NeuroImage **12,** 366–380 (2000)
doi:10.1006/nimg.2000.0628, available online at http://www.idealibrary.com on IDEAL®

A New Statistical Approach to Detecting Significant Activation in Functional MRI

Jonathan L. Marchini and Brian D. Ripley

Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG United Kingdom

Received February 10, 2000

There are many ways to detect activation patterns in a time series of observations at a single voxel in a functional magnetic resonance imaging study. The critical problem is to estimate the statistical significance, which depends on the estimation of both the magnitude of the response to the stimulus and the serial dependence of the time series and especially on the assumptions made in that estimation. We show that for experimental designs with periodic stimuli, only a few aspects of the serial dependence are important and these can be estimated reliably via nonparametric estimation of the spectral density of the time series, whereas existing techniques are biased by their assumptions. The linear model with (stationary) serially dependent errors can be analyzed entirely in frequency domain, and doing so provides many insights. In particular, we introduce a technique to detect periodic activations and show that it has a distribution theory that enables us to assign significance levels down to 1 in 100,000, levels which are needed when a whole brain image is under consideration. Nonparametric spectral density estimation is shown to be self-calibrating and accurate when compared to several other time-domain approaches. The technique is especially resistant to high frequency artefacts that we have found in some datasets and we demonstrate that time-domain approaches may be sufficiently susceptible to these effects to give misleading results. The method is easily generalized to handle event-related designs. We found it necessary to consider the trends in the time series carefully and use nonlinear filters to remove the trends and robust techniques to remove "spikes." Using this in connection with our techniques allows us to detect activations in clumps of a few (even one) voxel in periodic designs, yet produce essentially no false positive detections at any voxels in null datasets. © 2000 Academic Press

INTRODUCTION

Functional magnetic resonance imaging (fMRI) is a noninvasive technique used to measure the temporal hemodynamic response of the brain to a given stimulus. Datasets consist of 3-D grids of N volume elements, called voxels, repeatedly scanned at n intervals through time. So at each voxel there will be n measurements, $\{y_t: t = 0, \ldots, n-1\}$, which constitute the voxel time series of that point in space. The volume at each time point is acquired as a series of (normally) contiguous 2-D slices of voxels in the axial, sagittal, or coronal orientation. Each slice can be viewed as a 2-D image of pixels (see Figs. 12 and 17).

The stimulus, $\{x_t: t = 0, \ldots, n-1\}$, will be designed to activate the cognitive component under study and we will be interested in modelling the relationship between y_t and x_t and obtaining inference from that model concerning the significance of the response to the stimulus at each voxel.

A common stimulus design in fMRI experiments is one which is periodic in nature. The simplest example involves c repeats of a block which consists of b volumes scanned during a baseline stimulation, followed by b volumes scanned during a stimulus known to contain the cognitive component under study. The design can be characterised by its fundamental frequency, ω_c, measured in the number of cycles per second. Figure 1 shows a real time series for a voxel in an area of activation in response to the periodic "boxcar" stimulus shown at the bottom of the figure.

Hemodynamic response to the underlying neuronal activity and noise attributable to a variety of sources will cause the fMRI signal in areas of activation to be a blurred, delayed, and noisy version of the stimulus, as seen in Fig. 1. The time series also exhibit significant serial dependence in time, as noted by Friston *et al.* (1994a). The accuracy of any method proposed to detect activation will depend on the way in which these factors are accounted for in a model.

Various methods have been proposed in the literature (Bullmore *et al.*, 1996; Friston *et al.*, 1994a, 1995c; Worsley and Friston, 1995; Lange and Zeger, 1997; Locascio *et al.*, 1997; Purdon and Weisskoff, 1998; Zarahn *et al.*, 1997), which model the underlying correlation structure of the time series in either the time-domain or the spectral-domain. Most time-domain ap-

1053-8119/00 $35.00

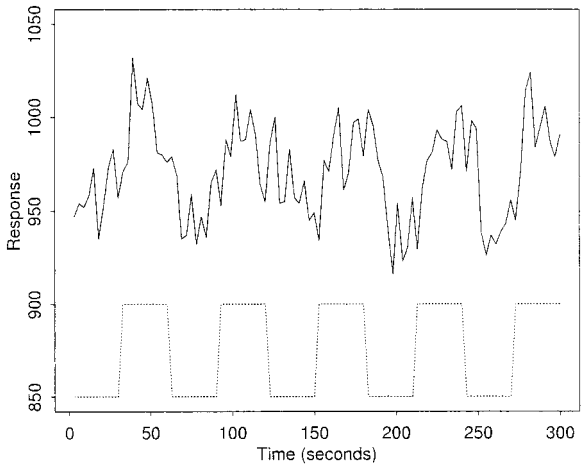

FIG. 1. A real response (solid line) from the IOP dataset in an area of activation from the visual experiment. The periodic boxcar shape of the visual stimulus is shown below (dotted line).

proaches attempt to prewhiten the time series using estimates of the correlation structure, thus enabling an analysis to proceed under the assumption of independence. Prewhitening can be achieved by switching to the spectral-domain and using an appropriate spectral density estimate. The duality that exists between these two domains implies that a given method may be computed in either domain, although this may prove easier in one domain than in the other.

Assuming a periodic stimulus design, fMRI time series analysis in the spectral-domain is greatly simplified and is the more natural approach. The nature of periodic designs ensures that the components of variance attributable to the response to the stimulus will occur at a few discrete frequencies in the spectral-domain, these being the fundamental frequency of activation and its harmonics. More importantly, the fundamental frequency of activation contains the majority of information regarding the response to the stimulus and we can concentrate solely on estimating the accuracy of the variance component at this frequency. The problem therefore reduces from estimating the entire correlation structure to just a few (even one) components. The way in which these component(s) are estimated is then the crucial factor in determining the validity and power of the derived test.

In this paper, we make as few assumptions as possible by avoiding the use of parametric models at all stages of the analysis. Nonparametric techniques are applied to trend removal, estimation of the underlying correlation structure and validation of theoretical distributions. An approach for the analysis of periodic designs is described within the frame work of spectral time series theory and applied to activation and null experiments. Nonparametric spectral density estimation is shown to be self-calibrating and accurate when compared to several other time-domain approaches. The technique is especially resistant to high frequency artefacts that we have found in some datasets and we demonstrate that time-domain approaches may be sufficiently susceptible to these effects to give misleading results. The relationship to existing methods is discussed and extensions to event-related designs are described at the end of the discussion.

MATERIALS AND METHODS

Datasets

The analysis of two datasets is reported in this paper.

Audio-visual stimulation dataset. A dataset was obtained courtesy of Dr. Brammer at the Institute of Psychiatry (IOP) (Kings College, London). One hundred T_2^*-weighted MR images (TE 40 ms; TR 3 s) were acquired at a field strength of 1.5 Tesla. Acquisition consisted of 100 vol of 14 contiguous axial slices. Each slice consisted of a 64×64 grid of voxels with dimensions $3.1 \times 3.1 \times 7.7$ mm.

The experimental consisted of simultaneous but unrelated visual and auditory stimulation. The visual stimulation occurred in a complete periodic design consisting of 5 blocks of 20 scans. Each block consisting of 10 baseline scans followed by 10 scans acquired during visual stimulation (a flashing checkerboard). The auditory stimulation design was an incomplete periodic design. Blocks consisted of 26 scans, 13 baseline scans, followed by 13 activation scans (listening to a talking book). The last block thus contained only 9 activation scans.

Null dataset. A null dataset was obtained from our collaborators at the Oxford Centre for Functional Magnetic Resonance Imaging of the Brain. Two hundred T_2^*-weighted MR images (TE 30 ms; TR 3 s) were acquired at a field strength of 3.0 Tesla. Acquisition consisted of 200 vol of 21 contiguous axial slices. Each slice consisted of a 64×64 grid of voxels with dimensions $4 \times 4 \times 6$ mm. There was no stimulus; all scans were acquired under baseline conditions.

Motion correction was applied to the datasets using AIR (Woods *et al.,* 1992) within the MEDx analysis package. Although it has been suggested (Friston *et al.,* 1996) that rigid body realignment alone may not be sufficient to remove all motion artefacts, this is not the focus of this paper and has no bearing on the method of analysis we propose.

Trend Removal

Many voxel time series exhibit low frequency trend components. These may be due to aliased high frequency physiological components or drifts in scanner sensitivity. Whatever the cause, these trends tend to vary nonlinearly in time and may result in false-posi-

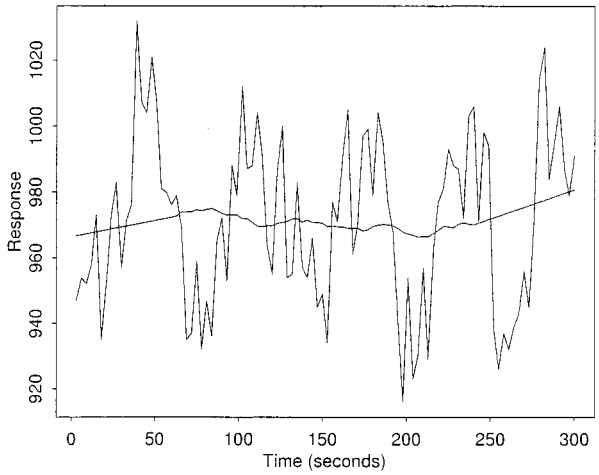

FIG. 2. A voxel time series from the IOP dataset in an area of activation by the visual stimulus.

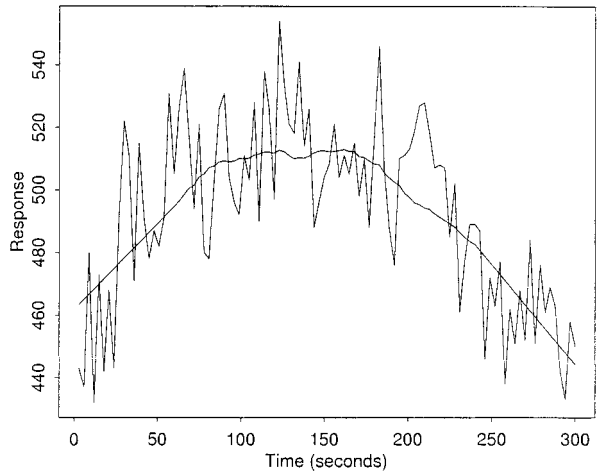

FIG. 3. A voxel time series from the IOP dataset showing an obvious nonlinear trend.

tive activations if they are not accounted for in the model.

Nonlinear trends can be removed in a variety of ways (see, for example, Venables and Ripley, 1999). The method must be able to remove obvious nonlinear trends and leave voxel time series with no obvious trend almost unchanged. Methods should also be evaluated by their ability to estimate trends at voxels for which there is a significant response to the stimulus. Some methods incorrectly infer trends at the ends of such series and this should be avoided. We have found a simple running-lines smoother to be a reliable method of trend removal.

A running-lines smoother fits a linear regression (by least-squares, or robustly) to the k nearest neighbors of a given point and then uses the line to predict the response at that point. For the analysis of periodic designed fMRI experiments we suggest setting k equal to twice the cycle length. Figures 2–4 show three voxel time series each with trends fitted by applying a running lines smoother using $k = 40$. Obvious trends have been removed and only minimal end effects occur for voxels exhibiting activation.

Nonlinear trend removal of this type is ineffective for voxel time series which exhibit large isolated spikes at some time points. These effects may be minimized by reducing all observations beyond a certain distance from the mean value to within some specified limit. This technique is called Winsorizing (Wilcox, 1998) and will protect against false positive responses to the stimulus caused by these large peaks. We have found such effects to occur mainly outside the brain.

Time Series Analysis

Time series is the study of correlated measurements that occur in time (Bloomfield, 1976, 1991; Diggle, 1990; Priestley, 1981). Typically, we observe one or more realisations $\{y_t: t = 0, \ldots, n-1\}$ of a stochastic process $\{Y_t\}$ and wish to describe this process and make inferences about it. The challenge in many applications is accurate description of the structure of the correlations between time points and comparison with the properties of known processes. The validity of any inference may depend critically on the accuracy of the model of these correlations.

In the time-domain, the correlation structure of a second-order stationary process is described by the autocovariance function (acf) γ_k defined as

$$\gamma_k = \mathrm{Cov}(Y_t, Y_{t-k}) \tag{1}$$

$$= \mathbb{E}((Y_t - \mu)(Y_{t-k} - \mu)) \tag{2}$$

This function measures the relationship of observations separated by a lag k in time.

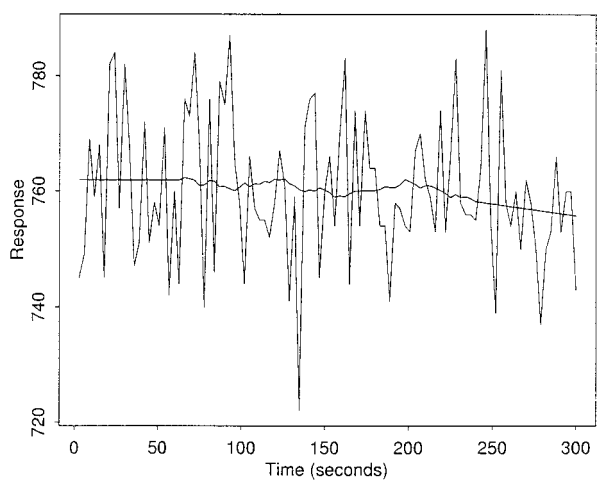

FIG. 4. A voxel time series from the IOP dataset in an area of no activation by either stimulus.

In the spectral-domain the correlation structure is represented by the spectral density, $f(\omega)$, $\omega \in [0, 1/2\delta]$, defined by

$$f(\omega) = \sum_{k=-\infty}^{\infty} \gamma_k \exp(-i2\pi\delta k\omega), \tag{3}$$

where δ is the sampling interval of the time series in seconds and ω is measured in cycles per second. The spectral density describes the long-run properties of the process in terms of the tendency for periodic components of different frequencies to occur in any given realization.

For the analysis of fMRI voxel time series we assume a linear model for the response in the time-domain

$$Y_t = X_t\beta + Z_t \quad t = 0, \ldots, n-1 \tag{4}$$

where Y_t represents the response in the voxel at time t, X_t is row t of an $(n \times p)$ design matrix X, whose columns span the space of the hemodynamically modulated stimulus and confounding effects like trends, β is a $(p \times 1)$ parameter vector and Z_t is a second-order stationary stochastic process with zero mean and unknown correlation structure. For a realisation of length n this correlation structure will be described completely by an $n \times n$ covariance matrix $\sigma^2 V = KK^T$, where

$$V_{ij} = \frac{\gamma_{|i-j|}}{\sigma^2} \tag{5}$$

and K is a (nonunique) matrix square root. If V is known, the correlations can be removed completely by applying the pre-whitening matrix K^{-1} to both sides of the matrix form of model (4), which gives

$$K^{-1}Y = (K^{-1}X)\beta + \eta, \tag{6}$$

where $\mathrm{Var}(\eta) = \sigma^2 I$. Ordinary least squares will then provide optimal parameter estimates and standard distribution theory can then be applied to assess the significance of the response to the stimulus. This method is exactly equivalent to generalized least squares with covariance matrix V. Since V is unknown, it must be estimated simultaneously with β, normally using an iterative procedure.

The equivalent model in the spectral-domain is obtained using the discrete Fourier transform at the Fourier frequencies

$$\omega_j = j/\delta n \quad j \in S = \{0, \ldots, \lfloor n/2 \rfloor\}. \tag{7}$$

For a series $\{h_t: t = 0, \ldots, n-1\}$ this is defined as

$$d_h(\omega_j) = n^{-1} \sum_{t=0}^{n-1} h_t \exp(-i2\pi\omega_j\delta t). \tag{8}$$

The model then becomes

$$d_Y(\omega_j) = d_X(\omega_j)^T\beta + d_Z(\omega_j), \tag{9}$$

for $j = 0, \ldots, \lfloor n/2 \rfloor$. It can be shown that $d_Z(\omega_j)$ and $d_Z(\omega_k)$ are approximately uncorrelated for large n and that (apart from $j = 0, n/2$)

$$\mathrm{E}(|d_Z(\omega_j)|^2) \approx \frac{f_Z(\omega_j)}{n}, \tag{10}$$

where $f_Z(\omega_j)$ represents the spectral density of the error process at the Fourier frequencies. If $f_Z(\omega)$ is known, weighted least squares will provide optimal estimates of β and if $f_Z(\omega)$ is unknown it must be estimated simultaneously with β, as in the time-domain.

The duality between the time and spectral-domains due to their explicit mathematical relationship implies that every proposed method of analysis may be computed in either domain, although this may prove considerably easier in one domain than in the other. The choice of domain will then depend only upon which provides the most natural interpretation of the problem under study. To detect activation fMRI voxel time series we prefer to work in the spectral-domain since the approach is greatly simplified in the case of periodic designs.

Spectral Time Series Analysis of Periodic Stimulus Designs

By assuming the design consists of c repeats of a symmetric baseline-activation block and that the brain's functional response can be approximated by a linear system, all the information concerning the response to the stimulus will lie at the fundamental frequency and harmonics of the stimulus design. These will be the exact Fourier frequencies, ω_j for $j \in S_F = \{c, 2c, 3c, \ldots, \lfloor n/2 \rfloor\}$. If we further assume that any confounding effects such as trends have been removed from the time series beforehand, then the component $d_X(\omega_j)$ in model (9) will be zero at all Fourier frequencies other than the fundamental and harmonics. Then model (9) becomes

$$d_Y(\omega_j) = \begin{cases} d_X(\omega_j)^T\beta + d_Z(\omega_j) & j \in S_F \\ d_Z(\omega_j) & j \notin S_F, \end{cases} \tag{11}$$

which implies that by taking the discrete Fourier transform of the time series at each voxel, the majority

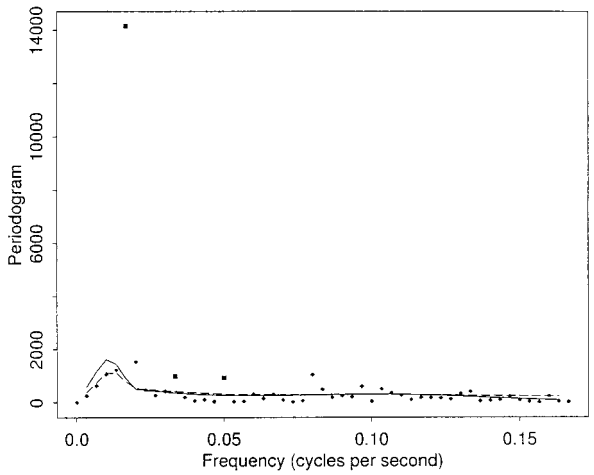

FIG. 5. The periodogram of a detrended voxel time series in an area of activation. The fundamental frequency and first two harmonics are plotted as squares. The large ordinate at the fundamental frequency is indicative of the response and to a lesser extent the first two harmonics also exhibit evidence of a response. Spectral density estimates before (thick line) and after (dashed line) spatial smoothing. The spectral density estimates do not apply to the extreme frequencies 0 and 1/2δ, which are not included in the estimation procedure.

of the frequencies will contain information solely about the correlation structure of the underlying stochastic process at the voxel. This can be illustrated by plotting the periodogram, $I(\omega_j)$, of the time series for a voxel in an area of activation, for all Fourier frequencies.

The periodogram is defined by

$$\begin{aligned} I(\omega_j) &= n|d_Y(\omega_j)|^2 \\ &= n(\Re(d_Y(\omega_j))^2 + \Im(d_Y(\omega_j))^2) \end{aligned} \tag{12}$$

and represents the variance of the time series attributable to an oscillation of frequency ω_j. Figure 5 shows the periodogram of the detrended voxel time series shown in Fig. 2. The extremely large ordinate at the fundamental frequency is indicative of the response to the stimulus. The first two harmonics are noticeably larger than we might expect but after that higher harmonics are indistinguishable from the underlying process. The fundamental frequency and first two harmonics accounted for 47.9, 3.3, and 3.1% of the total variation respectively and higher harmonics were found to contain little or no information regarding the response to the stimulus, agreeing with Bullmore *et al.* (1996). Even though there is a little information in the second and third harmonics, it is harder to extract from the noise and in this paper we concentrate on detecting activation via the fundamental frequency.

The phase of the response to the stimulus in any given voxel will be unknown and can be accommodated into model (11) by assuming the design matrix X consists of two columns containing orthogonal sinusoids at the fundamental frequency ω_c. The optimal estimator of β is found by minimizing the quantity

$$\sum_{j=0}^{\lfloor n/2 \rfloor} \frac{\|d_Y(\omega_j) - d_X(\omega_j)^T\beta\|^2}{f_Z(\omega_j)/n}$$

to obtain

$$\hat{\beta} = 2\begin{pmatrix} \Re(d_Y(\omega_c)) & 0 \\ 0 & -\Im(d_Y(\omega_c)) \end{pmatrix}, \tag{13}$$

which does not depend on the underlying spectral density $f_Z(\omega)$. Comparing (12) and (13) we see that there is an explicit relationship between $I(\omega_c)$ and the optimal estimate of β. It then becomes clear that $I(\omega_c)$ is the optimal estimate of the power at the fundamental frequency regardless of the form of the spectral density $f_Z(\omega)$. Thus, a test for the significance of the response to a periodic stimulus should be based on $I(\omega_c)$.

The asymptotic sampling properties of the periodogram are well-known. For increasingly long series:

(I). $I(\omega_j)/f(\omega_j) \sim$ E where E has a standard exponential distribution.

(II). $I(\omega_j)$ and $I(\omega_k)$ are independent for all $k \neq j$.

This is a very general result (Priestley, 1981) that requires only mild conditions on the form of the stationary stochastic process and includes AR, MA, and ARMA processes as special cases. The asymptotic distribution is known to be accurate for series of moderate length, such as those encountered in fMRI experiments. The result implies that $I(\omega_j)$ is an unbiased estimator of $f(\omega_j)$, but it is not a consistent estimator since its variance is constant and does not depend on n. If we assume that the underlying spectral density $f(\omega)$ is smooth then we can use a smoothed version of $I(\omega)$, which we denote $g(\omega)$, to estimate $f(\omega)$. This theory does not apply to the extreme frequencies 0 and 1/2δ, which are not included in the estimation procedure.

The spectral density estimate provides us with a baseline against which to test for significant departures from the underlying process. From (*I*),

$$\frac{I(\omega_j)}{f(\omega_j)} \sim E \tag{14}$$

Substituting $g(\omega_j)$ for $f(\omega_j)$, we define the ratio statistic, R_j, as

$$R_j = \frac{I(\omega_j)}{g(\omega_j)}, \quad \omega_j = j/\delta n \tag{15}$$

Large values of R_j indicate that the component of the voxel time series characterized by oscillations at the Fourier frequency ω_j, is larger than we would expect given our estimate of the spectrum. For a periodic

design only the R_j at the fundamental frequency of activation and its harmonics are of interest, and hence only the estimate of the spectral density (of the noise) at those frequencies.

For incomplete periodic designs the fundamental frequency of activation and its harmonics will not lie exactly at Fourier frequencies and the variance attributable to the response will occur in a spread of frequencies around the fundamental and harmonics. We can avoid this by padding the series (adding zeroes) to make up the length of a complete periodic design, so that for the auditory stimulus in the IOP dataset we add four observations to make 104 scans. The effects of the addition can be ameliorated by *tapering* (Bloomfield, 1976) the series before padding, that is smoothly reducing the magnitude of the series to zero at each end. As trend removal is more difficult at the ends, tapering may be a good idea even for complete periodic designs.

Estimation of the Spectral Density

We have seen that we are only interested in the spectral density at a few (often just one) Fourier frequencies, and we are interested in the spectral density of the error series Z_t not the observed series Y_t. Parametric models (such as low-order autoregressive processes) fit a restrictive overall shape to the spectral density and so may be (and in our experiments were) biased estimates of the spectral density at the frequencies of interest. Even if they are not appreciably biased, they will be using the information at much higher frequencies to determine the spectral density at the fundamental frequency of the design. A flexible nonparametric approach was tried and seems highly successful.

An estimate of the spectral density is obtained by smoothing the periodogram on log scale (Wahba, 1980), $\log(I(\omega))$, to ensure that the spectral density estimate is positive everywhere when we transform back from log-scale, and to stabilize the variance of the terms being smoothed. In this case,

$$\log I(\omega_j) \sim \log f(\omega_j) + \log \mathrm{E}, \tag{16}$$

where log E has a standard Gumbel distribution from (I). The mean of this Gumbel distribution is equal to the Euler–Mascheroni constant

$$\mathbb{E}(\log \mathrm{E}) = \gamma = -0.57721 \ldots \tag{17}$$

So by smoothing $Z_j = \log I(\omega_j) - \gamma$ we obtain an asymptotically unbiased estimate of $\log f(\omega_j)$. This is then inverted to obtain an estimate of the spectral density, denoted $g(\omega_j)$. A smoothing spline (Silverman, 1985) was chosen to smooth the log-periodogram.

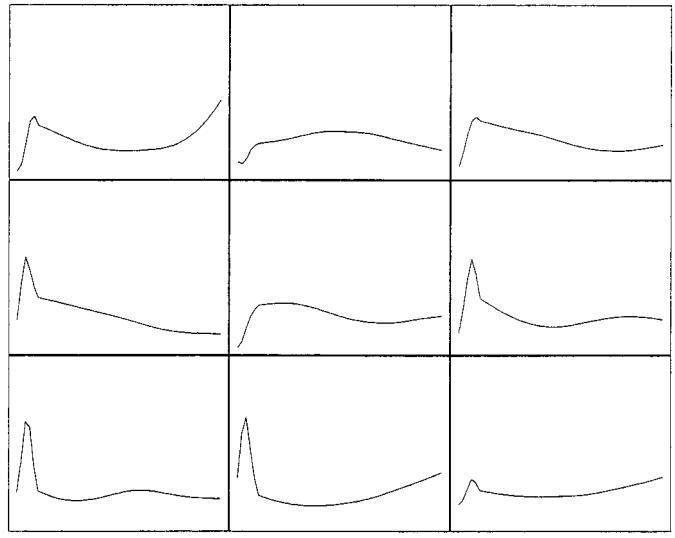

FIG. 6. Normalized spectral density plots on the same scale for a 3×3 grid of voxels in an area of no activation for the visual experiment.

To ensure the spectral density of the noise is estimated independently of any response to the stimulus, the periodogram ordinates at the fundamental frequency of stimulation and its first two harmonics are not included in this procedure.

The application of detrending will cause the first few periodogram ordinates to be small compared to higher ordinates and it has been observed by Zarahn *et al.* (1997) and by us that lower frequency ordinates tend to be more variable than those at higher ordinates. Thus, we will want to allow the spline more freedom at lower frequencies. This is done by transforming the x-axis of the periodogram before fitting the spline to give more space between lower frequency ordinates.

So far we have estimated the spectral density independently at each voxel, but we applied a small amount of spatial smoothing to the estimates. Figure 6 shows a 3×3 grid of normalized spectral density estimates plotted on the same scale, in an area of no activation. The spectra show differences mainly at lower frequencies where we would expect the spectral density estimate to be more variable. Local smoothing of the spectra in space will thus ensure a more stable estimate of the spectral density at each voxel. We used a discretized isotropic Gaussian filter to smooth normalized spectral density estimates. More advanced spatial filtering may be applied that takes into account differences in the correlation structures of different tissue types and may better protect against bias introduced by smoothing spectra whose form does not vary smoothly in space. In Fig. 5 the lines show the spectral density estimate before and after spatial smoothing is applied to the spectra.

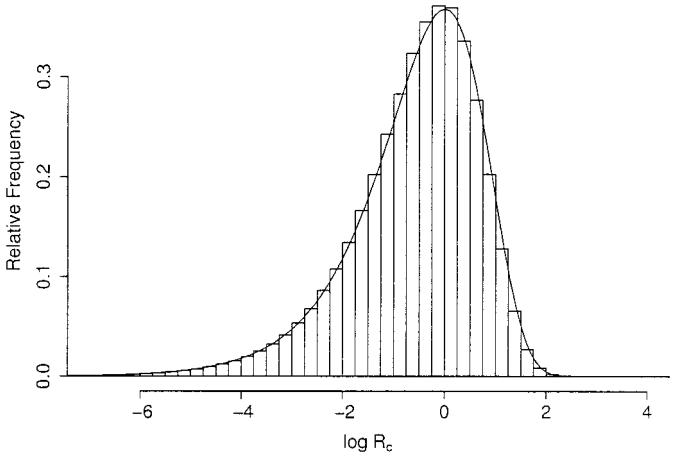

FIG. 7. Histogram of log R_j values from the analysis of the null dataset at all frequencies except the first few lowest, together with the Gumbel density (thick line).

Testing for a Response to the Stimulus

By calculating the statistic R_j at the fundamental frequency of activation, ω_c, we obtain a test statistic, R_c, for significant activation.

$$R_c = \frac{I(\omega_c)}{g(\omega_c)}, \quad \omega_c = c/\delta n \tag{18}$$

Large values of R_c indicate a large effect at the fundamental frequency.

The sampling properties of the periodogram (see Eq. (14)) imply that under the hypothesis of no activation the statistic R_c has a standard exponential distribution, asymptotically, and so log R_c has a standard Gumbel distribution. The accuracy of this asymptotic distribution may be determined by comparison to nonparametric estimates of the distribution of R_c. Equations (14) and (15) indicate that it is not just the statistic R_c (at the fundamental frequency of stimulation) that is asymptotically exponentially distributed, but rather all of the statistics R_j apart from $j = 0$ and $n/2$. Each R_j tests for a unexpectedly large effect with frequency ω_j, where we are only interested in a few of those frequencies. Even if the exponential distribution is not a good model, all of the statistics R_j will have the same asymptotic distribution under the null hypothesis of no activation, except the first few lowest frequencies which will be affected by the trend removal. As there will be little or no response to the stimulus at all frequencies except the fundamental and first few harmonics we can use the values of R_j at these other frequencies to obtain a large sample from the null distribution. This provides a benchmark against which to compare the theoretical distribution of R_c, at negligible computational cost at this stage of the analysis since these R_j values occur as a byproduct of the estimation of R_c. Thus the technique can be self-calibrated.

Figure 7 shows the histogram of log R_j values from the analysis of the null dataset at all frequencies except the lowest few. The Gumbel density is overlaid on the plot and indicates that this theoretical distribution fits the null distribution very well. Figures 8 and 9 show histograms of log R_j values from the analysis of the visual experiment in the IOP dataset. In Fig. 8 all frequencies are included, except the fundamental, its first two harmonics and the lowest few frequencies. In Fig. 9 only the values at the fundamental frequency are used. The difference in the right tail of the histogram compared to the Gumbel density are due to those voxels which exhibit activation.

Figure 10 compares the nonparametric distribution to the theoretical. The probability density of the ratio statistic was estimated using the local density estimation procedure described in Loader (1996) using the IOP data. The estimate was based on the R_j values at all frequencies except the fundamental, first two harmonics and the first few lowest frequencies. The estimated cdf $F(x)$ is transformed using the inverse of the theoretical cumulative density function (cdf) and plotted against x, which would give a straight line if the theoretical distribution was true. This plot is similar to a Q–Q plot in which ordered samples are plotted against each other, with a straight line indicating agreement in their distribution. Rather than plot discrete samples against each other we plot continuous densities, these being the estimate of the null distribution using the R_j values and the theoretical exponential distribution. The right hand scale indicates P values in powers of 10 for the theoretical distribution. This shows that the nonparametric density has a slightly shorter tail than the theoretical distribution but the

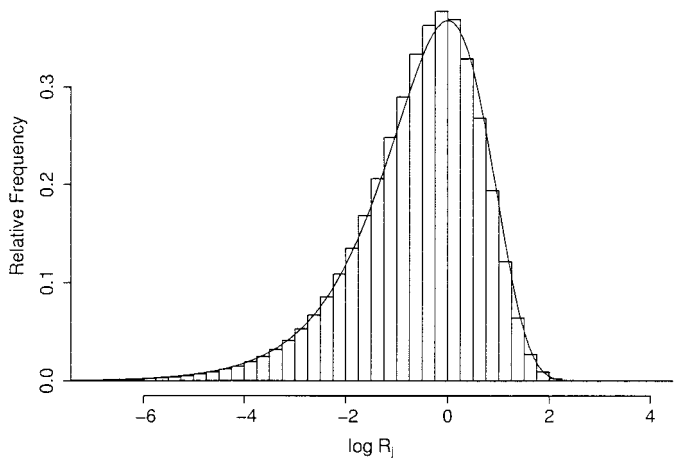

FIG. 8. Histogram of log R_j values from the analysis of the visual experiment in the IOP dataset, together with the Gumbel density (thick line). All frequencies were used, except the fundamental, first two harmonics, and the first few lowest frequencies.

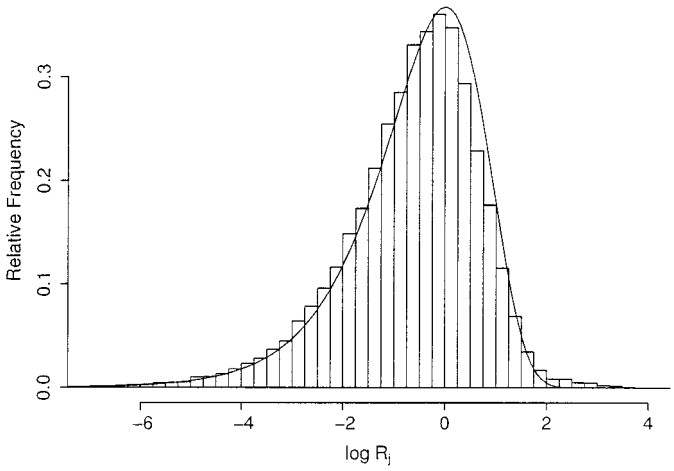

FIG. 9. Histogram of log R_j values from the analysis of the visual experiment in the IOP dataset, together with the Gumbel density (thick line). Only the values at the fundamental frequency were used. The differences between the right tails of the histogram and the Gumbel density are due to those voxels with activation.

agreement is very good at all values far into the tail of the distribution. Density estimates calculated using the null dataset showed similarly excellent agreement with the theoretical distribution, as indicated by Fig. 7.

Our tests provide very strong support for the validity of the theoretical distribution.

Detecting Activated Clusters

Because there are so many voxels (20,000 or more) in a whole brain image, fairly rare events for any voxel are quite likely to occur somewhere. If the procedure for a single voxel has probability α of a false positive, the expected number of false positives is $N\alpha$ when N voxels are considered, and the probability of any false positives is bounded above by this quantity (by what is known as Bonferroni's bound). Thus if we want to look for activations on one or a small cluster of voxels we need to consider events with false positive rates of 10^{-5} to 10^{-4}. Figure 10 shows our distribution theory is very accurate in that region.

The level at which to threshold the statistic or P value image to assess the validity of the global null hypothesis of no activation anywhere in the image will be determined by the form of the spatial correlations between voxels. For example, if the voxels were independent then a Bonferroni correction could be applied. Such a test will be sensitive to areas of activation that consist of small high peaks of activation but not for large low peaks of activation which will go undetected by a global threshold method of this type. This has been noted in the literature (Friston *et al.*, 1991, 1994b; Worsley *et al.*, 1996; Poline *et al.*, 1997) in which the theory of random fields has been applied to the analysis of statistic maps that correct for the presence of spatial correlations and account for the topology of clusters of activated voxels. Two crucial assumptions of these theoretical approaches are that the statistic map is a realization of some homogeneous random field and that the voxel dimensions are sufficiently small to provide an accurate representation of the underlying continuous process of the brain's response. These assumptions can be guaranteed if the original data is smoothed in space but this will also considerably blur the signal. In effect the analysis has been fitted to comply with the theory when it is surely better to apply a theory to the analysis which makes more realistic assumptions. This has been achieved in part by the use of randomization approaches such as those employed by Bullmore *et al.* (1996, 1999) and Locascio *et al.* (1997). These approaches allow a wider range of cluster characteristics to be measured and assessed for significance although the question of how to characterize each cluster is one on which there seems to be no clear consensus and is the focus of our current work in this area.

RESULTS AND DISCUSSION

In this section we show the results of our method when applied to the visual experiment of the IOP dataset and compare it to another approach. We give a theoretical discussion of previous approaches in the literature and compare some of them to our own approach using the null dataset. The ability of our method to detect high frequency artefacts is described and the effect of these artefacts on parametric time-domain models of the underlying correlation structure is illustrated and discussed. We finish with a discus-

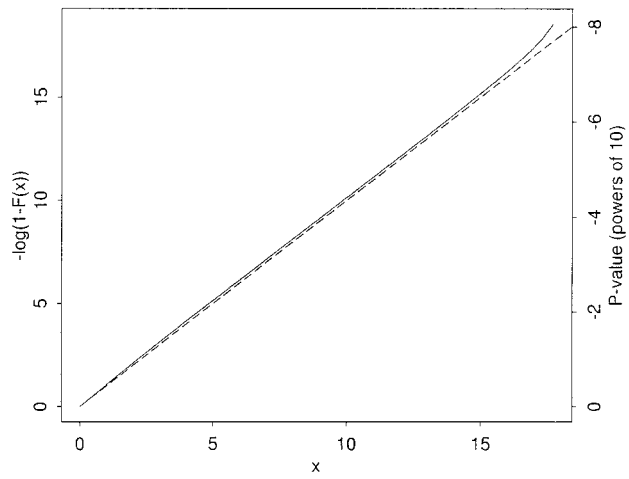

FIG. 10. Comparison of the theoretical density for R_c against the nonparametric density estimate. The estimated cdf $F(x)$ is transformed using the inverse of the theoretical cdf and plotted against x (thick line), which would give a straight line (dashed line) if the theoretical distribution was true. The right hand scale indicates P values in powers of 10 for the theoretical distribution and indicates the close agreement well into the tails of the two distributions.

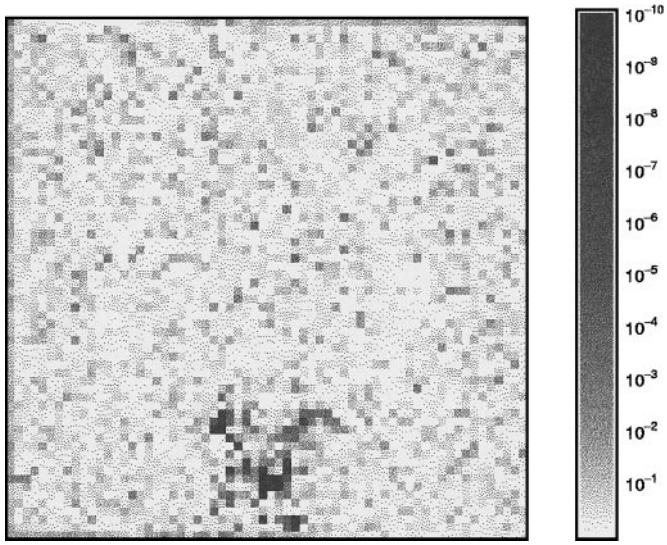

FIG. 11. *P* Value image for a slice of the IOP dataset analyzed for response to the visual stimulation.

sion of extensions of our approach to handle event-related designs.

Computations were carried out in the statistical systems S-PLUS and R. Spatial smoothing of the normalized spectral densities was used only within slices.

IOP Dataset Experiments

The spectral analysis was first applied to the visual experiment in the IOP dataset. We also analysed the auditory experiment with clear-cut results which are not shown here.

Figure 11 shows the *P* value image for a slice of the IOP dataset, analyzed for a response to the visual stimulus, using our proposed R_c statistic and its theoretical distribution. Figure 12 shows Fig. 11 thresholded to show only *P* values below 10^{-4} and overlaid on an image of the slice. (The threshold level was chosen purely for illustration purposes.) Color is used to indicate differences in activation within and between the clusters. The image obtained using a threshold derived from the empirical statistic distribution at higher frequencies was not visibly different to Fig. 12. This is as expected, since the theoretical and empirical distributions are very similar.

For comparison, this experiment was analyzed using the approach proposed in Bullmore *et al.* (1996) who use the Cochrane-Orcutt procedure to analyse periodic stimulus designs. The Cochrane-Orcutt procedure (Cochrane and Orcutt, 1949) is applicable to any autoregressive process of order *P* (Diggle, 1990) and simplifies the prewhitening equations; iteration can be used to estimate the parameters of the AR(*P*) process and β. In Bullmore *et al.* (1996) the response to the stimulus was characterised by sinusoidal terms at the fundamental frequency and first two harmonics and an AR(1) model was assumed and validated. A statistic called the Fundamental Power Quotient (FPQ) was then used to assess the significance of the response to the stimulus at each voxel. This statistic is defined to be the ratio of the power at the fundamental frequency (FP) to its estimated standard error SE(FP). Thus the FPQ statistic is analogous to our own R_c statistic, in which the numerator $I(\omega_c)$ represents the power at the fundamental frequency ω_c and $\widehat{f(\omega_c)}$ estimates its standard error. In the original paper trends were dealt with by including linear terms in the design matrix along with the sinusoidal terms, but in our own main implementation of this approach we dealt with trends using the nonlinear methods described before.

Figure 13 shows the *P* value image for the same slice considered in Fig. 11, using the AR(1) plus nonlinear detrending approach. The *P* values were derived using the theoretical unit-mean exponential distribution of the FPQ statistic. Figure 14 shows Fig. 13 thresholded to show only *P* values below 10^{-4} and overlaid onto an image of the slice. Compared to Fig. 12 there is considerably more activation detected. Assuming that we should only see activation within the visual cortex plus one in 10,000 false-positives, the amount of activation detected in other areas suggests that for this dataset the theoretical distribution is not an adequate fit.

In Bullmore *et al.* (1996) the inadequacy of the theoretical distribution was recognised and then adjusted for, using an auxiliary randomization experiment. We stress that the approach proposed in this paper does not require a randomization experiment to obtain the true null distribution. Our method is completely self-

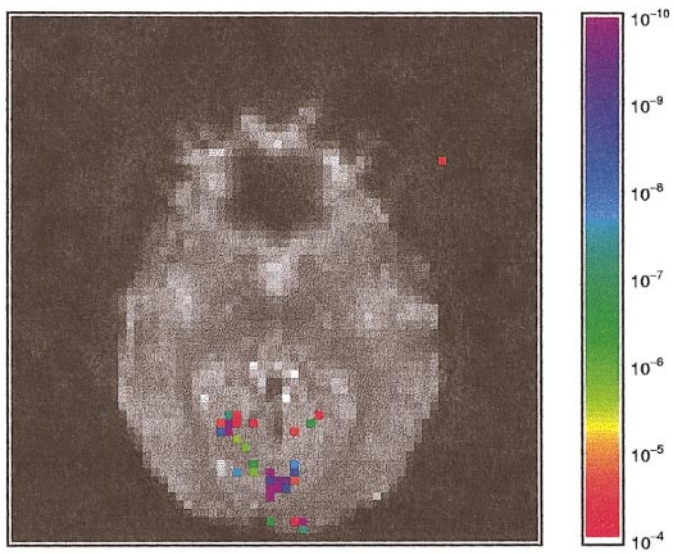

FIG. 12. *P* Value image (shown in Fig. 11) thresholded to show *P* values below 10^{-4} and overlaid onto an image of the slice. Colors indicate differential responses within each cluster. An area of activation is shown in the visual cortex, as well as a single "false-positive," that occurs outside of the brain.

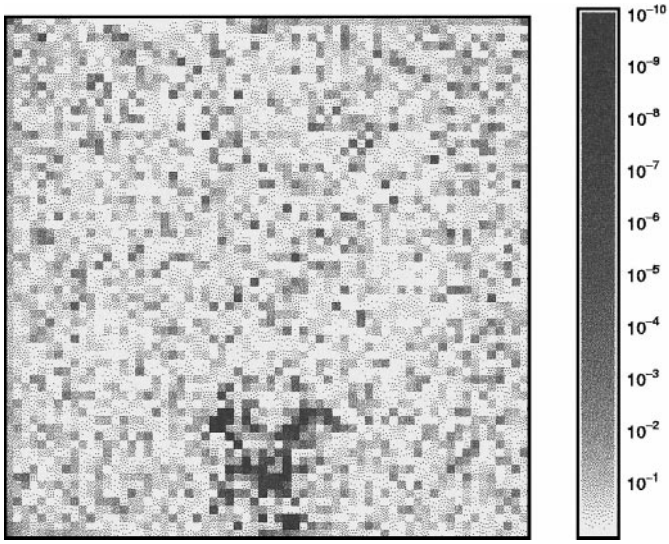

FIG. 13. Nominal *P* value image for a slice of the IOP dataset (shown in Fig. 11), analyzed for a response to the visual stimulation using an AR(1) model and the nonlinear detrending proposed in this paper.

calibrating since information at higher frequencies is utilised to estimate the true null distribution, which has been shown to agree extremely well with its theoretical form.

We attempted to recalibrate the AR(1) method by estimating the null distribution using all those pixels thought to lie well away from areas of activation. We used all voxels in the top half of each slice to construct an empirical distribution and then a recalibrated threshold corresponding to a *P* value of 10^{-4} was calculated. This method of recalibration could not normally be used since we will not know where to expect a response to the stimulus. Figure 15 shows this threshold applied to the *P* value map shown in Fig. 13 and overlaid onto an image of the slice. The areas of activation that remain are very similar to those detected by our own approach, shown in Fig. 12. After having corrected the null distribution the threshold was much higher than the one proposed by the theoretical distribution, which explains for the difference in the colors of the clusters when comparing Figs. 12 and 15.

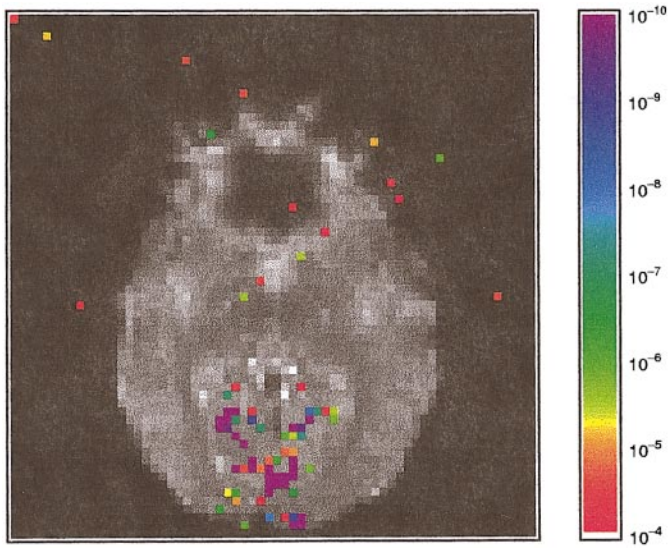

FIG. 14. *P* Value image (shown in Fig. 13) thresholded to show *P* values below 10^{-4} and overlaid onto an image of the slice. Colors indicate differential responses within each cluster.

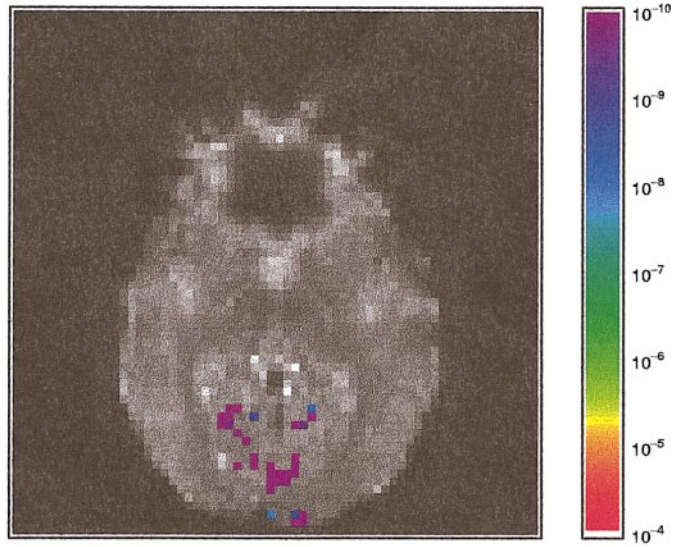

FIG. 15. *P* Value image (shown in Fig. 13) thresholded using a recalibrated estimate of the 10^{-4} *P* value point of true null distribution. This was calculated from the empirical distribution of all voxels in the top half of each slice, which were assumed to contain no activation in response to the stimulus. The surviving voxels are shown in color against the same nominal *P* value scale used in Fig. 14.

Comparison to Other Approaches

In this paper we have proposed a nonparametric model for the underlying correlation structure of the voxel time series, whereas most other approaches described in the literature assume a particular parametric form. We give a theoretical discussion of the current approaches in the literature and in subsequent sections illustrate how some of these approaches compare to our own.

The approach of (Bullmore *et al.*, 1996) has already been described. In the context of the present paper this is equivalent to fitting a parametric model to the periodogram to estimate the spectral density and may fail to estimate adequately the spectral density at the one frequency of interest in periodic designs. Also the AR(1) model fails to take into account any prior trend removal, in that low frequency terms are almost completely removed and the resulting shape of the spectral density will lie outside the range of the AR(1) model.

Failure to model adequately the spectral density has two effects on the distribution of the test statistic for $\beta \neq 0$. One, as observed by Bullmore *et al.* (1996) and illustrated in the previous section, is to produce a systematic bias in the null-hypothesis distribution, which can be corrected (at considerable computational cost) by a randomization experiment. The other effect is to increase the random variability of the test statistic and thereby produce a less powerful test.

Somewhat higher-order autoregressive-moving-average processes are used to model the correlation structure in Locascio *et al.* (1997) using a model selection approach, but these still fail to accurately model the correlation structure in up to 5% of voxels. Using high-order AR(p) processes (using p up to 30 or so) is a well-known method for estimating the spectral density of a time series. It is not a recommended one and some authorities (for example, Thomson, 1990) warn strongly against its use.

The Statistical Parametric Mapping (SPM) approach (Friston *et al.*, 1994a, 1995b,c; Worsley and Friston, 1995) assumes a linear model of the form in Eq. (4). In certain situations spatial filtering of the response at each time point is recommended. This is intended to reduce the "noise" in the data but will also blur any response to the stimulus, reducing the high resolution afforded by fMRI. Temporal smoothing is applied at each voxel time series. The difference between the resulting correlation structure and its estimate is then assumed to be small enough for valid inference using a generalised least squares procedure. An estimate of the true degrees of freedom for the null distribution is used to provide more accurate P values. This amounts to assuming a parametric form for the spectra density determined by the smoothing and estimating a single scale constant for the noise. The papers give various suggestions to estimate the scale of the noise, but more is targeted at the frequencies which determine the accuracy of the estimation of β (and hence the test statistics for $\beta \neq 0$). In the current implementation of this approach within the SPM99 software an option is also given to use an AR(1) model for the underlying correlation structure. In this approach the AR(1) coefficient is estimated at each voxel and then averaged over voxels.

Zarahn *et al.* (1997) suggested two parametric models for the underlying correlation structure, which were fitted to voxel-averaged power spectra using an iterative nonlinear least squares procedure. This global estimate was used at all voxels within the linear modeling framework of (Worsley and Friston, 1995) but produced results inconsistent with the assumption of a spatially stationary correlation structure. The use of temporal smoothing resulted in false-positive rates closer to the theoretical distribution, but this was reported only at the 5% significance level and thus provided no information about the agreement elsewhere in the distribution. Also, the results were sensitive to the use of high pass filter used.

Purdon and Weisskoff (1998) proposed an AR(1) plus white noise model for the underlying correlation structure for each voxel time-series. The AR(1) component is used to model large low-frequency effects (trends), while the white-noise component represents scanner noise. This model is fitted using an iterative nonlinear least squares procedure to a voxel averaged power spectrum from a real dataset. The parameter estimates are then used to simulate datasets which are used to illustrate the consequences of assuming different models for the correlation structure. It is demonstrated that prewhitening using the model that generated the datasets performs much better than methods which assume a white-noise model and marginally better than a proposed extension to the approach of Worsley and Friston (1995), although results are only shown in detail for P values between 10^{-1} and 10^{-2}. The spectral density of an AR(1) plus white noise model is determined by three parameters and is not much more flexible than an AR(1) model.

Lange and Zeger (1997) apply a spectral time series model in the form (9). By expressing X_t in (4) as the convolution of the periodic stimulus design with a parameterized delay function, the form of the hemodynamic modulation is estimated simultaneously with β at each voxel, using an iterated generalised least squares procedure. The spectral density is estimated using a simple average of squared residuals around each of the fundamental and harmonic frequencies, which is also smoothed in space. Thus the correlation structure is estimated nonparametrically, but this aspect of the approach was not the focus of the paper and more robust approaches were not considered. A simple average filter may be reasonable at high frequencies, but not at lower frequencies near the fundamental frequency, where there is likely to be larger variation in the underlying spectral density of detrended voxel time series. Our experiments using this method agree with the discussants of the paper who question the identifiability of the parameters. This means that different combinations of the parameters may provide an equally good fit to the data, leading to redundancy in the model. The optimization required to provide estimates of the parameters in the delay function ensures that this method runs very slowly.

Null Dataset Experiments

We applied three different parametric time-domain approaches and our own nonparametric spectral approach to the null dataset. A periodic stimulus consisting of 10 complete cycles in the 200 scans was assumed. The fundamental frequency of this design will be 1 cycle per minute and is the same as that of the IOP dataset visual experiment. We compared the empirical

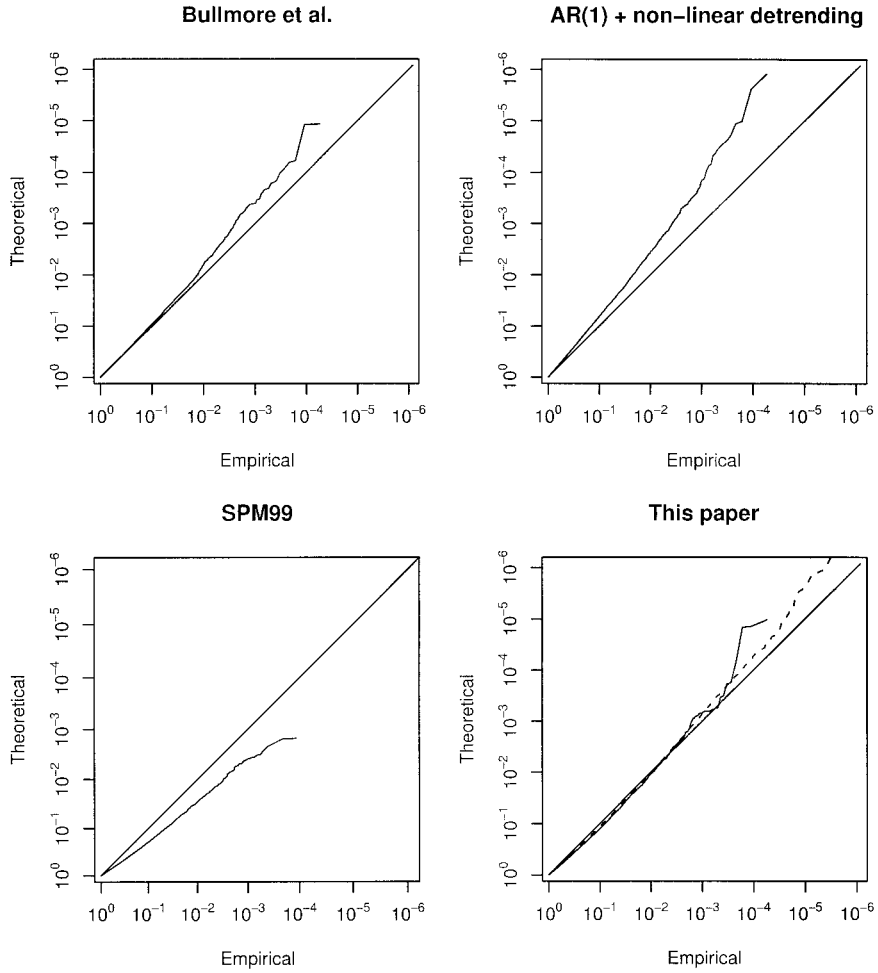

FIG. 16. Four methods are compared by plotting empirical *P* values obtained from the analysis of the null dataset against their theoretical values. In each case the straight line indicates exact agreement with the theoretical distribution. Top left: the approach proposed by (Bullmore *et al.*, 1996); top right: the AR(1) model proposed by (Bullmore *et al.*, 1996) with linear detrending replaced by the nonlinear detrending proposed in this paper; bottom left: the global AR(1) approach applied using SPM99; bottom right: the nonparametric spectral density estimation approach proposed in this paper. The solid line is for the assumed stimulus frequency and the dashed line is the calibration across all frequencies.

statistic distribution to its theoretical in each case using *PP*-plots shown in Fig. 16. The straight line at 45° indicates exact agreement with the theoretical distribution. As *P* value decreases from left to right the lines will tend to become increasingly more erratic as there are fewer observations in the tail of the distribution to accurately represent that part of the distribution. The plot in the top-left shows *PP*-plot obtained using the AR(1) approach proposed by (Bullmore *et al.*, 1996). It shows that the empirical distribution is significantly longer tailed than the theoretical and that the AR(1) model is systematically underestimating the spectral density of the underlying correlation structure. The plot in the top-right shows how the situation is worse in this example when the linear detrending employed in (Bullmore *et al.*, 1996) is replaced by the more resistant nonlinear detrending that we have proposed in this paper. The plot in the bottom left shows the results of applying the AR(1) approach implemented within the SPM99 software as described above. In this case, the empirical distribution has much shorter tails than the theoretical due to the global estimate systematically overestimating the spectral density of the underlying correlation structure. The plot in the bottom right shows the results of applying our own nonparametric spectral approach. The solid line represents the empirical distribution obtained at the fundamental frequency of activation and the dashed line represents the empirical distribution obtained from a range of higher frequencies. The dashed line effectively represents a very large sample from the null distribution and it can be seen that both lines are very similar to each other and in good agreement with the theoretical distribution.

The ability of the method we propose to self-calibrate and the above evidence indicating that the method is

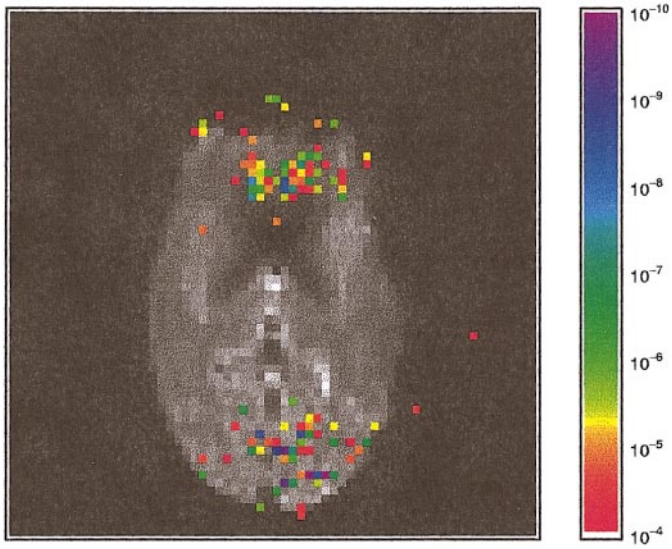

FIG. 17. *P* Value image of the statistic R_{88} in a slice from the null dataset, thresholded at 10^{-4}, and overlaid onto an image of the slice.

well calibrated with the theoretical distribution provide strong support for its use.

High Frequency Artefacts

In the null dataset used for this paper and in some others we have found artefacts which have been attributed to Nyquist ghosting. We have found these effects to occur in narrow bands of high frequencies and have been detected using the R_c statistic at these frequencies. Figure 17 shows the *P* value image of the statistic R_{88} in a slice from the null dataset, thresholded at 10^{-4} and overlaid onto an image of the slice. This indicates that very significant high frequency effects occur in this dataset and that these effects can occur well inside the brain image.

Time-domain models like AR(1) models involve specification of both the form of the matrix V and the innovations variance σ^2 and are susceptible to inaccuracies in the estimation of either. If no steps are taken to remove or protect against high frequency artefacts these will significantly bias the estimation of both V and σ^2.

To illustrate this point we selected an individual voxel from the null dataset containing a significant high frequency artefact and applied three different methods of detecting significance of a response to an imaginary periodic stimulus. Figure 18 shows the periodogram of the selected time series after nonlinear trend removal with a spectral density estimate shown as a thick line. The high frequency artefact can be seen as the two large periodogram ordinates at Fourier frequencies 87 and 88, which correspond to a period of approximately 6.8 s.

The first approach (Method I) was our implementation of the AR(1) model proposed in (Bullmore *et al.*, 1996) using nonlinear detrending. The second approach (Method II) was the same as the first but with the high frequency artefact at Fourier frequencies 87 and 88 included in the design matrix to stop the artefact biasing the estimate of the underlying correlation structure. The approach of this paper was the third method used (Method III). The results are summarized in Table 1. By comparing Methods I and II we see that removing the high frequency artefacts from the time series causes a large difference in both the estimated AR(1) coefficient and the estimated innovations variance. The high frequency artefact masks the true correlation structure and inflates the innovations variance. After removal the FPQ statistic is reduced and becomes closer to that obtained using Method III. Even then the difference is quite significant and is caused by the inability of the parametric AR(1) model to accurately fit the periodogram of the detrended voxel time series. From this we can conclude that nonparametric spectral estimation is a fast, accurate, and resistant method of estimating the underlying correlation structure of each voxel time series.

Conclusion

Parametric models are in general very sensitive to the assumptions they employ and significant loss in statistical efficiency can occur if these assumptions are invalidated. In contrast, if the assumptions of a parametric approach are valid then the use of a nonparametric approach will result in a comparatively small loss in efficiency. When considering voxel time series from fMRI datasets there can be no guarantees that the correct parametric model has been chosen and a nonparametric approach seems the most prudent in this context. We have demonstrated that the assumptions of commonly proposed time-domain models are not resistant to high frequency artefacts.

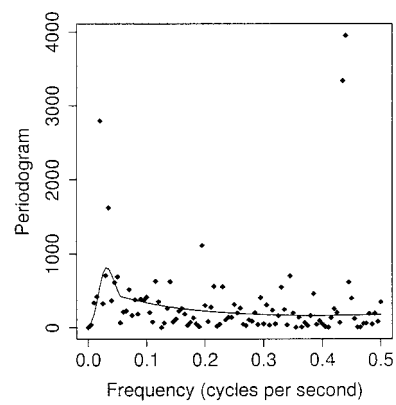

FIG. 18. The periodogram of a voxel time series within an area of the brain exhibiting the high frequency artefact. The trend was removed and a spectral density estimate is shown as a thick line. The high frequency artefact can be seen as the two large periodogram ordinates at Fourier frequencies 87 and 88.

TABLE 1

	AR(1) coefficient	$\widehat{\sigma^2}$	Numerator	Denominator	Ratio statistic
Method I	−0.0027	15765	591.88	315.29	1.877
Method II	0.2839	11276	624.98	419.46	1.490
Method III	—	—	689.30	513.88	1.341

Note. Comparison of three different methods of detecting a response to a periodic stimulus applied to an individual voxel time series from the null dataset containing a significant high frequency artefact. Method I: Our implementation of the AR(1) model proposed in (Bullmore *et al.,* 1996) using nonlinear detrending; Method II: As Method I but with high frequency artefacts removed; Method III: Our own spectral-domain approach. The AR(1) coefficient and $\widehat{\sigma^2}$ are not applicable to Method III which effectively estimates these quantities through the spectral density estimate. Columns four, five, and six contain the numerator, denominator, and ratio statistic, respectively, for each of the three methods. For methods I and II these three quantities correspond to FP, SE(FP), and FPQ, respectively. For method III they correspond to $I(\omega_c)$, $\widehat{f(\omega_c)}$, and R_c.

We believe that any approach proposed within the time-domain may have difficulty providing resistant estimates of both V and σ^2. There is also no guarantee that the parametric model chosen will be sufficiently flexible to capture the true form of the correlation structure even if artefacts are removed and a model selection procedure is employed. The flexible nonparametric approach proposed here acts as an insurance policy against the results being badly affected by artefacts, and is guaranteed to be near-optimal under all realistic operational conditions. It also offers a quick and accurate way to check the calibration of the procedure.

Extensions to Event-Related Designs

The direct analysis of nonperiodic designs will not be as simple as that of the periodic designs, since the response due to the stimulus will be spread over a range of frequencies. The use of iteratively reweighted least squares in the spectral domain, as described above, provides the generalization to cope with nonperiodic designs. The computational complexity that results in the use of the parameterized delay function in Lange and Zeger (1997) can be simply overcome by using a suitable set of basis functions Friston *et al.* (1995a) convolved with the stimulus $\{x_t: t = 0, \ldots, n - 1\}$ before transformation into the spectral-domain. Nonparametric spectral density estimation from the residuals of the iteratively reweighted least squares procedure could then be used and P values for significant activation at each voxel could be obtained using the standard distribution theory of linear models. Initial results of this approach are very promising and provide the generalization afforded by the linear model together with the accurate and resistant method of nonparametric spectral density estimation.

An alternative approach to nonperiodic designs is to treat them as a series of time series which can be time-aligned. For example, we might have treated the visual stimulus in the IOP dataset as five separate 20-scan blocks, with the stimulus occurring in the middle of each block. Fourier-transforming each block, averaging and then taking the squared magnitude would result in an analysis very similar to that presented here. This idea can be applied to any experiment in which stimuli occur at sampling points, and by phase adjustment of the Fourier transforms before averaging can be extended to any experiment with known stimulus times.

ACKNOWLEDGMENTS

The authors are grateful to Dr. Stephen Smith (Oxford Centre for Functional Magnetic Resonance Imaging of the Brain, funded by the Medical Research Council) for advice and datasets. J.M. was supported by an EPSRC Research Studentship.

REFERENCES

Bloomfield, P. 1976. *Fourier Analysis of Time Series: An Introduction.* Wiley, New York.

Bloomfield, P. 1991. Time series methods. In *Statistical Theory and Modelling* (D. V. Hinkley, N. Reid, and E. J. Snell, Eds.), pp. 152–176. Chapman and Hall, London.

Bullmore, E., Brammer, M., Williams, S. C. R., Rabe-Hesketh, S., Janot, N., David, A., Mellers, J., Howard, R., and Sham, P. 1996. Statistical methods of estimation and inference for functional MR image analysis. *Magn. Reson. Med.* **35:** 261–277.

Bullmore, E. T., Suckling, J., Overmeyer, S., Rabe-Hesketh, S., Taylor, E., and Brammer, M. J. 1999. Global, voxel, and cluster tests, by theory and permutation, for a difference between two groups of structural mr images of the brain. *IEEE Trans. Med. Img.* **18**(1): 32–42.

Cochrane, D., and Orcutt, G. H. 1949. Application of least squares regression to relationships containing autocorrelated error terms. *J. Am. Stat. Soc.* **44:** 32–61.

Diggle, P. J. 1990. *Time Series: A Biostatistical Introduction.* Oxford Univ. Press, Oxford.

Friston, K., Frith, C., Frackowiak, R., and Turner, R. 1995a. Characterising evoked hemodynamics with fMRI. *Neuroimage* **2:** 157–165.

Friston, K., Frith, C., Liddle, P., and Frackowiak, R. 1991. Comparing functional (PET) images: The assessment of significant change. *J. Cereb. Blood Flow Metab.* **11:** 690–699.

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

Friston, K., Holmes, A., Worsley, K., Poline, J.-B., Grasby, P., Williams, S., Frackowiak, R., and Turner, R. 1995c. Analysis of fMRI time series revisited. *Neuroimage* **2:** 45–53.

Friston, K., Jezzard, P., and Turner, R. 1994a. Analysis of functional MRI time-series. *Hum. Brain Map.* **1:** 153–171.

Friston, K., Worsley, K., Frackowiak, R., Mazziotta, J., and Evans, A. 1994b. Assessing the significance of focal activations using their spatial extent. *Hum. Brain Map.* **1:** 214–220.

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

Lange, N., and Zeger, S. L. 1997. Non-linear time series analysis for human brain mapping by functional magnetic resonance imaging. *Appl. Stat.* **46:** 1–29.

Loader, C. R. 1996. Local likelihood density estimation. *Ann. Statist.* **24:** 1602–1618.

Locascio, J. L., Jennings, P. L., Moore, C. I., and Corkin, S. 1997. Time series analysis in the time domain and resampling methods for studies of functional magnetic resonance brain imaging. *Hum. Brain Map.* **5:** 168–193.

Poline, J.-B., Worsley, K., Evans, A., and Friston, K. 1997. Combining spatial extent and peak intensity to test for activations in functional imaging. *Neuroimage* **5:** 83–96.

Priestley, M. B. 1981. *Spectral Analysis and Time Series.* Academic Press, London.

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

Silverman, B. W. 1985. Some aspects of the spline smoothing approach to nonparametric regression curve fitting. *J. R. Stat. Soc. B* **47:** 1–52.

Thomson, D. J. 1990. Time series analysis of Holocene climate data. *Phil. Trans. R. Soc. A* **330:** 601–616.

Venables, W. N., and Ripley, B. D. 1999. *Modern Applied Statistics with S-PLUS*. Springer, New York.

Wahba, G. 1980. Automatic smoothing of the log periodogram. *J. Am. Stat. Assoc.* **75:** 122–132.

Wilcox, R. 1998. Trimming and winsorization. In *Encyclopedia of Biostatistics* (P. Armitage and T. Colton, Eds.), Vol. 6, pp. 4588–4590. Wiley, Chichester.

Woods, R. P., Cherry, S. R., and Mazziotta, J. C. 1992. Rapid automated algorithm for aligning and reslicing PET images. *J. Comp. Assist. Tomog.* **16:** 620–633.

Worsley, K., and Friston, K. 1995. Analysis of fMRI time series revisited—Again. *Neuroimage* **2:** 173–181.

Worsley, K., Marrett, S., Neelin, P., and Evans, A. 1996. Searching scale space for activation in PET images. *Hum. Brain Map.* **4:** 74–90.

Zarahn, E., Aguirre, G. K., and D'Esposito, M. 1997. Empirical analyses of BOLD fMRI statistics. *Neuroimage* **5:** 179–197.