ELSEVIER

NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 22 (2004) 492–502

Commentary

Statistical parametric mapping for event-related potentials: I. Generic considerations

Stefan J. Kiebel* and Karl J. Friston

Functional Imaging Laboratory, Wellcome Department of Imaging Neuroscience, Institute of Neurology, WC1N 3BG, London, UK

Received 16 October 2003; revised 7 February 2004; accepted 12 February 2004

In this paper, we frame the strategy and motivations behind developments in statistical parametric mapping (SPM) for the analysis of electroencephalogram (EEG) data. This work deals specifically with SPM procedures for the analysis of event-related potentials (ERP). We place these developments in the larger context of integrating electrophysiological and hemodynamic measurements of evoked brain responses through the fusion of EEG and fMRI data. In this paper, we consider some fundamental issues when selecting an appropriate statistical model that enables diverse questions to be asked of the data and at the same time retains maximum sensitivity. The three key issues addressed in this paper are as follows: (i) should multivariate or mass univariate analyses be adopted, (ii) should time be treated as an experimental factor or as a dimension of the measured response variable, and (iii) how to form appropriate explanatory variables in a hierarchical observation model. We review the relative merits of the different options and explain the rationale for our choices. In brief, we motivate a mass univariate approach in terms of sensitivity to region-specific responses. This involves modeling responses at each voxel or space bin separately. In contradistinction, we treat time as an experimental factor to enable inferences about temporally distributed responses that encompass multiple time bins.

In a companion paper, we develop statistical models of ERPs in the time domain that follow from the heuristics established here and illustrate the approach using simulated and real data.

Keywords: Electroencephalogram; Event-related potentials; Statistical parametric mapping

Introduction

There is a clear consensus that the most promising applications of neuroimaging rest upon the integration of different modalities. It has been shown recently that multimodal data acquisition and fusion are useful for gaining additional insight into the neuronal causes of observed hemodynamic and electrophysiological brain responses (Czisch et al., 2002; Goldman et al., 2002; Lemieux et al., 2001; Salek-Haddadi et al., 2002, 2003; Trujillo-Barreto et al., 2001).

In this paper, we work towards one aspect of a particular combination of modalities, electroencephalography (EEG), and functional magnetic resonance imaging (fMRI). An integration of these modalities is promising because of the complementary spatiotemporal resolution of fMRI and EEG. Furthermore, both techniques are the most accessible modalities in research and clinics. An integration of EEG and fMRI is not only potentially useful from a theoretical point of view, but also in practice. One important component of any integrative initiative is the ability to model both types of data in the same mathematical framework to make inferences that are mutually informed. This first component can be seen as a prelude to a full data fusion based on an integrated model for fMRI and EEG.

A candidate for a such a framework is statistical parametric mapping (SPM) (Friston, 2004), which is a mass univariate approach to modeling spatiotemporal neuroimaging data. SPM was originally developed to deal with metabolic or hemodynamic imaging time series, that is, PET, SPECT, and fMRI data. A similar spatiotemporal model can be derived for EEG data. Other groups have already illustrated the usefulness of SPM techniques through applications of SPM to EEG data. For example, Bosch-Bayard et al. (2001) have described an SPM approach to source reconstructed Fourier transformed EEG data. Park et al. (2002) have implemented a procedure that produces statistical parametric maps with source reconstructed EEG data. Barnes and Hillebrand (2003) have applied SPM to source reconstructed MEG data. These developments demonstrate the applicability of SPM to many kinds of neuroimaging data.

This paper deals specifically with the characterization of event-related potentials (ERPs) as measured with the EEG using the same SPM concepts developed for metabolic imaging (Friston, 2004, Friston et al., 2002b). The extension of SPM procedures, to cover ERPs, entails a number of critical choices, which are the subject of this paper. In a companion paper, we describe a temporal model for averaged ERPs, which can be used to test hypotheses about localized effects in peristimulus time or in the peristimulus time or frequency domain. These hypotheses can be tested using the same model. Inference is made in a classical sense based on the t or F statistic. A future communication will extend the model to deal with spatiotemporal data based on the principles described below.

* Corresponding author. Functional Imaging Laboratory, Wellcome Department of Imaging Neuroscience, Institute of Neurology, 12 Queen Square, WC1N 3BG, London, UK. Fax: +44-20-7813-1420.
E-mail address: skiebel@fil.ion.ucl.ac.uk (S.J. Kiebel).
Available online on ScienceDirect (www.sciencedirect.com.)

1053-8119/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2004.02.012

This paper is structured as follows. We will first review, briefly, multimodality integration and outline the overall strategy that we are pursuing. The second section focuses on the different observation and statistical models that could be used for the analysis of ERP data in the light of two key distinctions. These distinctions are between multivariate and mass univariate analyses over space and between treating time as a continuous dimension of the response variable versus a discrete replication factor. The implications of these different approaches for estimation and inference will be described and the motivation for the choices we have made is presented. In the third and final section, we describe the general analysis procedures that ensue. These procedures are based on a hierarchical linear observation model. In a companion paper, the specific operational details and model assumptions for ERP data are presented. These are then applied to synthetic and real data to establish their construct validity in relation to established approaches.

Integration of fMRI and EEG data

Over the past years, there has been an enormous interest in the fusion or integration of electrophysiological and hemodynamical measurements of evoked neuronal responses. The integration of these data (usually fMRI and EEG or MEG) can be classified into three sorts:

- integration through temporal prediction;
- integration through spatial constraints; and
- integration through fusion.

The simplest approach, integration through temporal prediction, is to use one modality to predict the other. A significant mutual information or correlation between the two modalities gives one modality access to the greater spatial or temporal resolution of the other. Perhaps the most compelling example of this approach is the use of EEG seizure activity as predictors or explanatory variables for concurrently recorded fMRI responses (Lemieux et al., 2001; Salek-Haddadi et al., 2002). In this instance, the underlying cause of the EEG measures such as spikes or epochs of seizure activity can be characterized with the spatial precision of fMRI if the EEG metrics predict regional hemodynamic responses significantly.

The second approach, integration through spatial constraints, is inherently asymmetrical in the sense that one modality provides priors or constraints on the estimation of the generators or causes of the other modality. The best example of this approach is the use of spatial information in fMRI activation profiles as spatial constraints on equivalent dipole or distributed estimates of EEG sources, for example, Baillet and Garnero (1997); Dale et al. (2000); Phillips et al. (2002a); Toma et al. (2002). This allows the explicit use of the spatial precision of fMRI where the mixing or integration of the data occurs within a Bayesian estimation framework (Dale et al., 2000). Here the fMRI data are usually treated as fixed and known priors, enabling conditional estimates of EEG responses. Unlike integration through prediction, the temporal information in the fMRI is usually discounted when constructing the spatial priors.

Neither integration through prediction or through constraints represents a true integration of the data, in the sense that there is no common temporal forward model that links the underlying neuronal dynamics of interest to both the measured hemodynamic and electrical responses. Approaches that use forward or generative models fall into the integration or bifusion class. In this framework, a forward model is developed that can explain the electrophysiological and blood oxygen level dependent (BOLD) consequences of any change in synaptic activity. The fMRI and EEG data are then used, conjointly, to estimate the parameters of this model. The estimation is usually Bayesian in nature; however, the priors are not based upon the response variables in either modality (Trujillo-Barreto et al., 2001).

Within this rough taxonomy, we have chosen to pursue integration by fusion and have been working on simple forward models for the hemodynamic response (Friston et al., 2002a) that could be placed alongside (temporal) forward models for EEG to provide a complete generative model for both fMRI and EEG data. We are currently exploring the use of mean field treatments of neuronal ensemble dynamics as the basis for both the hemodynamic and EEG forward models (David and Friston, 2003). The compound forward model is framed in terms of differential equations and constitutes a dynamical model. We already have in place Bayesian estimation procedures for these sorts of models (Friston et al., 2002b). In the next subsection, we discuss the work presented in this paper in the light of developing a simple, robust, and easily useable multimodality fusion framework.

Background work

The strategy that we are pursuing for the integration by fusion has three phases. The first phase will be summarized briefly here and is elaborated elsewhere (Phillips et al., 2002a). This and a companion paper detail the second phase, which is to bring the anatomically reconstructed sources into the same estimation and inference machinery currently used for fMRI data analysis. The third phase, the development of an invertible forward model for EEG or ERPs, will be described in subsequent communications. This final phase will entail a framework in which EEG and fMRI data are combined in the same temporal forward model.

The first phase was to transform raw multichannel EEG data into distributed source estimates that conform to the same standard anatomical space used for fMRI data acquisition and analysis. This has been accomplished through the introduction of two techniques to the source reconstruction problem. It should be noted that our objective precludes the use of equivalent dipole models irrespective of their utility in other applications. Rather we use distributed source solutions to the inverse problem. Our estimation procedure uses a Bayesian framework to compute conditional estimates of current density at each voxel and each peristimulus time point. This conditional or maximum a posteriori (MAP) estimate obtains from a relatively simple weighted or regularized least squares estimator under Gaussian assumptions about the observation error. The two techniques that have enabled this approach are (i) the use of anatomically informed basis functions and (ii) the use of restricted maximum likelihood (ReML) to identify the regularization parameters or variance parameters that moderate the relative importance of likelihood and multiple prior terms.

Informed basis functions were introduced to the neuroimaging literature for fMRI and PET in Kiebel et al. (2000) and taken into the EEG field by Phillips et al. (2002a). Informed basis functions represent a method of imposing priors on the source estimators that is computationally very efficient and has a clear motivation from a point of view of information theory (hence informed). The basic idea is to express the spatially distributed source estimate, at any point in time, in terms of a linear mixture of anatomical basis functions. The basis functions have generally a much smaller

dimensionality than the number of voxels and are chosen to preserve the most information after the prior source distribution is projected onto the basis set. In practice, the basis functions are obtained by taking the major eigen vectors or principal components of the prior covariance matrix of the sources. This prior covariance matrix embodies simple but tenable assumptions. For example, sources must arise in grey matter and show local spatial coherence. The importance of informed basis functions for the distributed source estimation problem in EEG is that the implicit dimension reduction enables the use, in practice, of iterative algorithms that are required for the ReML estimate of the prior covariance components (or equivalently regularization parameters). This is the second new technique mentioned above.

Informed basis functions impose hard constraints in the sense that solutions are disallowed if they are not spanned by the basis set. This is equivalent to using priors that enforce the estimates to be zero in these regions through a mean of zero specified with infinite precision (precision is the inverse of the variance.) In EEG, there are some other potential priors that can be embodied in the conditional estimation whose relative importance may not be easy to establish. Examples of these prior constraints, usually expressed in terms of prior variances, include the grey matter and spatial coherence constraints used to form the basis functions, depth priors, or indeed spatial priors from fMRI experiments used to implement the 'integration through constraints approach' mentioned in the previous section. The relative magnitudes of these prior covariance constraints are controlled by a set of variance parameters. These can be estimated with ReML using an expectation–maximization (EM) algorithm. These procedures can be construed as a parametric empirical Bayesian (PEB) approach to the conditional estimators. Interestingly, the relative values for the variance parameters can change dramatically depending upon the source configurations and experimental design as discussed in Phillips et al. (2002b). As such, the use of empirically determined variance parameters may confer a face validity on the conditional estimators that would have been lost with the use of predetermined and arbitrary regularization or hyperparameters.

The remainder of this paper is about how the source reconstructed EEG data can be analyzed using the same procedures currently applied to fMRI and covers the key issues that attend the second phase of our program.

Statistical models for source reconstructed EEG time series

There are some fundamentally different alternatives that present themselves when choosing an appropriate statistical model for the analysis of EEG time series. We restrict ourselves to the analysis of (averaged) ERP data (Fig. 1). By ERP data, we mean averaged event-related time courses (Rugg and Coles, 1995), where each of these time courses has been averaged within subject and trial type (condition) to provide one peristimulus time series for each trial type and each subject.

Reconstruction using informed basis functions and ReML variance parameter estimates, the usual artifact removal or correction and averaging are all considered here to be preprocessing or reconstruction procedures that parallel the reconstruction, realignment, and spatial normalization of fMRI data. After reconstruction, the ERP data constitute a time series of three-dimensional images over peristimulus time bins. These images may be scalar images corresponding to current source density or three-variate images

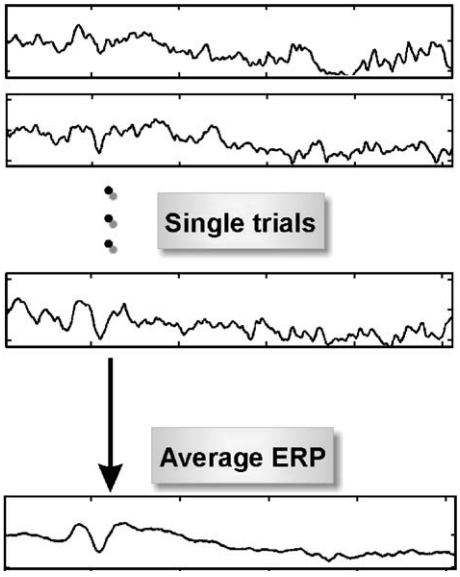

Fig. 1. An average ERP is estimated over single trials for each trial type and each subject.

retaining the source orientation information. Here we implicitly assume that we are dealing with univariate or scalar response variables at each voxel and time bin. Note that this discussion does not, in principle, depend on using any specific method or implementation of source reconstruction or preprocessing. Any EEG or ERP software package, which transforms EEG data to source reconstructed ERP data, can be used in conjunction with the methods we describe below. This includes spatiotemporal approaches that exploit temporal constraints, for example, Darvas et al. (2001); Galka et al. (2003); Phillips et al. (2004).

The approaches we consider can also be applied to ERP data, which have been projected onto the scalp surface. Of course, this two-dimensional representation does not allow for a full integration with fMRI data but might be the only way to proceed when source reconstruction is not feasible (Fig. 2).

The three key questions addressed in this paper are as follows:

1. Whether to use mass univariate or multivariate models over space.
2. How does peristimulus time enter the model.
3. How to form appropriate explanatory variables in a hierarchical observation model used to estimate and make inferences about experimentally induced effects.

Notation

At this point, we introduce the notation and some of the variables used in the following sections. We assume that the source reconstructed data consists of three-dimensional images with M voxels. Each image contains data for one time bin. In other words, each voxel in three-dimensional space contains one time series over peristimulus time. We assume that we have measured the same trial types in each subject and all ERP data have the same number of time bins.[1] The number of subjects is N_{subjects}, the

[1] These assumptions are not strictly necessary but simplify our notation.

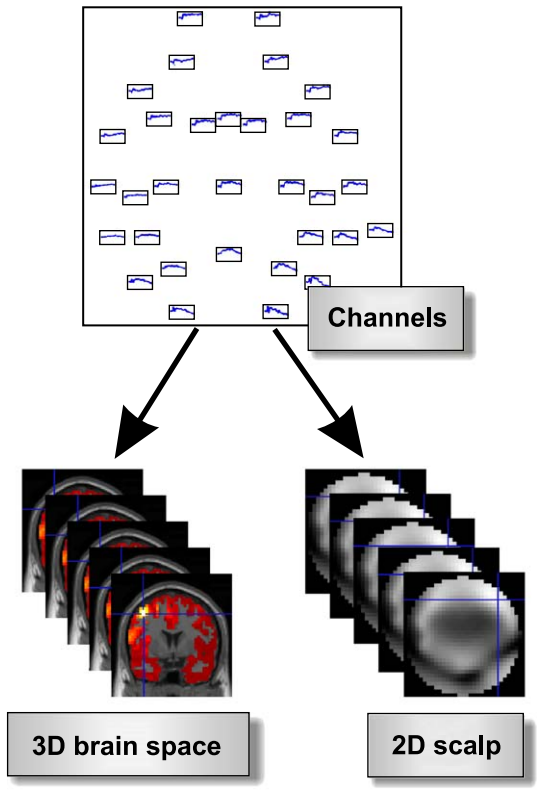

Fig. 2. For each trial type and subject, the average ERP, for each channel, is projected to either three-dimensional brain space (source reconstruction) or interpolated on the scalp surface. This results in either 3-D and/or 2-D image time series.

number of trial types is N_{types}, and the number of time bins per ERP is N_{bins}. The total number of images is given by $N = N_{subjects}N_{types}N_{bins}$, see also Fig. 3.

Multivariate versus mass univariate

In this subsection, we discuss potential options available to model spatial correlations of the error (nonsphericity). Accounting for these correlations is important for valid inference. To avoid confusing this issue with the temporal issues, we will assume that the data compose one point in peristimulus time. The difference between a multivariate and a mass univariate approach is the difference between treating the data as a single M-dimensional response or M univariate observations. In other words, do we consider each image as a single observation or as a family of single voxel observations?

Statistical parametric mapping (SPM) represents a mass univariate approach while something like a multivariate analysis of variance (MANOVA) would constitute a multivariate approach (cf., Yandell, 1997, pp. 275 ff.). Although there are some fundamental differences between the two approaches, they are more closely related than may appear at first glance. This is because we can convert a multivariate observation model into a mass univariate observation model by simply rearranging the matrix formulation. Consider the multivariate linear model with a response variable y comprising $N_d = N_{subjects}N_{types}$ images (e.g., images of current source density at 100 ms after presentation of a visual stimulus) over M voxels:

$$y = X\beta + \epsilon \quad (1)$$

where y is the $N_d \times M$ data matrix, X is an $N_d \times P$ design matrix, β is a $P \times M$ parameter matrix and ϵ is an $N_d \times M$ error matrix, where each row of ϵ is assumed to be sampled independently from the same multivariate normal distribution with zero mean. The classical analysis of this model (MANOVA) proceeds by computing sample covariance matrices of the transposed data y^T and the residuals. Wilk's lambda (Chatfield and Collins, 1980, p. 148) is used to test for the covariance of the treatment effects, relative to the covariance of the residuals and after transformation, compared with an F distribution. The important point about the MANOVA is

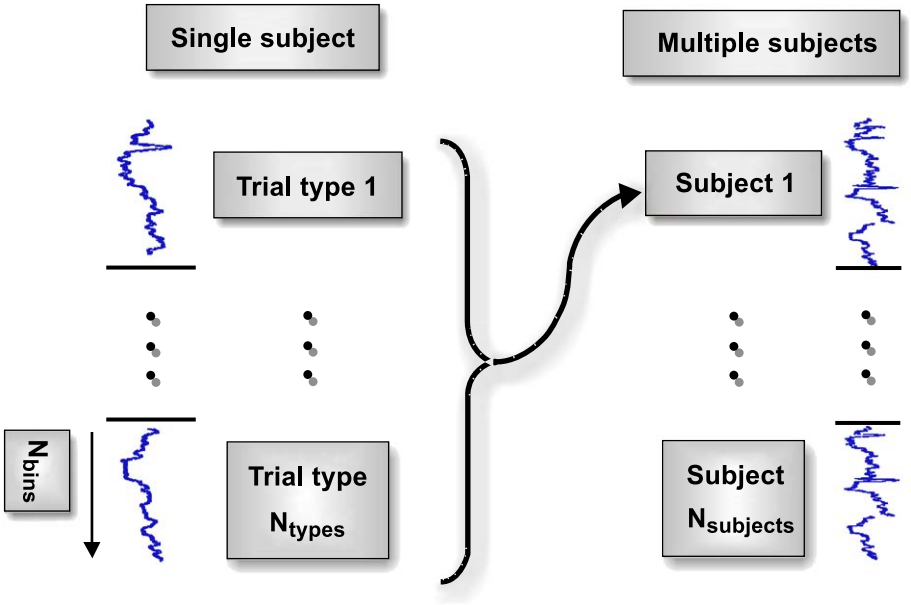

Fig. 3. Measured data. For each voxel, one has $N_{subjects}N_{types}$ ERPs, giving $N = N_{subjects}N_{types}N_{bins}$ data points.

that the errors are assumed to be correlated over voxels and that this correlation is taken into account when deriving the statistic.

It is helpful to understand the implicit assumptions about spatiotemporal nonsphericity in MANOVA by considering a more general univariate formulation: Eq. (1) can be rearranged into a univariate model by stacking the columns of the response matrix on top of each other to form a response vector and forming an augmented design matrix using a Kronecker tensor product. The parameter and error matrices are similarly vectorized.

$$\mathrm{vec}(y) = (I_M \otimes X)\ \mathrm{vec}(\beta) + \mathrm{vec}(\epsilon) \tag{2}$$

where $\otimes$ denotes the Kronecker tensor product and vec(·) is the operator that stacks a matrix columnwise to produce one column vector. The matrix I_M is the $M \times M$ identity matrix.

The essential difference between Eqs. (1) and (2) lies in the, hitherto, unspecified assumptions about the error terms on the right hand side. Generally, when using MANOVA, the covariance matrix of the transposed error has $M \times M$ elements. A covariance matrix is symmetrical and therefore contains $M(M + 1)/2$ unknown elements or, in our case, variance parameters. Each variance parameter controls the (co)variance between the error at voxel i and the error at voxel j. These variance parameters must be estimated. In MANOVA, this is done using the residuals of the fitted model.

Similarly, in Eq. (2), the error covariance matrix has dimensions $N_d M \times N_d M$ and is fully specified by $M(M + 1)/2$ variance parameters (remember that we assume that each row of ϵ is sampled from the same distribution):

$$\mathrm{Cov}[\mathrm{vec}(\epsilon)] = \sum_{i=1,\ldots,M} \sum_{j=1,\ldots,M} \lambda_{ij} Q_{ij}$$

$$Q_{ij} = \tilde{Q}_{ij} \otimes I_{N_d} \tag{3}$$

where $\tilde{Q}_{ij}$ is an $M \times M$ matrix with $\tilde{Q}_{ij}(i,j) = \tilde{Q}_{ij}(j,i) = 1$ and zeros elsewhere. The λ_{ij} is the variance parameter that can be estimated using restricted maximum likelihood (ReML).

However, one does not need to estimate all variance parameters in an unconstrained way. The point made by Eq. (3) is that it can accept constraints on the variance parameters. Such constraints allow us to use (and estimate) fewer variance parameters. For instance, we could assume that covariances depend only on the spatial distance between voxels:

$$\mathrm{Cov}[\mathrm{vec}(\epsilon)] = \sum_{d=0,\ldots,k} \lambda_d Q_d$$

$$Q_d = \tilde{Q}_d \otimes I_{N_d} \tag{4}$$

where $\tilde{Q}_d$ is an $M \times M$ matrix with $\tilde{Q}_d(i,j) = 1$, $\forall\ i,j{:}abs(i - j) = d$ and zero elsewhere. Variable k is some maximum distance between any pair (i,j) of voxels. This specific constraint would reduce the number of variance parameters dramatically from a few hundred (depending on M) to a handful $(k + 1)$.

The use of constraints is critical because in neuroimaging, the number of images N is typically much smaller than the number of voxels M, that is, $N \ll M$. It would be impossible to estimate all the variance parameters (Eq. (3)) from the data without using constraints. This is the reason why one cannot apply a MANOVA to neuroimaging data directly. Instead, one reduces its dimensionality by using a principal component analysis (PCA) or a similar device (Friston et al., 1996; Worsley et al., 1997).

In contradistinction to multivariate approaches, mass univariate approaches consider the data at each voxel i in isolation, that is,

$$y_i = X\beta_i + \epsilon_i \tag{5}$$

ignoring the spatial correlations (at this stage). Note that ordinary least squares (OLS) estimates of β are identical for (Eqs. (1), (2), and (5). This enables us to estimate, for each voxel i, P parameters β_i and one variance parameter λ_i independently of other voxels.

The critical issue for mass univariate approaches is how to deal with the spatial covariances that have been ignored in Eq. (5). The impact of spatial covariances is accommodated at the inference stage through adjusting the P values associated with the SPMs (images of statistics formed from the parameter and variance parameter estimates). This adjustment or correction uses random field theory (RFT) and assumes that the error terms conform to a good lattice approximation to an underlying continuous spatially extended process (Worsley et al., 1996). In other words, it assumes that the errors are continuous and spatially extended. The RFT correction plays the same role as a Bonferroni correction (Yandell, 1997, pp. 93 ff.) for discrete data. The power of the RFT approach is that valid inference needs only one spatial covariance parameter for each voxel. This is the smoothness, which is the determinant of the covariance matrix of the spatial first partial derivatives of the error fields (Worsley et al., 1999). As with the MANOVA, these are estimated using the residuals about the fitted model. The RFT correction does not assume spatial stationarity of the errors or that the spatial autocovariance function is Gaussian. All it assumes is that the error fields are continuous (i.e., smooth). The important distinction between the SPM mass univariate approach with RFT correction and the equivalent MANOVA approach, with a full covariance matrix, is that the former only requires 2 M (M spatial and M temporal) variance parameters whereas the latter requires $M(M + 1)/2$ variance parameters.

A further difference between SPM and multivariate approaches is that SPM inferences are based on regionally specific effects as opposed to spatially distributed modes. In SPM, classical inference proceeds using the voxel-specific t or F value, having adjusted the P values using an RFT correction to accommodate the fact that M voxels have been tested, whereas in multivariate statistics inference is made about effects over all voxels. Rejection of the null hypothesis in MANOVA allows one to infer that there is a treatment effect in some voxel(s) but it does not tell one where. In principle, if the treatment effect was truly spatially distributed, SPM would be much less sensitive than MANOVA. However, the aim of functional neuroimaging is to establish regionally specific responses. By definition, diffuse spatially distributed responses are not useful in trying to characterize functional architectures. Furthermore, the goal of fMRI or EEG integration is to endow electrophysiological measures with a spatial precision. This goal is met sufficiently by mass univariate approaches.

In conclusion, the special nature of neuroimaging data and the nature of the regionally specific questions that are asked of these spatially extended continuous observations point clearly to the adoption of mass univariate approaches and the use of RFT to accommodate spatial nonsphericity. This conclusion is based upon the fact that for spatially continuous data, we only need the covariances of the first partial derivatives of the error at each voxel, as opposed to the spatial error covariances among all voxels.

Secondly, the nature of the hypotheses we wish to test is inherently regionally specific.

The temporal dimension

Having established the utility of an SPM-like approach to the analysis of each voxel time series, we now have to consider whether time is a fourth dimension of the response variable or a discrete series of observations over time bins. The RFT correction has been generalized to any arbitrary number of dimensions by Worsley et al. (1996). Many interesting applications of high-dimensional SPMs exist, for example, augmenting space with scale–space dimensions (Siegmund and Worsley, 1995) or even more abstract dimensions in computational neuroanatomy where one has to deal with issues of statistical flattening on high-dimensional manifolds (Worsley et al., 1999). In principle, one could simply treat time as another dimension of the response variable to produce four-dimensional SPMs that span the standard anatomical space and peristimulus time. These SPMs would have activations or regions above the threshold (excursion sets) that covered a cortical region and a temporal domain following the stimulus. This would allow both for anatomical and temporal specificity of inferences using adjusted P values. The appeal of this approach echoes the points made in the previous section. The nice thing about creating four-dimensional (over space and time) SPMs is that temporal correlations or nonsphericity among the errors over time can be dealt with in a parsimonious way, at the inference stage, using random field corrections. This means that one only needs to estimate the temporal smoothness at each time bin as opposed to the temporal correlations over all time bins. The assumption underpinning the RFT is clearly tenable.

The alternative to treating time as a dimension is to assume that it is an experimental factor with the same number of levels as there are bins in peristimulus time. In this instance, one has to estimate the temporal variance parameters by analogy with the spatial variance parameters in Eq. (3). In other words, one has to estimate the temporal correlations of the error to make an appropriate nonsphericity adjustment to ensure valid inference.[2] This is a solved problem in the context of fMRI where serial correlations have been the subject of recent work (Friston et al., 2002a; Penny et al., 2003; Purdon and Weisskoff, 1998; Woolrich et al., 2001). ReML estimation procedures based upon expectation–maximization (EM) allow the temporal variance parameters to be specified in terms of separable covariance components (Friston et al., 2002b). This technique is used in the current version of the SPM software (SPM2). It allows for a flexible model of serial correlations, where the estimated nonsphericity is used either to whiten the data (ML estimates) or in the estimation of the degrees of freedom (OLS estimates).

In fMRI, one typically assumes that the temporal error process is stationary but the algorithms used by the SPM software can accommodate nonstationariness. These might arise, for example, if between subjects, the early or endogenous components of evoked EEG responses were less or more variable than late components. This would lead to different frequency structures in the error early, relative to late, in peristimulus time. This would induce non-stationarity over time.

If it is possible to implement either approach given current methodologies, which would be the most suitable? The answer here is motivated by the sorts of questions that are asked of ERP data.

In the application of SPM to fMRI data, it is not possible to make any inferences about the spatial extent of activation foci in SPMs. This shortcoming translates, in the context of space–time SPMs, into precluding inferences about the temporal extent of evoked responses. This is a severe limitation for ERP studies that would preclude, for example, inferences about differential latencies among trial types or groups. In short, inferences about ERPs pertain to evoked transients, which have a temporally extended form. To enable these inferences, it is necessary to specify hypotheses that encompass many time bins. This precludes the use of space–time SPMs where contrasts can only be specified for each voxel and time bin. In the spatial domain, we are interested in region- or voxel-specific inferences because an activation in one part of the brain does not have any quantitative meaning in relation to activation in a different structure. Conversely, the relative sizes of responses over time bins at a given voxel are meaningful because they define the form of the evoked transient. Therefore, it is necessary to compare responses over time explicitly.

This can only be done by treating time as an experimental factor. There are also practical reasons to favor an explicit temporal model. For example, it would be rather difficult and inconvenient to visualize significant effects using four-dimensional SPMs.

In conclusion, given the sorts of questions asked in ERP research, we have chosen to treat time as a factor and forgo the natural appeal of four-dimensional SPMs.

Modeling the correlations

The preceding two sections addressed two key issues: (i) whether to use multivariate or mass univariate models and (ii) which model to use in the temporal domain. It may help the reader to observe that both issues touch upon the underlying question of how to model the error covariance matrix of spatiotemporal data.

By assuming a factorization of the spatial and temporal domain, we were able to separate modeling of the spatial and temporal correlations. The discussion about multivariate versus mass univariate approaches is about modeling spatial correlations among voxels. Here, we chose the mass univariate approach. Using this approach, at the estimation stage, the spatial correlations are irrelevant. To take the spatial correlations into account, at the subsequent inference stage, we use the Gaussian random field approach, which means spatial correlations do not enter into the estimation. However, temporal correlations between ERP time bins are an integral part of the observation model and can have an effect on the estimates. In Fig. 4, we summarize our modeling approach.

The remainder of this paper focuses on the temporal models used in spatial mass univariate approaches, in particular their hierarchical nature and special issues of nonsphericity.

The form of temporal observation models

By electing to treat time as a factor, we create an interesting distinction between explanatory variables that model time effects and experimental design variables that model the treatment effects of other experimental factors (e.g., group differences). As men-

[2] This applies if one uses OLS parameter estimates. For maximum likelihood (ML) estimates, temporal correlations have to be estimated to whiten the data.

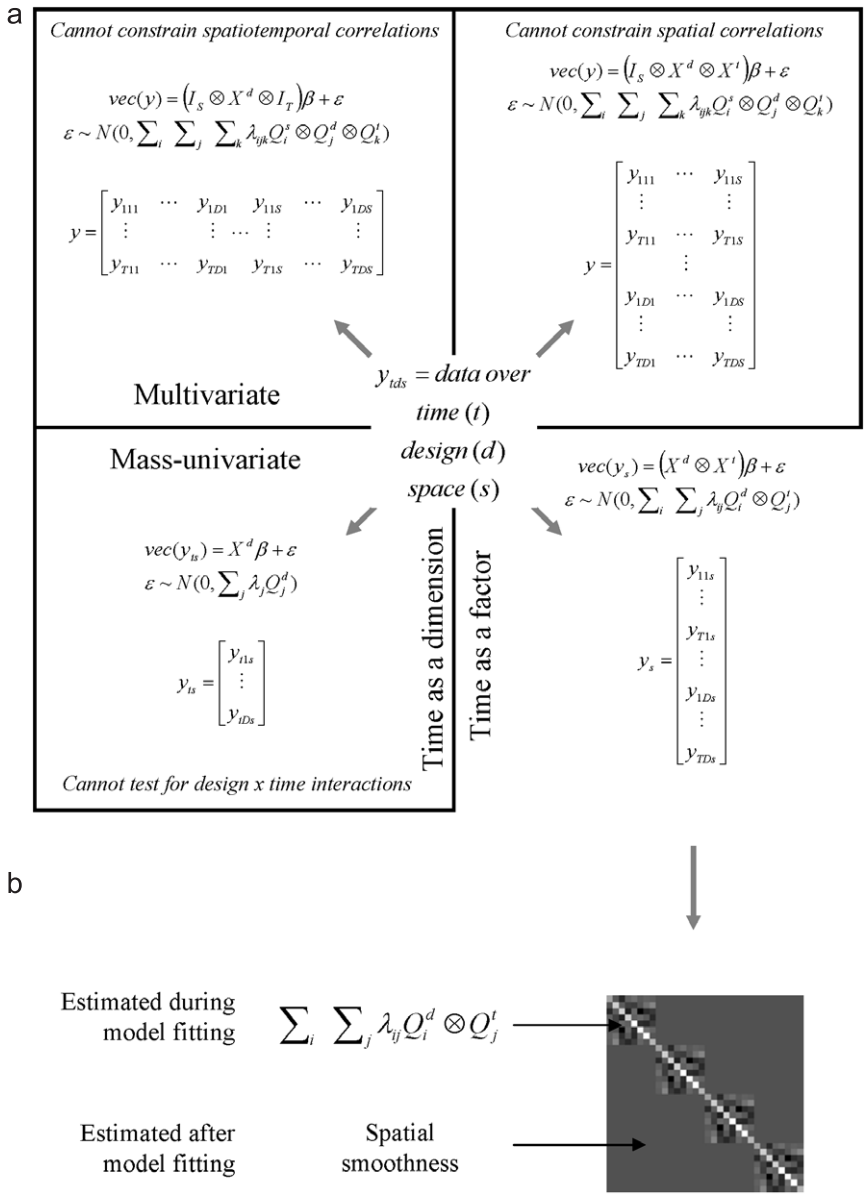

Fig. 4. Schematic demonstrating the different formulations of observation models implied by the multivariate mass univariate and temporal dimension-factor distinctions. (a) The upper panels represent multivariate formulations in which there is no opportunity to place continuity constraints on the spatial or (upper left) spatiotemporal correlations. By treating each time bin or brain location as a separate component of the observed response, one effectively creates a model, which is not informed about the continuity of these responses over space and time. The mass univariate model in the lower left panel could, in principle, embody continuity constraints in both space and time through the application of random field theory. However, there would be no opportunity to assess design by time interactions. This means one could not test explicitly for differences in evoked ERP wave-forms or changes in evoked oscillations. The lower right panel represents the formulation we have chosen to adopt, in which time is treated as a factor but space is not. (b) Correlations among the errors, induced by design and expressed over time, are estimated during model fitting. Correlations over space are estimated post hoc and are used to adjust P values for inferences about spatially or regionally specific effects.

tioned above, we are assuming that each peristimulus time series represents one particular trial type within an EEG session. The temporal explanatory variables model evoked responses in each subject- and trial type-specific ERP. Further explanatory variables model treatment effects among trial types and/or sessions or subjects. We will refer to the temporal explanatory variables as temporal effects and to the experimental design variables as experimental effects. These are encoded by the design matrices X^t and X^d, respectively.

This natural distinction points to a framework in which ERP data can be modeled effectively: the linear hierarchical model (Friston et al., 2002b). A hierarchical model can be used to decompose the data into within-ERP (temporal effects) and between-ERP components (experimental effects). There is an important difference between the sorts of inferences that can be made with ERP data. This distinction rests upon the form of the hierarchical observation model and the level, in this hierarchy, at which the inference is made. Usually, these hierarchical observa-

tion models have two levels, engendering the distinction between fixed and random effects analyses. In two-level hierarchical observation models, the response at the first level is caused by first-level parameters that themselves are modeled as random or stochastic variables at the second level, that is,

$$y = X^{(1)}\beta^{(1)} + \epsilon^{(1)}$$

$$\beta^{(1)} = X^{(2)}\beta^{(2)} + \epsilon^{(2)} \quad (6)$$

where the data y consists of one column vector of length $N_{\text{subjects}}N_{\text{types}}N_{\text{bins}}$ (Fig. 3). The ERP data are ordered such that trial type-specific ERPs of an individual subject are next to each other. That is, $X^{(1)} = I_{N_{\text{subjects}}} \otimes I_{N_{\text{types}}} \otimes X^t$, where X^t is a first-level design matrix, embodying temporal effects, for a single ERP. The second-level design matrix is given by $X^{(2)} = X^d \otimes I_{N_{pt}}$, where, for example, $X^d = 1_{N_{\text{subjects}}} \otimes I_{N_{\text{types}}}$ (an averaging matrix), N_{pt} is the number of columns in X^t, and 1_N denotes a column vector of ones of length N. In a companion paper, we will actually specify some matrices X^t and apply them to the analysis of simulated and real ERP data. Here, we focus on general properties of the linear hierarchical model and discuss, below, the specification of X^t. The model in Eq. (6) reflects the natural hierarchy of observed ERP data: At the first level, the observations are modeled in a subject- and trial type-specific fashion, that is, within ERP peristimulus time, where $\epsilon^{(1)}$ is the observation error. At the second level, we model the parameters $\beta^{(1)}$ over subjects and trial types. In other words, the ensuing hierarchy models temporal within-ERP effects at the first and between-ERP effects (over subjects and trial types) at the second level. At this level, the error $\epsilon^{(2)}$ represents between-subject variability not modeled in X^d.

Hierarchical observation models

The structure of the next four subsections is as follows. First, we discuss the reasons for choosing hierarchical models and outline the equations for a two-level model. We then briefly consider the estimation of model parameters and making inferences. Finally, we illustrate the modeling of temporal effects using wavelets. In a companion paper, we specify, in detail, the model, that is, the form of the design matrices and the nonsphericity, and discuss useful applications.

There are several reasons why hierarchical linear models are useful for characterizing ERP data. To start with, they afford a substantial latitude for modeling and hypothesis testing ERP differences. Note that the first-level design matrix $X^{(1)}$ defines a projection onto some (sub)space of the data. Each parameter is associated with a specific dimension of this subspace. One can consider all kinds of transforms for ERP data, where the Fourier transform or the wavelet transform is just two examples. The Fourier transform is useful for making inference about power in some frequency range. The wavelet transform is an appropriate choice when making inferences about localized time frequency effects. By using contrasts of the first- or second-level parameter estimates, we can test for effects localized in peristimulus time and within certain frequency ranges. The important point about the two-level hierarchical model is that it enables different (linear) transforms at the first level. It will also be shown, in a companion paper, that the conventional ERP model, an analysis of variance (ANOVA) on peristimulus time window averages, can be formulated in the context of a two-level model, although this is a rather restricted application.

A further motivation for hierarchical models is that they finesse the parameterization of nonsphericity. In this context, the error is decomposed into level-specific components. From Eq. (6), we see that the first-level error $\epsilon^{(1)}$ is the error about the fitted response, that is, the observation error. The second-level error $\epsilon^{(2)}$, with an averaging design component X^d, is the deviation of each first-level parameter from the average value for a particular trial type in a specific subject. This error arises from between-subject variability. The distinction between these two components allows the decomposition of the overall error into two partitions, the within- and between-subject variability. These partitions have distinct nonsphericity structures, which can be modeled using level-specific variance components. These facilitate robust and accurate error estimation, which in turn is necessary for valid statistics.

An advantage of level-specific error components is that one can make inferences at either level. This relates to the distinction between fixed and random effects analyses. For example, in Eq. (6), $\beta^{(2)}$ corresponds to the average response over subjects for a particular trial type. The variability of $\beta^{(2)}$, that is, the covariance matrix of $\beta^{(2)}$, can be derived by first collapsing the two-level model to one level:

$$y = X^{(1)}X^{(2)}\beta^{(2)} + X^{(1)}\epsilon^{(2)} + \epsilon^{(1)} \quad (7)$$

Using an ordinary least squares estimator, the covariance matrix of the estimated $\beta^{(2)}$ ($\hat{\beta}^{(2)}$) is

$$\text{Cov}(\hat{\beta}^{(2)}) = (X^{(1)}X^{(2)})^{-}\text{Cov}(X^{(1)}\epsilon^{(2)} + \epsilon^{(1)})(X^{(1)}X^{(2)})^{-T} \quad (8)$$

where X^{-} denotes the generalized inverse of X. Eq. (8) says that the covariance of the estimated second-level parameters is given by the projected error covariance of the collapsed model (Eq. (7)). The variability of the $\hat{\beta}^{(2)}$ is a mixture of the variability of the errors from both levels. Therefore, $\hat{\beta}^{(2)}$ not only varies because of intersubject, but also because of within-subject variability.

Inference at the first level involves a contrast of the first-level parameters $\beta^{(1)}$. Its covariance matrix is a function of the covariance matrix of $\beta^{(1)}$, that is, $\text{Cov}(\beta^{(1)}) = X^{(1)-}\text{Cov}(\epsilon^{(1)})X^{(1)-T}$. The resulting statistic can be used to test whether the estimated responses were significant in relation to the precision with which we observed them. This statistic is used to implement a fixed effects analysis. Fixed effects analyses are typically applied to test the estimated size of a subject-specific effect against its variance. These sorts of analyses are useful in single case studies, for example, in neuropsychology.

In neuroimaging, the conventional sort of inference is based upon a random effects analysis where one makes an inference about the population from which subjects were sampled. Contrasts at the second level can be used to test for effects across subjects. To form a statistic, one needs to test the contrast in relation to the variability of the second-level parameters (Eq. (8)). One speaks of a random effects analysis because the $X^{(1)}\epsilon^{(2)}$ term in Eq. (8) is treated as a random variable.

Estimation

Having specified the hierarchical form of the observation model, one wants to estimate model parameters and make inferences about effects. Inferences are based on a t or F statistic, which is formed from contrasts of the parameter and variance parameter estimates.

The parameters in Eq. (6) can be estimated using ordinary least squares (OLS) or maximum likelihood (ML). The variance parameters are obtained with restricted maximum likelihood (ReML) method (Harville, 1977). ReML is an iterative approach that provides an unbiased estimator of the variance parameters. We refer to Friston et al. (2002b) for a detailed account of the ReML method.

Contrasts and inference

Given the model and estimated parameters, we can make classical inferences about effects in peristimulus time or in the peristimulus time or frequency domain. With a linear model, one can use the standard t or F statistic (Worsley and Friston, 1995). The ReML variance parameter estimation allows one to either estimate the effective degrees of freedom for OLS estimators or to prewhiten the data y to form ML estimates. As we will show in a companion paper, contrasts in combination with t and F statistics can be used for testing a broad range of hypotheses. Among these are conventional tests for amplitude changes located in peristimulus time or, more generally, tests for evoked power localized in the peristimulus time or frequency domain.

Temporal basis functions

Basis function selection is a critical issue for ERP data because the response to a stimulus can vary considerably between subjects and the nature of that variation must be encompassed by the basis functions in X^t. These functions define signal relative to observation noise. Their specification entails prior assumptions about what constitutes an evoked response and how it is generated. The basis function coefficients $\beta^{(1)}$ summarize the trial type- and subject-specific response, where their variation over subjects is encoded by $\mathrm{Cov}(\epsilon^{(2)})$. From an empirical Bayesian perspective $\mathrm{Cov}(\epsilon^{(2)})$ represents the prior covariance of the parameters generating responses. The expression of these parameters in the response is defined by $X^{(1)}$. These considerations point to the importance of $X^{(1)}$, the temporal effects, and the utility of a two-level model that disambiguates observation noise $\epsilon^{(1)}$ from between-subject variation in physiological parameters that lead to the error $\epsilon^{(2)}$.

A lower bound on the number of basis functions in X^t is afforded by the dimensionality of the dynamical system generating evoked transients. In general, this will be much less than the number of time bins. There are several heuristic arguments that suggest the dimensionality of ERP generators may be quite small. For example, in nonlinear dynamical systems, a correlation dimension of >10 is usually indistinguishable from stochastic noise and has no interesting structure. Typically, the correlation dimension of generators estimated from continuously recorded EEG suggests that dimensionality ranges from two to three in severe pathology up to about seven or eight in the normal EEG. There are theoretical arguments

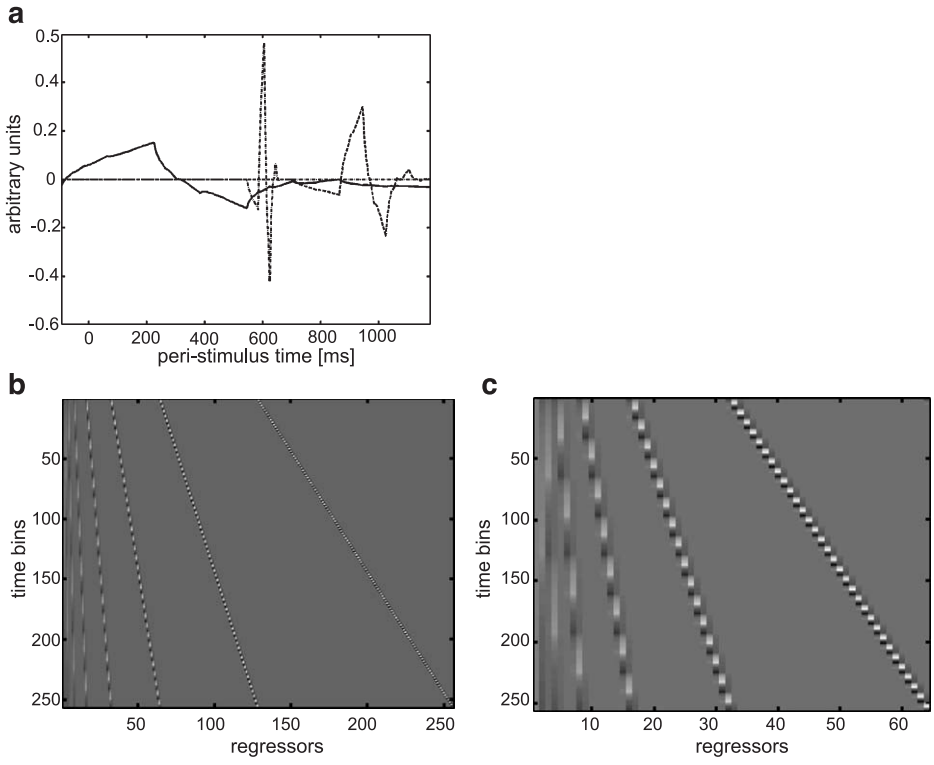

Fig. 5. Illustration of the wavelet model. We used Daubechies wavelets of order 4. (a) Three wavelet basis functions of different scale (frequency range) levels at different peristimulus times, (b) complete wavelet transform in design matrix form (256 regressors), and (c) truncated wavelet design matrix with the two highest scales removed (64 regressors).

that point towards a smaller dimensionality of dynamical neural systems. One of these is the existence of low-dimensional synchronization manifolds that arise when nonlinear dynamical systems are coupled into an ensemble (Breakspear, 2002). For the sake of argument, if we assumed that the dimensionality of the cell assemblies mediating ERPs was eight, then the impulse response function to any perturbation could, to first order, be described by a mixture of eight complex basis functions. However, this number is only a lower bound on the appropriate number of basis functions whose form will be generally unknown. We will defer a discussion of prior assumptions about how responses are generated and use either complete or nearly complete transforms.

Ideally, the data transformation implicit in $X^{(1)}$ should finesse the comparison of time frequency components at the second level. The wavelet transform is an obvious candidate. One of the mathematically attractive features of wavelets is that they provide orthogonal design matrices.[3] Orthogonality affords greater computational efficiency. This is because an orthogonal design matrix at the first level tends to lead to diagonal covariance matrices at the second level (assuming the wavelet coefficients are a priori independent). This means variance component and parameter specification and estimation is simpler than in the nondiagonal case.[4] For an introductory text to wavelets, see Press et al. (1992, chap. 13) or Gershenfeld (1998, chap. 11). For a mathematically more stringent description of wavelets in the context of filter theory, see Strang and Nguyen (1996).

Other groups have applied the wavelet transform to EEG and ERP data to detect changes in a specific time frequency window (Basar et al., 1999; Thakor et al., 1993; Trejo and Shensa, 1999). It has also been shown that the wavelet transform is useful when making inferences about power localized in the peristimulus time frequency domain, for example, Tallon-Baudry et al. (1998). Importantly, the wavelet transform is a linear model and can be represented as a matrix (Fig. 5b) so that we can use it in Eq. (6) to construct the first-level design matrix $X^{(1)}$. As an example of an orthogonal discrete wavelet basis function set, Fig. 5 shows design matrices that represent Daubechies wavelets of order 4 (Daubechies, 1992). We will discuss in more detail the advantages of the wavelet transform in relation to other transforms in a companion paper.

A wavelet transform is usually implemented as a lossless data transform (Fig. 5b), that is, we have as many wavelet basis functions as data points. However, we are using the wavelets to model neurophysiological responses that lie in some low-dimensional subspace of y. Therefore, we can use fewer basis functions than there are time bins at the first level, that is, a truncated wavelet transform (Fig. 5c). The truncation enables us to estimate observation error at the first level. In turn, this allows us to make inferences at the first level. Truncation also reduces the number of wavelet coefficients ($\beta^{(1)}$) at the second level, which is useful for computational reasons. Truncation corresponds to the imposition of prior constraints on the expected components of the signal. In other words, removing wavelets from a complete set is equivalent to assuming a priori the corresponding values of $\beta^{(1)}$ are zero with infinite precision (zero variance). These priors reflect our assumptions about the probability of time frequency components contributing to the ERP. For example, strong contenders for redundant basis functions are high frequencies at prestimulus time points. Note that a truncation imposes hard constraints on the model and can never be optimal for detecting any arbitrary signal (Abramovich et al., 2000). However, we argue that the ERP is not arbitrary but has a distinct and structured cause. This motivates a truncated wavelet transform. In a companion paper, we will use the truncated wavelet transform for modeling simulated and real ERP data. Additionally, we will compare the wavelet basis function set with the conventional analysis.

[3] Wavelets share this feature with other transforms like the Fourier or the discrete cosine transform.

[4] However, note that a further requirement for diagonal covariance matrices at the second level is a stationary error at the first level.

Conclusion

In this paper, we have set out the choices guiding the development of analytic procedures for ERP data using the statistical parametric mapping framework. We have focussed on motivation and justification, particularly in relation to the different sorts of statistical models and analyses that could have been used. Guided largely by the sorts of questions that are asked of the data, we conclude that a mass univariate approach is appropriate. In the temporal domain, the linear two-level hierarchical model is a natural choice to model ERP data over trial types and subjects. In this general framework, we have motivated the wavelet transform as an appropriate choice to model multisubject ERP data.

Acknowledgments

The Wellcome Trust funded this work. We would like to thank Marcia Bennett for help in preparing the manuscript and Rik Henson for valuable discussions.

References

Abramovich, F., Bailey, T., Sapatinas, T., 2000. Wavelet analysis and its statistical applications. Statistician 49, 1–29.

Baillet, S., Garnero, L., 1997. A Bayesian approach to introducing anatomo-functional priors in the EEG/MEG inverse problem. IEEE Trans. Biomed. Eng. 44, 374–385.

Barnes, G., Hillebrand, A., 2003. Statistical flattening of MEG beamformer images. Hum. Brain Mapp. 18, 1–12.

Basar, E., Demiralp, T., Schuermann, M., Basar-Eroglu, C., Ademoglu, A., 1999. Oscillatory brain dynamics, wavelet analysis, and cognition. Brain Lang. 66, 146–183.

Bosch-Bayard, J., Valdes-Sosa, P., Virues-Alba, T., Aubert-Vazquez, E., John, E., Harmony, T., Riera-Diaz, J., Trujillo-Barreto, N., 2001. 3D statistical parametric mapping of EEG source spectra by means of variable resolution electromagnetic tomography (VARETA). Clin. EEG Electroencephalogr. 32, 47–61.

Breakspear, M., 2002. Nonlinear phase desynchronization in human electroencephalograhic data. Hum. Brain Mapp. 15, 175–198.

Chatfield, C., Collins, A., 1980. Introduction to Multivariate Analysis. Chapman & Hall, London.

Czisch, M., Wetter, T., Kaufmann, C., Pollmacher, T., Holsboer, F., Auer, D., 2002. Altered processing of acoustic stimuli during sleep: reduced auditory activation and visual deactivation detected by a combined fMRI/EEG study. NeuroImage 16, 251–258.

Dale, A., Liu, A., Fischl, B.R., Buckner, R., Belliveau, J., Lewine, J., Halgren, E., 2000. Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. Neuron 26, 55–67.

Darvas, F., Schmitt, U., Louis, A., Fuchs, M., Knoll, G., Buchner, H., 2001. Spatiotemporal current density reconstruction (stCDR) from EEG/MEG-data. Brain Topogr. 13, 195–207.

Daubechies, I., 1992. Ten Lectures on Wavelets. SIAM, Philadelphia, PN, USA.

David, O., Friston, K.J., 2003. A neural mass model for MEG/EEG: coupling and neuronal dynamics. NeuroImage 20, 1743–1755.

Friston, K.J., 2004. Human Brain Function, second ed. Elsevier, San Diego, CA, USA. Chapter 31.

Friston, K.J., Poline, J.-B., Holmes, A.P., Frith, C.D., Frackowiak, R.S., 1996. A multivariate analysis of PET activation studies. Hum. Brain Mapp. 4, 140–151.

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

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

Galka, A., Yamashita, O., Ozaki, T., Biscay, R., Valdes-Sosa, P., 2003. A dynamical solution to the inverse problem of EEG generation using spatiotemporal kalman filtering. NeuroImage 19, S837.

Gershenfeld, N., 1998. The Nature of Mathematical Modeling. Cambridge Univ. Press, Cambridge, MA, USA.

Goldman, R., Stern, J., Engel, J., Cohen, M., 2002. Simultaneous EEG and fMRI of the alpha rhythm. NeuroReport 13, 2487–2492.

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

Kiebel, S., Goebel, R., Friston, K., 2000. Anatomically informed basis functions. NeuroImage 11, 656–667.

Lemieux, L., Salek-Haddadi, A., Josephs, O., Allen, P., Toms, N., Scott, C., Krakow, K., Turner, R., Fish, D., 2001. Event-related fMRI with simultaneous and continuous EEG: description of the method and initial case report. NeuroImage 14, 780–787.

Park, H., Kwon, J., Youn, T., Pae, J., Kim, J., Kim, M., Ha, K., 2002. Statistical parametric mapping of LORETA using high density EEG and individual MRI: application to mismatch negativities in schizophrenia. Hum. Brain Mapp. 17, 168–178.

Penny, W., Kiebel, S.J., Friston, K.J., 2003. Variational Bayes inference for fMRI time series. NeuroImage 19, 727–741.

Phillips, C., Rugg, M., Friston, K., 2002a. Anatomically informed basis functions for EEG source localization: combining functional and anatomical constraints. NeuroImage 16, 678–695.

Phillips, C., Rugg, M., Friston, K., 2002b. Systematic regularization of linear inverse solutions of the EEG source localization problem. NeuroImage 17, 287–301.

Phillips, C.L., Mattout, J., Maquet, P., Friston, K.J., 2004. Dynamic ReML solution for the EEG source reconstruction problem. Submitted abstract for HBM conference.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 1992. Numerical Recipes in C, second ed. Cambridge Univ. Press, Cambridge, MA, USA.

Purdon, P., Weisskoff, R., 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.

Rugg, M.D., Coles, M.G., 1995. Electrophysiology and Mind. Oxford Univ. Press, Oxford, UK.

Salek-Haddadi, A., Merschhemke, M., Lemieux, L., Fish, D., 2002. Simultaneous EEG correlated ictal fMRI. NeuroImage 16, 32–40.

Salek-Haddadi, A., Lemieux, L., Merschhemke, M., Duncan, K.F.J., Fish, D., 2003. Functional magnetic resonance imaging of human absence seizures. Ann. Neurol. 53, 663–667.

Siegmund, D., Worsley, K., 1995. Testing for a signal with unknown location and scale in a stationary gaussian random field. Ann. Stat. 23, 608–639.

Strang, G., Nguyen, T., 1996. Wavelets and Filter Banks. Wellesley-Cambridge Press, Wellesley, MA, USA.

Tallon-Baudry, C., Bertrand, O., Peronnet, F., Pernier, J., 1998. Induced gamma-band activity during the delay of a visual short-term memory task in humans. J. Neurosci. 18, 4244–4254.

Thakor, N.V., Xin-Rong, G., Yi-Chun, S., Hanley, D.F., 1993. Multiresolution wavelet analysis of evoked potentials. IEEE Trans. Biomed. Eng. 40, 1085–1094.

Toma, K., Matsuoka, T., Immisch, I., Mima, T., Waldvogel, D., Koshby, B., Hanakawa, T., Shill, H., Hallett, M., 2002. Generators of movement-related cortical potentials: fMRI constrained EEG dipole source analysis. NeuroImage 17, 161–173.

Trejo, L.J., Shensa, M.J., 1999. Feature extraction of event-related potentials using wavelets: an application to human performance monitoring. Brain Lang. 66, 89–107.

Trujillo-Barreto, N., Martinez-Montes, E., Melie-Garcia, L., Valdes-Sosa, P., 2001. A symmetrical Bayesian model for fMRI and EEG/MEG neuroimage fusion. Int. J. Bioelectromagn. 3 (http://www.ee.tut.fi/rgi/ijbem/volume3/number1/valdesosa/index.htm).

Woolrich, M.W., Ripley, B.D., Brady, M., Smith, S.M., 2001. Temporal autocorrelation in univariate linear modeling of fFMRI 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., Marett, S., Neelin, P., Vandal, A., Friston, K., Evans, A., 1996. A unified statistical approach for determining significant signals in images of cerebral activation. Hum. Brain Mapp. 4, 58–73.

Worsley, K.J., Poline, J.-B., Friston, K.J., Evans, A.C., 1997. Characterizing the response of PET and fMRI data using multivariate linear models. NeuroImage 6, 305–319.

Worsley, K., Andermann, M., Koulis, T., MacDonald, D., Evans, A., 1999. Detecting changes in nonisotropic images. Hum. Brain Mapp. 8, 98–101.

Yandell, B.S., 1997. Practical Data Analysis for Designed Experiments, first ed. Chapman & Hall, London, UK.