ELSEVIER
NeuroImage
www.elsevier.com/locate/ynimg
NeuroImage xx (2005) xxx – xxx

Detecting and adjusting for artifacts in fMRI time series data

Jörn Diedrichsen* and Reza Shadmehr

Department of Biomedical Engineering, Laboratory for Computational Motor Control, Johns Hopkins University School of Medicine, Baltimore, 720 Rutland Ave, 416 Traylor Building, MD 21205-2195, USA

Received 16 November 2004; revised 22 April 2005; accepted 28 April 2005

We present a new method to detect and adjust for noise and artifacts in functional MRI time series data. We note that the assumption of stationary variance, which is central to the theoretical treatment of fMRI time series data, is often violated in practice. Sporadic events such as eye, mouth, or arm movements can increase noise in a spatially global pattern throughout an image, leading to a non-stationary noise process. We derive a restricted maximum likelihood (ReML) algorithm that estimates the variance of the noise for each image in the time series. These variance parameters are then used to obtain a weighted least squares estimate of the regression parameters of a linear model. We apply this approach to a typical fMRI experiment with a block design and show that the noise estimates strongly vary across different images and that our method detects and appropriately weights images that are affected by artifacts. Furthermore, we show that the noise process has a global spatial distribution and that the variance increase is multiplicative rather than additive. The new algorithm results in significantly increased sensitivity in the ability to detect regions of activation. The new method may be particularly useful for studies that involve special populations (e.g., children or elderly) where sporadic, artifact-generating events are more likely.

Keywords: Restricted maximum likelihood; Functional MRI; Noise; Estimation; Weighted least squares

Introduction

While functional magnetic resonance imaging (fMRI) is an important method to investigate neural activity in vivo, it suffers from a low signal to noise ratio. A prominent source of noise is motion of the participant. Head movements induce spin-history artifacts (Friston et al., 1996) and motion-by-susceptibility interactions (Wu et al., 1997). Eye movements, as well as movements of the tongue or lower jaw, can alter the homogeneity of the magnetic field and introduce distortions in the images (Beauchamp, 2003; Birn et al., 1998, 1999). While realignment procedures can correct for head motion after acquisition, significant residual effects often remain (Grootoonk et al., 2000). A second source of noise is due to physiological processes like breathing or heart beat (Frank et al., 2001; Kruger and Glover, 2001). It is possible to partially correct for some of these artifacts (Andersson et al., 2001; Frank et al., 2001; Friston et al., 1996; Grootoonk et al., 2000), however, it is likely that the sources of noise vary from data set to data set. Often, the investigator is forced to visually inspect the raw data and exclude images that contain obvious artifact. Here, we present a simple and new approach to detect and correct for noise and artifacts in functional MRI time series data.

In fMRI, we typically measure the signal intensity from N voxels at acquisition time $t = 1 \ldots T$. Each of these T measurements constitutes an image. We assume that the time series of voxel n is an arbitrary linear function of the design matrix $\mathbf{X}$ plus a noise term:

$$\boldsymbol{y}_n = \mathbf{X}\boldsymbol{\beta}_n + \varepsilon_n \tag{1}$$

$\boldsymbol{y}_n$ and ε_n are $T \times 1$ column vectors, $\mathbf{X}$ is the $T \times p$ design matrix, and $\boldsymbol{\beta}_n$ is a $p \times 1$ vector of regression parameters. Generally, fMRI analysis either assumes independent and identically distributed noise or noise that has a specific temporal autocorrelation structure. In many widely used methods, the variance structure of the noise is first estimated from the data and then utilized in the estimation of the regression parameters (Aguirre et al., 1997; Friston et al., 1995; Worsley and Friston, 1995; Worsley et al., 2002). In all of these approaches, the noise term is assumed to arise from a stationary process; that is, the variance of $\varepsilon_n(t)$ as well as the autocovariance function is assumed to be independent of time.

How reasonable is this assumption? If one source of noise is due to random discrete events, for example, artifacts arising from the participant moving their jaw, then only some images will be influenced, violating the assumption of a stationary noise process. To relax this assumption, a simple approach is to allow the variance of noise in each image to be scaled by a separate parameter. Because we are mainly concerned with the question of stationarity, let us for now assume that the noise is temporally uncorrelated. We will later address the influence of temporal autocorrelation. Under the temporal independence

* Corresponding author. Fax: +1 410 614 9890.
E-mail address: jdiedric@bme.jhu.edu (J. Diedrichsen).

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

1053-8119/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2005.04.039

assumption, the variance–covariance matrix of the noise process ε_n would be:

$$\operatorname{var}(\varepsilon_n) = \begin{bmatrix} s_1 & 0 & \cdots & 0 \\ 0 & s_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_T \end{bmatrix} \sigma_n^2 = \mathbf{V}\sigma_n^2 \tag{2}$$

In Eq. (2), $\mathbf{V}$ is a diagonal $T \times T$ matrix, while σ_n^2 is a scalar. Thus, the expected variance of a voxel at a certain time is a function of the relative amount of variance in that image s_t and the overall noise σ_n^2 observed in that voxel (for a similar multiplicative variance model, see Worsley et al., 1996). In our model, the sum of all s_t is constrained to be T. Note that the variance scaling parameters s_t are the same across all voxels in one image. This implies that, if an image shows increased noise for some voxels, the noise for all other voxels should be increased to a similar degree. Effectively, Eq. (2) assumes that the artifacts that lead to increased noise levels are spatially extended and encompass considerable portions of the brain. We will test this assumption in the Results section.

Discrete events (e.g., swallowing) will impact only those images that were acquired during the event. What should be done with these images, once they are identified? A typical approach would be to discard images based on some fixed threshold. If we knew var(ε_n), however, the optimal approach would be to weigh the images by the inverse of their variance. The maximum likelihood estimate of $\boldsymbol{\beta}_n$ under the assumption that var(ε_n) = $\mathbf{V}\sigma_n^2$ is the generalized least square (GLS) estimate:

$$\hat{\boldsymbol{\beta}}_{n,GLS} = \left(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}_n \tag{3}$$

$$\mathbf{V}^{-1} = \begin{bmatrix} s_1^{-1} & 0 & \cdots & 0 \\ 0 & s_2^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_T^{-1} \end{bmatrix} \tag{4}$$

Because $\mathbf{V}^{-1}$ in Eq. (4) is a diagonal matrix, Eq. (3) is often called a weighted least squares (WLS) estimate of $\boldsymbol{\beta}_n$.

Use of this method, however, requires that we obtain a valid estimate of the noise level for each image. How could this be achieved? As a first step, consider using the residuals from the ordinary least squares (OLS) regression to estimate the variance parameters. The residuals are:

$$\mathbf{r}_n = \mathbf{y}_n - \mathbf{X}\hat{\boldsymbol{\beta}}_n = \left(\mathbf{I} - \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\right)\mathbf{y}_n = \mathbf{R}\mathbf{y}_n \tag{5}$$

where $\mathbf{R}$ is the residual-forming matrix for the OLS regression. Now, we could use $\mathbf{r}_n\mathbf{r}_n^T$ as an estimate of $\mathbf{V}\sigma_n^2$. If we assume that the effect of noise in fMRI data is spatially extended, we may approximate the coefficients $\hat{s} = [\hat{s}_1, \ldots, \hat{s}_T]^T$ by averaging these variance estimates, weighted by $\hat{\sigma}_n^2$, over the whole brain. Thus, our estimator would become:

$$\hat{s} = \frac{1}{N}\sum_{n=1}^{N}\operatorname{diag}\left(\mathbf{r}_n\mathbf{r}_n^T/\hat{\sigma}_n^2\right)$$

$$\hat{\sigma}_n^2 = \mathbf{r}_n^T\mathbf{r}_n/(T - \operatorname{rank}(\mathbf{X})) \tag{6}$$

where the operator diag transforms the diagonal of a square matrix into a column vector. While this estimator has intuitive appeal, it has the problem of being biased and will not result in valid t values (see Simulation results—uncorrelated noise section). The bias arises because we based our estimates of the variance scaling parameters on the residuals of the OLS regression. As we show in Appendix A, the expected sum of squares of the residuals E ($\mathbf{r}_n\mathbf{r}_n^T$) is not $\mathbf{V}\sigma_n^2$ but rather

$$E\left(\mathbf{r}_n\mathbf{r}_n^T\right) = E\left(\mathbf{R}\mathbf{y}_n\mathbf{y}_n^T\mathbf{R}^T\right) = \mathbf{RVR}\sigma_n^2 \tag{7}$$

In particular, time points at which many of the regressors are non-zero will have a lower expected residual than time points at which only the mean regressor is non-zero. In the extreme, if a regressor was non-zero only for one time point, the residual for that image would always be zero. One method to correct this bias is restricted maximum likelihood (ReML) estimation, an iterative method that maximizes the modified likelihood (for an overview, see Speed, 1997). When trying to estimate T variance components, however, the computational costs can quickly get out of hand. To help us out, we can use the fact that the calculations for each iteration can be considerably simplified due to the shape of the variance structure in Eq. (2) (see Appendix B).

Having obtained an unbiased estimate of $\mathbf{V}$, we can now replace $\mathbf{V}$ with $\hat{\mathbf{V}}$ in Eq. (3) and obtain the WLS estimate $\hat{\boldsymbol{\beta}}_n$. Images that have strongly increased residuals will now be given appropriately less weight.

In the following, we will use empirical data from a typical fMRI experiment with a blocked design to show that the noise estimates strongly vary across the sequence of images and that our method is useful in detecting artifacts. We will provide evidence that the artifacts act in a multiplicative rather than additive fashion and test our assumption that the noise process has a global spatial distribution across an image. Having tested the assumptions of the model, we will then show through Monte Carlo simulations that our correction method is unbiased and results in valid significance values. We then apply the method to the real data and test whether the approach improves the sensitivity of hypothesis testing. Finally, we consider the effects of temporally autocorrelated noise and suggest a way to incorporate temporal autocorrelation into the WLS approach.

Experimental methods

Data acquisition

Data were acquired with a Philips 3.0 T scanner. We used an echoplanar pulse sequence with Sensitivity-Encoded MRI (Pruessmann et al., 1999) and a SENSE-factor of 2. The whole brain was covered in 37 axial slices (3 mm thickness, 0.5 mm gap, TR = 2 s), each of which was acquired with in an 80 × 80 Matrix (FOV was 24.0 × 24.0 cm) and reconstructed to a 128 × 128 image, resulting in a voxel size of 1.9 × 1.9 × 3.5 mm. Each scan consisted of 6 dummy images and 144 images.

Experimental procedure

Participants held an fMRI compatible pneumatic-actuated two-joint robotic arm. Hand position was presented as a cursor on the back projection screen behind the participant and viewed through a mirror. The task of the participant was to move the cursor as quickly as possible to targets that appeared in 2 s intervals at positions 4 cm distant from each other. Periods of 10 movements

(20 s) alternated with periods of rest (14 s), during which the participant positioned the cursor in the center between the two targets. Four different types of movements were tested; the exact nature of these is not of special interest within the scope of this paper. Each scan consisted of 8 movement phases, 2 per condition, and the sequence of conditions was randomized. We ran a total of 8 scans in a single session, resulting in 1152 images per participant. Fifteen healthy participants, age ranging from 21–29, volunteered for the experiment. To minimize head movements, participants used a custom-fitted bite bar that was rigidly attached to the head coil. Experimental procedures were approved by the Johns Hopkins Institutional Review Board.

Data preprocessing and analysis

Analysis was performed using SPM 2 (Friston et al., 1999). Images were realigned to the first volume using a 6-parameter fixed body transformation. The data for each scan were scaled such that the overall mean of all voxels over 144 images was 100. After a possible weighing of the observations, a high-pass filter with a cutoff frequency of 128 TRs was used to remove slowly varying trends. All analyses were performed on spatially unsmoothed data. For the linear model, we used a boxcar function for each movement phase, convolved with the "SPM-canonical" hemodynamic response function (Friston et al., 1999) and an intercept term for each scan. This resulted in 9 regressors per scan. The estimate of the regression parameter $\hat{\boldsymbol{\beta}}$ served as a measure for the amount of signal change caused by that group of 10 movements. Estimates were obtained either using ordinary least squares (OLS) or weighted least squares (WLS) regression. For OLS, we used Eq. (3) with $\mathbf{V}$ set to identity. This regression was also used to compute the residual images (Eq. (5)). For WLS, we computed an unbiased version of the variance scaling parameter s_t for each image t, using ReML (Appendix B). This computation was performed on all voxels that showed a significant omnibus F test with $P = 0.05$. The estimates $\hat{s}_1, \ldots, \hat{s}_T$ were then used to form $\hat{\mathbf{V}}$ for Eq. (3).

Results

Temporal non-stationarity of the noise

We investigated the temporal characteristics of the noise process in our data. An example from a representative participant is shown in Fig. 1A. The variance scaling parameter for most images is low, and the residual images (e.g., Image 873, Fig. 1C) appear to be uniform with little or no spatial structure. However, note the sudden increase in variance for some images, particularly the first image in each scan (indicated by the vertical lines in Fig. 1A). Clearly, the assumption of a stationary noise process with time-homogeneous variance is violated in the current data set.

In Fig. 1A, the first image of each scan has a variance scaling parameter that is approximately 14 times the variance scaling parameter of a typical image. Visual inspection of the residuals from the OLS regression for these first images (e.g., residual image 1009, Fig. 1C) indicates that the artifact consists of wide-spread "ringing" across the image. Importantly, while the artifact was easily detected in the residual images, it could not be readily seen in the raw images (in fact, until this analysis was performed, we had overlooked the artifact). The reason for the artifact is currently unclear. To let the magnetization of the volume reach equilibrium, six dummy scans were included in the EPI acquisition sequence. Inclusion of more dummy scans did not alter the strength of the observed artifact. Within the scope of this paper, we merely wish to make the point that the artifact is easily detected by inspecting the residual image from the linear model. Furthermore, weighted regression would practically exclude such an image from further data analysis. By weighing the image by the inverse of the variance, i.e. by factor 1/14, influence of this image on the analysis would be negligible. However, because this artifact is rather atypical for fMRI data sets generated in other settings, we excluded the first image from all further analyses.

Two brief increases in error variance can be seen in the second half of the scan (Fig. 1A, about image 957 and 966). Each of these phases was accompanied by head motion of the participant. Although every participant used a custom-fitted bite bar, the realignment algorithm estimated movement in the order of 0.5 mm along the z direction (Fig. 1B). The slow drift in the y direction over the period of each scan, resulting in a slow "sinking" of the image, is related to thermal heating of the gradient coils and not actual head motion. Visual inspection of the residual images (Fig. 1C, residual image 966) indicates that the increased residual variance is not due to an incorrect realignment of the image. Incomplete or faulty realignment results in a rim at the outer edge of the brain and around the ventricles, leading to residuals that are positive on one side and negative on the other. Rather, the ventricles (see slice 22) show consistently increased signal. This suggests that the culprit is a spin-history artifact. Movement in the slice selection direction (z) can result in an increase in net magnetization as previously non-excited protons are brought into the slice, causing an increase in signal (Friston et al., 1996).

To investigate whether increases in the error variance of single images are exclusively due to head movement, we plotted the variance scaling parameter for each image against the amount of movement since the last image (Fig. 2). There is generally an increase in variance with increased movement of the head. However, there are also increases in variance that are unaccompanied by head motion. One possible cause is swallowing and movement of the laryngeal muscles. These kinds of movements can directly cause strong artifacts in the fMRI data through changes to the static magnetic field (B0) without causing overt head movements (Birn et al., 1998).

Increased residuals can, not only arise from motion artifacts, but also from true properties of the hemodynamic response that are not captured in the design matrix $\mathbf{X}$. For some participants, this could be seen in slightly increased variance scaling parameters just after the start and end of each task episode. This indicates that the model might not have captured the exact timing of the increase and decrease of the hemodynamic response or that some of the voxels responded transiently to the onset or offset of the task. The inspection of the variance scaling parameters aligned to the task episodes can thus illuminate aspects of the hemodynamic response that are poorly modeled. A solution would be to change the design matrix in such cases to better account for the shape of the true response. This can be achieved by first estimating the hemodynamic response function for each individual and then using this estimate to construct a subject-specific design matrix (Aguirre et al., 1998; Handwerker et al., 2004). Alternatively, one could

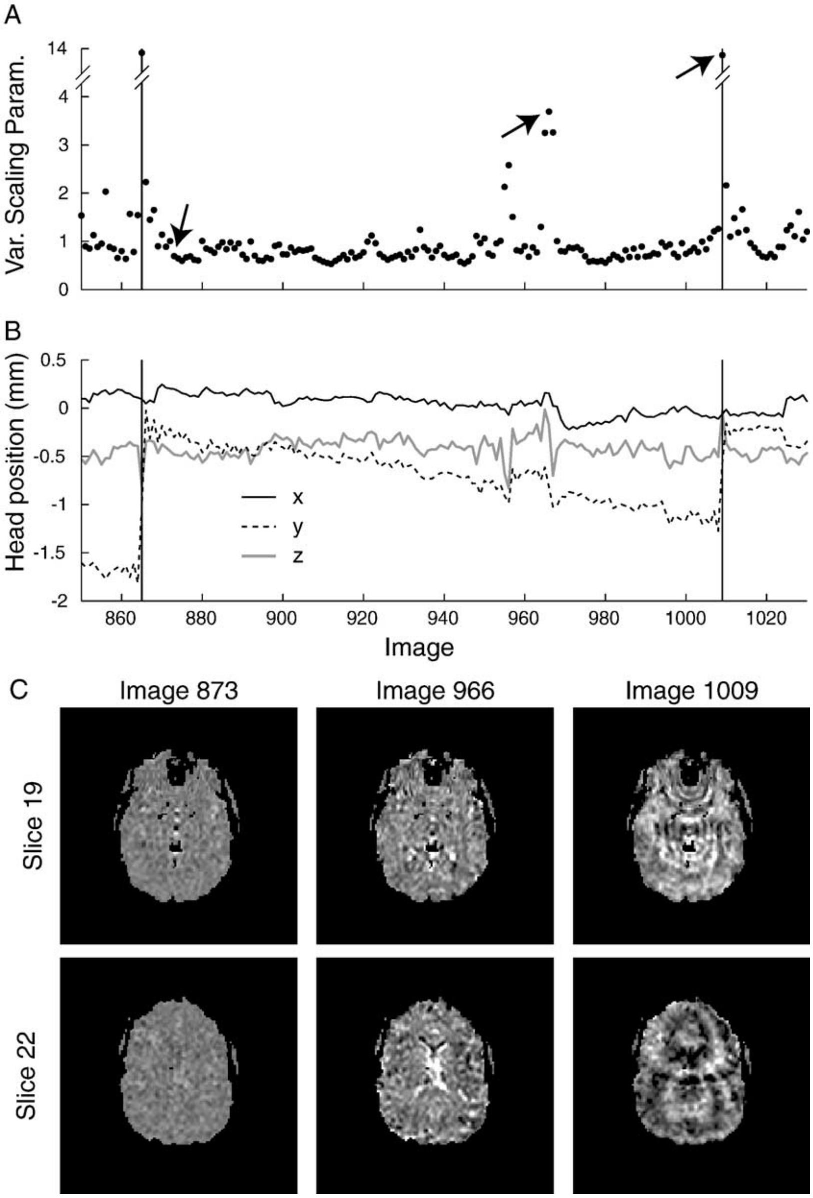

Fig. 1. (A) The variance scaling parameter s_t for images 850–1030. Vertical lines indicate the first image within each scan. There are large increases in variance at the beginning of each scan (vertical lines). There are also increases in variance coincident with head movements. (B) Realignment parameters for translation, relative to first image of the experiment. The drift in the y direction over each scan is related to thermal heating of the gradient coils. (C) Representative slices from residual images from the OLS regression: image 873 (normal residual), image 966 (movement related artifact), and image 1009 (first image in scan). Arrows in panel (A) indicate the variance estimates for the three scans shown here.

add the temporal derivative of the predicted response into the design matrix to absorb unwanted variance (Friston et al., 1998). In contrasts, the WLS method would suppress the influence of poorly modeled aspects of the response on the fit. While this may be less optimal than a change of the design matrix, the overall parameter estimates are not biased by WLS (see Fig. 4B).

Multiplicative vs. additive noise

A critical assumption of the model is that increases in noise variance are multiplicative. This predicts that, in absolute terms, voxels with a high σ_n^2 should show a bigger increase in variance due to the same artifact than voxels with a low σ_n^2. While noise sources such as spin-history artifacts, movement-by-susceptibility interactions, or incomplete correction for head motion may scale with the mean brightness of the voxel, at least some of the noise processes in fMRI are likely to be additive, for example, noise in the receiver coils or amplifiers. Thus, rather than a multiplicative noise model (Eq. (2)), an additive noise model might have been more appropriate:

$$\mathrm{var}(\mathbf{y}_n) = \begin{bmatrix} a_1 + \sigma_n^2 & 0 & \cdots & 0 \\ 0 & a_2 + \sigma_n^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_T + \sigma_n^2 \end{bmatrix} \tag{8}$$

To contrast the predictions of these two models, imagine a data set that includes a subset of images that are particularly

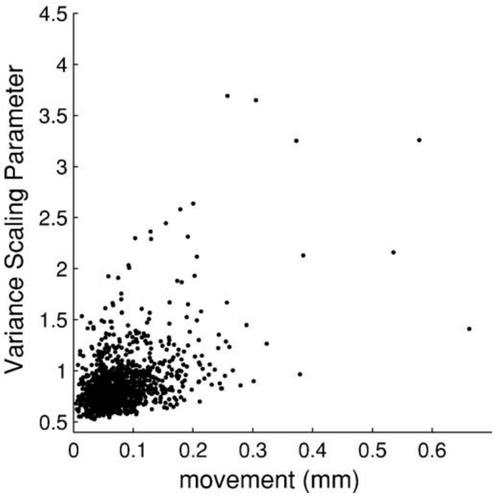

Fig. 2. Relationship between the variance scaling parameter and the estimated movement in millimeters since the last volume. While increased motion generally leads to increased noise, not all artifacts are related to motion of the head.

noisy, with a high image variance parameter (s_H or a_H) and a subset of low-noise images with a lower image variance parameter (s_L or a_L). Further assume that our image consists of K homogenous subsets of voxels, with each subset having a different voxel-specific variance σ_k^2. The variance of data from the voxel subset k for the high-noise images would be $\mathrm{var}(\mathbf{y}_{H,k}) = \sigma_k^2 s_H$ under the multiplicative model (Eq. (2)) and $\mathrm{var}(\mathbf{y}_{H,k}) = \sigma_k^2 + a_H$ under the additive model Eq. (8). Now, critically, the multiplicative model predicts that the increase in variances from low-noise to high-noise images should depend on the voxel-specific variance:

$$\mathrm{var}\left(\mathbf{y}_{H,k}\right) - \mathrm{var}\left(\mathbf{y}_{L,k}\right) = \sigma_k^2 s_H - \sigma_k^2 s_L = \sigma_k^2 (s_H - s_L)$$

In contrast, an additive model would state that the increase in variance from low-noise to high-noise images should be the same for all subset of voxels.

$$\mathrm{var}\left(\mathbf{y}_{H,k}\right) - \mathrm{var}\left(\mathbf{y}_{L,k}\right) = \sigma_k^2 + a_H - \left(\sigma_k^2 + a_L\right) = a_H - a_L$$

To test these predictions, we performed an OLS regression on a representative data set and used the residuals from all 1152 images to compute the variance scaling parameters $\hat{s}$ and the voxel-wise residual-mean-square $\hat{\sigma}_n^2$ Eq. (6).[1] We then selected 81 images with a high estimate of the variance scaling parameter, $\hat{s}_t > 1.5$ (mean = 2.6), and randomly sampled 81 examples from the remaining images (mean of $\hat{s}_t = 0.9$). Based on the voxel-wise mean-squared residual $\hat{\sigma}_n^2$, we then divided the data into 50 subsets of voxels. For each subset k, we calculated the average square residual from the OLS regression for the high-noise images and the low-noise images as an estimate for the noise variances $\mathrm{var}(\mathbf{y}_{H,k})$ and $\mathrm{var}(\mathbf{y}_{L,k})$ (Fig. 3A). If noise artifacts impact image variance additively, the difference between these estimates should be constant across different subsets of voxels. The data show, however, that the variance estimates for high- and low-noise images rise with different slopes, while the intercepts are similar. That is, the difference between the high-noise and low-noise images increases with increasing voxel-specific variance, arguing strongly for a multiplicative noise model.

Spatial distribution of artifacts

The other critical assumption of the proposed method is the spatial uniformity of the change in variance of the noise process. If noise variance was increased occasionally but only in restricted parts of the volume, the proposed method would lead to a loss of efficiency because we would ignore good data in some parts of the volume because other remote areas showed an increased residual variance.

To test the spatial uniformity of the noise process, we estimated the variance scaling parameters $\hat{s}_1, \ldots, \hat{s}_\mathrm{T}$ over the whole brain using ReML (Appendix B). Furthermore, we used the time series of the individual squared residuals from the OLS regression as a rough estimate for the variance scaling parameters in each voxel (because the $n \times n$ matrix $\mathbf{y}_n \mathbf{y}_n^T$ has rank 1, it is not possible to acquire the unbiased ReML estimate on the time series of just one voxel). If the variance of the noise process would indeed fluctuate in a global manner, then the time series of variance scaling

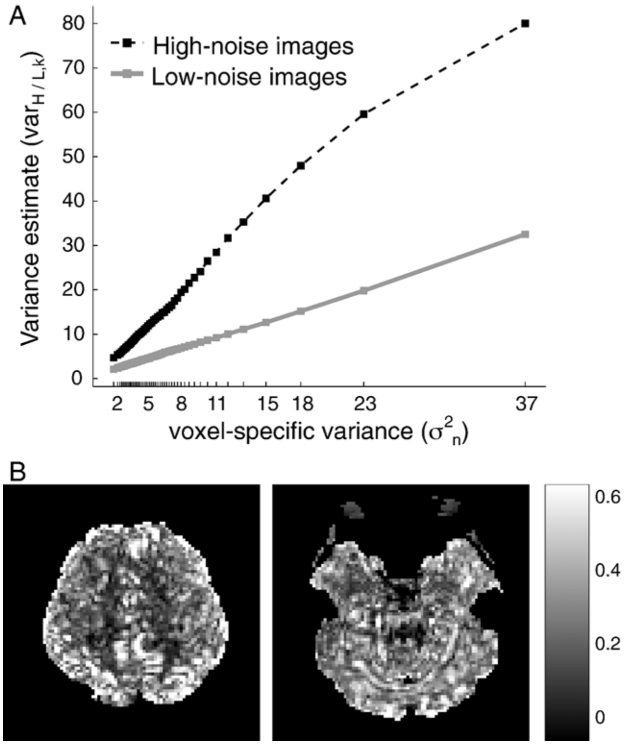

Fig. 3. (A) Multiplicative scaling of noise. The variance estimate for subsets of voxels with increasing voxel-specific variances, separately calculated on images with high noise (black dashed line) and low noise (gray line). (B) Spatial distribution of the noise process. Correlation coefficient of the time series of the squared residuals ($r^2_{n,1},\ldots, r^2_{n,T}$) with the time series of the overall estimated variance scaling parameters ($\hat{s}_1 ,\ldots, \hat{s}_T$).

[1] A possible argument against this approach is that we picked the high-noise images based on an estimate of the variance-scaling parameter following the multiplicative model. We therefore repeated the analysis by selecting the images with the highest mean-squared residual based on the unweighted residuals $\frac{1}{N}\sum_{n=1}^{N}\mathrm{diag}\left(\mathbf{r}_n\mathbf{r}_n^T\right)$. The results of this analysis were nearly identical to the results reported in Fig. 3A.

parameters from each single voxel should correlate substantially with the time series of variance scaling parameter from the whole brain.

To assess the size of the correlation that would be expected if the noise process was global, we conducted a Monte Carlo study (see next section for details). To model the noise-pikes observed in the real data, we increased the standard deviation of 5% of the images by factor 2. In one case, we increased the variance for all voxel on the same images, corresponding to a global noise process. The median correlation between the individual squared residual time series and the variance scaling parameters estimated on all voxels was 0.22. This value depends slightly on the exact assumptions about the noise process and the spatial smoothness of the data, but for realistic parameter settings, remains between 0.15 and 0.3. In the other case, we increased the variance for each voxel independently for different images, corresponding to a noise process that impacts each voxel individually. In this case, the median correlation was close to zero.

We then calculated the correlation between the individual and collective variance scaling parameter time series for each voxel of one exemplary data set. The median correlation was 0.25, and the correlation was slightly higher in the gray matter than in the white matter (Fig. 3B). This may reflect the fact that some of the variance fluctuations in the noise process were caused by physiological factors or the fact that motion-related artifact was more severe at the border of sulci. Most importantly, the correlation was equally high through the gray matter of the whole brain and within the range that we would expect if the noise process was global. In summary, the estimated variance scaling parameters appear to reflect a global noise process that affects the whole brain.

Simulation results—uncorrelated noise

We established that the variance structure of our fMRI data shows spikes for some images resulting from artifacts, that these artifacts mainly act in a multiplicative fashion, and that they have a widespread spatial distribution. We now use a Monte Carlo simulation to show that our suggested method leads to more efficient (lower variance) estimators than OLS and that the fixed effect t values are unbiased. We generated 288 images with 1000 voxels under the Null hypothesis (all $\beta = 0$), using independent and identically distributed noise. For the simulation with noise spikes, we increased the standard deviation of 5% of the images by factor 2. These data were fit with a linear model equivalent to the one used in the behavioral study, albeit only for 2 instead of 8 scans. Thus, there were 16 movement phases in total. Fitting was performed using OLS, WLS with the variance estimates from Eq. (6), and WLS with the ReML estimates as outlined in Appendix B. Standard deviation of $\hat{\boldsymbol{\beta}}$—and the false-detection rates α were computed across all 1000 voxels. The process was repeated 40 times.

Low variances of the regression weights are desirable because these estimators typically are submitted to a second-level analysis with subjects as a random factor (Searle et al., 1992). If our guess about the variance structure is closer to the truth than the assumption of independent and identically distributed (i.i.d.) noise made by the OLS analysis, then the variance of $\hat{\boldsymbol{\beta}}_{\mathrm{WLS}}$ should be lower than of $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$, and our second-level inference should become more sensitive.

Unbiased fixed-effect t values become important if we wish to make inferences on the single subject level while ignoring possible true variability in the βs from repetition to repetition. Under this assumption, the variance–covariance matrix of the parameter estimates $\hat{\boldsymbol{\beta}}$ can be estimated from the residuals of the regression (Friston et al., 1995).

$$\mathrm{var}\left(\hat{\boldsymbol{\beta}}_n\right) = \left(\mathbf{X}^T\hat{\mathbf{V}}^{-1}\mathbf{X}\right)^{-1}\hat{\sigma}_n^2$$

$$\hat{\sigma}_n^2 = \frac{\mathbf{r}_n^T\mathbf{r}_n}{T - \mathrm{rank}(\mathbf{X})}$$

$$t_{\mathrm{fixed}} = \frac{\boldsymbol{c}^T\hat{\boldsymbol{\beta}}_n}{\sqrt{\boldsymbol{c}^T\mathrm{var}\left(\hat{\boldsymbol{\beta}}_n\right)\boldsymbol{c}}} \tag{9}$$

The vector $\boldsymbol{c}$ denotes a linear combination of the $\hat{\boldsymbol{\beta}}$s.

The standard deviation of the resulting $\hat{\boldsymbol{\beta}}$s and the average (false) rejection rates of the Null hypothesis are shown in Table 1. When the data are free of noise spikes (i.i.d.), the three estimation methods result in comparable variances for the regression parameters. The first WLS, however, has a false-rejection rate of 5.5%, while valid t values should always lead to a false-rejection rate that is equal to the theoretically chosen α (5%). This is due to a bias in the estimation of the variance structure (Worsley and Friston, 1995) and can be avoided by using the ReML estimation for the variance scaling parameters (third column).

For data containing noise spikes, we separated the regressors into two groups, those that have 2 noise spikes (high noise) and those who have no noise spikes (low noise) during their task period (results for regressors with one noise spike fall in between these extremes). The standard deviation of the $\hat{\boldsymbol{\beta}}$s is substantially lower for both WLS methods than for OLS. The false-rejection rate resulting from OLS regression is too high for high-noise regressors, while it is too low for low-noise regressors. When the variance parameters are estimated using ReML, the false-rejection rates are again close to the theoretically desired level. In summary, both suggested WLS methods lead to consistently higher estimation efficiency on noisy

Table 1
Standard deviation (SD) and false-rejection rates (α) on data simulated without temporal covariance structure

Measure	OLS	WLS (biased)	WLS (ReML)
i.i.d.			
SD ($\hat{\boldsymbol{\beta}}$)	0.310	0.311	0.311
α (%)	5.05	5.53	5.10
i.i.d. + Noise spikes			
SD ($\hat{\boldsymbol{\beta}}$)—high	0.384	0.334	0.334
SD ($\hat{\boldsymbol{\beta}}$)—low	0.309	0.309	0.309
α—high(%)	7.85	5.19	5.08
α—low(%)	3.98	5.59	5.05

Column indicates the method of fitting. OLS: Ordinary least squares, WLS: Weighted least squares using either the biased variance estimates from Eq. (6) or using the restricted maximum likelihood (ReML) estimate. The simulation with noise spikes is evaluated depending on whether two (high noise) or no (low noise) noise spikes occurred during the respective task phase. False-rejection rates are the percentage of voxels that exceeded the critical t value for $P = 0.05$ (uncorrected).

data. However, only the estimation of the variance parameters using ReML results in valid fixed-effects t values.

Estimation efficiency on real data

We applied the ReML method to estimate the variance parameters from experimental data. Because the artifact on the first image may not be typical for other fMRI studies, we excluded this image from analysis. The general linear model of Eq. (3) was computed twice for each data set. In the first run, we performed ordinary least squares (OLS) regression with $\mathbf{V}$ set to the identity matrix. We then computed from the residuals the estimates of the variance scaling parameters $s_1 \ldots s_t$ according to the ReML method (see Appendix B). In a second step, we used this estimate $\hat{\mathbf{V}}$ to re-compute the regression parameters Eq. (3), producing a weighted least squares (WLS) estimate $\hat{\boldsymbol{\beta}}$.

For each of the four conditions, there are 16 repetitions in the experiment, and we use the notation $\hat{\boldsymbol{\beta}}_{n,i,j}$ to denote the resulting regression parameter estimate (nth voxel, ith condition, jth repetition). Due to this design, we gained information, not only about the mean signal change in that condition, but also about the variance of $\hat{\boldsymbol{\beta}}$ for a specific voxel and condition. Thus, we can compare the resulting variance of the $\hat{\boldsymbol{\beta}}$s and the resulting t values following either a mixed-effects or fixed-effects model difference between OLS and WLS regression.

To compare the change in these statistics in homogenous subpopulation of voxels, we sorted the voxels in bins based on their involvement in the corresponding condition. We inspected the upper 20th percentile of the most significantly activated voxels, assuming that these are the voxels of main interest to the investigator. We based this binning of the data on the average t values from OLS and WLS to avoid a selection bias in favor of one of the methods.

Fig. 4A shows the average percent change of the variance of $\hat{\boldsymbol{\beta}}$ for the 20% most activated voxels. The variance of the parameter estimates is 1.5–3% lower using WLS than using OLS. Because some of the variance of $\hat{\boldsymbol{\beta}}$ for both methods will reflect the true variance from repetition to repetition of the task, the achieved change in variance constitutes a lower limit for the improvement in estimation efficiency. While the variance of the $\hat{\boldsymbol{\beta}}$ coefficients changed between methods, the mean did not change in any systematic direction for any of the bins (Fig. 4B).

Impact on the inference in mixed- and fixed-effects models

We next considered the effect of the higher efficiency of the WLS estimator onto inference in the context of a within-subject mixed-effects model (Searle et al., 1992). In the mixed-effects analysis, we consider the $\boldsymbol{\beta}$ of each repetition to be sampled from a (theoretical) population of possible values. Thus, the variance of the estimates for condition i and voxel n should be equal to sum of the variance of the “true” $\boldsymbol{\beta}$ and the variance that arises from our error in estimating the true $\boldsymbol{\beta}$.

$$\mathrm{var}\left(\hat{\boldsymbol{\beta}}_{n,i,\cdot}\right) = \mathrm{var}\left(\boldsymbol{\beta}_{n,i,\cdot}\right) + \mathrm{var}\left(\hat{\boldsymbol{\beta}}_{n,i,\cdot} - \boldsymbol{\beta}_{n,i,\cdot}\right) \tag{10}$$

McGonigle et al. (2000) have shown that the variance of parameter estimates between scans or sessions can be substantial. Thus, a mixed-effects model with repetition as a random factor would be appropriate.

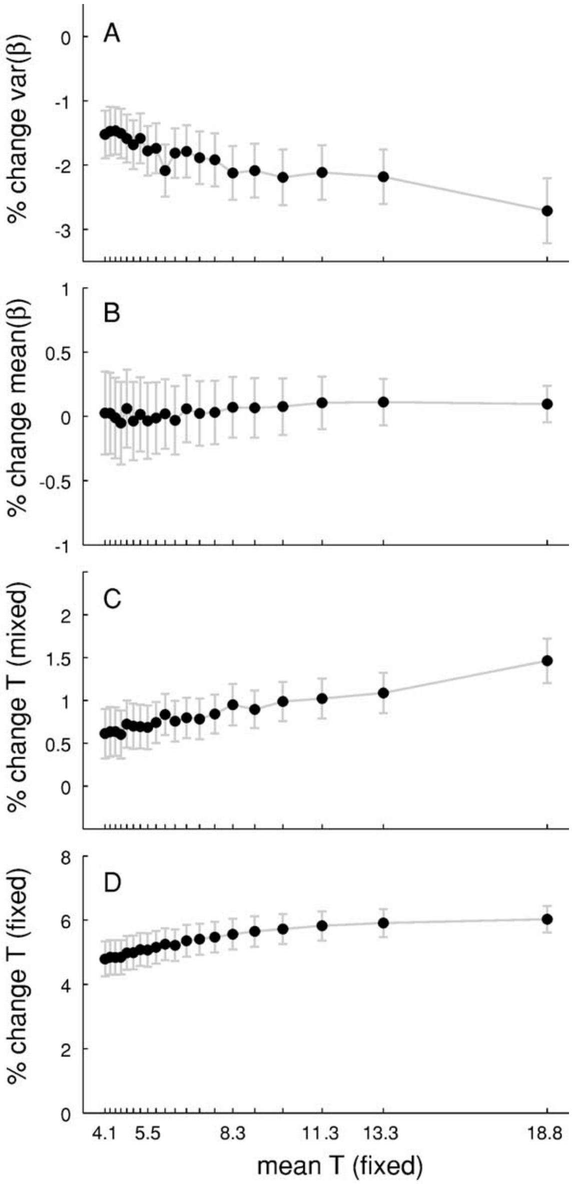

Fig. 4. Percentage change in the statistics from OLS to WLS: (WLS − OLS) / OLS * 100. Results are shown separately for the upper 20% of the most significantly activated voxels. The x axis shows the mean t value (averaged over OLS and WLS) for that percentile. (A) Change in the variance of $\hat{\boldsymbol{\beta}}$, estimated from the 16 occurrences of that condition across the experiment. (B) Change in the mean of $\hat{\boldsymbol{\beta}}$. (C) Change in the t value of a mixed-effects model. (D) Change in the t value of a fixed-effect model.

$$t_{\text{mixed}} = \frac{\sum \hat{\boldsymbol{\beta}}_{n,k,\cdot}/16}{\sqrt{\mathrm{var}\left(\boldsymbol{\beta}_{n,k,\cdot}\right)/16}} \tag{11}$$

The t values based on the WLS method are 0.5–1.5% higher than the OLS method (Fig. 4C). While this increase may seem modest, the increase in task-related volume is more substantial. If the statistical map is thresholded at a P value of 0.0001 (uncorrected), the OLS method shows 8.1% of the volume as being above threshold, while the WLS method indicates that 8.3% of the volume is above threshold. We might expect a similar gain for a between-subject mixed-effects model. In this case, however, functional and anatomical between-subject variability additionally increases the variance of the parameter estimates.

We also calculated the fixed-effects t value for each voxel and condition for both the OLS and the WLS method. The resulting t values are 4.5–6% higher for all bins in the WLS vs. OLS analysis (Fig. 4D). If we again threshold the volumes at P = 0.0001, we find a super-threshold volume of 20.3% of the total volume for the OLS analysis and 21.4% for the WLS analysis.

In summary, our method leads to higher estimation efficiencies and higher detection power in a mixed-effects analysis and to an even larger degree in a fixed-effects analysis. Because we excluded the artifactual first image from the analysis and because our subjects used a bite bar to avoid head movements, we believe that the observed improvement may be quite typical for fMRI studies and will be even more substantial if the scans include more artifacts.

Temporally autocorrelated noise

The above analysis assumed temporally uncorrelated noise. However, it is known that fMRI time series typically shows temporal autocorrelation, even after a high-pass filter has been applied to remove slow varying trends (Woolrich et al., 2001; Worsley et al., 1996; Zarahn et al., 1997). Ignoring this autocorrelation can lead to lower estimation efficiency and impact the validity of fixed-effects t values. To address this issue in the frame work of the suggested WLS method, we combined the variance model in Eq. (2) with a term that simulates a temporal autocorrelation of a fixed shape. In particular, we added a temporal autocorrelation term as it would arise from an autoregressive noise process with a regression coefficient of $a = 0.2$, as is the usual approach in SPM2. The weights of the diagonal terms ($s_1,\ldots, s_T$) and the weight of the autoregressive term (s_{T+1}) can again be estimated using ReML (see Appendix C). A covariance model that is composed of an autoregressive term plus uncorrelated white noise has been suggested by Purdon and Weisskoff (1998). The novel contribution here is that the diagonal elements of the covariance matrix are allowed to differ such that varying levels of noise for different images can be captured.

We used a Monte Carlo simulation to compare this method to the “standard” algorithm in SPM2, which uses ReML to estimate the weights for the temporal autocorrelation for a fixed autoregressive process, as well as for the derivative of the variance matrix in respect to the autoregressive coefficient a. We performed a simulation identical to the one described in the last simulation section, except that we used noise generated by an autoregressive process with a = 0.2. The data were fit with OLS, WLS (ReML), and two ReML methods that take into account temporal autocorrelation. The first was the standard procedure in SPM2, which estimates the contribution of two variance components: the variance structure of an autoregressive process with a = 0.2 and the derivative of that structure in respect to a. The second method was a combination of WLS and the autoregressive model as outlined in Appendix C.

For data without noise spikes, all algorithms lead to comparable estimation efficiency (Table 2). In contrast, t values are valid only for the two models that take into account the temporal autocorrelation structure. Methods that assume independent noise show false-rejection rates of higher than 5% (see also Friston et al., 1995; Purdon and Weisskoff, 1998). When noise spikes are added to the data, the standard deviation of the $\hat{\beta}$s is lower for both WLS methods (independent whether these account for temporal autocorrelation or not) than for the two methods that assume stationary variances. But only the method that combines WLS and temporal autocorrelation leads to valid t values under this condition.

Table 2
Standard deviation (SD) and false-detection rates (α) for data simulated with an autoregressive temporal covariance structure (order 1, coefficient = 0.2)

Measure	OLS	WLS	AR	WLS + AR
AR (1)				
SD ($\hat{\beta}$)	0.376	0.376	0.375	0.376
α (%)	9.13	9.45	5.05	5.06
AR (1) + Noise spikes				
SD ($\hat{\beta}$)—high	0.441	0.398	0.441	0.395
SD ($\hat{\beta}$)—low	0.375	0.376	0.375	0.375
α—high(%)	11.15	9.10	7.40	5.23
α—low(%)	7.65	9.44	4.55	5.08

Column indicates the method used for fitting. OLS: Ordinary least squares, WLS: Weighted least squares using restricted maximum likelihood (ReML) estimate. AR: Autoregressive model, as used in SPM2. WLS + AR: combination of weighted least squares and autoregressive model. The simulation with noise spikes is evaluated depending on whether two (high noise) or no (low noise) noise spikes occurred during the respective task phase.

Discussion

In model fitting, it is always a good idea to inspect the residuals. This is non-trivial endeavor in fMRI studies due to the sheer amount of data. While many analysis packages (e.g., SPM2, AFNI) only provide the residual-mean-squares image averaged over all time points, tools for the graphical inspection of residuals have been developed (Luo and Nichols, 2003). In this spirit, we propose here to inspect the time course of the squared residuals averaged across the brain and then to examine the individual residual images that show increased variance. In the present study, this strategy led to the detection of artifacts that corrupted the fMRI data on the first images of each scan, a fact that was not apparent from the inspection of the raw data files. Furthermore, movement-related artifacts could easily be detected. Modified SPM2 Matlab routines that allow for the extraction of the temporal statistics of the residuals and the implementation of WLS using ReML can be downloaded from http://www.bme.jhu/~reza/imaging/SPMj.html.

Ideally, one would use these techniques to identify unwanted sources of noise and consequently eliminate them. However, some artifacts are only detected after the data have been completely collected. Others cannot be avoided because they are caused by overt eye, mouth, or arm movements that are essential parts of the experimental paradigm. While specific methods have been suggested to correct for artifacts arising from head movements (Andersson et al., 2001; Friston et al., 1996; Grootoonk et al., 2000), we proposed here a general technique that allows the detection and correction of artifacts independent of origin and form. We showed that most noise-generating artifacts are quite spatially extended (see also, Frank et al., 2001), justifying the estimation of a variance scaling parameter for each image from the residual sums-of-squares across the volume of interest. These estimates can then be used to weigh each observation inversely to the estimated variance of that image, resulting in optimal estimation efficiency.

The validity of *t* values in a fixed-effects analysis depends on the accuracy of the estimated covariance structure of the errors (Friston et al., 2000). Our Monte Carlo simulations showed that, if the real covariance structure differs from the assumed structure, large biases in the significance values can occur. We therefore used ReML estimation to arrive at reasonably accurate estimates of the variance components. We derived a computationally feasible algorithm to implement this idea both for the assumption of independent and autocorrelated noise.

Applied to the current data set, the resulting estimators showed a higher efficiency than the traditional OLS estimators. Using a mixed-effects model, the gain was a 0.5–1.5% increase of the *t* value for voxels of interest, which translated to a 2.5% increase in the super-threshold volume, using a statistical threshold in the typical range for imaging studies. For a fixed-effect model, the gains were substantially higher, indicating that the noise variance is severely overestimated when noisy images are fully included into the analysis.

Because our study was conducted with a bite bar and with trained and healthy cooperative volunteers, we believe that our data had comparably few artifacts. However, in studies with children or special populations, we expect our procedure to result in higher improvements in estimation. In summary, the suggested method significantly improves the ability to draw sensitive and valid inferences from potentially noisy fMRI data.

Acknowledgments

The work was supported through grants from the NIH (NS37422) and the Human Frontiers Science Program. We are grateful to the staff of the F.M. Kirby Research Center for Functional Brain Imaging at Kennedy Krieger Institute for facilitating this study. This Center is funded by an NIH/NCRR resource grant (RR15241). We thank Matthew Brett, James J. Pekar, John v. Horn, Yasmin L. Hashamboy, and two anonymous reviewers for helpful comments.

Appendix A. Expected sum of squares of the residuals

$$\begin{aligned} E\{\mathbf{r}_n\mathbf{r}_n^T\} &= E\{\mathbf{R}\mathbf{y}_n\mathbf{y}_n^T\mathbf{R}^T\} \\ &= E\{\mathbf{R}(\mathbf{X}\beta+\varepsilon_n)(\mathbf{X}\beta+\varepsilon_n)^T\mathbf{R}^T\} \\ &= E\{\mathbf{R}\mathbf{X}\beta\beta^T\mathbf{X}^T\mathbf{R}^T\} + E\{\mathbf{R}\varepsilon_n\varepsilon_n^T\mathbf{R}^T\} \end{aligned}$$

$$\begin{aligned} \mathbf{R}\mathbf{X}\beta\beta^T\mathbf{X}^T\mathbf{R}^T &= \left(\mathbf{I}-\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\right)\mathbf{X}\beta\beta^T\mathbf{X}^T \\ &\quad \times \left(\mathbf{I}-\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\right) = 0 \end{aligned}$$

$$\begin{aligned} E\{\mathbf{r}_n\mathbf{r}_n^T\} &= E\{\mathbf{R}\varepsilon_n\varepsilon_n^T\mathbf{R}^T\} = \mathbf{R}\left(\mathrm{var}\{\varepsilon_n\} + E\{\varepsilon\}E\{\varepsilon^T\}\right)\mathbf{R}^T \\ &= \mathbf{RVR}\sigma_n \end{aligned} \tag{A.1}$$

Appendix B. Restricted maximum likelihood estimation

Assume a linear model as in Eq. (1), with a model of the noise covariance structure denoted by the matrix $\mathbf{V}(\boldsymbol{s})$, a function of a vector of parameters $\boldsymbol{s}$**.** Unbiased estimators of $\boldsymbol{s}$ can be obtained by maximizing the restricted log-likelihood of $\boldsymbol{s}$ (Speed, 1997). Given that $\mathbf{y}{\sim}N$ $(\mathbf{X}\beta, \mathbf{V}(\boldsymbol{s}))$, we have:

$$l_r(s|\mathbf{y}) = -\frac{1}{2}\left(\ln|\mathbf{V}(\boldsymbol{s})| + \ln|\mathbf{X}^T\mathbf{V}(\boldsymbol{s})^{-1}\mathbf{X}| + \mathbf{y}^T\mathbf{R}\mathbf{V}(\boldsymbol{s})^{-1}\mathbf{R}\mathbf{y}\right) + \text{const} \tag{B.1}$$

This likelihood can be maximized using an iterative method. We start with an arbitrary guess $\boldsymbol{s}^{(1)}$ for the variance parameters. For each iteration u, we compute the variance–covariance matrix **V**, following Eq. (2), and the corresponding residual-forming matrix **R**.

$$\mathbf{V}^{(u)} = \mathbf{V}\left(\boldsymbol{s}^{(u)}\right)$$

$$\mathbf{R}^{(u)} = \left(\mathbf{I}-\mathbf{X}\left(\mathbf{X}^T\mathbf{V}^{(u)-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{V}^{(u)-1}\right) \tag{B.2}$$

Using the derivation described by Speed (1997), we now compute the first derivative of the restricted log-likelihood with respect to the parameters s_i. Using the following identity:

$$\frac{\mathrm{d}\ln|\mathbf{V}(\boldsymbol{s})|}{\mathrm{d}s_i} = \mathrm{tr}\left(\mathbf{V}^{-1}\frac{\mathrm{d}\mathbf{V}}{\mathrm{d}s_i}\right)$$

where tr represents the trace operator, we arrive at:

$$\begin{aligned} \frac{dl_r}{\mathrm{d}s_i} &= -\frac{1}{2}\left(\mathrm{tr}\left(\mathbf{V}^{-1}\frac{\mathrm{d}\mathbf{V}}{\mathrm{d}s_i}\right) + \mathrm{tr}\left((\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\frac{\mathrm{d}\mathbf{V}^{-1}}{\mathrm{d}s_i}\mathbf{X}\right)\right. \\ &\qquad \left. + \; \mathbf{y}^T\mathbf{R}\frac{\mathrm{d}\mathbf{V}^{-1}}{\mathrm{d}s_i}\mathbf{R}\mathbf{y}\right) \end{aligned}$$

Using the following two identities:

$$\frac{\mathrm{d}(\mathbf{V}(\boldsymbol{s}))^{-1}}{\mathrm{d}s_i} = -\mathbf{V}^{-1}\frac{\mathrm{d}\mathbf{V}}{\mathrm{d}s_i}\mathbf{V}^{-1}$$

$$\mathrm{tr(ABC)} = \mathrm{tr(CAB)}$$

we simplify the derivative:

$$\begin{aligned} \frac{\mathrm{d}l_r}{\mathrm{d}s_i} &= -\frac{1}{2}\left(\mathrm{tr}\left(\frac{\mathrm{d}\mathbf{V}}{\mathrm{d}s_i}\mathbf{V}^{-1}\right)\right. \\ &\qquad -\mathrm{tr}\left(\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\frac{\mathrm{d}\mathbf{V}}{\mathrm{d}s_i}\mathbf{V}^{-1}\right) \\ &\qquad \left. -\mathrm{tr}\left(\mathbf{y}^T\mathbf{R}\mathbf{V}^{-1}\frac{\mathrm{d}\mathbf{V}}{\mathrm{d}s_i}\mathbf{V}^{-1}\mathbf{R}\mathbf{y}\right)\right) \\ &= -\frac{1}{2}\mathrm{tr}\left(\frac{\mathrm{d}\mathbf{V}}{\mathrm{d}s_i}\mathbf{V}^{-1}\left(\mathbf{I}-\mathbf{X}(\mathbf{X}^{\mathbf{T}}\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{V}^{-1}\right)\right. \\ &\qquad \left. -\,\mathbf{y}\mathbf{y}^T\mathbf{R}\mathbf{V}^{-1}\frac{\mathrm{d}\mathbf{V}}{\mathrm{d}s_i}\mathbf{V}^{-1}\mathbf{R}\right) \\ &= -\frac{1}{2}\mathrm{tr}\left(\frac{\mathrm{d}\mathbf{V}}{\mathrm{d}s_i}\mathbf{V}^{-1}\mathbf{R} - \mathbf{y}\mathbf{y}^T\mathbf{R}\mathbf{V}^{-1}\frac{\mathrm{d}\mathbf{V}}{\mathrm{d}s_i}\mathbf{V}^{-1}\mathbf{R}\right) \end{aligned}$$

Therefore, on iteration u, we have a vector $\mathbf{g}$ that represents the derivative of the log-likelihood:

$$\mathbf{g}_i^{(u)} = \frac{\partial l_r}{\partial s_i} = -\frac{1}{2}\mathrm{tr}\left(\frac{\partial \mathbf{V}(\mathbf{s})}{\partial s_i}\mathbf{V}^{(u)-1}\mathbf{R}^{(u)} - \mathbf{y}\mathbf{y}^T\mathbf{V}^{(u)-1}\mathbf{R}^{(\mathbf{u})}\frac{\partial \mathbf{V}(\mathbf{s})}{\partial s_i}\mathbf{V}^{(u)-1}\mathbf{R}^{(u)}\right) \tag{B.3}$$

Next, we compute the negative expected second derivative of the log-likelihood with respect to the parameters, the Fisher-information matrix $\mathbf{F}$:

$$\begin{aligned}\mathbf{F}_{i,j}^{(u)} &= -\left\langle \frac{\partial^2 l_r}{\partial s_i \partial s_j} \right\rangle \\ &= \frac{1}{2}\mathrm{tr}\left(\mathbf{V}^{(u)-1}\mathbf{R}^{(u)}\frac{\partial \mathbf{V}(\boldsymbol{s})}{\partial s_i}\mathbf{V}^{(u)-1}\mathbf{R}^{(u)}\frac{\partial \mathbf{V}(\boldsymbol{s})}{\partial s_j}\right)\end{aligned} \tag{B.4}$$

We update the guess of $\boldsymbol{s}$ using the Newton–Raphson method

$$\boldsymbol{s}^{(u+1)} = \boldsymbol{s}^{(u)} + \mathbf{F}^{(u)-1}\mathbf{g}^{(u)} \tag{B.5}$$

and iterate Eqs. (B.2–B.5) until convergence.

Calculating the Gradient and Fisher-scoring matrix can be computationally expensive, especially if each image has a separate parameter. However, the derivatives $\frac{\partial \mathbf{V}(\boldsymbol{s})}{\partial s_i}$ are all sparse $T \times T$ matrices with the only non-zero element being a one at the ith row and ith column. We define $\mathbf{P}^{(u)} = \mathbf{V}^{(u)-1}\mathbf{R}^{(u)}$ and introduce the notation $\mathbf{P}_{i,*}$ for the ith row and $\mathbf{P}_{*,j}$ for the j^{th} column of $\mathbf{P}$ and thus can simplify the computation of the gradient:

$$\begin{aligned}\mathbf{g}_i &= -\frac{1}{2}\mathbf{P}_{i,i} + \frac{1}{2}\mathrm{trace}\left(\mathbf{y}\mathbf{y}^T\mathbf{P}_{*,i}\mathbf{P}_{i,*}\right) \\ &= -\frac{1}{2}\mathbf{P}_{i,i} + \frac{1}{2}\left(\mathbf{P}_{i,*}\mathbf{y}\mathbf{y}^T\mathbf{P}_{*,i}\right)\mathbf{g} \\ &= -\frac{1}{2}\mathrm{diag}(\mathbf{P}) + \frac{1}{2}\mathrm{diag}\left(\mathbf{P}\mathbf{y}\mathbf{y}^T\mathbf{P}\right)\end{aligned} \tag{B.6}$$

as well as the computation of the Fisher-information matrix:

$$\begin{aligned}\mathbf{F}_{i,j} &= \tfrac{1}{2}\mathbf{P}_{i,j}\mathbf{P}_{j,i} \\ \mathbf{F} &= \tfrac{1}{2}\mathbf{P}\bigcirc\mathbf{P}^T\end{aligned} \tag{B.7}$$

where $\bigcirc$ is the element-by-element multiplication of two matrices. When using Eqs. (B.6) and (B.7) in the iterative algorithm outlined above, the iteration can be computed rapidly even for fairly long time series of data.

To compute the estimate for the variance parameters for a set of N voxels, we replace the term $\mathbf{y}\mathbf{y}^T$ in Eq. (B.6) with the weighted squared data over those voxels:

$$\frac{1}{N}\sum_{n=1}^{N}\mathbf{y}_n\mathbf{y}_n^T/\hat{\sigma}_n^2 \tag{B.8}$$

It is important to note that ReML will not converge in our case when the above matrix Eq. (B.8) is ill-conditioned, i.e. close to singular. To prevent this, it is necessary to ensure that enough independent voxel time series are averaged in Eq. (B.8). Therefore, it is advisable to use this method on spatially unsmoothed data.

Appendix C. Inclusion of autocorrelation terms in the variance

For the inclusion of an autocorrelation term, we augment the vector $\boldsymbol{s}$ by additional elements that stand for the covariance structure arising from an autoregressive process.

$$\begin{aligned}\mathrm{var}(\varepsilon_n) &= \begin{bmatrix} s_1 & 0 & 0 & \cdots \\ 0 & s_2 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & s_T \end{bmatrix}\sigma_n^2 \\ &\quad + s_{T+1}\begin{bmatrix} 1 & a & a^2 & \cdots \\ a & 1 & a & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ a^{T-1} & a^{T-2} & \cdots & 1 \end{bmatrix}\sigma_n^2 \\ &= \left(\mathrm{diag}^{-1}(s_{1\ldots T}) + s_{T+1}\mathbf{A}\right)\sigma_n^2\end{aligned} \tag{C.1}$$

where diag^{-1} takes a column vector and transfers it into the diagonal of a square matrix. We can now use ReML to estimate $\boldsymbol{s}$. We can use Eqs. (B.6) and (B.7) to compute the 1^{st}–T^{th} element of $\mathbf{G}$ and 1st–Tth row of $\mathbf{F}$ and invoke Eqs. (B.3) and (B.4) to compute element $T + 1$ of $\mathbf{G}$ and row $T + 1$ of $\mathbf{F}$ by noticing that $\frac{\partial \mathbf{V}(\boldsymbol{s})}{\partial s_{T+1}} = \mathbf{A}$.

References

Aguirre, G.K., Zarahn, E., D'Esposito, M., 1997. Empirical analyses of BOLD fMRI statistics: II. Spatially smoothed data collected under null-hypothesis and experimental conditions. NeuroImage 5, 199–212.

Aguirre, G.K., Zarahn, E., D'Esposito, M., 1998. The variability of human, BOLD hemodynamic responses. NeuroImage 8, 360–369.

Andersson, J.L., Hutton, C., Ashburner, J., Turner, R., Friston, K., 2001. Modeling geometric deformations in EPI time series. NeuroImage 13, 903–919.

Beauchamp, M.S., 2003. Detection of eye movements from fMRI data. Magn. Reson. Med. 49, 376–380.

Birn, R.M., Bandettini, P.A., Cox, R.W., Jesmanowicz, A., Shaker, R., 1998. Magnetic field changes in the human brain due to swallowing or speaking. Magn. Reson. Med. 40, 55–60.

Birn, R.M., Bandettini, P.A., Cox, R.W., Shaker, R., 1999. Event-related fMRI of tasks involving brief motion. Hum. Brain Mapp. 7, 106–114.

Frank, L.R., Buxton, R.B., Wong, E.C., 2001. Estimation of respiration-induced noise fluctuations from undersampled multislice fMRI data. Magn. Reson. Med. 45, 635–644.

Friston, K.J., Holmes, A.P., Poline, J.B., Grasby, P.J., Williams, S.C., Frackowiak, R.S., Turner, R., 1995. Analysis of fMRI time-series revisited. NeuroImage 2, 45–53.

Friston, K.J., Williams, S., Howard, R., Frackowiak, R.S., Turner, R., 1996. Movement-related effects in fMRI time-series. Magn. Reson. Med. 35, 346–355.

Friston, K.J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M.D., Turner, R., 1998. Event-related fMRI: characterizing differential responses. NeuroImage 7, 30–40.

Friston, K., Holmes, A.P., Ashburner, J., 1999. Statistical Parameter Mapping (SPM).

Friston, K.J., Josephs, O., Zarahn, E., Holmes, A.P., Rouquette, S., Poline, J., 2000. To smooth or not to smooth? Bias and efficiency in fMRI time-series analysis. NeuroImage 12, 196–208.

Grootoonk, S., Hutton, C., Ashburner, J., Howseman, A.M., Josephs, O., Rees, G., Friston, K.J., Turner, R., 2000. Characterization and correction of interpolation effects in the realignment of fMRI time series. NeuroImage 11, 49–57.

Handwerker, D.A., Ollinger, J.M., D'Esposito, M., 2004. Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. NeuroImage 21, 1639–1651.

Kruger, G., Glover, G.H., 2001. Physiological noise in oxygenation-sensitive magnetic resonance imaging. Magn. Reson. Med. 46, 631–637.

Luo, W.L., Nichols, T.E., 2003. Diagnosis and exploration of massively univariate neuroimaging models. NeuroImage 19, 1014–1032.

McGonigle, D.J., Howseman, A.M., Athwal, B.S., Friston, K.J., Frackowiak, R.S., Holmes, A.P., 2000. Variability in fMRI: an examination of intersession differences. NeuroImage 11, 708–734.

Pruessmann, K.P., Weiger, M., Scheidegger, M.B., Boesiger, P., 1999. SENSE: sensitivity encoding for fast MRI. Magn. Reson. Med. 42, 952–962.

Purdon, P.L., Weisskoff, R.M., 1998. Effect of temporal autocorrelation due to physiological noise and stimulus paradigm on voxel-level false-positive rates in fMRI. Hum. Brain Mapp. 6, 239–249.

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

Speed, T.P., 1997. Restricted maximum likelihood (ReML). In: Kotz, S., et al. (Eds.), Encyclopedia of Statistical Sciences. Wiley-Interscience, New York, pp. 472–481.

Woolrich, M.W., Ripley, B.D., Brady, M., Smith, S.M., 2001. Temporal autocorrelation in univariate linear modeling of FMRI data. NeuroImage 14, 1370–1386.

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

Worsley, K.J., Marrett, S., Neelin, P., Vandal, A.C., Friston, K.J., Evans, A.C., 1996. A unified statistical approach for determining significant voxels in images of cerebral activation. Hum. Brain Mapp. 12, 900–918.

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

Wu, D.H., Lewin, J.S., Duerk, J.L., 1997. Inadequacy of motion correction algorithms in functional MRI: role of susceptibility-induced artifacts. J. Magn. Reson. Imaging 7, 365–370.

Zarahn, E., Aguirre, G.K., D'Esposito, M., 1997. Empirical analyses of BOLD fMRI statistics. I. Spatially unsmoothed data collected under null-hypothesis conditions. NeuroImage 5, 179–197.