentional component of the task (detection of velocity changes). The brain regions and connections comprising the model are shown in the upper panel. The lower panel shows a characterization of the effects of V2 inputs on V5 and their modulation by posterior parietal cortex (PPC) using simulated inputs at different levels of PPC activity. It is evident that V2 has an activating effect on V5 and that PPC increases the responsiveness of V5 to these inputs. The insert shows all the voxels in V5 that evidenced a modulatory effect ($p < 0.05$ uncorrected). These voxels were identified by thresholding a SPM{F} testing for the contribution of second order kernels involving V2 and PPC while treating all other components as nuisance variables. The estimation of the Volterra kernels and statistical inference procedure is described in Friston and Büchel (2000c).

Acknowledgements

The Wellcome Trust funded this work.

References

Absher JR and Benson DF. (1993) Disconnection syndromes: an overview of Geschwind's contributions. *Neurology* **43**:862-867

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

Aertsen A and Preißl H. (1991) Dynamics of activity and connectivity in physiological neuronal Networks. in *Non Linear Dynamics and Neuronal Networks*. Ed Schuster HG VCH publishers Inc. New York NY USA p281-302

Aguirre GK Zarahn E and 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**:500-505

Andersson JL, Hutton C, Ashburner J, Turner R, Friston K. (2001) Modeling geometric deformations in EPI time series. *NeuroImage*. **13**:903-19

Ashburner J Neelin P Collins DL Evans A and Friston K. (1997) Incorporating prior knowledge into image registration. *NeuroImage* **6**:344-352

Ashburner J, and Friston KJ.(1999) Nonlinear spatial normalization using basis functions. *Hum Brain Mapp*. **7**:254-66.

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

Bandettini PA Jesmanowicz A Wong EC and Hyde JS. (1993) Processing strategies for time course data sets in functional MRI of the human brain. *Mag. Res. Med*. **30**:161-173

Baumgartner R Scarth G Teichtmeister C Somorjai R and Moser E. (1997) Fuzzy clustering of gradient-echo functional MRI in the human visual cortex Part 1: reproducibility. *J Mag. Res. Imaging* **7**:1094-1101

Biswal B Yetkin FZ Haughton VM and Hyde JS. (1995) Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Mag. Res. Med.* **34**:537-541

Boynton GM Engel SA Glover GH and Heeger DJ. (1996) Linear systems analysis of functional magnetic resonance imaging in human V1. *J Neurosci.* **16**:4207-4221

Büchel C Wise RJS Mummery CJ Poline J-B and Friston KJ. (1996) Nonlinear regression in parametric activation studies. *NeuroImage* **4**:60-66

Büchel C and Friston KJ. (1997) Modulation of connectivity in visual pathways by attention: Cortical interactions evaluated with structural equation modeling and fMRI. *Cerebral Cortex* **7**:768-778

Buckner R Bandettini P O'Craven K Savoy R Petersen S Raichle M and Rosen B. (1996) Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging. *Proc. Natl. Acad. Sci. USA* **93**:14878-14883

Bullmore ET Brammer MJ Williams SCR Rabe-Hesketh S Janot N David A Mellers J Howard R and Sham P. (1996) Statistical methods of estimation and inference for functional MR images. *Mag. Res. Med.* **35**:261-277

Burock MA Buckner RL Woldorff MG Rosen BR and Dale AM. (1998) Randomized Event-Related Experimental Designs Allow for Extremely Rapid Presentation Rates Using Functional MRI. *NeuroReport* **9**:3735-3739

Buxton RB and Frank LR. (1997) A model for the coupling between cerebral blood flow and oxygen metabolism during neural stimulation. *J. Cereb. Blood. Flow Metab*. **17**:64-72.

Clark VP Maisog JM and Haxby JV. (1998) fMRI study of face perception and memory using random stimulus sequences. *J Neurophysiol.* **76**:3257-3265

Dale A and Buckner R. (1997) Selective averaging of rapidly presented individual trials using fMRI. *Hum Brain Mapp.* **5**:329-340

Friston KJ Frith CD Liddle PF and Frackowiak RSJ. (1991) Comparing functional (PET) images: the assessment of significant change. *J. Cereb. Blood Flow Metab.* **11**:690-699

Friston KJ Frith C Passingham RE Liddle PF and Frackowiak RSJ. (1992a) Motor practice and neurophysiological adaptation in the cerebellum: A positron tomography study. *Proc. Roy. Soc. Lon. Series B* **248**:223-228

Friston KJ Grasby P Bench C Frith CD Cowen PJ Little P Frackowiak RSJ and Dolan R. (1992b) Measuring the neuromodulatory effects of drugs in man with positron tomography. *Neurosci. Lett.* **141**:106-110

Friston KJ Frith C Liddle P and Frackowiak RSJ. (1993) Functional Connectivity: The principal component analysis of large data sets. *J Cereb. Blood Flow Metab.* **13**:5-14

Friston KJ Worsley KJ Frackowiak RSJ Mazziotta JC and Evans AC. (1994a) Assessing the significance of focal activations using their spatial extent. *Hum. Brain Mapp.* **1**:214-220

Friston KJ Jezzard PJ and Turner R. (1994b) Analysis of functional MRI time-*series Hum. Brain Mapp*. **1**:153-171

Friston KJ Ashburner J Frith CD Poline J-B Heather JD and Frackowiak RSJ. (1995a) Spatial registration and normalization of images. *Hum. Brain Mapp.* **2**:165-189

Friston KJ Holmes AP Worsley KJ Poline JB Frith CD and Frackowiak RSJ. (1995b) Statistical Parametric Maps in functional imaging: A general linear approach *Hum. Brain Mapp.* **2**:189-210

Friston KJ Frith CD Turner R and Frackowiak RSJ. (1995c) Characterizing evoked hemodynamics with fMRI. *NeuroImage* **2**:157-165

Friston KJ Ungerleider LG Jezzard P and Turner R. (1995d) Characterizing modulatory interactions between V1 and V2 in human cortex with fMRI. *Hum. Brain Mapp*. **2**: 211-224

Friston KJ Williams S Howard R Frackowiak RSJ and Turner R. (1996a) Movement related effects in fMRI time series. *Mag. Res. Med.* **35**:346-355

Friston KJ Holmes A Poline J-B Price CJ and Frith CD. (1996b) Detecting activations in PET and fMRI: levels of inference and power. *NeuroImage* **4**:223-235

Friston KJ Price CJ Fletcher P Moore C Frackowiak RSJ and Dolan RJ. (1996c) The trouble with cognitive subtraction. *NeuroImage* **4**:97-104

Friston KJ Poline J-B Holmes AP Frith CD and Frackowiak RSJ. (1996d) A multivariate analysis of PET activation studies. *Hum. Brain Mapp.* **4**:140-151

Friston KJ Frith CD Fletcher P Liddle PF and Frackowiak RSJ. (1996e) Functional topography: multidimensional scaling and functional connectivity in the brain. *Cerebral Cortex* **6**:156-164

Friston KJ. (1997) Testing for anatomical specified regional effects. *Hum. Brain Mapp*. **5**:133-136

Friston KJ Büchel C Fink GR Morris J Rolls E and Dolan RJ (1997) Psychophysiological and modulatory interactions in neuroimaging. *NeuroImage* **6**:218-229

Friston KJ Josephs O Rees G and Turner R. (1998a) Non-linear event-related responses in fMRI. *Mag. Res. Med.* **39**:41-52

Friston KJ Fletcher P Josephs O Holmes A Rugg MD and Turner R. (1998b) Event-related fMRI: Characterizing differential responses. *NeuroImage* **7**:30-40

Friston KJ, Zarahn E, Josephs O, Henson RN, Dale AM (1999a) Stochastic designs in event-related fMRI. *NeuroImage.* **10**:607-19.

Friston KJ, Holmes AP, Price CJ, Buchel C, Worsley KJ. (1999b) Multisubject fMRI studies and conjunction analyses. *NeuroImage*. **10**:385-96.

Friston KJ, Mechelli A, Turner R, Price CJ.(2000a) Nonlinear responses in fMRI: the Balloon model, Volterra kernels, and other hemodynamics. *NeuroImage.* **12**:466-77.

Friston KJ, Josephs O, Zarahn E, Holmes AP, Rouquette S, Poline J. (2000b) To smooth or not to smooth? Bias and efficiency in fMRI time-series analysis. *NeuroImage*. **12**:196-208.

Friston KJ, Buchel C.(2000c) Attentional modulation of effective connectivity from V2 to V5/MT in humans. *Proc Natl Acad Sci U S A*. **97**:7591-6.

Gerstein GL and Perkel DH. (1969) Simultaneously recorded trains of action potentials: Analysis and functional interpretation. *Science* **164**: 828-830

Girard P and Bullier J. (1989) Visual activity in area V2 during reversible inactivation of area 17 in the macaque monkey. *J Neurophysiol.* **62**:1287-1301

Goltz F. (1881) in "*Transactions of the 7th international medical congress*" (W MacCormac Ed) Vol. I JW Kolkmann London. p218-228

Grafton S Mazziotta J Presty S Friston KJ Frackowiak RSJ and Phelps M. (1992) Functional anatomy of human procedural learning determined with regional cerebral blood flow and PET. *J Neurosci.* **12**:2542-2548

Grootoonk S, Hutton C, Ashburner J, Howseman AM, Josephs O, Rees G, Friston KJ, Turner R. (2000) Characterization and correction of interpolation effects in the realignment of fMRI time series. *NeuroImage.* **11**:49-57.

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

Hirsch JA and Gilbert CD. (1991) Synaptic physiology of horizontal connections in the cat's visual cortex. *J. Neurosci.* **11**:1800-1809

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

Josephs O Turner R and Friston KJ. (1997) Event-related fMRI *Hum. Brain Mapp.* **5**:243-248

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

Lange N and Zeger SL. (1997) Non-linear Fourier time series analysis for human brain mapping by functional magnetic resonance imaging (with discussion) *J Roy. Stat. Soc. Ser C.* **46**:1-29

Liddle PF Friston KJ Frith CD and Frackowiak RSJ. (1992) Cerebral Blood-flow and mental processes in schizophrenia *J Royal Soc. of Med.* **85**:224-227

Lueck CJ Zeki S Friston KJ Deiber MP Cope NO Cunningham VJ Lammertsma AA Kennard C and Frackowiak RSJ. (1989) The color centre in the cerebral cortex of man. *Nature* **340**:386-389

McIntosh AR and Gonzalez-Lima F. (1994) Structural equation modeling and its application to network analysis in functional brain imaging. *Hum. Brain Mapp*. **2**: 2-22

McKeown M Jung T-P Makeig S Brown G Kinderman S Lee T-W and Sejnowski T. (1998) Spatially independent activity patterns in functional MRI data during the Stroop color naming task. *Proc. Natl. Acad. Sci.* **95**:803-810

Mørch N Kjems U Hansen LK Svarer C Law I Lautrup B and Strother SC. (1995) Visualization of neural networks using saliency maps. IEEE International Conference on Neural Networks Perth Australia. 2085-2090

Petersen SE Fox PT Posner MI Mintun M and Raichle ME. (1989) Positron emission tomographic studies of the processing of single words. *J Cog. Neurosci*. **1**:153-170

Phillips CG Zeki S and HB Barlow HB. (1984) Localisation of function in the cerebral cortex Past present and future. *Brain* **107**:327-361

Purdon PL and 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

Price CJ Wise RJS Ramsay S Friston KJ Howard D Patterson K and Frackowiak RSJ. (1992) Regional response differences within the human auditory cortex when listening to words. *Neurosci. Lett*. **146**:179-182

Price CJ and Friston KJ. (1997) Cognitive Conjunction: A new approach to brain activation experiments. *NeuroImage* **5**:261-270

Sychra JJ Bandettini PA Bhattacharya N and Lin Q. (1994) Synthetic images by subspace transforms I Principal component images and related filters. *Med. Physics* **21**:193-201

Talairach P and Tournoux J. (1988) "*A Stereotactic coplanar atlas of the human brain*" Stuttgart Thieme

Treue S and Maunsell HR. (1996) Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature* **382**: 539-41

Vazquez AL and Noll CD. (1998) Nonlinear aspects of the BOLD response in functional MRI. *NeuroImage* **7**:108-118

Worsley KJ Evans AC Marrett S and Neelin P. (1992) A three-dimensional statistical analysis for rCBF activation studies in human brain. *J Cereb. Blood Flow Metab*. **12**:900-918

Worsley KJ and Friston KJ. (1995) Analysis of fMRI time-series revisited - again. *NeuroImage* **2**:173-181

Worsley KJ Marrett S Neelin P Vandal AC Friston KJ and Evans AC. (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**:179-197

Zeki S. (1990) The motion pathways of the visual cortex. in "*Vision: coding and efficiency*" C Blakemore Ed. Cambridge University Press UK p321-345