# Comparing groups of subjects in fMRI studies: a review of the GLM approach.

FMRIB Technical Report TR00DL1

**Didier G. Leibovici and Stephen Smith**

Oxford Centre for Functional Magnetic Resonance Imaging of the Brain (FMRIB),
Department of Clinical Neurology, University of Oxford, John Radcliffe Hospital,
Headley Way, Headington, Oxford, UK

**Abstract**

A review of the different approaches using the General Linear Model (GLM) to analyse multi-subject fMRI studies is presented. The first part of the paper attempts to expose the approaches with the least amount of statistic knowledge whilst the second part embeds those approaches in the GLM framework necessitating more statistical mathematical awareness, but enabling more advanced applications. Fixed and Random subject analysis are then re-expressed within the GLM, also taking into account autocorrelation of the measures. This occurs because of the time series aspect of the data, and also if the groups come from repeated sessions (*e.g.* different conditions) ; the spatial autocovariance is usually treated at the end, as the model is voxelwise. General hypothesis testing makes use of unidimensional contrasts, as well as multidimensional contrasts leading to different statistical maps (*e.g.* $t$-maps, $F$-maps).

## 1 Introduction

The aim of this paper is to give an overview of the existing methods based on the General Linear Model (GLM) to assess activity in an fMRI study, resulting in a statistical parametric map. The successive steps of every method are: modelling the response at each voxel by a GLM, then testing a hypothesis (about the parameters of the model) and representing the observed statistic map thresholded at a given level according to the point distribution of the statistic (uncorrected levels) or according to the field distribution of the statistic (corrected levels for local maxima) [1]. The GLM model used can refer to a single subject, one group of subjects or more groups of subjects which can represent different subjects (*e.g.* male, female) or the same subjects (*e.g.* placebo, drug A, drug B in a cross-over design).
Note that the modelling part is univariate, i.e. is separated for each voxel[2], and is in fact usually separated for each subject. The paradigm applied is usually well balanced (*e.g.* the same number of "OFF" and "ON" scans in a block-design). The simplest approach uses only a $t$-test, and Friston et.al [7] embedded this analysis in a General Linear model to be able to take into account covariates in the analysis. Autocorrelation in fMRI time series requires some changes necessitating also the use of GLM [1, 16, 17]. Multi-subject fMRI experiment can also be expressed in a GLM framework with different forms according to the approach taken: fixed or random subject analysis. Estimation can be improved using spatial consideration (smoothing) in order to regain some quality of the estimation (spatially smoothed autocorrelation [3]). In the context of multi-subject analysis a similar spatial consideration is investigated by Worsley [18] to "diminish between-subject variability".
Section 2 redescribes the classical single subject analysis, for an ON and OFF (block-design) experiment in its simplest approach. Then the multi-subjects (section 3) approaches are explained with fixed-effect and random-effect analyses, and alternatives such as "conjunction analysis" [5] and "variance ratio smoothing" [18]. One group and two groups analyses are summarised in section 4. These procedures were related to $t$ distributed statistics; section 5 returns to GLM, to redescribe them briefly under general hypothesis testing. It takes also into account time-series autocorrelation, and investigates the case of more than two groups which involves $F$

[1]The point distribution is the distribution at a given voxel, *i.e.* one random variable, and the field distribution is the continuous version of multivariate distribution *i.e.* a distribution of a vector of random variables.

[2]The only spatial consideration comes when thresholding the statistic map.

[3]Looking for robust estimates of the autocorrelation can be achieved by using a robust estimate for each time series, and by smoothing spatially the autocorrelation obtained.

distributed statistics. The repeated measures aspect of fMRI experiments can also be illustrated in a multivariate model. This point is discussed in the last section of this article introducing also a multivariable multivariate general linear model written to be able to take into account spatial correlation aspects in the model. This last model will be fully described in a later paper.

## 2 Single-Subject Analysis

In a single-subject fMRI study, one collects, at each voxel, a time series of responses (intensities of the BOLD signal) to a stimulus, e.g. an ON (condition B) and OFF (condition A) experiment (box-car design).
Let $y_t$, be the time series observed from $t = 1 \cdots T$, at a given voxel. Among the $T$ values observed at this voxel, $T_A$ of them were recorded while under condition A (OFF or rest condition) and $T_B$ of them were recorded while under condition B (ON or stimulation condition) according to the paradigm.
The observations are assumed to be independent and identically distributed (i.i.d.)[4] and to come from a Normal distribution with the same variance $\sigma^2$ and different means $\mu_A$ and $\mu_B$ for each sample. To decide if there was an activation (at this voxel) during the experiment, one has to compare the means in the two conditions. If the difference of the means is *big enough* relatively to its dispersion, one will assume activation. For that purpose is used the following statistic (where the denominator is an estimate of the standard deviation of the numerator estimating the difference of the means) with its derived distribution (see appendix A) under the null hypothesis (of no activation $\mu_A = \mu_B$):

$$t_o = \frac{\bar{y_B} - \bar{y_A}}{\sqrt{\hat{\sigma}^2(1/T_A + 1/T_B)}} \sim t_{dist}(T_A + T_B - 2) \tag{1}$$

which means that $t_o$ comes from (or follows) a Student distribution ($t$) with $(T_A + T_B - 2)$ degrees of freedom (df). $\bar{y_A}$ and $\bar{y_B}$ are the sample means (*i.e.* $\bar{y_A} = 1/T_A \sum_{t \in A} y_t$) and $\hat{\sigma}^2$ is an estimate of the common variance, which is usually the pooled estimate $\hat{\sigma}^2 = s_p^2 = \frac{(T_A-1)s_A^2 + (T_B-1)s_B^2}{(T_A+T_B-2)}$ with $s_A^2 = 1/(T_A - 1)\sum_{t \in A}(y_t - \bar{y_A})^2$. In fMRI the data is often balanced, i.e. $T_A = T_B = T/2$; thus:

$$t_o = \frac{\bar{y_B} - \bar{y_A}}{2\hat{\sigma}/\sqrt{T}} \sim t_{dist}(T - 2) \tag{2}$$

Now following classical hypothesis testing, if the probability $p(t_{dist}(T_A + T_B - 2) \geq t_o) < \alpha$ for a chosen level $\alpha$ (*e.g.* 0.05 or 0.01 or 0.001) one rejects the null hypothesis and decides that the voxel was activated at the level $\alpha$ (level of decision not of activation)[5].
*Remark:*
If you decide (test) that the variances are unequal, the denominator has to be estimated by $\sqrt{(s_A^2/T_A + s_B^2/T_B)}$ and the degrees of freedom are approximated by the Satterthwaite formula[10]:

$$\frac{[s_A^2/T_A + s_B^2/T_B]^2}{(s_A^2/T_A)^2/(T_A - 1) + (s_B^2/T_B)^2/(T_B - 1)} \tag{3}$$

but if the data is balanced and the sample size not too small the departure from equal variances is usually negligible.

## 3 Multi-subject analysis

Because of the separability aspect of the analysis, one easy implementation is to perform a two-stage procedure. At the first stage the analysis above is performed for each of the $n$ subjects: following the previous section (or, for the more general case, see the GLM section), at each voxel an estimation of activation parameter can be made: $\hat{b}_i = \bar{y}_{iB} - \bar{y}_{iA}$ with standard error $\hat{\sigma}_{b_{\epsilon_i}} = \sqrt{\hat{\sigma}^2_{\epsilon_i}(\frac{1}{T_B} + \frac{1}{T_A})}$ $\quad i = 1 \cdots n$.
So for each of the $n$ subjects can be built the statistic map

$$t_o(i) = \frac{\hat{b}_i}{\hat{\sigma}_{b_{\epsilon_i}}}$$

[4] Note that independence for a time series is not realistic and autocorrelation has to be taken into account, see further.
[5] As already indicated in the introduction, this level of decision is said to be uncorrected as based on the point distribution and not using the field distribution (multiple comparison problem).

to test $b_i = 0$. How do we combine these maps to conclude that there is activation in the population from which these subjects come? This is the purpose of multi-subject analysis, *i.e.* the second stage. From the experiment a different conclusion is drawn for the population according to the method chosen: fixed subject effect analysis allows a conclusion limited to the sample studied; a random subject effect allows a population conclusion. Because the latter can be "too conservative" due to large estimated subject-subject variability (partially due to small sample sizes generally studied ), some alternative approaches have also been developed: "conjunction analysis", and "variance ratio smoothing".

## 3.1 Fixed Subject Analysis

Following the assumption that each subject's estimated activation parameters are taken from the distributions:

$$\hat{b_i} \sim N(m_b\,,\,\hat{\sigma}^2_{b_{\epsilon_i}}) \quad i = 1 \cdots n$$

the population mean activation by $m_b$ is estimated $\widehat{m}_b = \bar{\hat{b}}$ with the following distribution $\bar{\hat{b}} \sim N(m_b\,,\,1/n\,\hat{\sigma}^2_{b_\epsilon})$ , $\hat{\sigma}^2_{b_\epsilon} = 1/n \sum_i \hat{\sigma}^2_{b_{\epsilon_i}}$.

Now, testing $m_b = 0$ (null hypothesis) is a one sample $t$-test:

$$t_o(fixed) = \frac{\bar{\hat{b}}}{1/\sqrt{n}\hat{\sigma}_{b_\epsilon}} \tag{4}$$

Note that deciding that the pooled variance is a "better" estimate of single-subject variance, that is, for every $i$, $\hat{\sigma} = \hat{\sigma}_{b_\epsilon}$ then:

$$t_o(fixed) = \frac{1/n \sum_i \hat{b_i}}{1/\sqrt{n}\hat{\sigma}} = 1/\sqrt{n} \sum_i t_o(i) \tag{5}$$

This simplified fixed-effects analysis is therefore virtually identical to the "Group Z" method: $Z_{group} = 1/\sqrt{n} \sum_i Z(i)$.

The main shortcoming of the fixed-effects approach is that it considers the errors of measurement estimated for the subjects as the only source of variation when estimating the population mean. That is to say, only within-subject variation is accounted for. No consideration of between-subject variation is considered; therefore it is valid only for the subjects chosen in this experiment (no sampling variation). Put another way, when looking at variation of an activation, it has to account for variation *during the experiment on a subject* (between measures or within-subject) and the choice of subjects (between subject).

## 3.2 Random Subject Analysis

The inadequacy of the fixed-effects approach is seen in the first assumption:[6]

$$\hat{b_i} \sim N(m_b\,,\,\hat{\sigma}^2_{b_{\epsilon_i}}) \quad i = 1 \cdots n \tag{6}$$

One must consider the estimated activation as what it is, an estimation of the activation for the given subject:

$$\hat{b_i} \sim N(b_i\,,\,\hat{\sigma}^2_{b_{\epsilon_i}}) \quad i = 1 \cdots n \tag{7}$$

and now consider the activation for a given subject as a random observation of the activation for the population: $b_i = m_b + \eta_i$, $b_i$ is random $\sim N(m_b\,,\,\sigma^2_{b_\eta})$.

It is a *two levels* variation (within subject and between subjects): before, one had

$\hat{b_i} = b_i + \epsilon_i$ , now:

$\hat{b_i} = m_b + \eta_i + \epsilon_i$, i.e. two error terms.

So, $\widehat{m}_b = \bar{\hat{b}} \sim N(m_b\,,\,1/n\,\sigma^2_{(b_\eta + b_\epsilon)})$ , $\hat{\sigma}^2_{(b_\eta+b_\epsilon)} = \hat{\sigma}^2_{b_\eta} + \hat{\sigma}^2_{b_\epsilon}$ and testing $m_b = 0$ is a one sample $t$-test:

$$t_o(random) = \frac{\bar{\hat{b}}}{1/\sqrt{n}\hat{\sigma}_{(b_\eta+b_\epsilon)}} \tag{8}$$

[6] It is valid as a conditional distribution (conditionally to the given subject). A fixed factor (ANOVA language) means that all levels of the factor (here the different subjects) encompass the possible levels one can encounter in the population studied; e.g. gender is a fixed factor with two levels.

Note that only the whole variation is needed, as one is not interested in the individual variance components. The whole variation is calculated directly using the set of estimated single-subject activation levels (the $\hat{b}_i$'s) as

$$\hat{\sigma}^2_{(b_\eta+b_\epsilon)} = \sum_i (\hat{b}_i - \bar{\hat{b}}_.)^2/(n-1) \tag{9}$$

which usually will reflect mainly the between-subject variation.
*Remarks:*
Notice here that introducing the random effects for the subject brings a constant autocorrelation between time measures (Compound Symmetry): let $i = 1 \cdots n$ and $j = 1 \cdots T$ identifying respectively the subjects and the time of measurements; the model can be written $y_{ij} = \beta_j + \eta_i + \epsilon_{ij}$ with $E(\eta_i) = 0$, $E(\epsilon_{ij}) = 0$, $var(\epsilon_{ij}) = \sigma^2_w$, and $var(\eta_i) = \sigma^2_s$ then $corr(y_{ij}, y_{i'j'}) = \frac{\sigma^2_s}{(\sigma^2_w+\sigma^2_s)}$ for $i = i'$.

## 3.3 Degrees of freedom and comments

Notice that $t_o(fixed) > t_o(random)$; the difference may be quite large, as the between-subject variation is normally much larger than the within-subject variation (scan-to-scan). Sample size must be considered here, as to be able to give an accurate decision about the population (as with the random-effects analysis) one needs a good estimate of the between-subject variation.
There is also a difference in the degrees of freedom:
$df(fixed) = n(T_A + T_B - 2)$, and $df(random) = n - 1$. As a rule, with more general models, the degrees of freedom for a fixed-effect approach will be the sum of the degrees of freedom from every single-subject analysis and the degrees of freedom for a random-effect approach will be the degrees of freedom of the second stage model (explained further in the GLM section). Notice the very low $df(random)$ compared with the fixed-effects approach (a typical $n$ is 10, and $T > 100$).
The random-effects approach is the valid one but is difficult to be confident in (power analysis) as the sample size used in general is very small compared to what is usually needed in estimation (say at least 30). For fMRI studies it has been suggested that 12 to 15 subjects would suffice. From a large sample ($n = 40$), Darrell et. al (1998)[4] presented a power analysis resulting in sufficient power observed for about 20 subjects (depending on the brain area under consideration). A possible alternative way of obtaining more degrees of freedom (and so a better estimation) would be to estimate the random variance needed using more data that is on the whole brain, for example by pooling or smoothing the variances. This is similar to the idea used by Worsley in the "variance ratio method" described in the next section.

## 3.4 Alternatives to the Random or Fixed choice

### 3.4.1 Conjunction analysis

Facing this dilemma of an invalid population analysis (fixed-effect approach) and a valid one (random-effect) but difficult to apply usually because of overestimation of variances due to small samples (random-effects), some alternatives have been investigated. One is the "conjunction analysis" (Friston et al. [5]), which uses all the subjects' maps to localise where all the subjects activated (at a chosen level; see single subject analysis, *i.e.* it is a thresholded map of the *min* map over the subjects). Then using simple probability theory, one can relate the "level of activation" (p-value calculated for example from Gaussian Random Field Theory - see Worsley(1999 submitted)) and the proportion of the population which shows this activation at the previously defined level (in the single-subject analysis). Let $(t)$ and $(a)$ mean respectively, *t*ested status of activation with the experiment and true status of *a*ctivation, while $+$ and $-$ is the status activated or not, simple probability calculus gives:

$$\begin{aligned} \alpha_c = P(all\ t_+) &= [P(t_+)]^n \\ &= [P(t_+/a_-)P(a_-) + P(t_+/a_+)P(a_+)]^n \\ &= [\alpha(1-\gamma) + \beta\gamma]^n \end{aligned} \tag{10}$$

Where $\alpha$ is the chosen single-subject level of activation (p-value level to threshold each map), $\beta$ is the power or sensitivity ($1 - \alpha$ is the specificity) of the experiment which is not known and can be set at 1 to provide a lower bound of the proportion $\gamma$ of the population showing the effect. Setting $\beta = 1$ gives

$$\gamma \geq \gamma_1 = \frac{\alpha_c^{1/n} - \alpha}{1 - \alpha}$$

Thus the conclusion about the population is qualitative, *e.g.* with a certainty of 0.95 $(1-\alpha_c)$, we can say that at least 80% $(\gamma_1)$ of the population would activate at level 0.001 $(\alpha)$. The results given above can also be used to decide on a sample size[6], *e.g.* for the above conclusion one would need at least 14 subjects.
*Remarks:*
The $\alpha_c$ could be calculated using permutation testing procedure instead of using random field theory. The conjunction could also be defined as a given proportion of subjects; this could cope better with problems such as poor localisation due to registration problems, as conjunction analysis is certainly very sensitive to subject outliers in terms of the locations of the activations. If one decides that $m \leq n$ must activate to define a conjunction then given $m$:

$$\begin{aligned} \alpha_{cm} &= P(m\ t_+) \\ &= \binom{n}{m} P(t_+)^m (1-P(t_+))^{n-m} \\ &= \alpha_c \binom{n}{m} \left(\frac{1-P(t_+)}{P(t_+)}\right)^{(n-m)} \end{aligned} \tag{11}$$

with $P(t_+) = [\alpha(1-\gamma)+\beta\gamma]$ as given above. One must notice that the amount of conjunction $m/n$ must be at least the expected $\gamma$, otherwise the multiplicative function introduced in the above equation is not monotonic with $m$. This problem is also linked with values of $\gamma$ and the sample size $n$. Roughly speaking when $\gamma$ is close to 1 one would need an $m$ very close to $n$ and so a large $n$ to achieve some gain in performing a conjunction of only $m$.

### 3.4.2 Variance ratio smoothing

As mentioned in the previous section, smoothing the observed variance over the whole brain would produce a better estimate of the random variance, but that assumes constant underlying variance over the brain (spatial stationarity), which does not seem to be the case (*e.g.* difference in white matter and grey matter[16]). But what Worsley et al.(2000)[18] supposes is that the ratio of random-effects and fixed-effects variance is locally constant, so that smoothing the ratio would produce a pooled estimate.
The method consists of performing random and fixed analysis in the first place, then of spatially smoothing the ratio of variances obtained with the two methods (random/fixed), then returning to "random-effects" variance by multiplying the smoothed ratio with the fixed-effects variance before performing the group $t$ test. They estimate the degrees of freedom for the test as:

$$df(Wratio) = 1/[1/df_{ratio} + 1/df(fixed)] = df_{ratio}/(1 + df_{ratio}/df(fixed))$$

where $df_{ratio} = df(random)[2(\frac{FWHM_s}{FWHM_{data}})^2 + 1]^{3/2}$ and $FWHM_s$ is the Gaussian smoothing parameter which enables one to move between a random analysis if set to 0 (no smoothing) and a fixed analysis if set to $\infty$ (smoothing the variance ratio to one everywhere). Sensible choices for $FWHM_s$ would be not to increase too much the degrees of freedom comparatively to its no smoothing situation ($df(random)$), an obvious limit being the degrees of freedom of a single subject experiment *i.e.* $df(fixed)/n$. The value recommended in [18] is 15mm for an original smoothness of 6mm ($FWHM_{data}$) which was actually reaching the upper bound for the experiment given in example. As the final degrees of freedom obtained are *estimated* Worsley et al. recommend to be high (they chose 100) for the result not to be dependent on the estimation of it.

## 4 One or two Groups of Subjects

A one-group analysis is a multi-subjects analysis with the obvious restriction that every subject of the random sample studied is a member of this group.
Two-group analysis is carried out with a two-sample $t$-test under the same framework:

$$\widehat{b1}_i \sim N(b1_i\,,\,\hat{\sigma}^2_{b1_{\epsilon_i}}) \quad i = 1 \cdots n1 \quad \widehat{b2}_i \sim N(b2_i\,,\,\hat{\sigma}^2_{b2_{\epsilon_i}}) \quad i = 1 \cdots n2$$

All the first level analysis has been done for every subject in each group. Note that this looks very similar to the original simple single-subject analysis, but here the sample size of groups may be quite different (a situation to avoid if possible). So one will test $\theta = m_{b1} - m_{b2} = 0$ with the statistics:[7]

$$t_o(fixed) = \frac{\overline{\widehat{b1}} - \overline{\widehat{b2}}}{\sqrt{(\widehat{\sigma 1}^2_{b_\epsilon}/n_1 + \widehat{\sigma 2}^2_{b_\epsilon}/n_2)}} \approx t_{dist}((n_1 + n_2)(T_A + T_B - 2)) \tag{12}$$

or

$$t_o(random) = \frac{\overline{\widehat{b1}} - \overline{\widehat{b2}}}{\sqrt{(\widehat{\sigma 1}^2_{(b_\eta + b_\epsilon)}/n_1 + \widehat{\sigma 2}^2_{(b_\eta + b_\epsilon)}/n_2)}} \approx t_{dist}((n_1 + n_2) - 2) \tag{13}$$

with the same definitions as in section 3 for each group. Note that here no conjunction analysis approach can be made unless the same subjects are in both groups (for example, before and after medical treatment) as pairing of subjects across the groups would be required - in fact this then ends up reverting to a one-group analysis (with a paired $t$-test). The "variance-ratio method" can, however, be performed.
When $g$ groups are studied one can compare them two by two, then introducing a multiple comparison (not as problematic as the statistical-map one). If equal variances are assumed, a pooled variance of all $g$ samples must be used for either method, instead of just the two considered. To do more advanced comparisons [8] one has to return to the GLM or ANOVA to be able to test, for example, if all the groups have the same activation, or if there is a trend in the groups, as one would expect for groups defined by increasing doses of a treatment. These would involve either F statistics and/or using a linear function of the parameters estimated in the model, i.e. contrasts.
*Remarks:*
The distributions for the statistics given here are approximations as the variance estimates are given under unequal variances assumption and in that case the Satterthwaite formula (3) should be used for the degrees of freedom (for the fixed effect one has to re-integrate first levels in the formula beforehand, see appendix). To use the degrees of freedom given one has to replace (either in the fixed or random) the variance in each group by their pooled

$$\widehat{\sigma_{pooled}}^2 = \frac{\sum_{u=1}^{g}(n_u - 1)\widehat{\sigma_u}^2}{((\sum_u n_u) - g)}$$

Notice that for equal sample sizes ($n_1 = n_2$) the pooled estimate of the variance would give the same result for $t_o$as the unequal variance and so the given distributions become exact.

# 5 General Linear Model

It is possible to do all of what has been done in the previous sections within a more general framework called the general linear model. It involves some matrix algebra that the reader should be familiar with. Firstly, the single-subject analysis is rewritten within this framework, which then enables one now to account for temporal auto-correlation. Then the general hypothesis testing is described, enabling models and hypothesis testing for a $g$-groups analysis.

## 5.1 The model for single-subject analysis

Let $y_t$, be the time series observed from $t = 1 \cdots T$, at a given voxel, this is represented as a vector on $\mathbb{R}^T$; $y$ is a $T \times 1$ vector. The relationship between the observed time series and the paradigm of the experiment ( with any other covariates) is expressed as:

$$y = X\beta + \epsilon$$

where $X$ is a $T \times q$ matrix representing the design and the covariates, $\beta$ is a $q \times 1$ vector of parameters (expressing the linear relationship)[9] and $\epsilon$ is a $T \times 1$ vector of random errors (departure from the model or residuals) assumed identically distributed (independent or not independent) with mean 0 and common variance $\sigma^2$. That will make the distribution of the vector $\epsilon$ with mean 0 and variance covariance matrix $\sigma^2 Id_T$ if independent[10] or $\sigma^2 V$, V being then the autocorrelation matrix (supposed to be known in this section).

[7] If one supposes equality of "estimated variances" $\hat{\sigma 1}^2 = \hat{\sigma 2}^2$ and $n_1 = n_2$, like a "Group Z", one obtains $t_o(fixed) = 1/\sqrt(2n)(\sum_i t_o 1(i) - \sum_i t_o 2(i))$ or $Z_{group12} = 1/\sqrt{2}(groupeZ_1 - groupZ_2)$ or$= \frac{1/\sqrt{n}_1 groupeZ_1 - 1/\sqrt{n}_2 groupeZ_2}{\sqrt{((n_1+n_2)/n_1 n_2)}}$ if $n_1 \neq n_2$.

[8] Although Hotelling's $T^2$ might be used at this stage

[9] As a trivial illustration of the link with the first section, let $X$ have two columns; one column containing 1 when the condition is A and 0 otherwise (this column identifies condition A), and one column identifying B, then $(\beta_2 - \beta_1)$ is going to be the activation looked for, as an estimate of the difference of the means.

[10] $E(\epsilon) = 0$ and $var(\epsilon) = \sigma^2 Id_T$ are called the Gauss-Markov conditions.

One usual way to estimate $\beta$ is by Least Squares of the "residuals"[11]:

$$\widehat{\beta} = arg \min_{\beta}[^t(y - X\beta)(y - X\beta)] \tag{14}$$

This problem can be solved by taking partial derivatives; this leads to the so called normal equation ${}^tXX\beta ={}^t XY$ and to a solution using the inverse or g-inverse[12] of ${}^tXX$:

$$\widehat{\beta} = ({}^tXX)^{-}\ {}^tXY \tag{15}$$

This is called the ordinary least squares (OLS) estimate, and if the errors are independent ($var(\epsilon) = \sigma^2 Id_T$) according to Gauss-Markov theorem[13] [11][3] it is the best linear unbiased estimate (BLUE) of $\beta$ which means that its mean (expectation) is $\beta$ and the estimate is of minimum variance (among the linear unbiased estimates[14]):

$$var(\ \hat{\beta}\ ) = \sigma^2({}^tXX)^{-} \tag{16}$$

If $var(\epsilon) = \sigma^2 V$, $V$ can be replaced with $V = K^tK$ (if $V$ is positive definite $K^{-1}$ exists[15]), applying $K^{-1}$ to the GLM gives the model:
$K^{-1}y = K^{-1}X\beta + K^{-1}\epsilon$ meeting the Gauss-Markov properties. The OLS estimate of $\beta$ is then written:

$$\hat{\beta} = ({}^t(K^{-1}X)K^{-1}X)^{-}\ {}^t(K^{-1}X)K^{-1}y = ({}^tXV^{-1}X)^{-}\ {}^tXV^{-1}y \tag{17}$$

and rewriting the optimisation problem shows that this is the solution of $\min_\beta[^t(y - X\beta)V^{-1}(y - X\beta)]$ called the Generalised Least Squares (GLS)[16]. One has also

$$var(\hat{\beta}_{GLS}) = \sigma^2({}^tXV^{-1}X)^{-} \tag{18}$$

*Remark:*
When $var(\epsilon) = \sigma^2 V$, $var(\hat{\beta}) = \sigma^2({}^tXX)^{-}\ {}^tXVX({}^tXX)^{-}$, then the OLS is BLUE if and only if $\mathcal{I}m(VX) \subset \mathcal{I}m(X)$ (the space generated by VX is included in the one generated from X); so in that case OLS$\equiv$ GLS [12].

## 5.2 Hypothesis testing and Estimability

Now assuming Normal distribution of errors one has a Normal distribution for the parameter vector $\beta$ which in this section is supposed to be the BLUE; then to test a hypothesis of the form $c\beta = 0$ where $c$ is a $1 \times q$ vector, the following statistic is used:

$$t_o = \frac{c\hat{\beta}}{\sqrt{c\ var(\ \hat{\beta}\ )\ {}^tc}} \sim t_{dist}(trace(P_{X\perp})) = t_{dist}(T - rank(X)) \tag{19}$$

with the unbiased estimate[17]

$$\hat{\sigma}^2 = \frac{{}^t\epsilon V^{-1}\epsilon}{T - rank(X)} \tag{20}$$

The reader can check that taking $X$ with 2 columns one of 1 everywhere and the other of 1 and -1 whether under condition B or A; $\beta ={}^t (\mu \quad b)$, choosing $c = (0 \quad 1)$ gives the same $t_0$ shown in the single-subject analysis section.
Only $c\beta$ which are *estimable* can be used in the testing procedure. A linear function of the parameter $c\beta$ is said to be *estimable* if a linear unbiased estimate exists:

$$\text{if } \ell \text{ exists, so that } E({}^t\ell y) = c\beta$$

[11]To obtain the best estimate one wants to minimise the variation between what is observed ($y_t$) and what is modelled ($X\beta$)
[12]Generalised inverse [12] is used when $X$ is not of full rank giving a non-invertible ${}^tXX$.
[13]The Gauss-Markov theorem is in fact established for any estimable function (see further) of $\beta$.
[14]It can be shown that for independent Normally distributed errors, the OLS is also the maximum likelihood BUE *i.e.* among all estimates.
[15]If $V$ is only semi-definite positive (and thus singular) everything is done with g-inverses, $V^{-} = {}^tK^{-}K^{-}$, the BLUE property remains if and only if $VV^{-}X = X$ [11][12].
[16]When $V$ is known $\widehat{\beta}_{GLS}$ is also BLUE and BUE under Normality.
[17]Or more generally with $V$ singular[3] $\hat{\sigma}^2 = \frac{{}^t(y - X\hat{\beta})V^{-}(y - X\hat{\beta}}{trace(AV)} \sim \chi^2(trace(AV))$ where $y \sim (\mu = X\beta, \sigma^2 V)$ and $A$ is the resulting from the quadratic form of the residuals, *i.e.* $A ={}^t P_{X\perp}V^{-}P_{X\perp} = V^{-}P_{X\perp}$ then giving $trace(AV) = trace(VV^{-}) - trace(P_X)$ giving $T - rank(X)$ if $V$ non-singular [12].

and this is if and only if ${}^t\ell X = c$. This makes $c\widehat{\beta}$ a unique BLUE of $c\beta$.
Contrasts are particular estimable linear functions with the additional property that $c1_q = 0$, where $1_q = {}^t(11\cdots 1)$, *i.e.* the sum of the entries (weights) in the contrast is zero.
This definition holds for more general hypothesis testing when $C$ is a $(q-k)\times q$ matrix of $(q-k)$ independent linear estimable functions. The following statistic (with distribution under $H_0$) called the Lawley-Hotelling trace is used to test the general hypothesis $H_0$; $C\beta = 0$:

$$\begin{aligned} LHt = \frac{(rank(P_{X\perp}))}{\nu} trace(HE^{-1}) &\approx F_{((q-k)p, Df)} \\ \text{with } H &= ({}^t(C\hat{\beta})[C({}^t(XV^{-1}X)^{-}\,{}^tC]^{-}C\hat{\beta} \\ E &= {}^t(P_{X\perp}Y)V^{-1}P_{X\perp}Y \end{aligned} \tag{21}$$

This statistic [18] is in fact for Multivariate GLM, *i.e.* when $Y$ is a matrix $n\times p$ of $p$ variables $y_1\ y_2\cdots\ y_p$. In our case $p=1$ and $LHt$ reduces to the traditional $F$ ratio statistic:

$$F = \frac{({}^t(C\hat{\beta})[C({}^t(XV^{-1}X)^{-}\,{}^tC]^{-}C\hat{\beta}}{(q-k)\hat{\sigma}^2} \sim F_{(q-k, rank(P_{X\perp}))} \tag{22}$$

Writing $C\hat{\beta} = 0$ as:
$C({}^tXV^{-1}X)^{-}({}^tXV^{-1}X)\hat{\beta} = C({}^tXV^{-1}X)^{-}\,{}^tXV^{-1}\hat{\mu}$
$= C({}^tXV^{-1}X)^{-}\,{}^tXV^{-1}P_Xy = C({}^tXV^{-1}X)^{-}\,{}^tXV^{-1}y = {}^tC_HV^{-1}y = 0$; denoting $C_H = X({}^tXV^{-1}X)^{-}\,{}^tC$; $C_H$ generates what is called in the appendix the *conditioned model*, then the numerator of the F statistic given in the appendix $\|P_{C_H}y\|^2$ is the numerator given here.
As an example the $F$ statistic can be used in a single subject analysis with a paradigm having more than two conditions (ON and OFF) but different levels for the ON conditions, like different audio stimuli.
*Remarks:*
Notice that the statistic (22) can be written:

$$\frac{({}^t(C\hat{\beta})[var(C\hat{\beta})]^{-}C\hat{\beta}}{(q-k)}$$

called Wald's type statistic. This form is simple and then advantageous to be used directly in any context. In the Lawley-Hotelling trace statistic, one must also note that $H = {}^t(P_{C_H}y)V^{-1}P_{C_H}y$, sometimes called the hypothesis statistic ($E$ being the error statistic) and can be derived also using the form $({}^t(C\hat{\beta})[var(C\hat{\beta})]^{-}C\hat{\beta}\times\sigma^2$: with OLS estimate, $C_H = VX({}^tXX)^{-}\,{}^tC$ generates the conditioned model. When not considering the BLUE of $\beta$ their distributions are then approximated (see next section for estimates of the degrees of freedom for the denominator).

## 5.3 Taking into account the autocorrelation

If we suppose $V \neq Id_T$ one will have to estimate it. If the estimate is "good enough", the estimates using GLS (sometimes called the EGLS for Estimated GLS or Empirical GLS, or plug-in estimated or two-stage least squares) will be good as well (nearly BLUE)[19]. It might be difficult to obtain a good estimate (estimating $V$ is estimating $T(T-1)/2$ parameters), so precluding the estimated-GLS[20] approach unless a model for the autocorrelation or $\sigma^2V$ in general can be made, reducing the number of parameters to estimate.
The pre-whitening approach is in fact the EGLS approach; usually $V$ is modelled, for example, as an autoregressive process to have fewer parameters to estimate, as in Bullmore et al.(1996)[1]. For example let $\epsilon_t$ be an autoregressive process of order 1; $\epsilon_t = \rho\epsilon_{t-1} + e_t$ with $e_t$ as a white noise. Then the correlation matrix $V$ takes the form (exponential model):

$$V = \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T-1} \\ \rho & 1 & \rho & \cdots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho^{T-1} & & & \cdots & 1 \end{pmatrix} \tag{23}$$

[18]The distribution given (exact is $min(q-k,p)=1$) is a traditional McKeon(1974)*Biometrika, 61:381-383* approximate distribution, where $Df = 4 + \frac{(q-k)p+2}{B-1}$ and $B = \frac{(rank(P_{X\perp})+(q-k)-p-1)(rank(P_{X\perp}-1)}{(rank(P_{X\perp})-p-3)(rank(P_{X\perp})-p)}$, and $\nu = ((q-k)p)[\frac{Df-2}{Df}][\frac{rank(P_{X\perp})}{(rank(P_{X\perp})-p-1)}]$
[19]The situation is less critical in a number of cases where the covariance structure is known and few parameters have to be estimated.
[20]It is nonetheless possible to do this for fMRI studies with the hypothesis of the same $V$ for all of the voxels, so giving a large sample to estimate $V$.

$$\text{then } V^{-1} = 1/(1-\rho^2)\begin{pmatrix} 1 & -\rho & 0 & \cdots & 0 \\ -\rho & 1+\rho^2 & -\rho & \cdots & 0 \\ 0 & -\rho & 1+\rho^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & & & \cdots & 1 \end{pmatrix} = 1/(1-\rho^2)^t\psi\psi$$

$$\text{with } \sqrt{1/(1-\rho^2)}\psi = \begin{pmatrix} \sqrt{(1-\rho^2)} & 0 & 0 & \cdots & 0 \\ -\rho & 1 & 0 & \cdots & 0 \\ 0 & -\rho & 1 & & \\ \vdots & & \ddots & \ddots & \\ 0 & \cdots & & -\rho & 1 \end{pmatrix} = K^{-1}$$

where $\rho$ is going to be estimated, for example, on the residuals after a first stage least squares ($\hat{\rho} = \sum_2 \epsilon_i \epsilon_{i-1} / \sum_1 \epsilon_i^2$) or by fitting an exponential model onto the autocorrelogram. Classical time series analysis will allow estimations for autoregressive processes with higher order of dependence.

GLS (or EGLS) uses $\hat{V}^{-1}$ and so can be very sensitive to a poor estimation of the correlation structure. Pointing out this problem, Worsley and Friston et.al(1995)[17] chose another way to take into account the autocorrelation. The GLM model can be written: $y = X\beta + K_I e = X\beta + \epsilon$, with $e$'s uncorrelated, so that $var(\epsilon) = \sigma^2 K_I\, {}^tK_I = \sigma^2 V_I$. Their "shaping approach" uses the assumption that their chosen matrix filtering $K_{ij} = e^{[-(i-j)^2/2\tau^2]}$ (with a chosen $\tau$) will swamp the autocorrelation, *i.e.* $KK_I \approx K$ where $K_I$ would give the true unknown one. Under this assumption the model is then written $y' = X'\beta + K\epsilon \approx X'\beta + Ke = X'\beta + \epsilon'$ with notations $y' = Ky$ and $X' = KX$. Deriving an OLS estimate gives an unbiased estimate but not BLUE anymore:[21]

$$\hat{\beta} = ({}^tX'X')^{-}\, {}^tX'y'$$

with variance

$$var(\hat{\beta}) = \sigma^2({}^tX'X')^{-}\, {}^tX'VX'({}^tX'X')^{-} \tag{24}$$

given that now

$$var(\epsilon') = \sigma^2 V = K var(\epsilon)\, {}^tK = KV_I\, {}^tK = \sigma^2 KK_I\, {}^tK_I\, {}^tK \approx \sigma^2 K\, {}^tK \tag{25}$$

To estimate $\sigma$ they use the same formula as described in the previous section (equation (20) but with an OLS optimisation): $\hat{\sigma}^2 = \frac{{}^t\epsilon'\epsilon'}{tr(P_{X'\perp}V)}$ where $P_{X'\perp} = Id_T - P'_X$ and $P'_X = X'({}^tX'X')^{-}\, {}^tX'$. The $t_o$ map derived in the previous section is then used with the *effective* degrees of freedom[22] calculated using classical results on quadratic form theorems [12]:

$$edf = \frac{2E(SS)^2}{var(SS)} = trace(P_{X'\perp}V)^2/trace(P_{X'\perp}VP_{X'\perp}V) \tag{26}$$

The distribution of $\hat{\sigma}^2$ is approximatively [23] distributed as $\chi^2_{edf}$ which then makes an approximate $t_{(edf)}$ distribution for $t_o$.

Notice in this presentation, without "swamping", setting $K = K_I^{-1}$ brings you back to GLS (see previous section). Within this debate of accounting for autocorrelation in fMRI, Woolrich et al. [16] propose a robust estimate $\widehat{V_I}$ of the autocorrelation of the time series. This robust estimate is then used either to pre-whiten the data (GLS) or to do an OLS estimation (*i.e.* classical OLS under a non Gauss-Markov assumption[24]) with filtering (as in Worsley and Friston's paper but without the swamping assumption) or without filtering (therein called variance correction) . The robustness is achieved in two ways, firstly by smoothing the non-parametric autocorrelation function for each voxel and secondly by smoothing the estimate spatially in a non-linear fashion to be able to preserve different patterns according to different matter types.

*Remark:*

Notice that a better estimate of $\sigma^2$ would have been using the form:

$$\frac{{}^t\epsilon' V^{-1}\epsilon'}{trace({}^tP_{X'\perp}V^{-1}P_{X'\perp}V)} \tag{27}$$

[21] Notice that GLS brings one back to OLS under a Gauss-Markov model assumption.

[22] if $SS/E(MS) \sim \chi^2_f$ then $var(SS)/E(MS)) = 2f$ hence $f = \frac{2E(SS)^2}{var(SS)}$

[23] It would be exact if $P_{X'\perp}V$ was idempotent [12] making $edf = trace(P_{X'\perp}V)$ as well.

[24] Without "swamping" the formulae given above for $\hat{\sigma}^2$ and (26) still hold if $V$ is replaced by $KV_I\, {}^tK$.

as given in equation (20) (here not with the BLUE), which uses the inverse of $V$ which Worsley and Friston wanted to avoid, but this time $V$ is given (Worsley and Friston) or is a robust estimate (Woolrich et al.) It seems that having a robust estimate of $V$ would incline one to use a GLS approach (pre-whitening). Note that in the last version of SPM'99 (Friston et al.) the "swamping" idea is dropped and an AR(1) model is used to estimate $V_I$. As this paper is written Friston[8] investigated the problem of estimating $V_I$ and choosing a filter in order to minimise the bias of (24). It is known that using an unbiased estimate $\widehat{V_I}$ for plug-in GLS estimation (EGLS) leads to good results for point estimation ( for $\beta$) but to underestimation for $var(\widehat{\beta} \setminus \widehat{V_I})$ (plug-in GLS variance), and more problematic underestimation of the true variance [2] of $\widehat{\beta}_{GLS}$, *i.e.* knowing $V_I$. So using OLS variances with a well chosen filter and a sensible estimate of $V_I$ would allow to obtain a more sensible estimate of the variance of $\widehat{\beta}$.
If $V = Id_T$, *i.e.* no auto-correlation, or within the GLS framework, *edf* is then equal to $T - rank(X)$, as seen before.

### 5.4 Applying GLM

Introducing GLM modelling involves the choice of the "design matrix" $X$ implying a model for the observed $y$, and the choice of the covariance structure (*e.g.* autocorrelation pattern) which relates to the sources of errors. In a simple way, $X$ models what is expected and $V$ models the errors[25]. What is not put in $X$ will be reflected in the errors and may need an appropriate structure to be properly taken into account for a "good" model. Inversely, what is put in $X$ is taken away from the errors. A good example of this is a confound covariate which will be put into the design $X$ to account for in the model as explaining what is expected, therefore adjusting the other fixed effects and not inflating error variation otherwise.
The subject effect can be thought to be part of the model ($X$) because one would expect different responses for different subjects. But that way the error variation will not take into account the sampling variation of the subjects (fixed approach), unless it is properly modelled as well in $V$ (implying random approach). Note that if one does not include the subjects in the model ($X$), the sampling variations will be considered in the errors, but will be pooled with the other sources of errors. We will come back to these points in the next section.
The problem of GLM with structured covariance lies in the estimation of $V$ and $\beta$ at the same time. Mixed models are GLM models of this type where part of the covariance structure comes from random effects. Part of the design describes fixed effects and part of it describes random effects:

$$y = X\beta + Z\gamma + \epsilon = X\beta + \eta \tag{28}$$

with $E(\gamma) = E(\epsilon) = 0$, $var(\epsilon) = R$, $var(\gamma) = G$ and $cov(\gamma, \epsilon) = 0$, or, $E(\eta = 0)$ and $V = var(\eta) = ZG\,{}^tZ + R$. Maximum likelihood techniques such as REML (REstricted ML see [3] for example) find $\hat{V}$ maximising

$$\ell_R(G, R) = -\frac{1}{2}[\log|V| + \log|{}^tXV^{-1}X| + {}^t(y - X\hat{\beta}_{GLS})V^{-1}(y - X\hat{\beta}_{GLS})] \tag{29}$$

and then plug it into the GLS estimator of $\beta$ to have a final $\hat{\hat{\beta}}_{GLS}$. For small samples, obtaining REML maximisation might be unreliable in the general case and still problematic for the unequal within-subject covariances hypothesis, but in a lot of situations the maximisation problem is simplified and can even lead to direct calculation.
In the general case the Mixed model can accept any form for $Z$, $G$ and $R$, then an algorithm [12] can give the REML. For balanced data "everything becomes simpler". This is because it implies $\mathcal{I}m(VX) \subset \mathcal{I}m(X)$ for many situations which makes the OLS equivalent to the GLS. This is particularly the case in fMRI (where continuous exploratory variables come from convolution of "balanced" dummy variables with a model of haemodynamic response thus preserving the balanced aspect) and that is the reason why the two-stage or two-level approach is here valid or optimal.
Using mixed model for fMRI studies considers the random effect as the subject effect. $Z$ will be a matrix of dummy variables identifying the subjects ($i = 1 \cdots n$), making $ZG\,{}^tZ$ a well-structured matrix of the form $G \otimes 1_T\,{}^t1_T$. $R$ is usually $\sigma_w^2 Id_{nT}$. As the subjects are supposed to be independent $G = \sigma_s^2 Id_n$ so that $ZG\,{}^tZ = Id_n \otimes \sigma_s^2 1_T\,{}^t1_T$ implying a constant correlation pattern for measures within subject, it is the one used in the section 3.2.
*Remark:*
In this model it is important to notice introducing a pattern of autocorrelation between time measures is not done through $Z$ (random effects) as $\gamma$ would be redundant with $\epsilon$; it is introduced in $R$.

[25] Note in all the text $V$ is sometimes expressing the covariance structure and sometimes only the correlation structure; no distinction has been made and it should be clear enough, noting that usually here common variance is assumed and so the same scalar $\sigma^2$ enables one to go from one to another.

# 6 Multi-subject analysis with GLM

The single-subject analysis seen at the beginning of this paper is a GLM with particular choices for $X$ and $V$ ($V = \sigma^2 Id_T$ and $X$ with 2 columns; one for the mean and one for the box-car, which is equivalent to the model where $y$ and $X$ are demeaned, reducing $X$ to just one demeaned box-car).
The GLM approach enables one to take into account in the model correlated errors, and also to provide different statistical maps according to the hypothesis being tested. These models can be used as a first step or first stage of a multi-subject analysis, *i.e* for every subject, as is done in section 3. In fact a GLM can also be used at a second level of the multi-subject analysis, but usually the second level model will be simply to estimate a mean or difference of means of the contrasts estimated at the first level. One must remark that it is at the second level where the distinction between the fixed subject effect approach and the random subject effect approach is made. The fixed approach will suppose a known error given by the pool of the errors from all the first stage models and thereby will not take into account the subject sampling variation, and the random approach will re-estimate the whole error variance.
The purpose of this section is to rewrite fixed and random analysis described in the previous section within the GLM model, to enable one to carry out the analysis in one step (although there may be practical reasons for not doing so).

## 6.1 Fixed subject effect

Firstly, we will rewrite the single-group study. One has for each $i = 1 \cdots n$ the GLM models:

$$y_i = X\beta_i + \epsilon_i \tag{30}$$

with $E(\epsilon_i) = 0$ and $var(\epsilon_i) = \sigma_i^2 Id_T$. The case of auto-correlation $var(\epsilon_i) = \sigma_i^2 V_i$ will be examined later on. It is rather straightforward to recognise the fixed effect in the following model:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} X & 0 & \cdots & 0 \\ 0 & X & \cdots & 0 \\ \vdots & & \ddots & \\ 0 & \cdots & & X \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix} \tag{31}$$

or $y = (Id_n \otimes X)\beta + \epsilon$ with obvious definitions for $y$, $\beta$ and $\epsilon$ with $var(\epsilon) = diag(\sigma^2_{\epsilon_i[1..n]}) \otimes Id_T$. One then derives easily the OLS for $\beta$ which gives the vector of the OLS's from each model (30):

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_n \end{pmatrix} \text{ and } var(\hat{\beta}) = diag(\sigma^2_{\epsilon_{i[1..n]}}) \otimes ({}^tXX)^- = diag(var(\hat{\beta}_i)_{[1..n]}) \tag{32}$$

Note that if one assumes that $var(\epsilon) = \sigma^2 Id_{nT}$ then:

$$\hat{\sigma}^2_\epsilon = \frac{{}^t\epsilon\epsilon}{trace(P_{(Id_n \otimes X)^\perp})} = \sum_i \frac{{}^t\epsilon_i\epsilon_i}{nT - nrank(X)} = \overline{\hat{\sigma}^2_{\epsilon_{i[1..n]}}} \tag{33}$$

which is the pooled estimate of variances, then:

$$var(\hat{\beta}) = \sigma^2_\epsilon Id_n \otimes ({}^tXX)^- \tag{34}$$

In either cases testing $L\beta = 0$, with $L = (1/n, \cdots, 1/n) \otimes c$, where $c$ is the contrast used on a single model to assess the activation looked for, with the statistic (19) will give the statistic used in (4), and if $c$ is a matrix ($C$) containing more than one contrasts the $F$ statistic (22) will be used.

## 6.2 Random subject effect

Firstly, notice that the random subject effect model could use this same model with a covariance structure (block-diagonal) which would need to be estimated in the first place. One would get a EGLS but this model cannot be used to estimate the covariance matrix (this is because writing this model as a mixed model makes

$\beta$ random and fixed at the same time). One has to rewrite the model, for example with the form:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} X \\ X \\ \vdots \\ X \end{pmatrix} \beta + Z\gamma + \begin{pmatrix} \epsilon'_1 \\ \epsilon'_2 \\ \vdots \\ \epsilon'_n \end{pmatrix} \tag{35}$$

or

$$y = (1_n \otimes X)\beta + (Id_n \otimes 1_T)\gamma + \epsilon'$$

where $Z = (Id_n \otimes 1_T)$ identifying the subjects (dummy variables)[26] , with also $E(\gamma) = 0$ and $var(\gamma) = G$ as random effects. One can easily see that OLS gives $\hat{\beta} = \overline{\hat{\beta_{(i)}}}_{[1..n]}$.
Depending on the structure of covariance chosen an estimate will be found directly (*e.g.* with compound symmetry) or algorithmically (for general REML estimation). For the random analysis with simple compound symmetry (*i.e.* with $var(y) = var(\epsilon') + ZG\,{}^tZ = \sigma_\epsilon^2 Id_{nT} + \sigma_s^2 Id_n \otimes J_T = Id_n \otimes (\sigma_\epsilon^2 Id_T + \sigma_s^2 J_T)$, where $\sigma_\epsilon^2$ is the scan-to-scan variation, supposed to be the same for all the subjects, and $\sigma_s^2$ is the random subject variation[27] introducing correlations into the errors) then with estimation for $\beta$ as OLS (here equivalent to GLS) "ANOVA" estimators (equivalent to REML) can be given by:

$$\hat{\sigma}_\epsilon^2 = MSE = {}^t y P_{[1_n \otimes X, Id_n \otimes 1_T]^\perp} y / [n(T-1) - rank(X) + 1] \tag{36}$$

and

$$\hat{\sigma}_s^2 = \frac{MSs - MSE}{T} \text{ where } MSs = {}^t y (P_{[1_n \otimes X, Id_n \otimes 1_T]} - P_{[1_n \otimes X]}) y / (n-1) \tag{37}$$

Our interest here is only on the subject error $\hat{\sigma}_\epsilon^2 + (T\hat{\sigma}_s^2) = MSs$ which is the two-stage approach estimation seen before (obviously dividing by $T$ to go on second level).

If all parameters in $X$ are considered as random, making $Z = (Id_n \otimes X)$, the estimation can be done in the same way with: $var(y) = var(\epsilon') + ZG\,{}^tZ = Id_n \otimes (\sigma_\epsilon^2 Id_T + \sum_{j=1}^p \sigma_j^2 X_j\,{}^tX_j)$ and considering the random parameters independent (non-correlated). Using Hendenson's Method [12, 3] 36 and 37 (for each random parameters) become:

$$\hat{\sigma}_\epsilon^2 = MSE = {}^t y P_{[1_n \otimes X, Id_n \otimes X]^\perp} y / [n(T - rank(X))] \tag{38}$$

and

$$\hat{\sigma}_j^2 = \frac{SSs - MSE\ (n-1)}{trace((P_{[1_n \otimes X, Id_n \otimes X]} - P_{[1_n \otimes X, Id_n \otimes X_{(-j)}]}) Id_n \otimes X_j {}^t X_j)} \tag{39}$$

where $X_{(-j)}$ is $X$ without the column $X_j$ and

$$SSs = {}^t y (P_{[1_n \otimes X, Id_n \otimes X]} - P_{[1_n \otimes X, Id_n \otimes X_{(-j)}]}) y \tag{40}$$

The error considered for a parameter is similarly $\hat{\sigma}_\epsilon^2 + (\theta\hat{\sigma}_j^2) = SSs/(n-1) = MSs$, with $\theta = trace((P_{[1_n \otimes X, Id_n \otimes X]} - P_{[1_n \otimes X, Id_n \otimes X_{(-j)}]}) Id_n \otimes X_j {}^t X_j)/(n-1)$ equal to $T$ under orthogonality of $X_j$'s.
For general correlation structure, or one derived from time series analysis, instead of using an REML estimation algorithm, a "two-stage strategy" could be to estimate the autocorrelation matrix $C_i$ from every subject's time series, then pool these to give an estimate of the common auto-correlation $C$, to be able to define the covariance structure of the form $V = Id_n \otimes \sigma_s^2 J_T + Id_n \otimes \sigma_w^2 \hat{C}$, then solve REML (or Henderson's method III [12]) for $\sigma_w^2$ and $\sigma_s^2$. Notice a redundancy in the previous formula as the covariance structure describes at the same time a constant covariance for the same subject ($\sigma_s^2$) and the auto-covariance coming from time series autocorrelation; this suggests a better covariance model of the form $V = \sigma^2 Id_n \otimes \hat{C}$ which then take subject the variance component off the model (no need of $Z$ as well). At this point, time series methods are used to estimate the $C_i$ in the first place, and robust estimation such as in Woolrich et al.[16] including the non-linear spatial smoothing would improve the result.

*Remark:*
The model (35) without $Z$ could be considered as an intermediary model (Pme) between the fixed model (Fix) ( the model (35) but $var(\epsilon') = \sigma^2 Id_{nT}$), *i.e.* $var(\gamma) = 0$) and the random subject model (Ran) as it can be shown that the natural estimate of $\sigma^2$ is the pooled error variance of the fixed version ($MSE$ given in (36)) of

[26] This is if only random intercepts are considered, one may consider all parameters in $X$ as random, making $Z = (Id_n \otimes X)$
[27] Compound Symmetry is the covariance model used in the section 3.2, $J_T = 1_T\,{}^t1_T = TP_{1_T}$

and random approaches (subject error $MSs$ (37), but because of the $df$ differences between the two approaches this will underestimate the error variance. This method called here"pooling model errors" or Pme may be interesting when too few subjects are in the study : the error variance is going to be larger than for a fixed analysis and the increase of degrees of freedom ($(n-1)$) not too large:

$$\begin{aligned}\text{model Pme: } y &= (1_n \otimes X)\beta + \epsilon_{\text{Pme}} \\ &= \mathcal{X}\beta + \epsilon_{\text{Pme}} \\ \text{model Fix: } y &= (\mathcal{X} \quad Z)B + \epsilon_{\text{Fix}} \\ &= FB + \epsilon_{\text{Fix}} \\ \text{model Ran: } y &= \mathcal{X}\beta + Z\gamma + \epsilon_{\text{Ran}} \text{ and } var(\gamma) \neq 0\end{aligned} \tag{41}$$

then from model Fix ,

$$\begin{aligned} df_{\text{Fix}} MSE_{\text{Fix}} &= {}^t y(Id - P_F y \\ &= {}^t y(Id - P_{\mathcal{X}} + P_{\mathcal{X}} - P_F y \\ &= df_{\text{Pme}} MSE_{\text{Pme}} - (n-1)MSs \end{aligned} \tag{42}$$

thereby

$$\widehat{\sigma}^2_{\text{Pme}} = MSE_{\text{Pme}} = \frac{df_{\text{Fix}} MSE_{\text{Fix}} + (n-1)MSs}{df_{\text{Pme}}} \tag{43}$$

and one can easily check $df_{\text{Pme}} = df_{\text{Fix}} + (n-1) = nT - rank(X)$.

### 6.3 2 and $g$ groups multi-subject analysis

For 2 groups or $g$ groups this framework is used and the models are the same. As we have already seen, to compare two groups, a two-sample $t$-test is used, and that will be similar here, as the use of a contrast to compare two groups will form the difference of the means (of the "activation contrasts" applied to each subject). If one has $g$ groups, multiple comparisons of pairs of groups might be used (with some multiple comparisons corrections if needed) but an $F$-test might be used for an overall (or more than two) group effect.
With $g$ groups with the same number of subjects $n$ in each group, the model will have the form:

$$y = (Id_g \otimes Id_n \otimes X)\beta + \epsilon \tag{44}$$

with ${}^t y = ({}^t y_1, {}^t y_2, \cdots, {}^t y_g)$ and similar for $\beta$ and $\epsilon$. The general form of the "design matrix" would be a diagonal matrix of $g$ blocks of the form $Id_{n_k} \otimes X$ where $n_k$ is the sample size of the $k$th group (notice that for the random model when estimating the variance components the blocks are $1_{n_k} \otimes X$). As the groups are independent $var(\epsilon) = Id_g \otimes V$ where $V$ will have the form of either the fixed or random approach in a 1 group analysis, or more generally $var(\epsilon) = diag(V_{k_{[k=1\cdots g]}})$ if the $V_k$ are not assumed equal or the groups are unbalanced.
For a two-group comparison in a $g$-group analysis, the applied hypothesis is:

$$L_{12}\beta = (0\cdots 0, L_1, 0\cdots 0, -L_2, 0\cdots 0)\beta = 0 \tag{45}$$

where $L_k$ is a $1 \times n_k T$ contrast, providing the mean for this group of the activation contrast used in a single-subject model as for a 1-group analysis. Then the $t$-map is derived using the same framework as in the 1-group analysis, using the estimates for $\hat{V}$ according to the approach and the model chosen. If one wants to have a global assessment of group differences (among more than 2 groups) the $F$ statistic given in (22) has to be used with a contrast matrix, *e.g.* to compare all the groups $2\cdots g$ with a control group, $c = 1$, the matrix L of $(g-1)$ contrasts ${}^t L = ({}^t L_{2c}, {}^t L_{3c}, \cdots, {}^t L_{gc})$ is involved.
*Remarks:*
The degrees of freedom (GLS approach) to apply for $t$-test (2 groups comparison), for the fixed approach will be $df(fixed) = \sum_{k=1}^{g} n_k(T - rank(X))$, will be $T\sum_{k=1}^{g} n_k - rank(X)$ for the pooled Random and fixed model, and for the random model with the compound symmetry covariance structure $df(random) = \sum_{k=1}^{g} n_k - g$.

# 7 Towards Multivariable Multivariate GLM

Repeated-measures GLM is also available to describe fMRI studies; the so-called growth curve model could be the best way of conceptually modelling a multi-subject experiment. It has the great advantage of being very close to the two-stage understanding of the random subject analysis, and of being clearly explicit for a $g$ groups analysis.
A multivariate GLM model is usually written like the general form we have seen before but $y = Y$, $\beta$ and $\epsilon$ are matrices. A multivariate linear model can be in fact rewritten as a classical GLM and for example resolved this way as well. The growth curve model is an extension of the multivariate GLM to allow a model of the curve response. It is in fact also equivalent to a classical GLM considering a tensor product of the designs:

$$\underset{N\times T}{Y} = \underset{N\times p}{X_N} \; \underset{p\times m}{\beta} \; \underset{m\times T}{{}^tX_T} + \underset{N\times T}{\epsilon} \tag{46}$$

For a 1-group analysis $N = n$ and $X_N$ reduces to the "constant" $1_n$ (or $Id_n$ for the fixed model), otherwise $N = \sum_k n_k$ and $X_N$ identifies the groups. Notice that $X_T$ is in fact the design we had before (then denoted $X$) to identify the paradigm and covariates (within the time series). The model (46) is in fact equivalent to the univariate model (where underlined means all of the columns stacked, *i.e.* $NT \times 1$ or $pm \times 1$ vectors)

$$\underline{Y} = (X_N \otimes X_T)\underline{\beta} + \underline{\epsilon} \tag{47}$$

from which one recognises the model (35) if one has only one group. Some algebra gives the result for the least squares estimate:

$$(X_N \otimes X_T)\underline{\hat\beta} = (P_{X_N} \otimes P_{X_T})\underline{Y} \equiv P_{X_N} Y \, {}^tP_{X_T} \tag{48}$$

(where $\equiv$ means algebraically equivalent, *i.e.* a vector $NT \times 1$ or a matrix $N \times T$ of the same thing) [28].

The normal distributional assumption comes with the restriction of "separability", *i.e.* of the form $var(\underline{\epsilon}) = V_N \otimes V_T$, with usually the form $var(\underline{\epsilon}) = (Id_N \otimes \sigma^2\Sigma)$ where $\Sigma$ is a $T \times T$ matrix expressing the common (to all rows) correlation of the columns (random variables) of $Y$. Historically presented by Potthof and Roy(1964) (see reference in [14]), the model (46) was named GMANOVA (Growth curve Multivariate ANalysis Of VAriance) and was "solved" with a two-level analysis: projecting first according to $X_T$ and then solving the second level model. Let the projector onto $X_T$ expressed with a metric $Q_0$, $P_{X_T} = X_T({}^tX_TQ_0X_T)^{-1}\,{}^tX_TQ_0$, then writing $Y\,{}^tP_{X_T} = Y_0\,{}^tX_T$ one obtains $E(Y_0) = X_N\beta$ as MANOVA model. A known statistical result is that, under normality, the least squares estimate of $\beta$ in the MANOVA model, which is in fact $\hat\beta$ (48), is also the maximum likelihood of $\beta$ in the MANOVA model for the choice of $Q_0 = \widehat{\sigma^2\Sigma}^{-1} = ({}^tYP_{X_N\perp}Y/(rank(P_{X_N\perp}))^{-1}$. In fMRI this estimation is not sensible as $N$ is small comparatively to $T$, one would prefer time series methods (as mentioned before) to estimate the correlation structure $\Sigma$.
Estimating $\sigma^2$ brings back to the choices between random, fixed subject analysis or the "pooling model errors" (43). In the random analysis as in the usual Growth curve analysis $\sigma^2$ relates to common variance (across subject population) for every time measurement, the estimate $1/T[trace(Q_0)]$ could be used i.e. the pooling (across time) of residual variances of the subject model ($y_t = X_N\beta_t + \epsilon_t$ at t fixed), or could be better estimated from the MANOVA model so with $1/m[trace(S_0)]$ (where $S_0 = {}^tY_0P_{X_N\perp}Y_0/(N - rank(X_N))$). Notice (if $X_N = 1_N$) the diagonal of $S_0$ contains the variances of the parameters [29] used for the random subject analysis (9). When $X_N = Id_N$ as in the fixed subject analysis, the previous estimate does not make any sense (it is zero) and the natural estimate is used:

$$\hat\sigma^2 = \frac{{}^t\underline{Y}[\,{}^t(Id_{NT} - P_{X_N} \otimes P_{X_T})Id_N \otimes \hat\Sigma^{-1}(Id_{NT} - P_{X_N} \otimes P_{X_T})]\underline{Y}}{trace(\,{}^t(Id_{NT} - P_{X_N} \otimes P_{X_T})Id_N \otimes \hat\Sigma^{-1}(Id_{NT} - P_{X_N} \otimes P_{X_T})Id_N \otimes \hat\Sigma)} \tag{49}$$

with $X_N = Id_N$ and is the pooled subject errors with autocorrelation taken into account. This last estimate is also the "pooling model errors" estimate of $\sigma^2$ when $X_N \neq Id_N$ (basically $rank(X_N) < N$). Model (47) could come with assumption $var(\underline{\epsilon}) = (Id_N \otimes D_{\sigma^2}\Sigma)$ and an estimation related to the later comment would be $D_{\sigma^2} = diag(Q_0)$ or $D_{\sigma^2} = diag(X_TS_0\,{}^tX_T) = diag(P_{X_T}Q_0\,{}^tP_{X_T})$.
To retrieve completely the mixed model, one would have to write the covariance with the form $Id_N \otimes \sigma^2\Sigma = Id_N \otimes (\sigma^2_\epsilon C + \sigma^2_s J_T)$, but after all it is not necessary in fMRI studies to split or separate the covariance structure as the interest is only on the fixed effects and not actually on the variance components.

[28] and the projectors are expressed with or without the $V^{-1}$ according BLUE or not BLUE.

[29] Instead of averaging the variances of the parameters one may median or max depending on how we want to be conservative or not.

The Lawley-Hotelling statistic already described (21) can be used and a general hypothesis takes the form $L\beta M = 0$ or $L \otimes M\underline{\beta} = 0$ where $L$ will contrast the groups and $M$ contrasts the paradigm.
Mathematically it is possible to extend further this representation to any "dimension" in the data, for example incorporating the spatial dimension:

$$\underline{Y} = (X_B \otimes X_N \otimes X_T)\underline{\beta} + \underline{\epsilon} \tag{50}$$

where $X_B$ would relate a model on the voxels (or $X_B = Id_v$ if spatial modelling is not carried out, where $v$ is the number of voxels), $Y$ can be interpreted as a tensor of order 3 with a $(vNT) \times 1$ vector representation. To be useful this model would have to take into account the spatial covariance structure.
*Remark:*
Note that if the groups constitute a repeated experiment (*e.g.* with different doses of a drug), $V_N$ can be chosen to be for the form $V_N = Id_n \otimes \sigma_g^2 J_g$ as in compound symmetry, and similar techniques described for mixed models could be derived. This can be also implemented as a three-stage model: seeing the random model as a GMANOVA model then resolved as a two-stage model allow you to carry on levels in presence of a repeated design on the subjects giving a third model. The estimation in multi-stage for multilevel models is valid as long as the designs are well balanced which is the case in fMRI. Iterative GLS [30] may be needed to improve estimation. Note a dense literature on multilevel modelling allows considerations of more complex situations [9].

# 8 Discussion and Conclusions

In the introduction to this review, it was stated that classically the analysis is separated for each voxel, and separated for each subject. This is true in terms of estimation of the fixed effect and given that OLS is desired (although may not be optimal), but the estimation of the covariance structure does not imply separation (even if on some occasions one simply pools over all the subjects and groups). It might be desirable as well not to completely separate the analysis for voxels (which is normally the case, apart from the final stage of cluster-based thresholding). This restriction was, for example, lifted in the "variance ratio smoothing" method, and mentioned in the previous section.
Thus, a point which should be addressed is the sample size required to do an fMRI-GLM analysis. It has been already pointed out that for a random subject effect analysis, this is crucial for good power. This fact is dependent on another point not discussed in this paper but crucial which can be summarised in a question: "are these voxels representing the same variables over different time and subjects?". A short answer would be: "this is a matter of data registration before multi-group subjects statistical analysis!" (see [15] for a good discussion of this). In the present paper when assessing activation two types of variation have been discussed: the *within subject variation* (repeated measures and registration error of the time course), and the *between subject variation* (population variation of the signal at a given voxel). A third type of variation has not been discussed, which is crucial when looking at multi-subject fMRI experiments, the *variation due to error of co-registration of subjects* confounded or not with the *natural anatomic-functional variation* from subject to subject. These three types of variation can be characterised as (i) *measurement error and time variation*, (ii) *point population variation*, (iii)*spatial population variation and error*, where here, error and variation are used to emphasise differences in variation due to the technical process of measurement (*i.e.* purely instrumental error but also within variation), and variation due to more natural causes(*i.e. population variation*).
The analysis presented in this paper did not address the third type of variation (measurement error and time variation were also confounded). This *spatial variation* has obviously a greater impact in random subject analysis than in fixed subject analysis, as it plays a role in the location or estimation of possible activation (fixed and random approach) but also in the decision of significant activation through the variance estimation (only the random approach). A simple way to account for it is to perform ROI analysis or spatial smoothing before second level analysis. According to the activation estimation, these introduce a loss of specificity but also in sensitivity if the smoothing "width" is larger than the activation "area". Is this loss going to be counterbalanced by smaller variances achieved by smoothing? A way of keeping sensitivity when smoothing is to use a max-min filter [13] retaining for every subject only a maximum or minimum value on a neighbourhood of a given voxel (though activation would give a larger absolute value for the maximum, and deactivation would give a larger absolute value for the minimum). This approach was performed successfully [13] with permutation testing for fixed subject analysis in combining the z-maps, but could be more appropriate for random subject analysis (the filter is performed on the parameter maps) as a valid population method less conservative than its unfiltered counterpart.

[30] Alternatively estimating $\beta$ given the whole covariance, and estimating covariances (different levels) given $\beta$, is known to converge to the maximum likelihood estimates under normality.

## *Appendix*

# A Z-scores and distributions

Some univariate distribution results that the reader should be familiar with: let $x_i \sim N(\mu, \sigma^2)$ for $i = 1 \cdots n$ then $\bar{x} = 1/n \sum_i x_i \sim N(\mu, \sigma^2/n)$ and so $\frac{\bar{x}-\mu}{\sqrt{\sigma^2/n}} \sim N(0,1)$. When $\sigma^2$ is unknown, a well known unbiased estimate is: $\hat{\sigma}^2 = 1/(n-1)\sum_i (x_i - \bar{x})^2 \sim \chi^2(n-1)$ and now $\frac{\bar{x}-\mu}{\sqrt{\hat{\sigma}^2/n}} \sim t(n-1)$.

# B Deriving $F$ distributions

Here are some well known distributional results introducing the $\chi^2$ and the $F$ distributions in relation to GLM using the geometrical approach of projectors [12][3].
Let $y \sim N(\mu, \sigma^2 V)$ where $\mu = X\beta$ with $X$ of full rank $n \times p$, $p < n$, *i.e.* $\mu \in [X]$ where $[X]$ means the space generated by the columns of $X$, then:

i $\|P_G y\|^2 / \sigma^2 \sim \chi^2(trace(V^- P_G V), \|P_G \mu\|^2 / (2\sigma^2))$,
where $P_G = G({}^tGV^-G)^-\, {}^tGV^-$, is the orthogonal projector[31] onto G, then $\|P_G y\|^2 = {}^t(P_G y)V^- P_G y$ is the Sum of Squares due to G (SSG) in the metric space $(\mathbb{R}^n; V^-)$.[32]

ii if $[G_1] \perp [G_2]$ then $\|P_{G_1} y\|^2$ and $\|P_{G_2} y\|^2$ are independent, then their ratio follows an $F$ distribution:

$$\frac{trace(V^- P_{G_2} V)}{trace(V^- P_{G_1} V)} \frac{\|P_{G_1} y\|^2}{\|P_{G_2} y\|^2} \sim F(trace(V^- P_{G_1} V), trace(V^- P_{G_2} V)$$

iii testing $H_0 : \mu \in [X_0] \subset [X]$ with $rank(X_0) = k$, is equivalent to testing $P_{[X]\cap[X_0]^\perp} y = 0$ (*i.e. the projection of* $y$ *onto* $[X]$ *orthogonally to* $[X_0]$ *is null*),
then under $H_0$

$$\frac{(trace(VV^-) - trace(P_X))}{(trace(P_X) - trace(P_{X_0})} \frac{\left\|P_{[X]\cap[X_0]^\perp} y\right\|^2}{\left\|P_{[X]^\perp} y\right\|^2}$$
$$\sim F((trace(P_X) - trace(P_{X_0}), (trace(VV^-) - trace(P_X)))$$

note that as $[X_0] \subset [X]$, $P_{[X]\cap[X_0]^\perp} = P_{[X]} - P_{[X_0]}$. Sometimes $[X] \cap [X_0]^\perp$ is noted $[X]/[X_0]$ and can be called a conditioned model, $[X_0]$ called the sub-model. The $F_0$ ratio can be written as $(SSX - SSX_0/SSE)$ or $(SSE_0 - SSE)/SSE$ times the ratio of $df$. If $V = Id$ one finds the traditional $F_0 \sim F(p-k, n-p)$.

# C Satterthwaite [10] degrees of freedom

For a $t - test$ or an $F - test$ and the denominator is a linear combination of Mean sum of Squares (MS), the Satterthwaite formula to approximate the degrees of freedom for the denominator is:

$$\frac{[\sum_s c_s MS_s]^2}{\sum_s (c_s MS_s)^2 / dl_s} \tag{51}$$

where $dl_s$ are the degree of freedom for the $MS_s$.

# References


[1] E.T Bullmore, M.J Brammer, S.C.R Williams, S Rabe-Hesketh, N Janot, A.S David, J.D.C Mellersand, R Howard, and P Sham. Statistical methods of estimation and inference for functional MR image analysis. *Magnetic Resonance in Medicine*, 35:261–277, 1996.

[2] R Christensen. Linear Models for Multivariate, Time Series, and Spatial Data. Springer, New York, 1990.

[3] R Christensen. Plane Answers to Complex Questions - 2nd ed . Springer, New York, 1996.


[31] Really, $P_G$ should be denoted $P_{[G]}$, the projector onto the space generated by the columns of $G$.

[32] Note that $trace(V^- P_G V) = trace(P_G V V^-)$ and if $P_G = P_{F\perp} = Id - P_F$, it is equal to $trace(VV^-) - trace(P_F) = rank(V) - rank(P_F)$

[4] J. Darrell Van Horn, T.M Ellmore, G Esposito, and K.F Berman. Mapping Voxel-Based Statistical Power on Parametric Images. *Neuroimage*, 7:97–107, 1998.

[5] K.J Friston, A.P Holmes, C.J Price, C Bchel, and K.J Worsley. Multisubject fMRI Studies and Conjunction Analysis. *NeuroImage*, 10:385–396, 1999.

[6] K.J Friston, A.P Holmes, and K.J Worsley. Comments and controversies: How many subjects constitute a study? *NeuroImage*, 10:1–5, 1999.

[7] K.J Friston, A.P Holmes, K.J Worsley, J.P Poline, C.D Frith, and R.S.J Frackowiak. Statistical Parametric Maps in Functional Imaging: A General Linear Approach . *Human Brain Mapping*, 2:189–210, 1995.

[8] K.J Friston, O Josephs, E Zarahn, A.P Holmes, S Rouquette, and J.P Poline. To Smooth or Not to Smooth? *NeuroImage*, 12:196–208, 2000.

[9] H Goldstein, M.J.R Healy, and J Rasbash. Multilevel time series models with applications to repeated measures data. *Statistics in Medicine*, 13:1643–1655, 1994.

[10] F.E Satterthwaite. An approximate distribution of estimates of variance components. *Biometrics Bull*, 2:110–114, 1946.

[11] S.R Searle. Linear Models . John Wiley & Sons, New York, 1971.

[12] S.R Searle. Variance Components . John Wiley & Sons, New York, 1992.

[13] S.M Smith, P.M. Matthews, J.M. Gurd, and A Slater. Statistic Map Combination from Multi-subject Functional MRI Experiments. Technical Report TR98SMS2, University of Oxford, fMRI centre, 1998.

[14] N.H Timm. The CGMANOVA Model. *Communications in Statistics*, 26(5):1083–1098, 1997.

[15] R.P Woods. Modeling for intergroup comparisons of imaging data. *Neuroimage*, 4:S84–S94, 1996.

[16] M. Woolrich, B.D. Ripley, J.M. Brady, and S Smith. Temporal Autocorrelation in Univariate Linear Modelling of fMRI Data. *Human Brain Mapping 2000*, Abstract:S610, 2000.

[17] K.J. Worsley and K.J. Friston. Analysis of fMRI time-series -Revisited again. *Neuroimage*, 2:173–181, 1995.

[18] K.J. Worsley, C Liao, M Grabove, V Petre, B Ha, and A.C Evans. A General Statistical Analysis for fMRI Data. *Human Brain Mapping 2000*, Abstract:S648, 2000.