Technical Note

Mixed-effects and fMRI studies

K.J. Friston,[a] K.E. Stephan,[a,*] T.E. Lund,[b] A. Morcom,[c] and S. Kiebel[a]

[a]The Wellcome Department of Imaging Neuroscience, Institute of Neurology, 12 Queen Square, London WC1N 3BG, UK
[b]Danish Research Centre for Magnetic Resonance, Copenhagen University Hospital, 2650 Hvidovre, Denmark
[c]Department of Psychiatry, Brain Mapping Unit, University of Cambridge, Cambridge CB2 2QQ, UK

This note concerns mixed-effect (MFX) analyses in multisession functional magnetic resonance imaging (fMRI) studies. It clarifies the relationship between mixed-effect analyses and the two-stage "summary statistics" procedure (Holmes, A.P., Friston, K.J., 1998. Generalisability, random effects and population inference. NeuroImage 7, S754) that has been adopted widely for analyses of fMRI data at the group level. We describe a simple procedure, based on restricted maximum likelihood (ReML) estimates of covariance components, that enables full mixed-effects analyses in the context of statistical parametric mapping. Using this procedure, we compare the results of a full mixed-effects analysis with those obtained from the simpler two-stage procedure and comment on the situations when the two approaches may give different results.
© 2004 Elsevier Inc. All rights reserved.

Introduction

It is now standard practice, in functional magnetic resonance imaging (fMRI), to distinguish between fixed and random (i.e., mixed) effects analyses of multisession or subject fMRI studies. Multisession analyses rest upon a two-level linear hierarchical observation model and the distinction between within-session and between-session effects. Fixed-effect analyses use models in which the interaction between the effect and session is treated as a fixed quantity and the effect (e.g., activation) is compared against within-session error. This corresponds to inferring that a subject's response is significant in relation to the precision with which it can be measured. Conversely, the session or subject-specific effect can be treated as a random variable, such that the observed response is a mixture of fixed and random effects; hence, mixed-

effects (MFX) analyses. Here, the inference is at the second level and pertains to significant effects that are large in relation to between-session variability.

These models can be expressed mathematically as

$$y = \mathbf{X}_0\beta_0 + \mathbf{X}^{(1)}\beta^{(1)} + \varepsilon^{(1)}$$

$$\beta^{(1)} = \mathbf{X}^{(2)}\beta^{(2)} + \varepsilon^{(2)} \tag{1}$$

where y is the response variable, $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ correspond to the first- and second-level design matrices containing effects of interest, and X_0 represents some confounds or nuisance variables (e.g., drift). We will assume that the effects of interest and nuisance effects are orthogonal $X_0^T\mathbf{X}^{(1)} = 0$. $\varepsilon^{(1)}$ corresponds to the first-level error with within-session covariance $\mathrm{Cov}\{\varepsilon^{(1)}\} = \mathbf{C}^{(1)}$. $\varepsilon^{(2)}$ corresponds to the second-level error with between-session covariance $\mathrm{Cov}\{\varepsilon^{(2)}\} = \mathrm{Cov}\{\beta^{(1)}\} = \mathbf{C}^{(2)}$. This is also the covariance of the first-level parameters $\beta^{(1)}$ that are, implicitly, random quantities. Technically, models that conform to Eq. (1) belong to the class of conditionally independent hierarchical models when the response variables and parameters are assumed to be independent across units, conditionally on the hyper-parameters controlling the error terms (Kass and Steffey, 1989).

Typically, the first-level design matrix $\mathbf{X}^{(1)}$ has a number of partitions modeling session-specific effects

$$\mathbf{X}^{(1)} = \begin{bmatrix} \mathbf{X}_1^{(1)} & \cdots 0 & 0 \\ 0 & \ddots & \vdots \\ \vdots & & 0 \\ 0 & \cdots 0 & \mathbf{X}_S^{(1)} \end{bmatrix} \tag{2}$$

for S sessions. In a two-level model, the response variable y has two error covariance components, one for each level. This can be seen easily by substituting the second level in the first, which gives us a non-hierarchical form for Eq. (1):

$$y = \mathbf{X}_0\beta_0 + \mathbf{X}^{(1)}\mathbf{X}^{(2)}\beta^{(2)} + \eta^{(1)}$$

$$\eta^{(1)} = \mathbf{X}^{(1)}\varepsilon^{(2)} + \varepsilon^{(1)}$$

$$\mathrm{Cov}\left(\eta^{(1)}\right) = \mathbf{X}^{(1)}\mathbf{C}^{(2)}\mathbf{X}^{(1)T} + \mathbf{C}^{(1)} \tag{3}$$

---

* Corresponding author. The Wellcome Department of Imaging Neuroscience, Institute of Neurology, 12 Queen Square, WC1N 3BG, London, UK. Fax: +44 20 7813 1445.
E-mail address: k.stephan@fil.ion.ucl.ac.uk (K.E. Stephan).

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

Thus, the mixed-effects error has two covariance components $\mathbf{X}^{(1)}\mathbf{C}^{(2)}\mathbf{X}^{(1)T}$ and $\mathbf{C}^{(1)}$.[1]

Fixed-effects analyses ignore the second line of Eq. (1) and proceed by estimating the first-level parameters and variance parameters corresponding to the covariance component $\mathbf{C}^{(1)}$. However, if both levels of the model are included, it is necessary to estimate both first- and second-level covariance components. This is the case both for Bayesian and classical (i.e., frequentist) approaches (see Harville, 1977). For example, when one wants to use an empirical Bayesian scheme to make conditional inferences about responses at the first level, $\mathbf{C}^{(2)}$ plays the role of an empirical prior on $\beta^{(1)}$ (Kass and Steffey, 1989). In a classical setting, when one wants to make inferences at the second level using a mixed-effects analysis, it is also necessary to estimate both components. The estimation of multiple covariance components presents a slight problem for neuroimaging because it is usually iterative (see Discussion). Generally speaking, iterative schemes at each and every voxel are, computationally, very expensive. Fortunately, there are special cases of Eq. (1) for which non-iterative covariance component estimation suffices.

The main aim of this paper is to introduce a simple approach to full mixed-effects models in statistical parametric mapping that can be used as a reference for the robustness of the conventional two-stage procedure when its underlying assumptions are violated. In what follows, we will describe briefly the estimation of parameters and hyperparameters of hierarchical models and consider how some special, but common, cases lead to the two-stage procedure (proposed in Holmes and Friston, 1998 and also known as the "summary statistics" procedure; see Penny and Holmes, 2004). We will then describe a simple mixed-effects analysis for neuroimaging data. The computational burden incurred by estimating voxel-wise multiple covariance components is finessed by assuming that the relative amounts of within- and between-session random effects is fixed but the expression of this mixture varies in degrees from voxel to voxel. In the final section, we will apply the two-stage procedure and the full mixed-effects analysis to the same multisession data to verify that the two-stage procedure is reasonably robust to violations of the assumptions upon which it rests.

Theory

In this section, we describe the estimation of parameters and hyperparameters[2] controlling fixed and random effects in a two-level hierarchical model. We then describe the conditions under which a conventional two-stage procedure can be used without iterative hyperparameter estimation.

Non-sphericity at the second level

We start by constructing a model of second-level effects that is expressed in terms of first-level estimators and properly incorporates the covariance components from both levels. This model obtains by multiplying the non-hierarchical form of the second-level model (see Eq. (3)) with the generalized inverse of the first-level design matrix (i.e., $\mathbf{X}^{(1)-}$):

$$\hat{\beta}^{(1)} = \mathbf{X}^{(1)-}y$$

$$= \mathbf{X}^{(2)}\beta^{(2)} + \varepsilon^{(2)} + \mathbf{X}^{(1)-}\varepsilon^{(1)}$$

$$= \mathbf{X}^{(2)}\beta^{(2)} + \eta^{(2)}$$

$$\eta^{(2)} = \varepsilon^{(2)} + \mathbf{X}^{(1)-}\varepsilon^{(1)}$$

$$\mathrm{Cov}\left(\eta^{(2)}\right) = \mathbf{C}^{(2)} + \mathbf{X}^{(1)-}\mathbf{C}^{(1)}\mathbf{X}^{(1)-T} \tag{4}$$

The errors of this model comprise two components that are the two components of Eq. (3) projected to the second level by $\mathbf{X}^{(1)-}$. The first component $\mathbf{C}^{(2)}$ is the intrinsic variability of the second-level parameters. The second component $\mathbf{X}^{(1)-}\mathbf{C}^{(1)}\mathbf{X}^{(1)-T}$ results from observation error being projected, through the inverse of the first-level design matrix, to the second level. Generally, this component will induce non-sphericity[3] at the second level because $\mathbf{X}^{(1)-}\mathbf{C}^{(1)}\mathbf{X}^{(1)-T}$ does not necessarily conform to the identity matrix.

The model in Eq. (4) enables a simple implementation of mixed-effects analyses using classical inference procedures. These procedures rest on forming t or F statistics with appropriate adjustments for non-sphericity using the Satterthwaite approximation (e.g., Friston et al., 2002b; Geisser and Greenhouse, 1958; Worsley and Friston, 1995). However, these procedures require the non-sphericity to be parameterised in terms of a single component V, such that $\mathrm{Cov}\{\eta^{(2)}\} = \sigma^2\mathbf{V}$. This requires the relative contribution of the covariance components to be estimated so that the non-sphericity of the errors η^2 can be properly specified.[4]

$$\mathbf{V} \propto \mathrm{Cov}\left\{\eta^{(2)}\right\} = \mathbf{C}^{(2)} + \mathbf{X}^{(1)-}\mathbf{C}^{(1)}\mathbf{X}^{(1)-T} \tag{5}$$

Given $\mathbf{V}$ and a weighting matrix $\mathbf{W}$, we can take weighted least squares estimators of the second-level parameters $\beta^{(2)}$ and proceed in the usual way. For example,

---

[1] Note that Eq. (3) requires no particular form for either $\mathbf{X}^{(1)}$ or $\mathbf{X}^{(2)}$ except that $\mathbf{X}^{(1)}$ has as many columns as $\mathbf{X}^{(2)}$ has rows. This is assured because the two levels in the model have a direct hierarchical relationship (see Eq. (1)).

[2] Hyperparameters are parameters of the distribution of lower level parameters. Strictly speaking, only the second-level variance parameter is a hyperparameter because this controls the prior covariance of $\beta^{(1)}$. In order to maintain a simple nomenclature, both first- and second-level variance parameters are referred to as hyperparameters in this paper.

[3] Non-sphericity denotes the violation of the assumption that error terms are identically and independently distributed (i.i.d). In the case of non-sphericity, the error covariance matrix $\mathbf{C}$ deviates from a scalar multiple of the identity matrix. Non-sphericity can arise from heterogeneous variances over different levels of a factor (identity assumption violated: the diagonal of $\mathbf{C}$ has different values) or by dependence among the levels (independence assumption violated: $\mathbf{C}$ has off-diagonal values), as in repeated measures designs (Keselman et al., 2001).

[4] Using the ReML estimates in this way ignores the relative uncertainty and correlations among these estimates. Although heuristics exist for computing the adjusted degrees of freedom for multiple covariance components, these schemes are not used widely. See Kiebel et al. (2003). This consideration applies only to empirical Bayes (as opposed to full Bayes).

$$\hat{\beta}^{(2)} = \mathbf{M}\hat{\beta}^{(1)}$$

$$\mathbf{M} = \left(\left(\mathbf{WX}^{(2)}\right)^T \mathbf{WX}^{(2)}\right)^{-1}\left(\mathbf{WX}^{(2)}\right)^T\mathbf{W}$$

$$T = \frac{c^T\hat{\beta}^{(2)}}{\sqrt{c^T\sigma^2 MVM^T c}}$$

$$v = \frac{tr\{R\mathbf{W}V\mathbf{W}\}^2}{tr\{R\mathbf{W}V\mathbf{W}R\mathbf{W}V\mathbf{W}\}}$$

$$R = I - \mathbf{X}^{(2)}M$$

$$\sigma^2 = \frac{\hat{\beta}^{(1)T}R^T\mathbf{W}^T\mathbf{W}R\hat{\beta}^{(1)}}{tr\{R\mathbf{W}V\mathbf{W}\}} \tag{6}$$

Here the *t* statistic is distributed with v degrees of freedom under the null hypothesis that the contrast $c^T\hat{\beta}^{(2)} = 0$. The weighting matrix $\mathbf{W}$ can be chosen to implement ordinary least squares $\mathbf{W} = 1$, in which case the effective degrees of freedom $v \leq$ trace $\{R\}$. Alternatively, one can make $\mathbf{W} = \mathbf{V}^{-1/2}$, in which case the estimators are maximum likelihood (ML) and $v =$ trace $\{R\}$. Notice that the single hyperparameter σ^2 has an analytic non-iterative solution, which is the usual sum of squared residuals estimator. The only outstanding problem is to estimate the covariance components so that $\mathbf{V}$ can be specified according to Eq. (6).

ReML estimation of covariance components

The error covariances of the first and second level can be expressed by covariance components Q that are scaled by hyperparameters λ:

$$\mathbf{C}^{(1)} = \lambda^{(1)}\mathbf{Q}^{(1)}$$

$$\mathbf{C}^{(2)} = \lambda^{(2)}\mathbf{Q}^{(2)} \tag{7}$$

To simplify the exposition we have assumed each level has only one covariance component.[5] It is relatively simple to identify the maximum likelihood estimates of the hyperparameters $\lambda^{(1)}$ and $\lambda^{(2)}$, given the response variable y and design matrix $[\mathbf{X}_0\mathbf{X}^{(1)}\mathbf{X}^{(2)}]$ from the model in Eq. (3), where the covariance components are $\mathbf{X}^{(1)}\mathbf{Q}^{(2)}\mathbf{X}^{(1)T}$ and $\mathbf{Q}^{(1)}$. Hyperparameter estimation is restricted to the null space of the design matrix to ensure they are not biased (see Friston et al., 2002a). This restriction leads to restricted maximum likelihood (ReML) estimates of the covariance components. ReML is a special case of expectation maximization (EM; Dempster et al., 1977), and in the context of linear hierarchical observation models under Gaussian assumptions the objective functions of ReML and EM are identical (see Appendix A and Friston et al., 2002a for details). Note that ReML estimation of the covariance components induced by first- and second-level error proceeds in observation space (using the non-hierarchical form of Eq. (3)).

The two-stage procedure

If we do not know the hyperparameters, then they have to be estimated using ReML. As mentioned above, this would usually be done at the first level using the model in Eq. (3) (see Fig. 1 and Appendix A). However, there is one situation in which this iterative estimation can be omitted. This is when both components of Cov$\{\eta^{(2)}\}$ are expressed in exactly the same way. This will be the case when:

$$\mathbf{Q}^{(2)} \propto \mathbf{X}^{(1)-}\mathbf{Q}^{(1)}\mathbf{X}^{(1)-T} \tag{8}$$

In other words, when the second- and first-level components have the same form, the relative contributions of first- and second-level variance components are irrelevant at the second level, and one can proceed directly to Eq. (6), by assuming $\mathbf{V} \propto \mathbf{Q}^{(2)}$. This procedure corresponds to taking parameter estimates from the first level and treating them as a new response variable in a second-level analysis. This is the basis of the two-stage procedure that has an established role in general statistics (Laird and Ware, 1982) and fMRI data analysis in particular (Holmes and Friston, 1998). It rests on the assumption that the (projected) first- and second-level components have the same form. Under i.i.d.[6] assumptions at the second level, there are three conditions that are sufficient to ensure this is true. First, within-session error covariance is the same for each session. Second, the first-level design partitions $\mathbf{X}_i^{(1)}$ are the same for all sessions; and third, these partitions comprise one explanatory variable.[7] In this case, the contribution of the within-session variability to the second level is exactly the same from session to session. This contribution is indistinguishable from the endogenous second-level variability, and the mixture can be estimated non-iteratively using Eq. (6).

In summary, under the above conditions, between-level non-sphericity, that is, non-sphericity at the second level that is induced by first-level variance components ($\mathbf{X}^{(1)-}\mathbf{C}^{(1)}\mathbf{X}^{(1)-T}$), can be ignored; this is the basis for the computationally efficient two-stage procedure. However, it should be noted that the two-stage procedure does allow one to model within-level non-sphericity, that is, non-sphericity separately at the first level (e.g., serial correlations) and the second level (e.g., correlations over repeated measures); see Empirical demonstration of this paper and Friston et al. (2002b).

The need for full mixed effects analyses

The two-stage procedure is now common practice in the context of balanced designs, where equality of variance is a tenable assumption. However, in some situations, it is not possible to use exactly the same first-level session-wise design partitions (see Eq. (2)). For example, in the post hoc classification of events, as "remembered" or "forgotten", or in the study of EEG-correlated fMRI (e.g., seizure or alpha activity), the explanatory variables comprising the partitions $\mathbf{X}_i^{(1)}$ may not be under complete experimental control. A natural and frequently asked question is "how robust is the two-stage procedure to departures from balanced designs?" The formulation of the mixed-effects analyses in this section shows that departures will induce non-sphericity at the second level. This non-sphericity will take the form of unequal

---

[5] Generally, there may be more than one $Q_i^{(1)}$ and $Q_i^{(2)}$ that allow non-sphericity to be modeled at the first level (e.g., serial correlations) or the second level (e.g., correlations over repeated measures) (see Friston et al., 2002a,b). In the empirical example below we used several components at both levels.

[6] Independently and identically distributed.

[7] More generally a one-dimensional contrast of first-level parameter estimates is taken to the second level.

ReML MFX procedure

$$Y = data$$

$$X = [X^{(0)} \quad X^{(1)}] \qquad X = [X^{(0)} \quad X^{(1)} X^{(2)}]$$
$$V = I \qquad\qquad Q = \{Q_1^{(1)}, ..., X^{(1)} Q_1^{(1)} X^{(1)T}, ...\}$$

Step 1

$$\boxed{\hat{\beta}^{(1)} = (X^T V^{-1} X)^{-1} X^T V^{-1} Y}$$

$$\boxed{\lambda = REML\{YY^T/n, X, Q\}}$$

$$Y = \hat{\beta}^{(1)}$$
$$X = X^{(2)}$$
$$V = \sum_i \lambda_i^{(1)} X^{(1)} Q_i^{(1)} X^{(1)-T} + \sum_j \lambda_j^{(2)} Q_j^{(2)}$$

Step 2

$$\boxed{\hat{\beta}^{(2)} = (X^T V^{-1} X)^{-1} X^T V^{-1} Y}$$
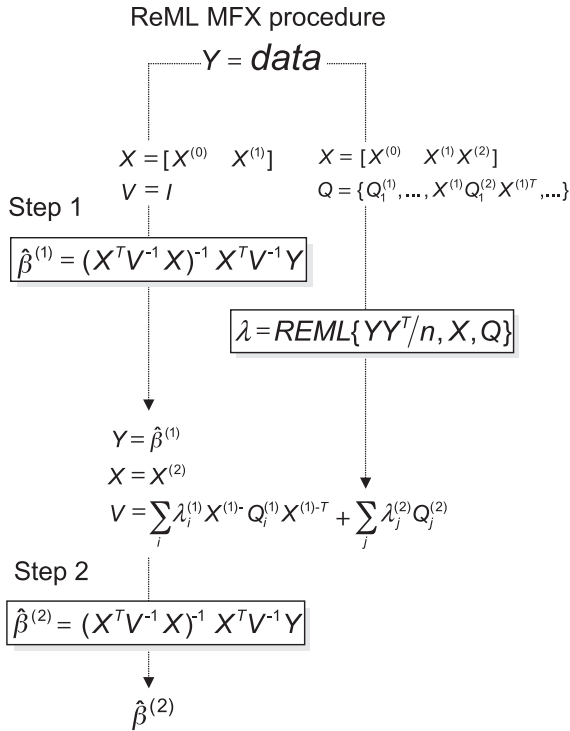
$$\hat{\beta}^{(2)}$$

Fig. 1. Schematic showing the architecture of the mixed-effects procedure in which ReML estimates of non-sphericity derived from a first-level model enter a second level model as a known quantity *V*.

variance, that is, heteroscedasticity. The question about unbalanced designs then reduces to the question "how robust is parametric inference to violations of homoscedasticity?" The general answer is that parametric statistics are notoriously robust to this particular form of violation. However, considerations of robustness are only meaningful when the exact violation can be specified. Consequently, it would be useful to have a simple and computationally tractable mixed-effects analysis to provide a reference for the two-stage procedure when its assumptions are violated.

Mixed-effects analyses in SPM2

In the next section we illustrate the approach that has been implemented in the current version of the SPM2 software. This procedure has been written to enable comparisons between full mixed-effects analyses and multistage approximations. Clearly, there are a variety of ways in which a full mixed-effects analysis could be implemented. We use a simple approach that is formally similar to the two-stage procedure but allows for ReML estimates and multiple variance components per level. Fig. 1 illustrates this scheme. In brief, second-level parameters are estimated using first-level parameter estimates in exactly the same way as in the two-stage procedure. The key difference is that the non-sphericity induced by unbalanced designs or non-sphericity at the first level enters a ReML estimate of *V* explicitly. This estimation removes any dependence on assumptions of sphericity at the second level. To enable efficient computation of the ReML estimates, they are pooled over "responsive" voxels. In our implementation, a responsive voxel is defined as surviving an *F* test for any effect of interest at an uncorrected threshold of *P* = 0.001. Operationally, as described

in Friston et al. (2002a), this involves replacing the outer product yy^T in the ReML algorithm (see Appendix A) with the equivalent sample covariance matrix over *n* voxels $\mathbf{YY}^T/n$. This means that the ReML estimate of non-sphericity can proceed given just the sample covariance matrix of the data. This is accumulated during estimation of the first-level parameters.

The key aspect of this procedure that distinguishes it from classical approaches for univariate data is the pooling over voxels. This is very similar to the device used by multistat (Worsley et al., 2002) and rests upon the assumption that the relative contribution of different error covariance components is the same in subsets of voxels. However, the absolute amount of this mixture is voxel specific and is estimated in the usual mass univariate fashion. Note that this device is peculiar to neuroimaging, where one has the opportunity to treat subsets of voxels as having the same covariance structure. The special benefit, in neuroimaging, is that one gets very precise hyperparameter estimates, allowing the non-sphericity to be treated as known. This is nice because the Satterthwaite approximation in Eq. (6) assumes *V* is known. In the next section we apply the two-stage procedure and this mixed effects analysis to the same data.

Empirical demonstration

The data set we chose to analyze comprised 1200 images that were acquired from the same volunteer in 10 contiguous sessions of 120 scans each. These data have been described elsewhere (Friston et al., 1998). The reason we chose these data was that each of the 10 sessions was different in terms of design. The experimental design involved 30-s epochs of single word streams and a passive listening task. The words were concrete, monosyllabic, and were presented at a number of different rates. Within each of the 10 sessions three epochs were presented at different rates. The word rate for each epoch was varied pseudorandomly over the 10 sessions. We modeled responses using an event-related model where the occurrence of each word was modeled with a delta function. The ensuing stimulus function was convolved with a canonical hemodynamic response function and its temporal derivative to give two regressors of interest for each of the 10 sessions. Examples of these regressors, for the first two sessions, are provided in Fig. 2 to illustrate that the number of words presented in each series varied considerably (word presentation rate was varied from 10 to 90 words/min). These effects were supplemented with confounding and nuisance effects. These comprised a mean and the first few components of a discrete cosine transform, removing drifts lower than 1/128 Hz.

The data were analyzed with a first-level model where $\mathbf{X}^{(1)}$ comprised a session-specific partition for each block of 120 scans. The design matrices are shown in Fig. 3 (the cosine terms have been omitted for clarity). We used 20 covariance components $\mathbf{Q}_i^{(1)}$ for the first level. These modeled session-specific variances and serial correlations using an AR(1) model (for the precise definition, see Eq. (2) in Friston et al., 2002b). The corresponding hyperparameter estimates are shown in Fig. 4. The variance components show fairly typical values ranging from about 0.8 to 1.7. Note that session-specific variance can vary by over a factor of two. The serial correlations, as indexed by their AR coefficients, were largely trivial, apart from session 8 with pronounced positive correlations (middle panel). This session also had the biggest error variance.

Time domain
regressors for session 1

Frequency domain
128 second High-pass filter

Time domain
regressors for session 2
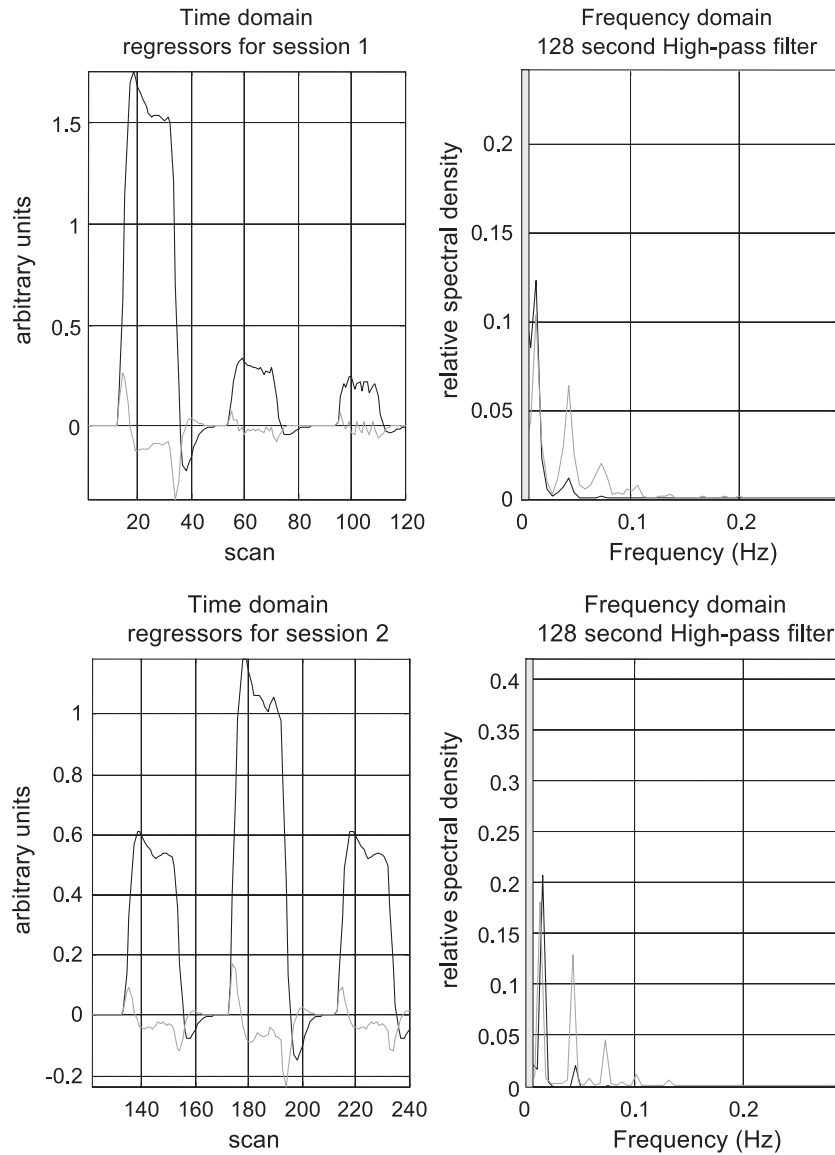
Frequency domain
128 second High-pass filter

Fig. 2. This is an illustration of the two regressors for two sessions that entered the first-level design matrix. These regressors show the variation in word presentation rates over the three epochs that comprised each session. The regressors were formed by convolving a stick function, encoding the occurrence of words, with a canonical hemodynamic response function and its temporal derivative. The left panel shows the regressors in the time domain and the right panel in the frequency domain (the grey bar represents the cut-off of the high-pass filter).

To demonstrate the flexibility of the analysis, we took both parameter estimates to the second level, enabling inferences about differences in either the canonical response form or its temporal derivative. This requires separate covariance components at the second level: for between-session variance in the canonical and the derivative parameters ($Q_1^{(2)}$ and $Q_2^{(2)}$) and for the covariance between them ($Q_3^{(2)}$)

$$\mathbf{Q}_1^{(2)} = I_S \otimes \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},$$

$$\mathbf{Q}_2^{(2)} = I_S \otimes \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \tag{9}$$

$$\mathbf{Q}_3^{(2)} = I_S \otimes \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

where S denotes the number of sessions, see Eq. (2), $\mathbf{I}_S$ represents the $S \times S$ identity matrix, and $\otimes$ represents the Kronecker product. The associated hyperparameters are shown in the lower panel of Fig. 4. The second-level design matrix implemented a simple one-way ANOVA with two levels (see Fig. 3) and allowed us to use an SPM{F} to test for any response in terms of the canonical response or its derivative.

Up to this point, the two analyses using the two-stage procedure and the full mixed-effects approach, respectively, do not differ. As explained above, the crucial difference lies in the assumptions about the non-sphericity of the second-level error that is induced by first-level error covariance components. The conventional two-stage procedure only models within-level non-sphericity. Thus, when entering the first-level parameter estimates (two for each session) into a second-level

1st and 2nd level design matrices

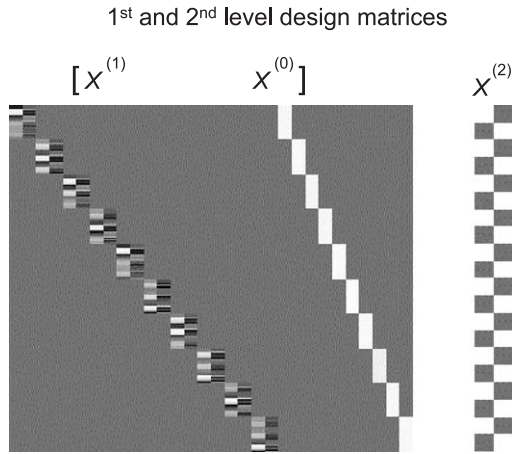$[X^{(1)} \qquad X^{(0)}] \qquad\qquad X^{(2)}$

Fig. 3. First- and second-level design matrices used in both exemplar analyses.

analysis, the ReML estimate for the second-level error covariance used

$$\mathbf{V} = \sum_j \lambda_j^{(2)} \mathbf{Q}_j^{(2)} \tag{10}$$

with $Q_j^{(2)}$ as specified by Eq. (9). In contrast, the full mixed-effects approach allows for the impact of first-level non-sphericity on parameter estimates at the second level. Here the ReML estimate used:

$$\mathbf{V} = \sum_j \lambda_j^{(2)} \mathbf{Q}_j^{(2)} + \sum_i \lambda_i^{(1)} \mathbf{X}^{(1)-} \mathbf{Q}_i^{(1)} \mathbf{X}^{(1)-T} \tag{11}$$

This is a generalization of Eq. (7) that includes i covariance components from the first level and j from the second. The results of these two analyses are presented in Fig. 5 and have been thresholded at $P = 0.05$ (corrected for the entire search volume). Details about the maxima and associated P values are provided in Table 1. It is evident that the inferences obtained from these two procedures are almost identical, with the MFX analysis being slightly more sensitive (which is compatible with analyses based on simulated data, see Beckmann et al., 2003). The results remain relatively unchanged despite the fact that we allowed for heteroscedasticity at the first level and the first-level designs were not balanced. Both factors contribute to non-sphericity at the second level. This non-sphericity is illustrated in Fig. 6 for the two-stage and MFX analyses (upper and lower panels, respectively). This figure shows the heteroscedasticity can vary by up to a factor of 4. In this figure, we have only shown non-sphericity for the first parameter estimate (corresponding to the session-specific canonical HRF). In case of the two-stage procedure (upper panel of Fig. 6), this is simply proportional to $\mathbf{I}_s$.

It should be noted that the first-level estimation and both second-level analyses allowed for within-level non-sphericity. As noted above, we modeled session-specific error variances and serial correlations using an AR(1) model at the first level. At the second level, we allowed for different variances in parameter estimates pertaining to the canonical response function, its time derivative, and for covariation between these estimators (see Eq. (9)). For the conventional two-stage analysis, this within-level non-sphericity was estimated using exactly the same pooling device and ReML estimation scheme as described for the mixed-effects analysis. The only difference was that the MFX estimate of

non-sphericity included extra covariance components $\mathbf{X}^{(1)} \mathbf{Q}_i^{(1)} \mathbf{X}^{(1)-T}$ induced by the first level (see Eq. (11)).

Discussion

In conclusion, we have described a simple procedure for implementing full mixed-effects analyses in SPM. This procedure uses restricted maximum likelihood estimates of different variance components induced by the hierarchical structure of observation models in multisession fMRI studies. These ReML estimates are based upon the sample covariance matrix of the response variable in, and only in, responsive voxels. The reason that this procedure was developed was to provide a reference for the commonly used two-stage procedure, when its assumptions are violated.
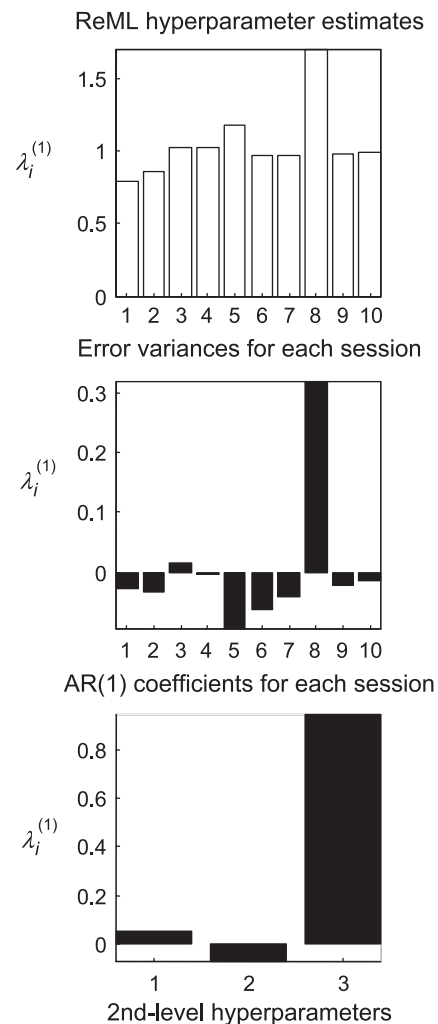
ReML hyperparameter estimates

Fig. 4. ReML estimates of the hyperparameters for within- and between-session error covariance components at the first level. The first 10 hyperparameters correspond to the error variance for each session at the first level and are shown in the upper panel. The middle panel shows hyperparameters controlling serial correlations (the first AR coefficient). The lower panel shows three second-level hyperparameters reflecting the components induced in the response variable by: (i) variation around the grand mean of the first parameter (the canonical HRF), (ii) variation in the second parameter (the temporal derivative), and (iii) covariation between the two parameters.
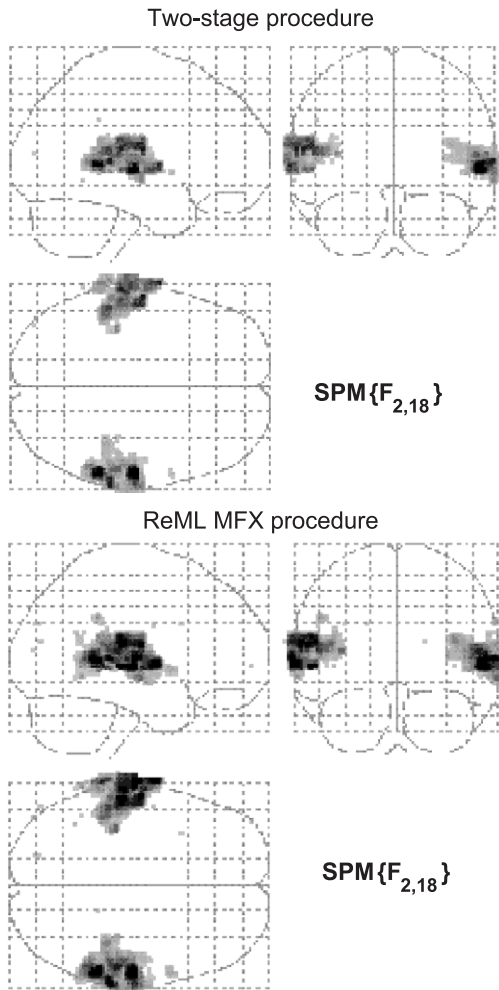
Two-stage procedure

$SPM\{F_{2,18}\}$

ReML MFX procedure

$SPM\{F_{2,18}\}$

Fig. 5. SPMs of the F contrast I_2 testing for the effect of words at the second level. This corresponds to a one-way ANOVA of the first-level parameter estimates. The upper panel shows the SPM for a conventional two-stage procedure. The lower panel shows the equivalent results for a full mixed-effects analysis. The only difference between these two analyses was in the form of the non-sphericity assumed for the residuals at the second level. See main text for details.

Why is iterative covariance component estimation necessary?

A frequent and natural question is about the need for iterative estimation of covariance components. The reason that, in most cases, one cannot solve directly for ReML estimates of covariance components is simple: these estimators are based on the variance of the residuals averaged over [integrating out] the conditional distribution of the parameters. However, the conditional distribution of the parameters depends on the ReML estimates. This circular dependence usually precludes a closed-form expression for the ReML estimates; in some special cases, of balanced data in mixed models, analytical solutions exist for the covariance components (see Laird et al., 1987; Searle et al., 1992; Szatrowski and Miller, 1980 for a formal treatment of this issue). However, iterative schemes like the EM algorithm, are required in most cases. In these schemes, the conditional expectation of the parameters is estimated in an E-Step using the current non-sphericity estimate, and then the ReML hyperparameters are estimated in an M-Step using the current parameter estimates

(see Dempster et al., 1977 and Appendix A in Friston et al., 2002a). Although parameter estimation can proceed non-iteratively at each hierarchical level, given the hyperparameters (and vice versa), one cannot estimate both the parameters and the hyperparameters without considering all components and all levels of a hierarchical model.

Why does the MFX analysis require two steps?

The implementation of the MFX analyses above includes two steps: the first step computes first-level parameters and ReML hyperparameters that are used by the second step to estimate second-level parameters (Fig. 1). Why not use the covariance component ReML estimates to assess contrasts of the first-level estimates directly? In other words, why bother re-estimating a further hyperparameter σ^2 at the second level? There are two reasons for this. First, recall that the ReML estimates are based on thousands of voxels and therefore correspond to expectations over voxels. They are not covariance component estimators from the point of view of any voxel. In fact, they are extremely precise estimates of non-sphericity, which is why they can be treated as known quantities in our application. This renders the neuroimaging application of MFX analyses more accurate than equivalent univariate procedures. Secondly, even if we used a ReML scheme at each and every voxel, classical statistics (e.g., the Satterthwaite procedure as used in Worsley and Friston, 1995) cannot accommodate joint uncertainty over multiple covariance component estimates. This is quite a complicated issue that is dealt with, in the context of fMRI data analysis, in a previously published note (Kiebel et al., 2003). In brief, the precision (degrees of freedom) of the standard error of any contrast depends on the form of the contrast and on the assumptions about the form of the likelihood $p(y|\lambda)$. Alternatively, techniques like Markov Chain Monte Carlo (MCMC) methods could be used, but these are computationally very expensive. The above implementation of MFX analyses represents a simple approximation that eschews these complications and allows one to use classical expressions for F and t statistics and their degrees of freedom.

Conclusion

The empirical example presented above suggests that the conventional two-stage procedure is robust to differences among the first-level design partitions and to unequal error variances at the first level. In our experience, the two-stage procedure with sphericity assumptions about the impact of first-level variance components on second-level parameter estimates (as in Eq. (10)) is a very reasonable approximation in the vast majority of experimental situations. The main contribution of this note is to highlight that the differences between a two-stage procedure and a full mixed-effects analysis reduce to non-sphericity assumptions about the second-level error. Furthermore, a simple procedure now exists for performing a full mixed-effects analysis in situations where one is worried about the robustness of the two-stage approach. This mixed-effects analysis uses completely standard restricted maximum likelihood techniques that were established in the 1970s (Dempster et al., 1977; Harville, 1977). The only difference between these classical ReML approaches and the neuroimaging application described here is that the sample covariance matrix over voxels

Table 1
Maxima and associated statistics for the SPMs presented in Fig. 5

Voxels in cluster	P value	Statistics	(F and Z)	(mm)
Two-stage procedure				
984	0.000	261.41	7.44	58 −18 2
		218.68	7.23	58 −44 4
		156.17	6.83	64 −34 6
1415	0.000	160.21	6.86	−60 −12 14
		154.49	6.82	−64 −26 14
		148.57	6.77	−54 −16 18
6	0.005	50.29	5.35	58 8 −2
1	0.008	47.76	5.28	−50 −42 34
1	0.016	43.29	5.15	−66 −52 −6
1	0.037	38.50	4.98	−40 −84 12
2	0.038	38.45	4.98	−40 −28 8
2	0.039	38.23	4.97	58 −28 0
ReML MFX procedure				
1536	0.000	289.84	7.56	58 −18 2
		247.94	7.38	56 −40 4
		229.08	7.29	64 −34 6
2039	0.000	239.44	7.34	−60 −12 14
		234.74	7.32	−70 −16 8
		224.43	7.26	−70 −2 2
14	0.000	87.62	6.10	−50 −42 34
33	0.000	72.61	5.85	58 8 −2
14	0.003	54.92	5.48	60 −54 22
7	0.004	52.27	5.41	64 −40 30
2	0.007	48.43	5.30	−40 −56 8
2	0.012	44.96	5.20	−52 −4 28
2	0.012	44.89	5.20	−52 −80 18
4	0.013	44.41	5.18	68 −42 16
1	0.016	43.38	5.15	−60 −54 18
1	0.016	43.06	5.14	−42 60 2
2	0.019	42.26	5.11	−20 −80 30
2	0.023	41.09	5.07	58 −8 −4

replaces the outer product of univariate data sequences (see Appendix A). It is hoped that this material will be useful to those researchers who cannot always guarantee fully balanced designs at the first level.

Acknowledgments

Appendix A

ReML estimates of the hyperparameters obtained by a Fisher scoring ascent on the ReML objective function that is exactly the same as the negative free energy in expectation maximization (see Friston et al., 2002a for further details).

Until convergence {

E-Step

$$V = \sum \lambda_i \mathbf{Q}_i$$

$$\mathbf{C} = \left(\mathbf{X}^T V^{-1} \mathbf{X}\right)^{-1}$$

M-Step

$$P = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}C\mathbf{X}^T\mathbf{V}^{-1}$$

$$g_i = -\frac{1}{2}tr\{P\mathbf{Q}_i\} + \frac{1}{2}tr\{P^T\mathbf{Q}_iP\mathbf{Y}\mathbf{Y}^T/n\}$$

$$H_{ij} = \frac{1}{2}tr\{P\mathbf{Q}_iP\mathbf{Q}_j\}$$

$$\lambda \leftarrow \lambda + H^{-1}g$$

}

Here, C is the conditional covariance of the parameters associated with the design matrix **X**, and $\mathbf{Y}\mathbf{Y}^T/n$ is the sample covariance matrix over n voxels (exceeding an uncorrected F threshold of $P < 0.001$ for effects of interest) over which ReML estimates are pooled. In the application considered in the main text **X** =
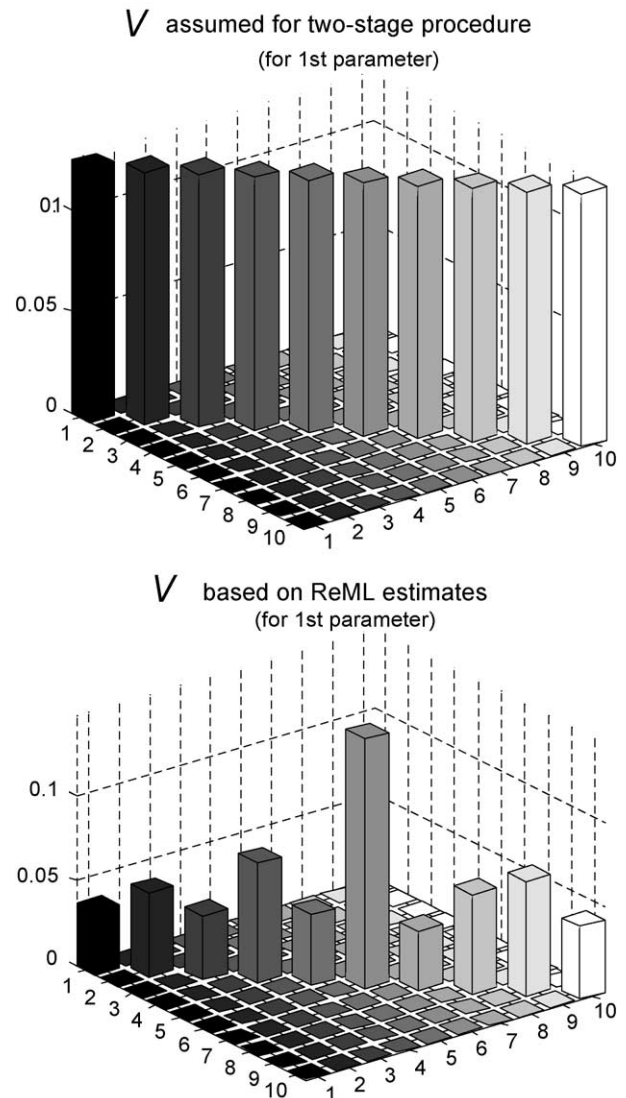

Fig. 6. Non-sphericity (of the session-specific canonical response) assumed for the second-level analyses under (i) sphericity assumptions (conventional two-stage procedure; upper panel) and (ii) embodying components from the first level (full MFX analysis; lower panel).

$[\mathbf{X}_0\mathbf{X}^{(1)}\mathbf{X}^{(2)}]$, and the covariance components are $\mathbf{Q} = \{\mathbf{Q}_1^{(1)}, \ldots, \mathbf{X}^{(1)}\mathbf{Q}_1^{(2)}\mathbf{X}^{(1)T}, \ldots\}$.

References

Beckmann, C.F., Jenkinson, M., Smith, S.M., 2003. General multilevel linear modelling for group analysis in fMRI. NeuroImage 20, 1052–1063.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc., Ser. B Stat. Methodol. 39, 1–38.

Friston, K.J., Josephs, O., Rees, G., Turner, R., 1998. Non-linear event-related responses in fMRI. Magn. Res. Med. 39, 41–52.

Friston, K.J., Penny, W., Phillips, C., Kiebel, S., Hinton, G., Ashburner, J., 2002a. Classical and Bayesian inference in neuroimaging: theory. NeuroImage 16, 465–483.

Friston, K.J., Glaser, D.E., Henson, R.N., Kiebel, S., Phillips, C., Ashburner, J., 2002b. Classical and Bayesian inference in neuroimaging: applications. NeuroImage 16, 484–512.

Geisser, S., Greenhouse, S.W., 1958. An extension of Box's results on the use of the F distribution in multivariate analysis. Ann. Math. Stat. 29, 885–891.

Harville, D.A., 1977. Maximum likelihood approaches to variance component estimation and to related problems. J. Am. Stat. Assoc. 72, 320–338.

Holmes, A.P., Friston, K.J., 1998. Generalisability, random effects and population inference. NeuroImage 7, S754.

Kass, R.E., Steffey, D., 1989. Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models). J. Am. Stat. Assoc. 407, 717–726.

Keselman, H.J., Algina, J., Kowalchuk, R.K., 2001. The analysis of repeated measures designs: a review. Br. J. Math. Stat. Psychol. 54, 1–20.

Kiebel, S.J., Glaser, D.E., Friston, K.J., 2003. A heuristic for the degrees of freedom of statistics based on multiple variance parameters. NeuroImage 20, 591–600.

Laird, N.M., Ware, J.H., 1982. Random-effects models for longitudinal data. Biometrics 38, 963–974.

Laird, N.M., Lange, N., Stram, D., 1987. Maximum likelihood computations with repeated measures: applications of the EM algorithm. J. Am. Stat. Assoc. 82, 97–105.

Penny, W., Holmes, A.P., 2004. Random-effects analysis. Human Brain Function. Elsevier, San Diego, pp. 843–850.

Searle, S.R., Casella, G., McCulloch, C.E., 1992. Variance Components. John Wiley & Sons, New York.

Szatrowski, T.H., Miller, J.J., 1980. Explicit maximum likelihood estimates from balanced data in the mixed model of the analysis of variance. Ann. Stat. 4, 811–819.

Worsley, K.J., Friston, K.J., 1995. Analysis of fMRI time-series revisited—Again. NeuroImage 2, 173–181.

Worsley, K.J., Liao, C., Aston, J., Petre, V., Duncan, G.H., Evans, A.C., 2002. A general statistical analysis for fMRI data. NeuroImage 15, 1–15.