A Global Estimator Unbiased by Local Changes

Jesper L. R. Andersson,[1] John Ashburner, and Karl Friston

The Wellcome Department of Cognitive Neurology, London, United Kingdom

The global activity is an important confound when analyzing PET data in that its inclusion in the statistical model can substantially reduce error variance and increase sensitivity. However, by defining global activity as the average over all voxels one introduces a bias that is collinear with experimental factors. This leads to an underestimation of true activations and the introduction of artefactual deactivations. We propose a novel estimator for the global activity based on the notion of finding a maximally nonlocal mode in a multivariate characterization of the data, while maximizing the locality of the remaining modes. The approach uses singular value decomposition (SVD) to find a provisional set of modes, which are subsequently rotated such that a metric based on the above heuristic is maximized. This metric is a version of the stochastic sign change (*SSC*) criterion that has been used previously for normalizing medical images with focal defects. The estimator was evaluated on simulated and real functional imaging (PET) data. The simulations show that the bias of the global mean, introduced by focal activations, is reduced by 80–90% with the new estimator. Comparison with a previous unbiased estimator, using the empirical data, yielded similar results. The advantage of the new estimator is that it is not informed of experimental design and relies only on general assumptions regarding the nature of the signal. © 2001 Academic Press

INTRODUCTION

Functional brain imaging data are typically analyzed using some form of statistical model in which the observed variance is explained in terms of effects of interest, confounds, and error (Friston *et al.,* 1995). The success of the model, in terms of detecting effects of interest with high sensitivity, depends critically on selecting a parsimonious set of confounds, which model the data well without sacrificing degrees of freedom.

One commonly used confound is the "global" variate, which is typically defined as the average intensity across all intracerebral voxels. The "globals" are of crucial importance and explain a huge proportion of the variance (Fox and Mintun 1989; Friston *et al.,* 1990) for positron emission tomography (PET). The picture is less clear for functional magnetic resonance imaging (fMRI) (Zarahn *et al.,* 1997; Aguirre *et al.,* 1997, 1998; Skudlarski *et al.,* 1999). However, the globals play a special role, compared to nuisance variables, in that they are typically not orthogonal to the effects of interest. This has the effect that their inclusion/exclusion in the model will affect not only the error term, but also the estimate of the effects of interest, i.e., it affects specificity as well as sensitivity (Holmes *et al.,* 1997). Furthermore, the way that the global variate is defined means that, typically, a bias is introduced since the sum of all activations will be included in the intensity average. This bias has been shown to significantly influence the results that obtain (Strother *et al.,* 1995, 1996; Andersson, 1997), and a method for circumventing this problem has been suggested (Andersson, 1997). The previous method is based on iteratively refitting the statistical model with increasingly unbiased globals, excluding significantly activated voxels from the global estimator through *F* map masking. The disadvantage of that approach is that the global estimator depends on the statistical model used, resulting in the potential for inferential bias. Furthermore, processing time scales with the number of iterations.

In the present paper we attempt a redefinition of the global variate and suggest an estimator for it. The estimator is based on rotating the matrices obtained from a singular value decomposition (SVD) of the data to make the eigenimages adhere to certain assumptions regarding the nature of the signal we are trying to characterize. The assumption is based on the notion that activations are localized to a limited number of distinct foci. A metric that implements this is the stochastic sign change (*SSC*) criterion (Venot *et al.,* 1983) that has been used for the normalization of medical images and is relatively insensitive to focal changes.

[1] To whom correspondence and reprint requests should be addressed at Karolinska MR Research Center, Karolinska Hospital N-8, 171 76 Stockholm, Sweden. Fax: 46 8 5177 6111. E-mail: jesper@mrc.ks.se.

A disadvantage of the *SSC,* as previously implemented, is the need for a high contrast-to-noise ratio (CNR) of the focal changes and its discontinuous nature. This makes it very tedious to locate the distributional modes, particularly in a multivariate setting. In the present paper we describe a modified version of the *SSC,* the approximate stochastic sign change criterion (*ASSC*). The *ASSC* is based on ideas borrowed from the field of neural networks. We implement this idea by using the *ASSC* to drive a factor rotation (Johnson and Wichern, 1998) of the eigenimages given by the SVD. We validate the new estimator by examining its behavior on simulated data and by comparing it to the previously suggested estimator (Andersson, 1997) using PET data. In what follows we provide the theoretical background to the different components of the approach and then present an empirical validation using real and simulated data.

METHODS

Theoretical Background

The general linear model. Functional imaging data are typically modelled by the general linear model (GLM) (e.g., Friston *et al.,* 1995), which describes data as a linear combination of explanatory variables and an additive error term. An example of such a model, for a given voxel, would be

$$y_i = \mu + \beta(G_i - \bar{G}) + \alpha_i + \gamma_i + \ldots + \epsilon_i, \qquad (1)$$

or in matrix terms,

$$\mathbf{y}_{m \times 1} = \mathbf{X}\beta + \epsilon, \qquad (2)$$

where the experimental design is defined by the indicator or explanatory variables, i.e., the columns of the design matrix $\mathbf{X}$, and m denotes number of scans. The important thing about Eq. (2) is that the observed data is modeled as a linear combination of the columns of the design matrix $\mathbf{X}$, where the weights of the linear combination are given by β.

An unbiased estimator of β is given by (ignoring serial correlations)

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \qquad (3)$$

which minimizes $\hat{\epsilon}'\hat{\epsilon}$ (e.g., Johnson and Wichern, 1998). $'$ denotes transpose. Consider the simplest experiment possible with one baseline condition and one "activated" condition only. We would typically model this experiment with

$$\mathbf{y}_{m \times 1} = [\mathbf{1} \quad \hat{\mathbf{g}} \quad \mathbf{a}][\mu \quad \gamma \quad \alpha]' + \epsilon, \qquad (4)$$

where γ and α denote the global effects and the effect of the experimental condition respectively, $\hat{\mathbf{g}}$ and $\mathbf{a}$ de-

note columns of the design matrix. Note that $\hat{\mathbf{g}}$ is an estimator of some underlying global effect $\mathbf{g}$, which we take to be the arithmetic mean of the data. In reality our model is closer to

$$\mathbf{y}_{m \times 1} = [\mathbf{1} \quad \mathbf{g} + \lambda\mathbf{a} \quad \mathbf{a}][\mu \quad \gamma \quad \alpha]' + \epsilon. \qquad (5)$$

The magnitude of λ depends on the spatial extent and magnitude of the activation described by the time course $\mathbf{a}$. Under certain assumptions it can be shown that Eq. (2) no longer yields unbiased estimators, but that the estimates are now

$$E(\hat{\beta}) = \begin{bmatrix} \mu \\ \gamma \\ \alpha - \lambda\gamma \end{bmatrix}, \qquad (6)$$

hence,

$$E(\hat{\alpha}) = \alpha - \lambda\gamma, \qquad (7)$$

which simply states the well known fact that bias is introduced by the arithmetic mean estimator whereby local activations lead to an underestimation of true activations ($\alpha > 0$) and to false deactivations in areas where the null hypothesis is true ($\alpha = 0$).

Desirable properties of a global estimator. Let us note that the global estimator does not necessarily have to be the average of all voxels and think about what properties a global confound should have. This confound can be construed as successful when it explains more variance, for a larger proportion of the voxels, than any other vector. The success of the global confound can be attributed to it modeling a variance component, or trend, that is shared by virtually all voxels and will consequently decrease the error variance in large portions of the brain. In other words a global confound explains the most variance over as much of the brain as possible. Clearly this "ideal" confound would, by definition, be the average time course over all voxels, were it not for the experimentally induced variance. Hence, what we really want is the variate whose spatial expression is as global as possible, while discounting contributions from any variates with a highly localized spatial expression. A confound of this sort would potentially reduce error variance on par with the global average, while preserving any variance due to spatially localized sources, notably experimentally induced activations.

The stochastic sign change (SSC) and the approximate stochastic sign change (ASSC) criteria. The *SSC* criterion was originally suggested precisely for the purpose of normalizing the intensity of nuclear medical images (Venot *et al.,* 1983, 1984) but has since been used mainly for image registration (Venot *et al.,* 1984; Herbin *et al.,* 1989; Minoshima *et al.,* 1992). It does not give a global measure as such, but given two stochastic
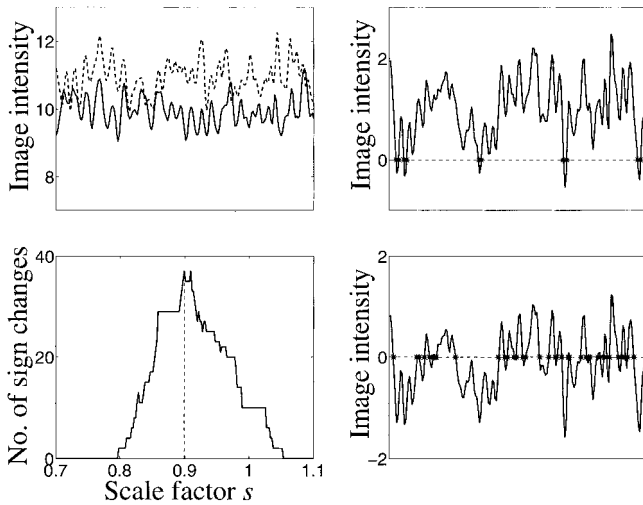
FIG. 1. Graphical explanation of the stochastic sign change criterion. The upper left (UL) panel shows two, $n = 200$, FWHM = 3, 1D processes (that may be thought of as a 200 voxels profile through an image of arbitrary dimensionality) with expectations 10 (solid line, denoted $\mathbf{y}_1$) and 11 (dashed line, denoted $\mathbf{y}_2$) and unit normal variance. The upper right (UR) panel shows the profile resulting from the linear combination $s\mathbf{y}_2 - \mathbf{y}_1$, where scaling factor s equals 1. The number of sign changes in UR equals 10. The lower left (LL) panel shows how the number of sign changes varies as s is varied between 0.7 and 1.1 (i.e., $SSC(s)$). Note the jagged appearance of the graph resulting from the SSC being an integer entity. The s corresponding to the maximum of SSC, indicated by the dashed line in LL, yields the profile shown in the lower right panel (LR) where the number of sign changes is 37. For this case the "true" s would have been 10/11 = 0.91, the global mean would have yielded $s = 0.910$, and the SSC indicates two equally good s, 0.899 and 0.909. The more "accurate" result obtained with the global mean reflects the fact that the SSC is a less efficient estimator than the global mean, this being an appreciable difference only for very small/smooth processes.

processes it yields a scale factor that relates them. The basic idea behind the SSC is best described graphically (a schematic explanation is provided in Fig. 1). In brief: given two image processes $y_1(x, y, z)$ and $y_2(x, y, z)$, containing additive noise, and the process resulting from their subtraction $y_3(x, y, z) = \mathbf{y}_1 - \mathbf{y}_2$ (where $\mathbf{y}_1$ and $\mathbf{y}_2$ are vector representations of y_1 and y_2, respectively, over all x, y, and z), the SSC is simply defined as the number of sign changes of $\mathbf{y}_3$ as counted along one, or several directions. Intuitively we would expect the SSC to be large when the intensities of $\mathbf{y}_1$ and $\mathbf{y}_2$ are similar, and small if, for example, the intensities in $\mathbf{y}_1$ are much larger than those in $\mathbf{y}_2$ (in which case $\mathbf{y}_3$ would be mainly positive). We would further expect the SSC to have a maximum when the intensities of $\mathbf{y}_1$ and $\mathbf{y}_2$ are maximally similar. In addition we would expect this measure to be relatively insensitive to focal changes of a limited extent, as demonstrated in Fig. 2.

Although having been used successfully for both normalization and registration the SSC is inherently discrete (integer) and nondifferentiable. That means that gradient or quasi-Newton methods cannot be used for its minimization and implementations have had to resort to very slow methods such as random (Venot *et al.*,

1984; Herbin *et al.,* 1989) or quasiexhaustive (Minoshima *et al.,* 1992) search methods. Furthermore its discreteness leads to multiple local maxima across a large range of search scales. This problem motivates a criterion (*ASSC*), which has the same basic properties as the *SSC,* but which is continuous and differentiable. Consider the following definition of the *SSC* criterion

$$SSC = \frac{1}{4}(\mathbf{D}\,\mathrm{sgn}(\mathbf{y}_1 - \mathbf{y}_2))'(\mathbf{D}\,\mathrm{sgn}(\mathbf{y}_1 - \mathbf{y}_2)), \quad (8)$$

where $\mathbf{y}$ is a column vector representation of the image, sgn is the sign operator that transforms any value v according to $\mathrm{sgn}(v) = v/|v|$. $\mathbf{D}$ is a differential operator matrix along any direction, or combination of directions, x, y, and z. It is readily apparent that the cause of the discontinuity is the sgn operator and that this is remedied by replacing it with a continuous differentiable squashing function such as arctan

$$ASSC = \frac{1}{\pi^2}(\mathbf{D}\,\arctan(c(\mathbf{y}_1 - \mathbf{y}_2)))'$$
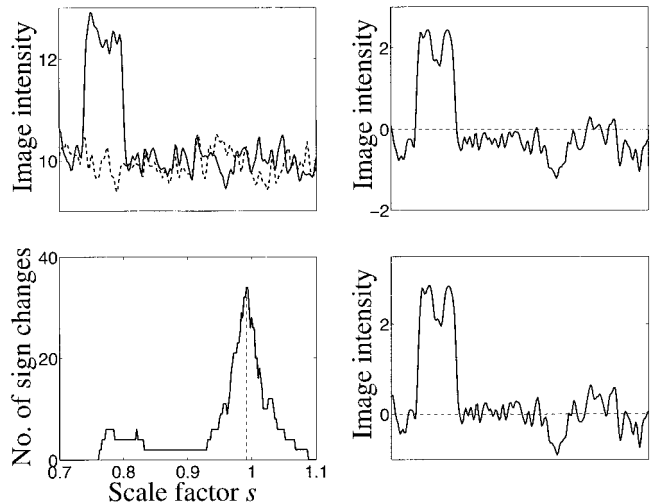$$\cdot(\mathbf{D}\,\arctan(c(\mathbf{y}_1 - \mathbf{y}_2))), \quad (9)$$

FIG. 2. Demonstration of the relative insensitivity of the *SSC* with respect to local changes. The upper left (UL) panel shows two, $n = 200$, FWHM = 3, simulated 1D processes with expectation 10 and $2^{-(1/2)}$ variance, one of which contains an "activation of magnitude" 2.5 and with a spatial extent of 15% of the process (solid line, denoted $\mathbf{y}_2$) and one that does not (dashed line, denoted $\mathbf{y}_1$). Scaling by the respective global mean (9.95 and 10.38 for unactivated and activated processes, respectively) and subtracting yields the process ($0.959\,\mathbf{y}_2 - \mathbf{y}_1$) shown in the upper right (UR) panel. This demonstrates nicely the effect encountered in actual PET studies, where the height of the activation is underestimated and where large portions of the "unactivated" parts of the brain have negative values. The lower left (LL) panel shows $SSC(s)$ and demonstrates how the activation produces a local maximum while not affecting the location of the global maximum which was found for $s = 0.992$. The lower right (LR) panel shows $0.992\,\mathbf{y}_2 - \mathbf{y}_1$, i.e., the linear combination given by the *SSC*. Note how (i) the magnitude of the activation has been increased and (ii) how the process changes seemingly randomly between positive and negative values in areas outside the "activation."

where c is simply a scale factor to adjust the range of the process to the domain of arctan. Figure 3 shows the behavior of the *ASSC* for a simulated 2-D process and compares it to the *SSC* as c of Eq. (9) is varied. The *ASSC* appears as a smoother version of the *SSC,* the smoothness decreasing as c increases. In fact the *SSC* (Eq. (8)) is the limiting case of *ASSC* (Eq. (9)) as $c \to \infty$.

We will preempt the results section by disclosing that a direct implementation of the *ASSC* for estimating the global variate does not work in practice due to the relatively poor CNR in regions of functional activation (Fig. 4). However, the *ASSC* can be used to characterize the "globalness" of linear combinations of scans, which have higher CNR. These combinations correspond to spatial modes. A global spatial mode will have a small number of zero-crossings and, once identified, can be used to compute the global variate. Local modes will have a higher number of zero-crossings. In brief, one can consider the problem of finding the global variate in terms of finding a global spatial mode. The global estimator for each scan obtains from a weighted average, over voxels, specified by this global mode. The question then is how to choose the linear combinations or modes without being informed of the experimental design. This question is addressed in the next section.

Singular value decomposition (SVD) and factor rotation. Consider the multivariate extension of Eq. (2) above

FIG. 4. The upper left (UL) panel demonstrates a simulation consisting of 64 scans, each of extent $64 \times 64 \times 1$, with a circular object varying according to one period of a sine wave. A smaller circular "activated" area varied according to a square wave orthogonal to the "global" signal and with a magnitude one-tenth that of the global signal. Gaussian random noise was added according to a multiplicative model (noise scales with global signal) to yield a contrast-to-noise ratio (CNR) of 5 or 0.5 followed by smoothing with a gaussian filter (FWHM = 3). The smoothing changes the effective CNR, but their relation should remain unchanged. The upper right panel shows the "global" signal as evaluated by the global mean on the CNR = 5 data (virtually identical for CNR = 0.5). The contamination from the local signal is immediately obvious. The lower left (LL) panel shows the global signal as evaluated by the *SSC* on the data with CNR = 5. Since the *SSC* only gives a delta measure, the curve was generated by multiplying the global mean of the first scan by the obtained scale factors. Note how the bias has vanished almost completely. The lower right panel shows the results when using the *SSC* on the CNR = 0.5 data.
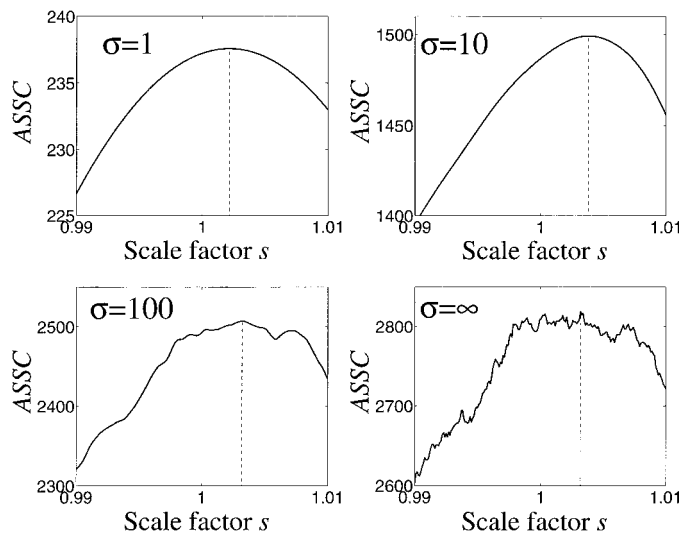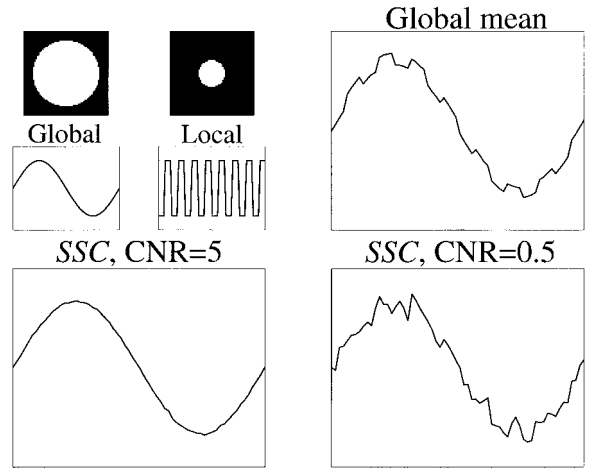

FIG. 3. The upper left panels through to the lower right shows the approximate stochastic sign change criterion (*ASSC*) evaluated according to Eq. (9) for a simulated 128×128 2D process with a FWHM of 3 and unit variance. In Eq. (9) c was varied to yield a standard deviation (σ) of the process $c(s\mathbf{y}_2 - \mathbf{y}_1)$ of 1, 10, 100, and infinity, the latter effectuating the stochastic sign change criterion (*SSC*). It can be easily appreciated how the *ASSC* becomes more and more like the *SSC* as c increases and also how local maxima start to appear as the standard deviation of the process goes above ~50. For this particular simulation the maxima of the *ASSC* were found at 1.0022, 1.0038, 1.0032, and 1.0032 for σ of 1, 10, 100, and ∞, respectively.

$$\mathbf{Y}_{n \times m} = \boldsymbol{\mu}_{n \times 1}\mathbf{1}'_{1 \times m} + \boldsymbol{\gamma}_{n \times 1}\mathbf{g}'_{1 \times m} \\ + \boldsymbol{\alpha}_{n \times 1}\mathbf{a}'_{1 \times m} + \ldots + \boldsymbol{\epsilon}_{n \times n}, \quad (10)$$

where n is the number of voxels, m the number of scans, $\mathbf{1}$, $\mathbf{g}$, and $\mathbf{a}$ are basis vectors, i.e., columns of the design matrix, and $\boldsymbol{\mu}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\alpha}$ are parametric images depicting the spatial distribution of the corresponding time courses. $\boldsymbol{\gamma}$ is the global mode we want to identify. An alternative factorization of the data $\mathbf{Y}$ is given by the SVD (dropping the size subscripts);

$$\mathbf{Y} = \mathbf{USV}' = s_1\mathbf{E}_1\mathbf{e}'_1 + s_2\mathbf{E}_2\mathbf{e}'_2 + \ldots + s_m\mathbf{E}_m\mathbf{e}'_m, \quad (11)$$

where the $\mathbf{e}$ are the eigenvectors of $\mathbf{Y}'\mathbf{Y}$, the $\mathbf{E}$ are the eigenimages or eigenmodes of $\mathbf{YY}'$, and the s are the square roots of the eigenvalues of either $\mathbf{Y}'\mathbf{Y}$ or $\mathbf{YY}'$. The formal equivalence of Eqs. (10) and (11) is evident, and is the basis for the use of SVD, or eigenimage analysis, in functional imaging (e.g., Friston *et al.,* 1993; Strother *et al.,* 1995). These methods rest on the implicit assumption that the eigenimages are "sensible," i.e., have some semblance to the parametric images alluded to in Eq. (10), and indeed their success
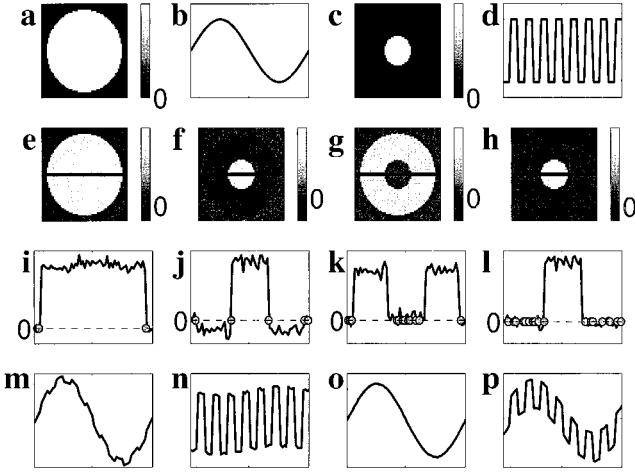
FIG. 5. (a and c) The spatial extent of the simulated object and the "activated" area and b and d show their respective "true" time courses. The two first eigenimages resulting from an SVD of the 64 "scans" are shown in e and f. It is clear that their eigenvariates, shown in m and n, are not those specified in the model, but rather linear combinations of them. (i and j) Activity profiles through the eigenimages demonstrating the same thing where the negative parts of j indicate that it is a has a negative component of the globals (much the same thing we would see in a subtraction map following scaling by the global mean). After rotation, such that the *CF* defined by equation 14 (with **w** = [0 1]) was maximized, the rotated eigenimages look like g and h, and profiles through them are shown in k and l. Note how the negative parts of the eigenimage depicting the "activation" (h and l) disappear and how there are now a large number of sign changes in the "unactivated" parts. Note also how the bias of the globals has disappeared to a large extent (cf. m and o). Furthermore, the rotation preserves the orthogonality of the eigenimages (k and l) but not that of the eigenvariates (o and p). This is advantageous since we know that the globals may not always be orthogonal to other effects. In short the factor rotation appears to partition data into a set of spatially orthogonal (but by no means independent) modes and the associated eigenvariates are the time-courses of those particular "objects."

indicates that this is somewhat true. Unfortunately, it can easily be demonstrated by simulating data according to Eq. (10) that the factorization offered by Eq. (11) is generally not the same. Indeed the SVD enforces orthogonality between the **E**s and the **e**s, which as we know cannot be guaranteed for the globals, and even when there is orthogonality the factorizations of Eqs. (10) and (11) are not generally equivalent. This can be seen, e.g., in Figs. 5m and 5n, where the observed eigenvariates are similar, but not identical, to the "true" time courses. Hence, the SVD is in itself not a plausible candidate to yield the "true" global time course. However, it may be used to formalize the intuitive notion of maximising the *ASSC* of a series of linear combinations of the original images alluded to above.

Factor rotation, or factor analysis, is a device by which the eigenvectors are manipulated to yield a set of rotated eigenvectors, or factors, which are more "sensible" according to ones presumptions (e.g., Johnson

and Wichern, 1998). This is achieved by introducing a rotation matrix **R** into Eq. (11) above such that

$$Y = USV' = URR'SV', \text{ where } RR' = I, \quad (12)$$

which yields a set of rotated eigenimages **UR**, and a set of rotated and scaled eigenvariates **R'SV'** that are no longer orthogonal. In addition the factor model often assumes that the data can be approximated by its first few components such that

$$Y_{n \times m} = U_{n \times m} S_{m \times m} V'_{m \times m} = U^*_{n \times m^*} S^*_{m^* \times m^*} V^{*'}_{m^* \times m} \\ + \epsilon_{n \times m} \approx U^* S^* V^{*'}, \quad (13)$$

where **U***, **S***, and **V*** indicate truncated versions containing the first m^* eigenimages, singular values, and eigenvariates, respectively. Henceforth we will use the unstarred notation to denote also the truncated matrices. Note that by using the truncated expression the size of **R** may be reduced from $m \times m$ to $m^* \times m^*$, and the rotational degrees of freedom from the binomial coefficient $\binom{m}{2}$ to $\binom{m^*}{2}$.

The trick in factor analysis is to find the rotation matrix **R**, a task that typically involves finding the **R** that maximizes/minimizes some function of **UR**, or of **R'SV'**. The choice of function to minimize is typically based on heuristic arguments and may be, for example, the kurtosis of the columns of **UR** (Johnson and Wichern, 1998). This is also related to independent component analysis (ICA), where **R** is chosen to maximize, e.g., the entropy of the columns of **UR**, and where **R** is no longer restricted to be orthogonal (e.g., Bell and Sejnowski, 1995; McKeown *et al.,* 1998).

The heuristic we use, for this particular application, goes as follows; for a truly global, spatial mode we expect all, or virtually all, voxels to have the same sign; i.e., we would not expect a global mode that correlates positively with some voxels to correlate negatively with others. Conversely, for a truly local spatial mode we would expect the sign to change randomly, except in restricted area/areas that constitute the mode. We therefore suggest that the "global" mode may be found by rotating the eigenimages **U**, such that the *ASSC* is minimized for the first rotated eigenimage, while at the same time maximized for the remaining ones. A cost function based on this idea is

$$CF(\mathbf{p}) = \mathbf{w}_{1 \times m} \text{diag}\left(\frac{1}{\pi^2} \left(\mathbf{D} \arctan(\mathbf{UR}(\mathbf{p}))\right)' \\ \cdot \left(\mathbf{D} \arctan(\mathbf{UR}(\mathbf{p})) \right) \right) \mathbf{1}_{m \times 1}, \quad (14)$$

where **R(p)** denotes the rotation matrix **R** as a function of the rotation angles **p**, and **w** is a weight vector. This

means that we will find $\mathbf{R}$ by finding the parameters $\mathbf{p}$ that minimize CF. According to the heuristic above our weight vector $\mathbf{w}$ could be a $1 \times m$ vector of the form $\mathbf{w} = [1 \; -1 \; -1 \ldots -1]$. However, we could also consider $\mathbf{w} = [0 \; -1 \; -1 \ldots -1]$ (see below). Figure 5 demonstrates the arguments for and results of factor rotation based on Eq. (14). Considering Figs. 5g, 5h, 5k, 5l, 5o, and 5p, we see that CF rotates areas with local effects away from the global mode giving spatially distinct modes, each expressing the time course of that particular mode. In particular we see that Fig. 5g corresponds to the unactivated parts and expresses mainly the global effects, and Fig. 5h corresponds to the activated part, and hence expresses all the effects (global and experimental). The rotated eigenimage in Fig. 5g can be thought of as a mask, or weights, with which the global variate is estimated through multiplication with the original time series. This is the same as the temporal expression or eigenvariate of a rotated mode. Using $\mathbf{w} = [0 \; -1 \; -1 \ldots -1]$ will rotate the global effects out of the local modes and effectively discount the global mode. On the other hand $\mathbf{w} = [1 \; -1 \; -1 \ldots -1]$ is equivalent to rotating the global effects away from the local modes, while at the same time retaining enough of the local effects in the global mode to avoid any sign changes in the "holes" that result from the local effects. Allowing those "holes" to extend all the way down to zero, by not punishing the zeros crossings they incur (by setting the first element of $\mathbf{w}$ to zero) means that a larger proportion of the bias may be removed. Although the two forms of $\mathbf{w}$ yield similar results, tests on both simulated and measured data indicate that $\mathbf{w} = [0 \; -1 \; -1 \ldots -1]$ removes a slightly larger proportion of the bias, and is used henceforth.

The minimum of CF can be found using a quasi-Newton method as described in Appendix A.

The procedure. The implementation of the ideas above, for estimating the global variate, follow:

1. Remove voxel means, or equivalently the block effects in a multisubject study. This is done by $\mathbf{Y}'_{adj} = (\mathbf{I} - \mathbf{H})\mathbf{Y}'$ where each column of $\mathbf{Y}'$ contains the values for one voxel, $\mathbf{H}$ is the "hat" matrix defined as $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ and $\mathbf{X}$ is the relevant design matrix (e.g., a single column of ones for each subject, modeling subject-specific mean effects). Also calculate the whole brain means of the voxel or block means from $\mathbf{g}^* = n^{-1}\mathbf{H}\mathbf{Y}'\mathbf{1}_n$, where $\mathbf{1}_n$ denotes an $n \times 1$ matrix of ones.

2. Calculate the SVD (Eq. (10)) by calculating $\mathbf{V}$ and $\mathbf{S}^2$ from the spectral decomposition of $\mathbf{Y}'_{adj}\mathbf{Y}_{adj}$ and compute a subset of the columns of $\mathbf{U}$ (eigenimages) from $\mathbf{U}^* = \mathbf{Y}\mathbf{V}^*(\mathbf{S}^*)^{-1}$, where the * denotes the dimension reduced matrices (Eq. (13)).

3. Find the rotation matrix $\mathbf{R}$ that maximizes the cost–function described by Eq. (14).

4. Calculate the "rotated" globals from $\mathbf{g} = \mathbf{g}^* + n^{-1}\boldsymbol{\xi}'\mathbf{1}_n\boldsymbol{\Psi}'$, where $\boldsymbol{\xi}$ is the first column of $\mathbf{U}^*\mathbf{R}$, i.e., the first rotated eigenimage, and where $\boldsymbol{\Psi}$ is the first ro-

tated eigenvariate, i.e., the first row of $\mathbf{R}'\mathbf{SV}'$. Note that $\mathbf{g}^*$ is simply an offset being added to the global vector and that $n^{-1}\boldsymbol{\xi}'\mathbf{1}_n$ is simply a scaling factor. Depending on how one models the global effects these factors may or may not be necessary. For example, when performing a within-subjects ANCOVA, with scaling between subjects, one could use $\boldsymbol{\Psi}'$ directly.

Practical considerations. The number of sign changes may be calculated along any one direction (x, y, or z as implemented by $\mathbf{D}$), or as the sum of sign changes along two or all directions. These measures are generally quite correlated, but there is a potential advantage of calculating the $ASSC$ along all directions, in particular for processes that are small in relation to their smoothness. This is easily performed by extending Eq. (9) such that

$$ASSC = \frac{1}{\pi^2} \mathbf{T}'\mathbf{T}, \text{ where } \mathbf{T}$$
$$= \begin{bmatrix} \mathbf{D}_x & 0 & 0 \\ 0 & \mathbf{D}_y & 0 \\ 0 & 0 & \mathbf{D}_z \end{bmatrix} (\mathbf{1}_3 \otimes c(\mathbf{y}_1 - \mathbf{y}_2)), \quad (15)$$

where $\mathbf{D}_x$, $\mathbf{D}_y$, and $\mathbf{D}_z$ are matrices implementing differentiation in the x-, y-, and z-directions, respectively, and $\otimes$ denotes tensor product. The price is execution speed, but our experience suggests there is a real advantage in using all directions. We do so in the rest of the paper.

Validation Studies

A large number of simulations were performed, an illustrative subset of which will be presented here. In general all simulations used 3-D processes of sizes ranging from $64 \times 64 \times 24$ to $128 \times 128 \times 64$. Activations were mostly spherical and always homogenous within the volume defined as "activated."

The global time course was either a sine wave ranging from a quarter of a period to several periods across the course of the "study" or a Markovian random walk. Categorical study designs ranged from simple two state designs through multilevel single factor to $2 \times 2 \times 2$ three-factor designs. Parametric linear designs were also considered, either as simple parametric designs, or as two-way categorical by parametric designs. The time courses of activations were *not* restricted to be orthogonal to the global time course.

Data were simulated according to either an additive model (noise and activation independent of global effects) or a multiplicative model (noise and activations scale with global effects) (Friston *et al.,* 1990). Noise was added as white (spatially and temporally) zero mean, normal distributed deviations.

CNR was varied across a wide range (0.1 to 10), either through varying activation magnitude or noise. The nominal CNR was specified prior to spatial

smoothing. Activations were mostly spherical and varied in size from very small to 25% of the simulated volume. One or more activated spheres were used independently of study design. For example, a 2×2 factorial could have a single "activated" area, which was activated by both tasks and which expressed an interaction, or it could have three "activated" areas, each expressing a different effect.

The simulated volumes were smoothed with a gaussian filter to a specified FWHM.

Experimental Data

We used a previously presented data set (Watson *et al.,* 1993) to demonstrate the effects of biased globals and the improvement achieved with the present method. In short 11 healthy volunteers were scanned, in a CTI-953B brain PET scanner operating in 3-D mode (Spinks *et al.,* 1992), alternating between two states; viewing stationary "Mondrians" (irregular arrangements of partially overlapping rectangles) and viewing moving "Mondrians." The scans were corrected for movements, spatially normalized using SPM99 and smoothed using a 10-mm FWHM isotropic gaussian filter.

Data Analysis

Simulations. A metric describing the bias in the globals was defined on a per contrast basis in the following way. If the design of a specific simulation is given by the design matrix $\mathbf{X}$, then the regressor describing a particular effect is given by $\mathbf{Xc}$, where $\mathbf{c}$ is a contrast weight vector describing the effect of interest. Let $\mathbf{g}$ denote the vector of "true" global effects and $\hat{\mathbf{g}}$ the vector of estimated globals, then the bias was defined as $bias = \mathbf{c}'\mathbf{X}'(\hat{\mathbf{g}} - \mathbf{g})$. For two different estimates of globals, e.g., $\hat{\mathbf{g}}_{gm}$ (global mean) and $\hat{\mathbf{g}}_{fr}$ (factor rotation), we define the percentage reduction in bias as

$$100\mathbf{c}'\mathbf{X}'(\hat{\mathbf{g}}_{gm} - \hat{\mathbf{g}}_{fr})/\mathbf{c}'\mathbf{X}'(\hat{\mathbf{g}}_{gm} - \mathbf{g}).$$

For each simulation globals were estimated as the global mean and by factor rotation and their respective biases were calculated along with the reduction resulting from the use of factor rotation.

Experimental data. No "true" answer is available for the experimental data and we were limited to demonstrating construct validity by comparing the results to those obtained with a previous method (Andersson, 1997). The data was analyzed using a multisubject, ANCOVA by subject design explicitly modeling condition by replication effects (one column per scan).

For one analysis the global activity was calculated in the "normal" way as the average across all intracerebral voxels.

For another set of analyses the globals were calculated as the average across all intracerebral voxels, excluding those voxels that exhibited a P value less than 0.05 based on the SPM$\{F\}$ from the previous analysis. This was iterated 10 times to obtain asymptotic estimators (Andersson, 1997).

The third set of analyses involved calculating the globals by the method suggested in the present paper, basing the *ASSC* on all three dimensions (Eq. (15)), starting with $\mathbf{UR}$ scaled to yield a standard deviation of 10, changing it for the final iterations of the factor rotation to yield a standard deviation of 100. The number of eigenimages included in the factor rotation was varied between 2 and 10.

Two contrasts showing areas with higher activity for moving vs stationary scenes and lower activity for moving vs stationary scenes were used to generate two SPM$\{t\}$ for each analysis. The maximum t value and the number of suprathreshold voxels ($P < 0.01$, uncorrected) were used to characterize each analysis.

RESULTS

Simulations

Direct application of ASSC to difference images. A straightforward application of the *SSC* or *ASSC* to estimate global difference between subsequent scans revealed, as expected, a profound sensitivity to CNR. Figure 4 shows an example of a simulation demonstrating this effect. To show that this is a general problem of *SSC,* and not specific to the *ASSC,* we used the *SSC* in this particular simulation (the results for the *ASSC* are almost identical). The explanation for this behavior is that as CNR decreases the two peaks seen in Fig. 2 increase their proximity until they merge and the peak attributed to the activation starts to skew the distribution from the unactivated areas. On simulated data one can always find a level of spatial smoothing such that the direct approach works reasonably well. However, the filter width has to be tailored and depends on CNR and activation extent. This approach always failed on experimental data and was not pursued further.

ASSC as cost function for rotating eigenimages. We changed the scaling of $\mathbf{UR}$ in Eq. (14) (equivalent to varying the scale factor c in Eq. (9)) to yield standard deviations of the eigenimages (σ) ranging from 0.1 to 100. The results were highly consistent across all simulations and indicated that σ should be at least 10 in order to obtain good results (Fig. 6). Furthermore there were slight additional gains for σ up to 100, but the cost-function (Eq. (14)) became "rougher," increasing the risk of local maxima. From this we established a multiscale factor rotation consisting of (i) a maximization of Eq. (14) with $\sigma = 10$, followed by (ii) a maximization with $\sigma = 100$, using the results from the first
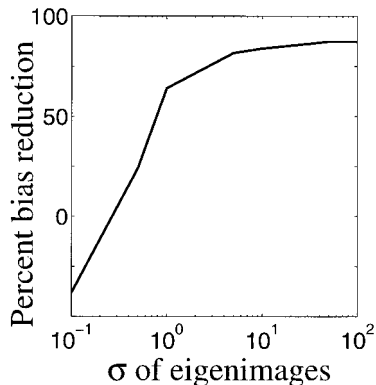
FIG. 6. Percentage bias reduction (as defined in main text) for different choices of σ for simulated data. The simulation consisted of 64 64 × 64 × 24 scans of a cylindrical object with 28 voxels radius and a spherical activation of 13 voxels radius. The global time course was one half period of a sine wave and the activation was a square wave with a fundamental frequency of 0.25 scans. Gaussian white noise was added to yield a CNR of 1, followed by smoothing with a 3 voxels FWHM gaussian kernel. The results represent the average of 10 realizations with different levels of noise. The curve starts from below zero for σ below ~0.2, which indicates the introduction of additional bias (a consistent finding) and rises steeply until σ equals ~1 and it reaches a plateau at σ = 100, at which 85–90% of the bias is removed.

maximization as starting estimates for the second. The second step changes the results only marginally, but on the other hand does converge quickly.

In terms of bias and bias reduction the results indicate a surprising consistency across the entire simulated range of activation magnitudes, activation extents, CNR, and filter widths. An example of a typical simulation can be seen in Fig. 7, which shows the bias resulting from using the global mean, and the small amount of bias remaining when using the *ASSC.* In order to retain a constant bias across CNR the activation magnitude was constant and the noise was varied. With the exception of unfiltered data with very low CNR (0.1) there is a consistent reduction in bias when using the *ASSC* across CNR and filter widths; 80–90% of the bias is removed irrespective of the original degree of the bias (a function of the extent of the activation). This finding is consistent provided enough eigenmodes are used in the rotation.

The number of eigenmodes one should retain in the rotation is, not surprisingly, equal to the number of effects entered into the simulation; i.e., if data were simulated using a design matrix containing global effects, two levels each of two factors and an interaction, then one should keep four eigenmodes. In general each eigenmode corresponds to an effect given by $\mathbf{Xb}$, where $\mathbf{X}$ is the design matrix used to create the data and $\mathbf{b}$ is any column vector (cf. a contrast vector $\mathbf{c}$). The inclusion of the eigenmode corresponding to $\mathbf{Xb}$ in the rotation will eliminate/reduce the bias from a contrast $\mathbf{c}$ equal to or similar to $\mathbf{b}$. Due to the nature of the SVD subsequent eigenmodes will correspond to "contrasts"

$\mathbf{b}$ with smaller and smaller effects. In short, the number of eigenmodes should be equal to the rank of the design matrix.

Experimental Data

An example of the *CF* described by Eq. (14) as a function of the two predominant rotation angles is shown in Fig. 8. When calculating the globals in the conventional way the SPMs exhibited massive activations in visual cortex and widespread, possibly artefactual, deactivations in other parts of the brain (Fig. 9). The *t* values of the "deactivations" decreased on average by 1.1–1.3 for both the *F* map masking method and
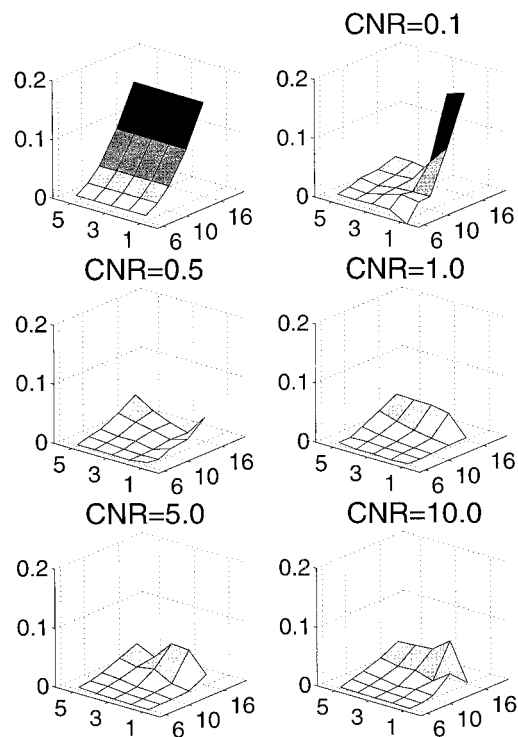

FIG. 7. Bias and bias reduction as a function of extent of activated area, filter width and CNR. The simulations consisted of 64 64 × 64 × 24 scans containing a cylindrical object with radius 28 voxels, where the globals varied as one half of a sine wave. For each simulation there was a single spherical activated area of radius 6, 8, 10, 13, or 16 voxels, which varied as a square wave with a frequency of 0.25 scans. White gaussian noise was added to the images to yield a CNR of 0.1, 0.5, 1, 5, or 10 prior to smoothing. The smoothing employed a Gaussian kernel yielding a FWHM of 1 (no smoothing) 2, 3, 4, or 5 voxels. Each combination of activation extent, CNR and FWHM were simulated ten times, and the points in the graphs represent the mean of these realizations. The upper left panel shows the bias (as defined in the main text) of the globals when using the global mean. It can be seen that, as expected, it is affected by the extent of the activation and is independent of smoothing. Panels 2 through 6 show the remaining bias after SVD and factor rotation with σ = 10, followed by a few iterations with σ = 100. It can be seen that for very poor CNR some smoothing is crucial, whereas otherwise the results are surprisingly consistent across the tested range of CNRs and FWHMs.
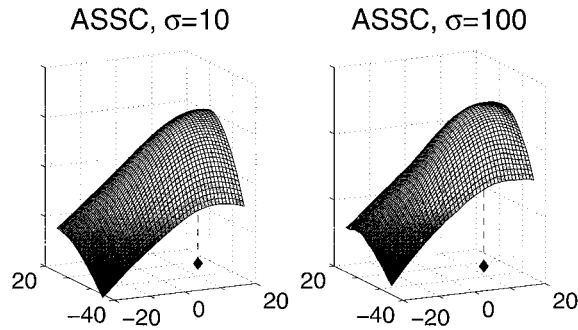
FIG. 8. The *CF* (the sum of the *ASSC* of the local modes) as a function of rotation angle (in degrees) around two axes while holding the angle around the remaining (eight) axes fixed at the values obtained by the full maximization. Note how the function appears quite well behaved (quadratic) over this rather large range, especially for $\sigma = 10$. Note further the slight hint at "rougher" behavior for $\sigma = 100$ and that the locations of the maxima differ slightly ([16°, −2°] and [14°, −5°] for $\sigma = 10$ and $\sigma = 100$, respectively).

TABLE 1

Results from *F* Map Masking of V5 Data Set

Iteration	Max pos t^*	Max neg $t^\dagger$	SptExt‡ pos	SptExt§ neg
1	20.19 (0.000)	6.20 (0.001)	27725 (0.000)	21011 (0.000)
2	20.76 (0.000)	5.34 (0.015)	33513 (0.000)	4037 (0.000)
3	20.99 (0.000)	5.13 (0.032)	35162 (0.000)	2936 (0.000)
4	21.04 (0.000)	5.07 (0.039)	35726 (0.000)	1953 (0.000)
5	21.06 (0.000)	5.06 (0.041)	35852 (0.000)	1930 (0.000)
6	21.06 (0.000)	5.05 (0.042)	35888 (0.000)	1922 (0.000)
7	21.06 (0.000)	5.05 (0.042)	35894 (0.000)	1922 (0.000)
8	21.06 (0.000)	5.05 (0.042)	35897 (0.000)	1921 (0.000)
9	21.06 (0.000)	5.05 (0.042)	35898 (0.000)	1921 (0.000)
10	21.06 (0.000)	5.05 (0.042)	35898 (0.000)	1921 (0.000)

Note. Values refer to highest positive (*) and negative (†) *t* values, and spatial extent (voxels) of largest clusters of connected voxels with *t* values > 2.36 ($P = 0.01$) in the positive (‡) and negative (§) SPM{*t*}'s. Values in parenthesis refer to the associated *P* values.

for the present method (Tables 1 and 2 and Figs. 9 and 10) and the "activations" increased by roughly the same amount. In fact, no deactivated voxels survived the corrected threshold when using the present method, in contrast to using the conventional globals where there were six significant maxima. The *F* map masking yielded one significant voxel located in the left prefrontal cortex. It is obviously not possible to say for certain that the deactivations encountered when using conventional globals are false, although given the results of the simulations one might suspect this to be the case.
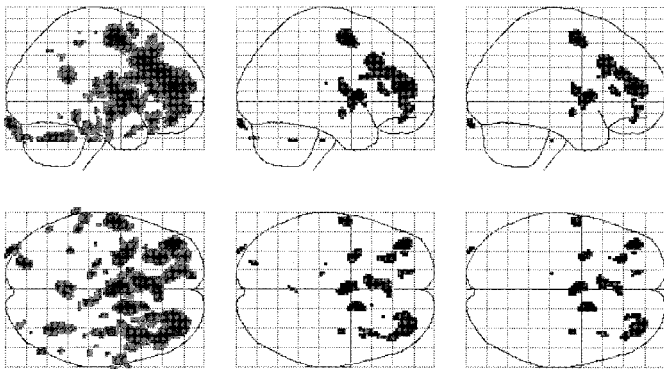
DISCUSSION

We have proposed a redefinition of the global variate in PET activation studies and developed an estimator based on that definition. Simulations indicate that this new estimator greatly reduces bias that ensues from conflating experimentally induced variance and global changes when using the global mean.

In the present paper we have not examined the use of the new global estimator on fMRI data. The reason for this is that the complex spatio-temporal behavior of fMRI signals warrants a separate analysis. It is clear, however, that the question of the global covariate is as pertinent for fMRI as for PET, given the conflicting reports of its usefulness (Zarahn *et al.,* 1997; Aguirre *et al.,* 1997; Skudlarski *et al.,* 1999) and of the potential errors introduced by failing to model the global effects (Aguirre *et al.,* 1998).

FIG. 9. Maximum intensity projections of *t* maps, displayed at $P < 0.001$ uncorrected, for the contrast depicting areas with less activity during moving stimuli, compared to stationary stimuli. This contrast was chosen because of its sensitivity to a bias in the global covariate induced by the "true" activation in the negative contrast. The global normalization was, from left to right, global mean, ten iterations of *F* map masking and *ASSC*-based factor rotation retaining five eigenmodes. Note the typical widespread areas of deactivation extending well into white matter resulting from using the global mean. The results obtained with *F* map masking and factor rotation are quite similar and contain considerably less deactivated areas. The areas that persist are located to the left frontal medial gyrus and the middle frontal gyrus bilaterally.

TABLE 2

Results from Factor Rotation of V5 Data Set

No. of eigenmodes	Max pos t^*	Max neg $t^\dagger$	SptExt‡ pos	SptExt§ neg
1	20.19 (0.000)	6.20 (0.001)	27725 (0.000)	21011 (0.000)
2	20.42 (0.000)	5.91 (0.002)	29502 (0.000)	13638 (0.000)
3	20.86 (0.000)	5.22 (0.023)	34913 (0.000)	3024 (0.000)
4	21.08 (0.000)	5.05 (0.043)	36270 (0.000)	1854 (0.000)
5	21.18 (0.000)	4.91 (0.068)	37648 (0.000)	1584 (0.000)
6	21.12 (0.000)	4.97 (0.055)	37009 (0.000)	1715 (0.000)
7	21.18 (0.000)	5.05 (0.042)	36196 (0.000)	1890 (0.000)
8	21.10 (0.000)	4.89 (0.073)	37370 (0.000)	1642 (0.000)
9	21.20 (0.000)	4.83 (0.090)	37699 (0.000)	1582 (0.000)
10	21.11 (0.000)	4.92 (0.066)	37073 (0.000)	1701 (0.000)

Note. Values refer to highest positive (*) and negative (†) *t* values, and spatial extent (voxels) of largest clusters of connected voxels with *t* values > 2.36 ($P = 0.01$) in the positive (‡) and negative (§) SPM{*t*}'s. Values in parenthesis refer to the associated *P* values.
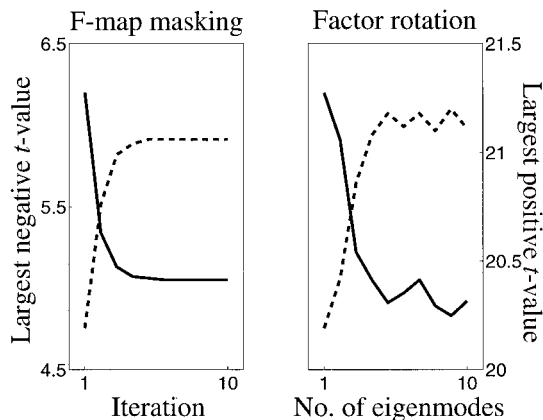
FIG. 10. The largest (most positive, solid lines) and smallest (most negative, dashed lines) t value from the contrast subtracting stationary from moving stimulus for F map masking and for factor rotation. It should be noted that despite the apparent similarity between the graphs, they really depict completely different things. The left panel shows how the t values change with subsequent iterations of the F map masking method. The right panel in contrast shows what the t values are after the full iterative factor rotation when retaining between 1 and 10 eigenmodes. Hence, in the left panel each point represents one iteration, whereas in the right panel each point represents the completion of one iterative procedure. The right panel shows how additional eigenmodes improve the results up to five eigenmodes, after which it reaches a plateau. It can also be seen that the final result of the F map masking is quite similar to those obtained by factor rotation when "enough" eigenmodes are used.

The suggested method may seem like a complicated way to calculate something fairly simple. It should be pointed out however that we have tried a number of obvious (e.g., the median) and not so obvious (e.g., ICA) methods in our pursuit of a valid global estimator. The method above is the first we have found to perform satisfactorily under a wide range of conditions. We set out to find a multivariate characterization of the data, and to identify one specific "global" mode of this characterization without being informed of the experimental design. With this perspective in mind, our approach becomes more intuitive.

One important question in factor analyses is the number of eigenmodes one should retain in the rotation. The simulations indicated that one mode should be allowed for the global activity, one for each regressor which would normally have been used to model each factor (i.e., number of levels minus one) and one for each interaction. Hence, a simple single factor (with two levels) experiment, like that used for the validation, should be adequately modeled by two eigenmodes, and for a 2×2 factorial four eigenmodes should suffice. In the experimental data better estimates of the globals were obtained as the number of eigenmodes was increased to five (Fig. 10). This is not surprising since a multisubject study will contain, in addition to condition effects, subject by condition interactions. In a multivariate characterization of the data, these effects will surface in unpredictable ways. In the example

used above, eigenmodes that modeled the condition effects for subpopulations of the entire study emerged. In addition, there are frequently other effects such as time effects and time by condition interactions embodied in real data. There are objective methods for determining the number of factors to retain in general principal component analysis (Johnson and Wichern, 1998), but there is no real consensus regarding these, and the eigenvalue spectra from PET data seldom lend themselves to clear conclusions in terms of, e.g., scree plots.

For the above reasons we suggest the rank of the design matrix (which should be an adequate model of the data) is used to specify the number of eigenmodes, providing this is not too large. Our experience suggests that about 5 is a reasonable number for a typical PET study. The factor rotation is a problem of maximization of a function with respect to multiple parameters. The number of parameters increases as $\binom{m}{2}$, where m is the number of eigenmodes, i.e., it increases almost as the square of m. In addition, the number of function evaluations necessary for finding an extremum of a multiparameter function is approximately proportional to the square of the number of parameters (Press *et al.,* 1992). Hence, the execution time, in a worst case scenario, is proportional to the number of eigenmodes raised to the power of four. Fortunately, it is not quite as bad as this since the use of partial derivatives in the minimization decreases the exponent.

Clearly by reducing the number of eigenmodes one risks excluding experimentally induced variance, and thereby not rotating that variance out of the global mode. However, as long as the dominant spatiotemporally coherent sources of variance are due to the experimental design, each subsequent eigenmode contributes less and less to the bias of the globals. Disregarding subject by condition interactions each eigenmode correspond to one possible "contrast," where subsequent "contrasts" produce the activations of lesser magnitude and/or extent. Because a contrast that produces large activations is also likely to bias global estimators, global bias is removed, effectively, for the contrasts that matter.

Considering the, sometimes profound, bias that results from using the global mean as a regressor the need for a way to calculate an unbiased measure of the global activity is clear. The reason for developing an alternative method to that previously suggested (Andersson, 1997) was twofold.

First and foremost, the previous method involved fitting the data with a tentative model, and then changed the model based on the outcome of that fit. Although the results of the first fit were used in a highly indirect way (as a mask) there is always a risk that bootstrapping a model in this way may introduce inferential bias. The present method is not informed of the experimental design and is driven solely by the

assumption that true activations have a limited spatial extent. Second, the previous method involved an iterative refitting of the entire model, which was potentially quite time consuming. The current method consists of fitting all data to a reduced model once, followed by an iterative procedure in a reduced subspace, yielding shorter execution times. However, since the number of iterations needed can be great, this may not always be true, especially for small sized studies. As computing power increases, speed considerations will become less important, while those of accuracy will prevail.

In conclusion, we have suggested a novel way of estimating the global variate. Compared to simple averaging, it yields considerably less biased values while not being informed of the experimental design. Simulations and validation on PET data indicates that the method serves its purpose, although further applications to PET studies will be needed to establish its ultimate usefulness.

APPENDIX A: MINIMIZATION OF *CF*

Finding $\mathbf{R}$ that minimizes *CF* is a multidimensional optimization problem and as such potentially sensitive to local minima as well as time consuming. As a general rule optimization methods that utilize derivative information are considerably faster than those that do not (Press *et al.,* 1992). This was the motivation for adopting the *ASSC.* Hence we need an expression for the derivative of *CF* with respect to the optimization parameters $\mathbf{p}$. The derivation is slightly tedious (see Appendix B), but it can be shown that the gradient of *CF* is given by

$$
\nabla CF(\mathbf{p}) = \frac{2}{\pi^2}\, \mathbf{w}\underbrace{\left(\left(\underbrace{(\mathbf{1}_m \otimes \mathbf{I}_{n-1})}_{m(n-1)\times n-1}\underbrace{\mathbf{D}\,\arctan(\mathbf{UR}(\mathbf{p}))}_{n-1\times m}\right) \times \underbrace{(\mathbf{I}_m \otimes \mathbf{1}_{n-1})}_{m(n-1)\times m}\right)'}_{m\times m(n-1)}\underbrace{(\mathbf{I}_m \otimes \mathbf{D})}_{m(n-1)\times mn}
$$

$$
\cdot\underbrace{\left\{\underbrace{\left(\mathbf{1}_I' \otimes \mathrm{vec}\left[\frac{1}{1+(\mathbf{UR}(\mathbf{p}))^2}\right]\right)}_{mn\times I} \times \underbrace{(\mathbf{I}_m \otimes \mathbf{U})}_{mn\times mm}\underbrace{\frac{d\mathbf{R}}{d\mathbf{p}}}_{mm\times I}\right\}}_{mn\times I},
$$

(A1)

where Appendix B may be consulted for an explanation of notation. Assuming that *CF* may be approximated by a quadratic function at some point $\mathbf{p}_0$ reasonably close to the true minimum we can write

$$
CF(\mathbf{p}) = CF(\mathbf{p}_0) + \nabla CF(\mathbf{p}_0)(\mathbf{p} - \mathbf{p}_0) \\
+ \frac{1}{2}\,(\mathbf{p} - \mathbf{p}_0)'\mathbf{H}(\mathbf{p}_0)(\mathbf{p} - \mathbf{p}_0),
$$

(A2)

where $\mathbf{H}(\mathbf{p}_0)$ denotes the Hessian (the matrix of second partial derivatives) evaluated at the point $\mathbf{p}_0$. For a minimum of a multivariate scalar function the derivative should be zero and the Hessian should be positive definite (Marsden and Tromba, 1981). Now, if we differentiate the right side of Eq. (A2) with respect to $\mathbf{p}$ and set it to zero we get

$$
\nabla CF(\mathbf{p}_0) + (\mathbf{p} - \mathbf{p}_0)'\mathbf{H}(\mathbf{p}_0) = \mathbf{0} \Rightarrow \\
(\mathbf{p} - \mathbf{p}_0) = -\mathbf{H}^{-1}(\mathbf{p}_0)\nabla CF(\mathbf{p}_0)',
$$

(A3)

which means that if Eq. (A2) is a good approximation the step $(\mathbf{p} - \mathbf{p}_0)$ according to Eq. (A3) will take us directly to the minimum. The Newton method depends on the ability to calculate the Hessian, preferably analytically, and iteratively solves Eq. (A3). Given the complexity of calculating the gradient in this case we use a quasi-Newton method, which gradually builds an approximation of the inverse Hessian by performing a rank-2 modification of the current estimate at each iteration. The basic idea is to start with a positive definite "guess" of $\mathbf{H}^{-1}$, typically $\mathbf{I}$, and then for each iteration find the minimum along the line segment $\mathbf{p}_0 - \lambda\mathbf{H}_0^{-1}\nabla CF_0$, where $0 < \lambda \le 1$, and where $\mathbf{H}_0^{-1}$ and ∇CF_0 denotes the inverse Hessian and the gradient, respectively, at the point $\mathbf{p}_0$. Next we seek to find a new estimate of the inverse Hessian ($\mathbf{H}_1^{-1}$) of the form $\mathbf{H}_1^{-1} = \mathbf{H}_0^{-1} + \mathbf{C}$, where $\mathbf{C}$ is a rank-2 updating matrix that satisfies

$$
-\lambda\mathbf{H}_0^{-1}\nabla CF_0' = \mathbf{H}_1^{-1}(\nabla CF_1 - \nabla CF_0)'. \qquad \text{(A4)}
$$

The specific form of $\mathbf{C}$ that we use in the present paper is that of Broyden–Fletcher–Goldfarb–Shanno (BFGS) and a description of it may be found in Press *et al.* (1992). We find that as long as positive definiteness of $\mathbf{H}^{-1}$ is explicitly enforced at each iteration, by simply checking the eigenvalues and reusing the old estimate if not, this is a fast and robust way of finding the minimum of Eq. (14).

APPENDIX B: PARTIAL DERIVATIVES OF THE *ASSC*

The advantage of the *ASSC,* compared to the *SSC,* is that it is a continuous differentiable function. This is of particular importance when attempting to maximize it with respect to more than one parameter since it allows the use of quasi-Newton methods as opposed to slow direct searches. The following is a derivation of expressions for the derivatives of the *ASSC* with respect to the rotation matrix. In the following $\mathbf{I}_n$ will denote an $n \times n$ identity matrix and $\mathbf{1}_n$ will denote an $n \times 1$ column vector of ones. We will use $\otimes$ to denote the Kronecker (tensor) product and $\times$ to denote Hadamard (element-wise) product and we will use the vec operator defined such that $\text{vec}(\mathbf{A}_{m \times n}) = [a_{11} \cdots a_{m1}\ a_{12} \cdots a_{mn}]'$, i.e., it stacks the columns of $\mathbf{A}$ on top of each other. When we write an operator as operating on the entire matrix then that will mean performing that operation on all elements of that matrix, e.g.,

$$\left[\frac{1}{1 + \mathbf{A}^2}\right] = \begin{bmatrix} \dfrac{1}{1 + a_{11}^2} & \cdots & \dfrac{1}{1 + a_{1n}^2} \\ \vdots & \ddots & \vdots \\ \dfrac{1}{1 + a_{m1}^2} & \cdots & \dfrac{1}{1 + a_{mn}^2} \end{bmatrix}.$$

The derivative of an $m \times n$ matrix valued function ($\mathbf{A}$) with respect to a $k \times l$ matrix ($\mathbf{B}$) will be represented as an $mn \times kl$ matrix arranged such that

$$\left[\frac{d\mathbf{A}}{d\mathbf{B}}\right]_{mn \times kl} = \begin{bmatrix} \dfrac{\partial a_{11}}{\partial b_{11}} & \cdots & \dfrac{\partial a_{11}}{\partial b_{k1}} & \dfrac{\partial a_{11}}{\partial b_{12}} & \cdots & \dfrac{\partial a_{11}}{\partial b_{kl}} \\ \vdots & & \vdots & \vdots & & \vdots \\ \dfrac{\partial \dot{a}_{m1}}{\partial b_{11}} & \cdots & \dfrac{\partial \dot{a}_{m1}}{\partial b_{k1}} & \dfrac{\partial \dot{a}_{m1}}{\partial b_{12}} & \cdots & \dfrac{\partial \dot{a}_{m1}}{\partial b_{kl}} \\ \dfrac{\partial a_{12}}{\partial b_{11}} & \cdots & \dfrac{\partial a_{12}}{\partial b_{k1}} & \dfrac{\partial a_{12}}{\partial b_{12}} & \cdots & \dfrac{\partial a_{12}}{\partial b_{kl}} \\ \vdots & & \vdots & \vdots & & \vdots \\ \dfrac{\partial \dot{a}_{mn}}{\partial b_{11}} & \cdots & \dfrac{\partial \dot{a}_{mn}}{\partial b_{k1}} & \dfrac{\partial \dot{a}_{mn}}{\partial b_{12}} & \cdots & \dfrac{\partial \dot{a}_{mn}}{\partial b_{kl}} \end{bmatrix}.$$

Given this definition of the derivative and given that $\mathbf{A}_{m \times n} = \mathbf{B}_{m \times l}\mathbf{C}_{l \times n}$, then

$$\left[\frac{d\mathbf{A}}{d\mathbf{B}}\right]_{mn \times ml} = \mathbf{C}' \otimes \mathbf{I}_m \tag{B1}$$

and

$$\left[\frac{d\mathbf{A}}{d\mathbf{C}}\right]_{mn \times \ln} = \mathbf{I}_n \otimes \mathbf{B}. \tag{B2}$$

Assume we have a matrix $\mathbf{U}_{n \times m}$ consisting of m eigenimages, each consisting of n voxels and a rotation matrix $\mathbf{R}_{m \times m}$, which is completely determined by $\binom{m}{2}$ (henceforth denoted l) parameters $\mathbf{p}$ such that $\mathbf{R} = \mathbf{R}(\mathbf{p})$. We may then calculate a row vector **assc** containing the *ASSC* measure for each of the rotated eigenimages from

$$\mathbf{assc}(\mathbf{p}|\mathbf{U})_{1 \times m}$$

$$= \mathbf{1}_m'\text{diag}\left(\frac{1}{\pi^2}\,(\mathbf{D}\,\arctan(\mathbf{UR}(\mathbf{p})))'(\mathbf{D}\,\arctan(\mathbf{UR}(\mathbf{p})))\right), \tag{B3}$$

where $\mathbf{D}$ is a spatial differentiation operator implemented as an $(n - 1) \times n$ matrix. What we require is the matrix $\left[\dfrac{d\mathbf{assc}}{d\mathbf{p}}\right]_{m \times l}$, which contains the derivatives of the *ASSC* of each rotated eigenimage with respect to each of the l parameters $\mathbf{p}$. We now use the chain rule and proceed in steps noting that

$$\frac{d\mathbf{assc}}{d\mathbf{p}} = \frac{2}{\pi^2}\underbrace{(((\underbrace{\mathbf{1}_m \otimes \mathbf{I}_{n-1}}_{m(n-1) \times n-1})\underbrace{\mathbf{D}\,\arctan(\mathbf{UR}(\mathbf{p}))}_{n-1 \times m}) \times (\underbrace{\mathbf{I}_m \otimes \mathbf{1}_{n-1}}_{m(n-1) \times m}))'}_{m \times m(n-1)}\underbrace{\frac{d(\mathbf{D}\,arctan(\mathbf{UR}(\mathbf{p})))}{d\mathbf{p}}}_{m(n-1) \times l}, \tag{B4}$$

$$\frac{d(\mathbf{D}\,\arctan(\mathbf{UR}(\mathbf{p})))}{d\mathbf{p}} = \underbrace{(\mathbf{I}_m \otimes \mathbf{D})}_{m(n-1) \times mn}\underbrace{\frac{d(\arctan(\mathbf{UR}(\mathbf{p})))}{d\mathbf{p}}}_{mn \times l}, \tag{B5}$$

$$\frac{d(\arctan(\mathbf{UR}(\mathbf{p})))}{d\mathbf{p}}$$

$$= \left(\underbrace{ \mathbf{1}'_l \otimes \mathrm{vec}\left[\frac{1}{1+(\mathbf{UR}(\mathbf{p}))^2}\right] }_{mn\times l} \right) \times \underbrace{ \frac{d(\mathbf{UR}(\mathbf{p}))}{d\mathbf{p}} }_{mn\times l}, \tag{B6}$$

$$\frac{d(\mathbf{UR}(\mathbf{p}))}{d\mathbf{p}} = \underbrace{(\mathbf{I}_m \otimes \mathbf{U})}_{mn\times mm}\underbrace{\frac{d\mathbf{R}}{d\mathbf{p}}}_{mn\times l}, \tag{B7}$$

where $\mathbf{R}$ is the product of l matrices $\mathbf{R}_i$ each implementing rotation round one axis such that $(r_i)_{jk}$ is 1

if $j = k$, 0 if $j \neq k$ except for a and b ($a < b$) for which $(r_i)_{aa} = (r_i)_{bb} = \cos(p_i)$ and $(r_i)_{ab} = -(r_i)_{ba} = \sin(p_i)$. The derivative matrix of $\mathbf{R}_i$ with respect to p_i is obtained by instead setting $(r_i)_{aa} = (r_i)_{bb} = -\sin(p_i)$ and $(r_i)_{ab} = -(r_i)_{ba} = \cos(p_i)$ and the ith column of $d\mathbf{R}/d\mathbf{p}$ is hence derived from

$$\left[\frac{d\mathbf{R}}{d\mathbf{p}}\right]_{:,i} = \mathrm{vec}\left(\mathbf{R}_1 \dots \mathbf{R}_{i-1}\frac{d\mathbf{R}_i}{dp_i}\mathbf{R}_{i=1} \dots \mathbf{R}_l\right). \tag{B8}$$

Hence in total the derivative of the *ASSC* with respect to the rotation parameters $\mathbf{p}$ may be written as

$$\frac{d\mathbf{assc}}{d\mathbf{p}} = \frac{2}{\pi^2} \underbrace{(\underbrace{((\mathbf{1}_m \otimes \mathbf{I}_{n-1})}_{m(n-1)\times n-1}\underbrace{\mathbf{D}\arctan(\mathbf{UR}(\mathbf{p})))}_{n-1\times m} \times \underbrace{(\mathbf{I}_m \otimes \mathbf{1}_{n-1}))'}_{m(n-1)\times m}}_{m\times m(n-1)}\underbrace{(\mathbf{I}_m \otimes \mathbf{D})}_{m(n-1)\times mn}$$

$$\cdot \underbrace{\left\{\underbrace{\left(\mathbf{1}'_l \otimes \mathrm{vec}\left[\frac{1}{1+(\mathbf{UR}(\mathbf{p}))^2}\right]\right)}_{mn\times l} \times \underbrace{(\mathbf{I}_m \otimes \mathbf{U})}_{mn\times mm}\underbrace{\frac{d\mathbf{R}}{d\mathbf{p}}}_{mm\times l}\right\}}_{mn\times l}, \tag{B9}$$

where $d\mathbf{R}/d\mathbf{p}$ is given by Eq. (B8). There are a couple of things to note regarding this expression; the first being that it is not as complicated as one might think upon a first glance. The purpose of the Kronecker products is simply to duplicate/rearrange the elements of the matrices. For example, the first part of the expression, which we may write as

$$\underbrace{((\mathbf{1}_m \otimes \mathbf{I}_{n-1})}_{m(n-1)\times n-1}\underbrace{\mathbf{A}}_{n-1\times m} \times \underbrace{(\mathbf{I}_m \otimes \mathbf{1}_{n-1}))'}_{m(n-1)\times m}$$

will simply take the columns of $\mathbf{A} = [\mathbf{a}_1 \ \mathbf{a}_2 \cdots \mathbf{a}_m]$ and rearrange them such that

$$\underbrace{((\mathbf{1}_m \otimes \mathbf{I}_{n-1})}_{m(n-1)\times n-1}\underbrace{\mathbf{A}}_{n-1\times m} \times \underbrace{(\mathbf{I}_m \otimes \mathbf{1}_{n-1}))'}_{m(n-1)\times m}$$

$$= \underbrace{\begin{bmatrix} \mathbf{a}'_1 & \mathbf{0}' & \cdots & \mathbf{0}' \\ \mathbf{0}' & \mathbf{a}'_2 & \cdots & \mathbf{0}' \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}' & \mathbf{0}' & \cdots & \mathbf{a}'_m \end{bmatrix}}_{m(n-1)\times m}. \tag{B10}$$

The second thing to note is that a direct implementation of Eq. (B9) is not practical since it involves creating enormous matrices (e.g., $\mathbf{I}_m \otimes \mathbf{D}$) that are nothing but very sparse matrices with multiple copies of the same data. For a typical PET data set with $m = 5$ and $n = 50,000$ $\mathbf{I}_m \otimes \mathbf{D}$ would be a $249,995 \times 250,000$ matrix, whereas the actual storage requirement, if capitalizing also on the sparsity of $\mathbf{D}$, is in the order of $4n$, i.e, 300,000 times smaller.

A computationally feasible approach is to calculate $d\mathbf{assc}/d\mathbf{p}$ one row at a time for which an efficient expression is

$$\frac{d\mathbf{assc}}{d\mathbf{p}}[i,:] = \frac{2}{\pi^2}\underbrace{(\mathbf{D}\arctan(\mathbf{UR}(\mathbf{p})[:,i]))'}_{1\times n-1}$$

$$\cdot \underbrace{\mathbf{D}}_{n-1\times n}\left(\left(\mathbf{1}_l \otimes \underbrace{\frac{1}{1+(\mathbf{UR}(\mathbf{p})[:,i])^2}}_{n\times l}\right)\right. \tag{B11}$$

$$\left.\cdot \underbrace{\mathbf{U}}_{n\times m}\underbrace{\frac{d\mathbf{R}}{d\mathbf{p}}[(i-1)m+1:im,:]}_{m\times l}\right),$$

where we have adopted a Matlab like notation such that $\mathbf{A}$[a:b, c:d] denotes the submatrix consisting of rows a to b of the columns c to d. Using Eq. (B11) the

derivatives may be calculated within reasonable memory constraints without the need to duplicate any calculations.

REFERENCES

Aguirre, G. K., Zarahn, E., and 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, doi:10.1006/nimg.1997.0264.

Aguirre, G. K., Zarahn, E., and D'Esposito, M. 1998. The inferential impact of global signal covariates in functional neuroimaging analyses. *NeuroImage* **8:** 302–306, doi:1006/nimg.1998.0367.

Andersson, J. L. R. 1997. How to estimate global activity independent of changes in local activity. *NeuroImage* **6:** 237–244, doi:10.1006/nimg.1997.0302.

Bell, A., and Sejnowski, T. 1995. An information-maximisation approach to blind separation and blind deconvolution. *Neural Comput.* **7:** 1129–1159.

Draper, N., and Smith, H. 1981. *Applied Regression Analysis,* 2nd ed. Wiley, New York.

Fox, P. T., and Mintun, M. A. 1989. Non-invasive functional brain mapping by change distribution analysis of averaged PET images of $H_2^{15}O$ tissue activity. *J. Nucl. Med.* **30:** 141–149.

Friston, K. J., Frith, C. D., Liddle, P. F., Dolan, R. J., Lammertsma, A. A., and Frackowiak, R. S. J. 1990. The relationship between global and local changes in PET scans. *J. Cereb. Blood Flow Metab.* **10:** 458–466.

Friston, K. J., Frith, C. D., Liddle, P. F., and Frackowiak, R. S. J. 1993. Functional connectivity: The principal component analysis of large (PET) data sets. *J. Cereb. Blood Flow Metab.* **13:** 5–14.

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-B., Frith, C. D., and Frackowiak, R. S. J. 1995. Statistical parametric maps in functional imaging: A general linear approach. *Hum. Brain Mapp.* **2:** 189–210.

Herbin, M., Venot, A., Devaux, J. Y., Walter, E., Lebruchec, J. F., Dubertret, L., and Roucayrol, J. C. 1989. Automated registration of dissimilar images: Application to medical imagery. *Comput. Vision Graphics Image Process.* **47:** 77–88.

Holmes, A. P., Poline, J.-B., and Friston, K. J. 1997. Characterising brain images with the general linear model. In *Human Brain Func-*

tion (R. S. J. Frackowiak, K. J. Friston, C. D. Frith, R. J. Dolan, and J. C. Mazziotta, Eds.), pp. 59–84. Academic Press, San Diego.

Johnson, R. A., and Wichern, D. W. 1998. *Applied Multivariate Statistical Analysis.* 4th ed. Prentice-Hall, New Jersey.

Marsden, J. E., and Tromba, A. J. 1981. *Vector Calculus,* 2nd ed. Freeman, San Fransisco.

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

Minoshima, S., Berger, K. L., Lee, K. S., and Mintun, M. A. 1992. An automated method for rotational correction and centering of three-dimensional functional brain images. *J. Nucl. Med.* **33:** 1579–1585.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. 1992. *Numerical recipes in C,* 2nd ed. Cambridge Univ. Press, Cambridge, UK.

Skudlarski, P., Constable, R. T., and Gore, J. C. 1999. ROC analysis of statistical methods in functional MRI: Individual subjects. *NeuroImage* **9:** 311–329, doi:10.1006/nimg.1999.0402.

Spinks, T. J., Jones, T., Bailey, D. L., Townsend, D. W., Grootonk, S., Bloomfield, P. M., Gilardi, M.-C., Casey, M. E., Sipe, B., and Reed, J. 1992. Physical performance of a positron tomograph for brain imaging with retractable septa. *Phys. Med. Biol.* **37:** 1637–1655.

Strother, S. C., Anderson, J. R., Schaper, K. A., Sidtis, J. J., Liow, J.-S., Woods, R. P., and Rottenberg, D. A. 1995. Principal component analysis and the scaled subprofile model compared to inter-subject averaging and statistical parametric mapping: I. "Functional connectivity" of the human motor system studied with [^{15}O]PET. *J. Cereb. Blood Flow Metab.* **15:** 738–753.

Strother, S. C., Sidtis, J. J., Anderson, J. R., Hansen, L. K., Schaper, K., and Rottenberg, D. A. 1996. [^{15}O]Water PET: More "noise" than signal? In *Quantification of Brain Function Using PET* (R. Myers, V. Cunningham, D. Bailey, and T. Jones, Eds.), pp. 378–383. Academic Press, San Diego.

Venot, A., Lebruchec, J. F., Golmard, J. L., and Roucayrol, J. C. 1983. An automated method for the normalisation of scintigraphic images. *J. Nucl. Med.* **24:** 529–531.

Venot, A., Golmard, J. L., Lebruchec, J. F., Pronzato, L., Walter, E., Frij, G., and Roucayrol, J. C. 1984. Digital methods for change detection in medical images. *Proc. Int. Conf. Inform. Processing Med. Imaging,* pp. 1–16. Nijhof, The Hague.

Watson, J. D. G., Myers, R., Frackowiak, R. S. J., Hajnal, J. W., Woods, R. P., Mazziotta, J. C., Shipp, S., and Zeki, S. 1993. Area V5 of the human brain: Evidence from a combined study using positron emission tomography and magnetic resonance imaging. *Cerebral Cortex* **3:** 79–94.

Zarahn, E., Aguirre, G. K., and D'Esposito, M. 1997. Empirical analyses of BOLD fMRI statistics. I. Spatially unsmoothed data collected under null-hypothesis conditions. *NeuroImage* **5:** 179–197, doi:10.1006/nimg.1997.0263.