

Wavelets and statistical analysis of functional magnetic resonance images of the human brain

Ed Bullmore Brain Mapping Unit and Wolfson Brain Imaging Centre, University of Cambridge, Addenbrooke's Hospital, Cambridge, UK and Institute of Psychiatry, King's College London, London, UK, **Jalal Fadili** GREYC CNRS UMR 6072, Caen, France, **Michael Breakspear** Brain Dynamics Centre (Westmead Hospital) and School of Physics, University of Sydney, Australia, **Raymond Salvador** Brain Mapping Unit and Wolfson Brain Imaging Centre, University of Cambridge, Addenbrooke's Hospital, Cambridge, UK, **John Suckling** Brain Mapping Unit and Wolfson Brain Imaging Centre, University of Cambridge, Addenbrooke's Hospital, Cambridge, UK and **Michael Brammer** Institute of Psychiatry, King's College London, London, UK

Wavelets provide an orthonormal basis for multiresolution analysis and decorrelation or 'whitening' of nonstationary time series and spatial processes. Wavelets are particularly well suited to analysis of biological signals and images, such as human brain imaging data, which often have fractal or scale-invariant properties. We briefly define some key properties of the discrete wavelet transform (DWT) and review its applications to statistical analysis of functional magnetic resonance imaging (fMRI) data. We focus on time series resampling by 'wavestrapping' of wavelet coefficients, methods for efficient linear model estimation in the wavelet domain, and wavelet-based methods for multiple hypothesis testing, all of which are somewhat simplified by the decorrelating property of the DWT.

1 Introduction

1.1 General motivations for wavelet analysis of fMRI data

A wavelet is a little wave, or a brief wave. Unlike sine or cosine waves, which extend infinitely with a particular frequency and phase, wavelets are finitely extended or compactly supported; their oscillations decay more or less rapidly to zero (Figure 1). Over the last 15 years or so, wavelets have emerged as powerful new mathematical tools for analysis of complex datasets.

The first orthonormal basis after Fourier was constructed by Alfred Haar around 1910 and time-frequency analysis was subsequently developed by Dennis Gabor and John von Neumann in the late 1940s. The modern era – and the use of the word 'wavelet' – begins with work by Jean Morlet and Alex Grossman in the 1970s. Widespread application to signal processing followed the work of Stéphane Mallat¹ and the construction by Ingrid Daubechies² of a family of compactly supported orthonormal bases with arbitrary regularity or number of vanishing moments. A review of the historical development of wavelets is provided by Jaffard *et al.*³

Address for correspondence: MJ Brammer, Department of Biostatistics & Computing, Institute of Psychiatry, De Crespigny Park, London SE5 8AF UK. E-mail: m.brammer@iop.kcl.ac.uk

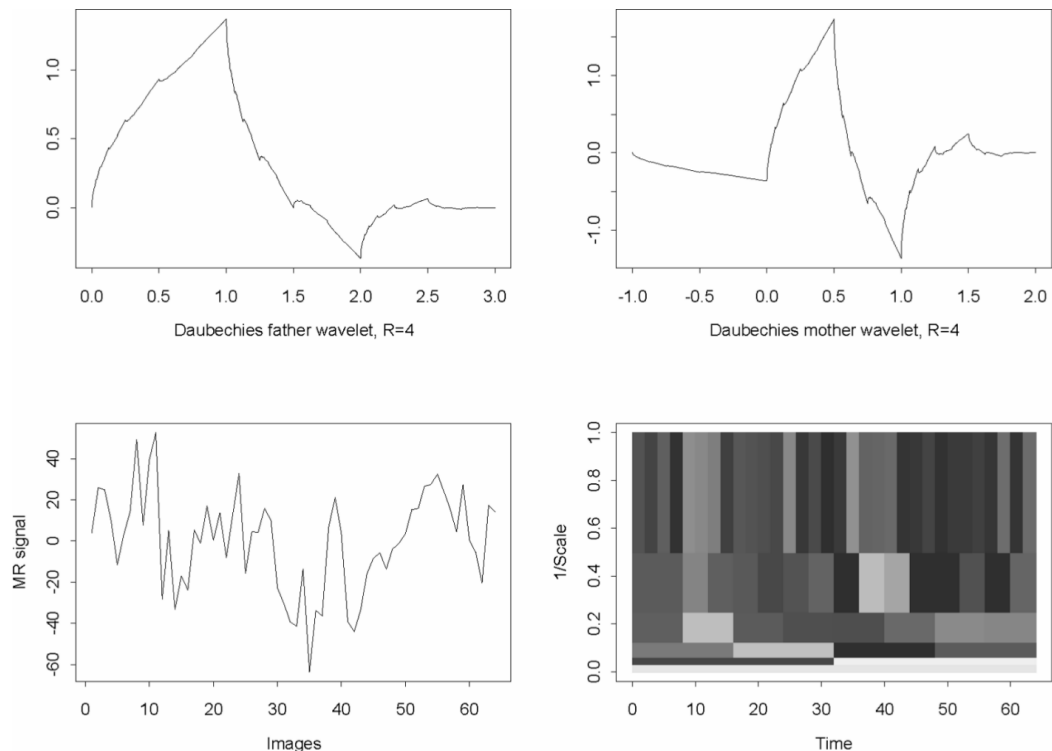


Figure 1 Wavelets and time-scale decomposition of a representative fMRI time series. (a), (b): Daubechies father and mother wavelets of order 4. (c) A 'raw' functional MRI time series extracted from the left postcentral gyrus, a cortical region important for somatosensory perception, in a single individual. (d) A time-scale plane tiling or scalogram illustrating the atomization or decomposition of the time series by the DWT using the Daubechies wavelets. The plane is covered by a set of rectangles of identical area, each of which is coloured in a shade of grey to indicate the sign and magnitude of the corresponding wavelet coefficient. Note that detail and approximation coefficients for the coarsest scale of the decomposition have the best resolution of scale (smallest height on the y-axis) but the worst resolution of time (greatest width on the x-axis) and vice versa for the detail coefficients for the finest scale of the decomposition.

Intuitively, wavelet analysis can be understood as a way of decomposing or atomizing the total energy or variance of a spatial process or time series by an orthonormal basis of wavelets, each of which is weighted by a coefficient representing the amount of energy in the data at a particular scale and location. If we think of the total energy in the data as a frequency–time or scale–space plane, then the discrete wavelet transform can be visualized as a tiling or tessellation of the plane in which each tile has the same area but tiles representing atoms of energy at fine scales have superior resolution in time or space compared to tiles representing atoms of energy at coarse scales (Figure 1).

This is evidently a multiresolutional analysis in that the energy of the data has been partitioned among a hierarchically organized set of scales. Low-frequency components of the energy will be represented by wavelet coefficients at coarse scales of the decomposition and higher frequency components will be represented by coefficients at finer scales. In this respect wavelet analysis is conceptually similar to Fourier analysis, which partitions the total energy of the data among an orthonormal basis set of sinusoidal

functions at different frequencies. However, wavelet analysis differs importantly from Fourier analysis by virtue of its natural adaptivity to local or nonstationary features of the data within scales of the decomposition. For example, a transient spike in a time series will be represented with difficulty by a set of stationary sinusoidal functions, but it will be captured quite deftly in terms of a few fine-scale wavelet coefficients located around the corresponding point in time. To quote Mallat¹: ‘If we are interested in transient phenomena – a word pronounced at a particular time, an apple located in the left corner of an image – the Fourier transform becomes a cumbersome tool’.

These two aspects of wavelet analysis – its multiresolutional nature and its adaptivity to nonstationary or local features in data – are sufficient to indicate that it will be of interest in analysis of functional magnetic resonance (fMRI) data, which we can expect will include possibly nonstationary features of interest at several scales. However, there are arguably at least three additional aspects of wavelet analysis that are advantageous for fMRI data analysis.

First, the wavelet transform is often a whitening or decorrelating transform of autocorrelated data, and this may prove to be statistically convenient in various ways. For example, as we show below, whitening of an autocorrelated time series by taking its wavelet transform can facilitate resampling or efficient linear model parameter estimation. Secondly, the wavelet transform has proven to be a useful basis for nonparametric regression, denoising or compression of large imaging datasets in many other applications. The signal-to-noise ratio in fMRI is often not much greater than one or two, so any techniques for enhancing representation of signal components are potentially valuable. Thirdly, the discrete wavelet transform implemented by Mallat’s pyramid algorithm is remarkably quick to compute: the algorithm has $O(N)$ complexity compared to $O(N \log(N))$ complexity of the fast Fourier transform. Computational speed is clearly of operational value in dealing with the large volumes of data (typically in the order of gigabytes) generated by a single fMRI study.

1.2 Wavelets, fractals and the brain

In addition to these general technical advantages of statistical analysis in the wavelet domain there is a related argument favouring the use of wavelet methods specifically in analysis of brain imaging data, which is founded on the expectation that such data may often demonstrate *fractal* properties. The fractal nature of the brain has a bearing on the suitability of wavelets for brain mapping because it has been recently recognized that wavelets are particularly apt for analysis and synthesis of fractal processes.^{4–6} The key feature of wavelet decomposition that makes it suitable for analysis of fractal data is that each level is a scaled (by a factor of 2) version of the next smallest scale. Hence a family of wavelet functions is itself a fractal.

Three particular ways in which wavelets are suitable or simplifying for statistical analysis of fractal processes are: i) wavelets effect a multiresolutional decomposition that is advantageous for analysis of fractal processes that, by definition, will demonstrate self-similar structure on several scales of measurement; ii) wavelets are theoretically optimal whitening or decorrelating filters for $1/f$ -like processes and many issues in estimation and hypothesis testing are simplified by independence; and iii) wavelets can be used to construct good estimators for fractal dimensions and other measures of

complexity.⁷ We will return to these advantageous aspects of wavelet analysis in greater detail below; first we define more thoroughly what we mean by ‘fractal’ and review some of the prior literature on fractal properties of brain mapping data that are illustrated in Figure 2.

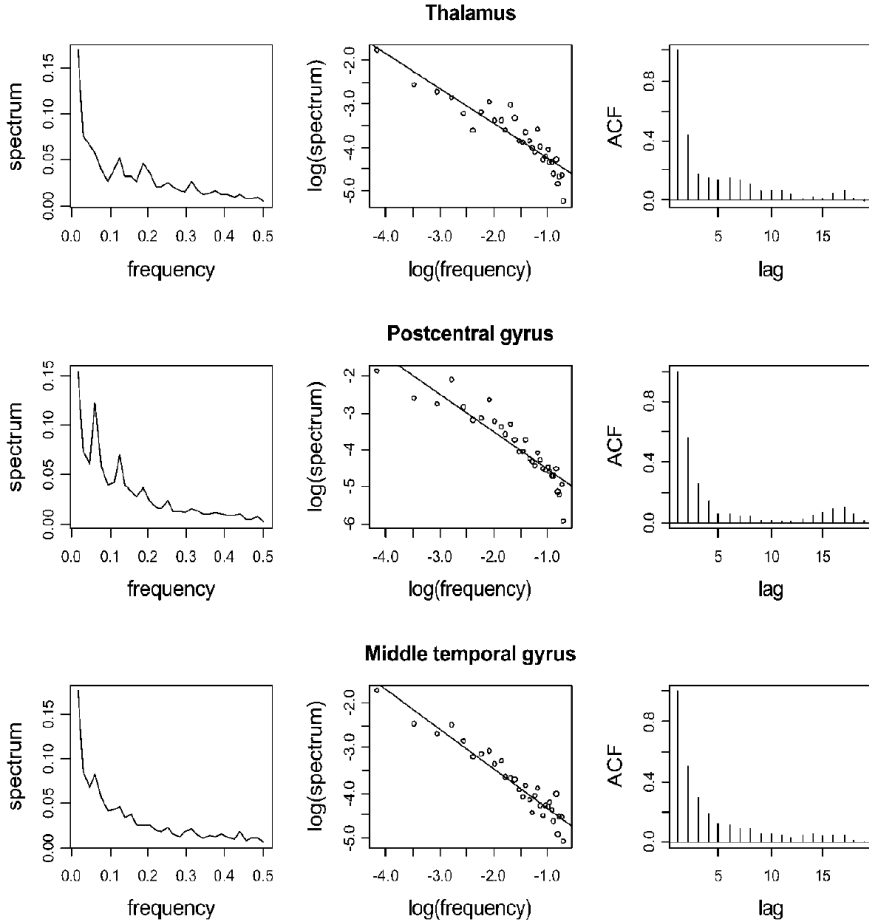


Figure 2 Long memory or $1/f$ -like characteristics of fMRI time series. Regional mean time series were extracted from three left-sided brain structures in each of 50 individual images: thalamus (top row), postcentral gyrus (middle row) and middle temporal gyrus (bottom row). The individually estimated power spectra and autocorrelation functions were averaged over all 50 images to obtain the group mean spectrum (left column) and autocorrelation function (right column) for each region. It is clear that there is disproportionate power at low frequencies in all regions, and the series are positively autocorrelated in time. The spectrum and ACF for the postcentral gyrus are distinguished by evidence for power at frequencies corresponding to the (fundamental) frequency of contralateral, peripheral somatosensory stimulation and its first few harmonics: this is a well-localized, experimentally driven effect. Thalamus and middle temporal gyrus were not significantly activated by this experimental stimulation and their data provide some insight into the properties of fMRI time series under the null hypothesis. The middle column shows a plot of $\log(\text{spectrum})$ vs. $\log(\text{frequency})$ for each region, and a straight line fitted by least squares to estimate the spectral exponent γ . The resulting parameter estimates for each region are: Thalamus $\hat{\gamma} = -0.80$, 95% CI $[-0.67, -0.93]$; Postcentral gyrus $\hat{\gamma} = -1.02$, 95% CI $[-0.86, -1.18]$; Middle temporal gyrus $\hat{\gamma} = -0.88$, 95% CI $[-0.79, -0.97]$.

The word ‘fractal’ was coined by Mandelbrot⁸ to define a class of objects with certain interesting properties in common. Typically, fractals demonstrate self-similarity (or self-affinity), meaning that the statistics describing the structure in time or space of a fractal process remain the same as the process is measured over a range of different scales. Exact self-similarity means that an object will look the same after rescaling by a single factor in all dimensions (zooming in or out); statistical self-similarity means that it will look approximately the same. Self-affinity means that the process will look the same when it is nonuniformly rescaled by a different factor for different dimensions. For example, self-affinity of the fractional Brownian motion $B(t)$, with Hurst exponent H , means that the rescaled series in time $B(st)$ will have the same statistical distribution as the rescaled signal on the original time scale $s^H B(t)$.

An informal, familiar example of self-similarity is provided by the complex branching structure of a tree, which is approximately preserved on examination of a single branch or twig, that is, under examination at progressively finer scales of resolution. The complexity of self-similar structures can be quantified in terms of their (usually noninteger) fractal dimensions: for example, a fractal surface will have a fractal dimension (D) in the range $2 < D < 3$, with more complex or space-occupying surfaces approaching the limit $D = 3$ and simpler, more nearly Euclidean planar surfaces having D closer to 2. Fractal time series that have $1 < D < 2$, like the human electrocardiogram (ECG), typically have long-term autocorrelations (long memory) in time and power laws for spectral density $S(f)$ as a function of frequency

$$S(f) \propto \frac{1}{f^\gamma}$$

or

$$\log \{S(f)\} = c + \gamma \log (f)$$

See Figure 2 for an example in fMRI. The power law exponent γ defining the slope of the linear relationship between log power and log frequency for a $1/f$ process is simply related to the fractal dimension of the process in time

$$D = T + \frac{3 - \gamma}{2}$$

where T denotes the topological dimension ($T = 1$ for a time series). There is also a simple relationship between the fractal dimension and the Hurst exponent $0 < H < 1$ which parameterizes the covariance matrix of fractional Brownian motion:^{9,10}

$$D = T + 1 - H$$

Here we are referring to the ‘box-counting’ or Hausdorff dimension – which captures an intuitive idea of how much space an object occupies. More technically, we can define the family of dimensions D_q by

$$D_q = \lim_{r \rightarrow 0} \frac{1}{q-1} \frac{\log \sum_k p_k^q}{\log r}$$

where the space has been divided into cells of size r and p_k is the relative density of points inside cell k . The value $q=0$ defines the Hausdorff dimension; $q=1$ defines the information dimension; and $q=2$ the correlation dimension approximated by the famous Grassberger–Procaccia algorithm.¹⁰

Following Mandelbrot,⁸ it has become increasingly clear that fractal properties are shared by a wide variety of natural phenomena, ranging from internet traffic¹¹ and econometric time series¹² through DNA base sequences¹³ and ecosystems¹⁴; see <http://linkage.rockefeller.edu/wli/1fnoise> for a comprehensive bibliography on $1/f$ noise. The pathophysiology of the heart is arguably the biomedical system most productively investigated to date using mathematical tools from fractal geometry and wavelet theory. It has been shown that the branching structures of the coronary arterial and His-Purkinje fibre trees are self-similar and have fractal dimensions. Moreover, the fractal geometry of cardiac anatomy has been related to the power law dynamics of the ECG and various fractal and wavelet-based measures of complexity of ECG data have been developed and shown to add value in diagnosis of coronary artery disease and prognosis of otherwise sudden arrhythmias;¹⁵ see <http://reylab.bidmc.harvard.edu/> for a bibliography of related work. More generally, the fractal geometry of metabolite exchange surfaces and vascular transport systems has been used persuasively to explain the widespread prevalence of non-Euclidean allometric scaling laws in biology.¹⁶

There have been some comparable applications of fractal concepts to analysis of neuroscience data. Dendritic branching patterns of single neurons have been quantified in terms of fractal dimensions.¹⁷ Other groups have measured fractal dimensions and $1/f$ spectral properties of electroencephalographic (EEG) signals.^{18–20} The fractal properties of anatomical surfaces and boundaries segmented from human MRI data have been measured.^{21–25} Fractal methods have been applied to analysis of radio-ligand SPET and PET images²⁶ and imaging-orientated models for cerebral blood flow have been proposed on the basis of the probably fractal geometry of cerebrovascular architecture.²⁷ There have also been some preliminary investigations of $1/f$ spectral properties in fMRI time series.^{28–30}

In short, fractal processes are virtually ubiquitous in nature and there has already been some successful work applying ideas from fractal geometry to analysis of several modalities of human brain mapping data. This provides a case for consideration of wavelets as more than ‘just another basis’, one among many possible mathematical domains, for the statistical analysis of fMRI data.

It is probably also relevant to note that wavelets are increasingly invoked in the theoretical and numerical study of complex dynamical systems. For example, wavelets have been shown parsimoniously to capture the rich dynamics of morphological phenomena such as microbial growth and nonequilibrium chemical reactions;³¹ to display the flow of information between scales in nonequilibrium fluid flows;³² and to predict the behaviour of spatially extended nonlinear dynamical equations.³³ These aspects of wavelets may be leading indicators of future applications to fMRI and electrophysiological data because they show how wavelets can shed light on the underlying mechanisms of pattern formation and information flow in complex systems like the brain.

1.3 Overview of wavelet-based methods for fMRI data analysis

Previous general reviews of wavelet applications in biomedical image processing, including some early work on fMRI, are provided by Aldroubi and Unser³⁴ and

Laine.³⁵ Statistical issues in wavelet analysis of time series are addressed comprehensively by Percival and Walden.³⁶ Wornell^{4,5} makes a detailed case for the general optimality of wavelet representations for analysis of fractal signals. Bruce and Gao³⁷ describe implementation of wavelet methods in S-PLUS.

Several research groups have pioneered applications of wavelets to various issues in fMRI data analysis. The most popular application to date has been image *compression or denoising*.^{38–44} *Multiresolutional analysis* of spatial maps of fMRI time series statistics has been explored by Ruttimann *et al.*⁴⁵ Brammer⁴⁶ and Desco *et al.*⁴⁷ *Linear model estimation* in the wavelet domain has been described by Fadili and Bullmore,³⁰ Meyer⁴⁸ and Müller *et al.*⁴⁹ *Resampling* of fMRI data in the wavelet domain has been developed by Bullmore *et al.*,⁵⁰ Hossien-Zadeh *et al.*,⁵¹ and Breakspear *et al.*⁵² There have also been applications of wavelets to the image processing problems of registering individual fMRI datasets in a standard anatomical space⁵³ and correcting unidirectional geometric distortions in echoplanar imaging data.⁵⁴

There have also been a number of interesting applications of wavelet methods to analysis of human brain mapping data in other modalities. Turkheimer *et al.*^{55,56} have developed methods for multiresolution analysis and linear modeling (e.g. ANOVA) of multisubject positron emission tomography (PET) studies in the wavelet domain; Cselenyi *et al.*⁵⁷ explored two- and three-dimensional wavelet transforms as spatial filters of radioligand binding potential maps measured using PET; Raz *et al.*⁵⁸ used wavelet packet analysis to decompose auditory evoked potentials (electrophysiological data) into component waveforms; and Barra and Boire⁵⁹ reported a technique for brain tissue classification or segmentation of structural MRI based on fuzzy clustering of wavelet coefficients.

In the rest of this paper, we provide a brief formal introduction to some key properties of the discrete wavelet transform (DWT). We will focus predominantly on dyadic, orthonormal wavelets in the knowledge that related methods for time-invariant or undecimated wavelet transformation, which effect a redundant or nonorthogonal multiresolutional decomposition, are also available and may have some advantages for aspects of time series analysis; see Ref. 36 for details. We will discuss in more detail its application to three aspects of statistical analysis of fMRI data: resampling of fMRI time series; time series modeling and nonparametric regression; and multiple hypothesis testing. In all three of these areas of application, the whitening or decorrelating property of the DWT can be exploited for statistical convenience.

2 The discrete wavelet transform (DWT)

2.1 Notation and definitions

Wavelets can be formally defined as families of functions that form an orthonormal basis for a large class of physically relevant (square integrable) functions. A wavelet family is obtained by dilation and translation of a compactly supported ‘mother’ wavelet ψ with zero integral over time $\int \psi(t) dt = 0$:

$$\psi_{j,k}(t) = \frac{1}{\sqrt{2^j}} \psi\left(\frac{t - 2^j k}{2^j}\right)$$

and by dilation and translation of a ‘father’ wavelet or scaling function ϕ with unit integral over time $\int \phi(t) dt = 1$

$$\phi_{j,k}(t) = \frac{1}{\sqrt{2^j}} \phi\left(\frac{t - 2^j k}{2^j}\right)$$

where $j = 1, 2, 3, \dots, J$ indexes the scale $S_j = 2^j = 2, 4, \dots, 2^J$ to which the wavelet has been dilated and $k = 1, 2, 3, \dots, K = N/2^j$ indexes the location in time or space to which it has been translated.

Wavelets can be distinguished by their smoothness or regularity, which is closely related to the number of vanishing moments R . The number of vanishing moments of a mother wavelet ψ is defined to be the largest integer R that satisfies $\int t^r \psi(t) dt = 0$, where $r = 0, 1, \dots, R - 1$.

2.2 Multiresolutional analysis

The discrete wavelet transform of a spatial process or time series achieves a multi-resolutional analysis in which the total variance or energy of the data is distributed over a hierarchy of scales, S_j . At each scale the data is decomposed into two orthogonal components – the detail coefficients $d_{j,k}$ and the approximation coefficients $a_{j,k}$. These coefficients are respectively defined as the inner products of the data and the correspondingly scaled and dilated mother $\psi_{j,k}$ or father $\phi_{j,k}$ wavelets. The detail coefficients contain information about variation in the data at a particular scale and the approximation coefficients represent the residual of the data after the information on this and all finer scales has been removed. The original data y can be losslessly recovered by adding the approximation at the coarsest scale and the details at this and all finer scales:

$$y = \sum_k a_{J,k} \phi_{J,k} + \sum_{j \leq J} \sum_k d_{j,k} \psi_{j,k}$$

2.3 Whitening properties of the discrete wavelet transform

Generally, the correlation between wavelet coefficients – both within and between scales of the decomposition – will be small even if the data are highly autocorrelated. This is sometimes called the whitening or decorrelating property of the DWT and it was first understood theoretically for the class of signals known as fractional Brownian motion.^{60–62} Wornell^{4,5} later showed that the DWT has optimally decorrelating or Karhunen–Loève properties for the wider class of signals with $1/f$ -like power spectral density functions. More specifically, we can say that the correlation between any two detail coefficients $d_{j,k}$ and $d_{j',k'}$

$$\langle d_{j,k}, d_{j',k'} \rangle \sim |2^j k - 2^{j'} k'|^{2(H-R)}$$

or that the correlation between any two detail coefficients at the same scale $d_{j,k}$ and $d_{j,k'}$

$$\langle d_{j,k}, d_{j,k'} \rangle \sim |k - k'|^{2(H-R)}$$

where R denotes the regularity of the wavelet and H is the Hurst exponent of the process. Hence, provided that the number of vanishing moments of the chosen wavelet

basis functions is sufficiently large, $R > 2H + 1$, the correlations will decay rapidly and indeed may be negligible even for adjacent coefficients.

2.4 Wavelet estimators of fractal noise parameters

The Hurst exponent (and therefore the fractal dimension and spectral exponent) of a time series can be estimated from its wavelet transform. If a $1/f$ -like noise with Hurst exponent H is projected onto a wavelet basis with R vanishing moments, and if $0 < (2H + 1) < 2R$, then the sample variance of the wavelet coefficients at the j th scale

$$\text{Var}\{d_k\}_j = \frac{1}{N-1} \sum_k (d_{k,j} - \bar{d}_{\cdot j})^2$$

is related to the scale by a power law with exponent $2H + 1$ ⁶⁰

$$\text{Var}\{d_k\}_j \approx (2^j)^{2H+1}$$

From this expression various estimators of H can be derived,^{4,63} of which the simplest is a least-squares fit of the linear model⁶⁰

$$\log_2(\text{Var}\{d_k\}_j) = c + (2H + 1)j + \epsilon_j$$

See Figure 3 for an example in fMRI.

3 Data resampling in the wavelet domain or ‘wavestrapping’

Data resampling by permutation or bootstrap offers many advantages for inference on functional neuroimaging data – in particular it obviates the need to make probably unrealistic assumptions about spatial autocovariance and other distributional attributes of the data and, perhaps for these reasons, an appropriate nonparametric test can have superior sensitivity compared to a parametric alternative.^{64,65} Moreover, there are many statistics of potential interest in brain mapping, for example, spatial and multivariate statistics, that do not have theoretically tractable or well-established distributions under the null hypothesis and therefore cannot properly be tested parametrically. By contrast, almost any statistic of interest may be accessible to inference on the basis of an appropriate resampling scheme (see Bullmore *et al.*⁶⁶ and Welchew *et al.*⁶⁷ for examples, respectively, of resampling spatial and multivariate statistics in brain mapping).

However, designing an appropriate resampling scheme for statistics estimated by analysis of a time series is of course complicated by nonindependence of the observations under the null hypothesis. If we write the linear model

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

where y is the fMRI time series observed at a single voxel, X is the design matrix summarizing experimental effects of interest, β is the parameter vector to be estimated, and ϵ is the vector of residuals, then we must allow that the covariance matrix of the

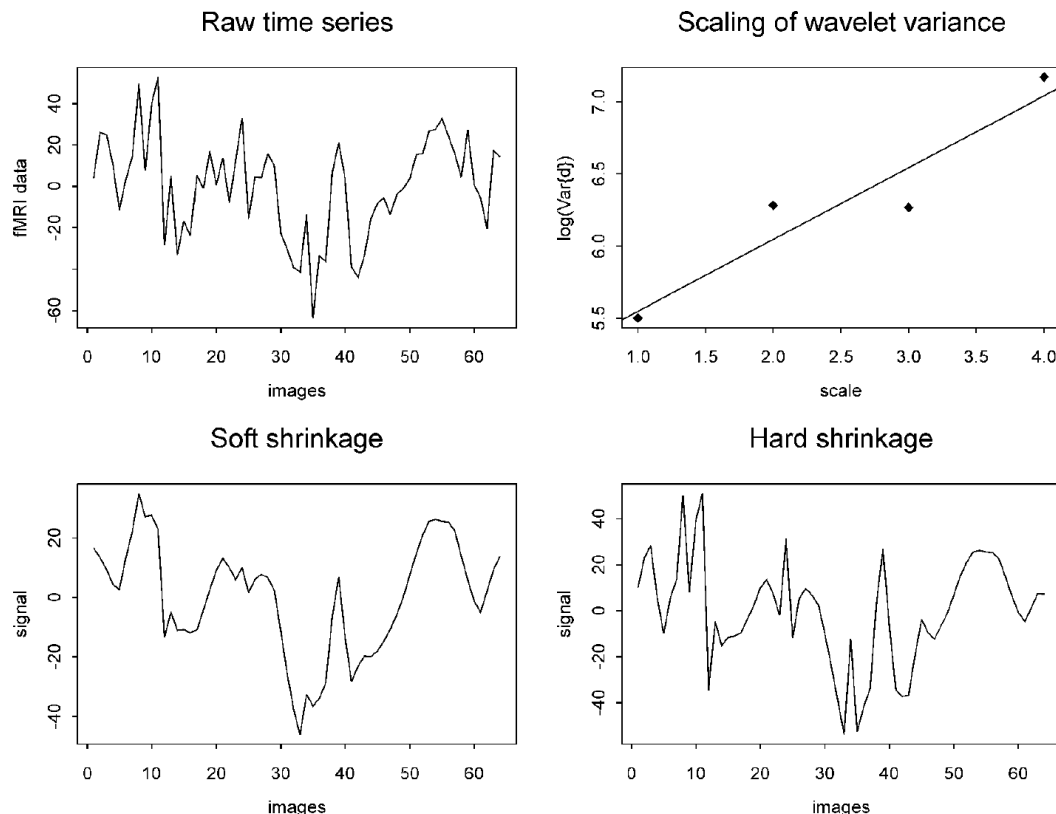


Figure 3 Functional MRI time series: wavelet shrinkage and wavelet estimator of Hurst exponent. (a), (b) A 'raw' functional MRI time series extracted from the left postcentral gyrus, a cortical region important for somatosensory perception, in a single individual. (c), (d) Results of wavelet shrinkage or nonparametric regression using soft (c) and hard (d) thresholding rules and Donoho and Johnstone's universal threshold. The approximately periodic structure of the signal is clearer after denoising by the soft shrinkage rule. (b) Wavelets provide convenient estimators of the Hurst exponent H (or fractal dimension D) of a time series. A simple estimator shown here is the slope of a straight line fitted to a plot of the log of the variances of the wavelet (detail) coefficients at different scales $\log(\text{Var}(d_k)_j)$ vs. the relevant scales of the decomposition; by this method, $H = 0.48$ or $D = 1.52$.

residuals σ will not generally be diagonal – there will be appreciable autocorrelation or serial dependency of the residuals in the time domain.⁶⁸

The causes of residual or endogenous autocorrelation in fMRI time series are not yet certainly known: they may include instrumental noise, head movement in the scanner, and aliased cardiac or respiratory cycle pulsation. In addition to these 'nuisance' sources of coloured noise, there may also be contributions from substantively more interesting neurophysiological processes. The problem for resampling is clearly that it will be inappropriate in this context simply to 'reshuffle' the data points in the time domain. This will destroy serial dependency or whiten the data and differentially bias the estimation of any standardized test statistic in the resampled series. It may also destroy physiologically important properties of the data. In short, random resampling

in time is not a valid basis for a test because the reshuffled units of observation are not *exchangeable*. (A set of n units of observation of the random variable X is termed exchangeable if the joint probability distribution $p(X_1, X_2, X_3, \dots, X_n)$ is invariant under permutation of the units; see Lindley and Novick⁶⁹ for details.) Various methods have been proposed to circumvent this problem (and are more completely reviewed by Davison and Hinkley⁷⁰). *Block resampling* involves defining the resampled unit as a block of consecutive time points; provided the length of each block is long enough to encompass all (possibly long memory) dependencies in the time series the unspecified correlational structure of the data may be preserved under resampling.⁷¹ *Model-based prewhitening* involves specifying a parametric form for the dependency in ϵ and correcting the data and design matrix so that the residuals of the corrected model are whitened. For example, if we specify a first-order autoregressive AR(1) model for the residual series

$$\epsilon_t = \eta \epsilon_{t-1} + \rho_t, \quad \rho \sim N(0, \sigma^2 I)$$

and transform the data and design matrix by the estimated AR(1) coefficient $\hat{\eta}$ so that

$$\begin{aligned} y_t^* &= y_t - \hat{\eta} y_{t-1} \\ X_t^* &= X_t - \hat{\eta} X_{t-1} \end{aligned}$$

then the residuals ϵ^* of the transformed model $y^* = X^* \beta + \epsilon^*$ will be white, provided the AR(1) model is adequate to account for the autocorrelation in the observed time series under the null hypothesis. This method can of course be generalized to accommodate a higher-order autoregressive model for the observed autocorrelation. The key point is that the data (or model residuals) are rendered exchangeable, and a permutation test on this basis is therefore valid, if the serial dependency in the residual process ϵ is accurately modeled by the specified AR model.

A third strategy is to resample the observed time series after *orthogonal transformation to another domain*. A well-known example of this approach is to take the Fourier transform of the observed time series, randomly permute the phases of the complex valued transform over all Fourier frequencies, then take the inverse transform of the ‘phase-scrambled’ data to obtain a resampled time series that preserves the spectral density of the observed data.⁷² Exchangeability is conferred by the independence of the Fourier transform at different frequencies, and by the independence of the real and imaginary components of the transform at each frequency. However, if the observed data are nonstationary, or include long memory dependencies, then this phase-scrambling algorithm may fail adequately to preserve their second-order stochastic properties;⁷⁰ it may also have undesirable effects on the amplitude distribution of the data.

The discrete wavelet transform provides an alternative device for this strategy of resampling a time series after orthogonal transformation. In its simplest form, ‘wavestrapping’ involves:

- 1) computing the discrete wavelet transform of the observed series;
- 2) randomly permuting the decorrelated detail and approximation coefficients within each scale of the decomposition; and

3) computing the inverse wavelet transform to recover the resampled series in the time domain.

As illustrated in Figure 4, this method can faithfully preserve the autocorrelational structure of a $1/f$ -like time series because its wavelet coefficients will be whitened and therefore exchangeable in the wavelet domain. It is also notable that, because of their spatio-temporal localization, wavelets provide a suitable basis for resampling of

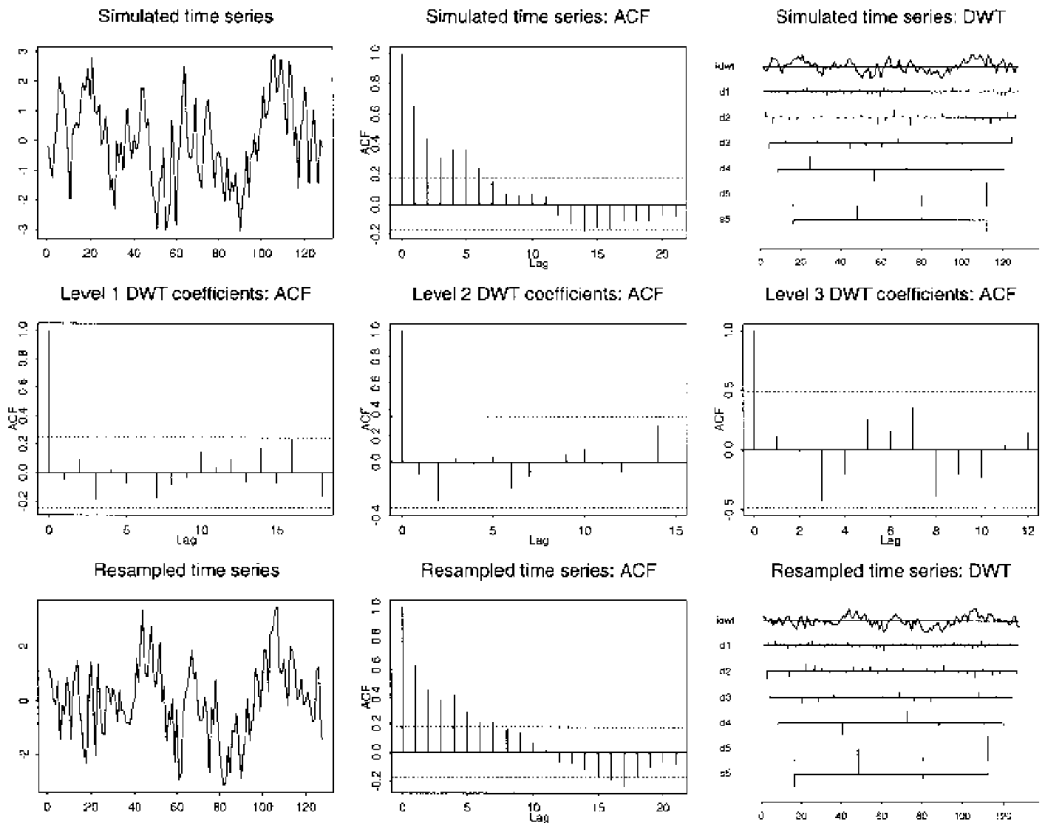


Figure 4 Resampling in the wavelet domain or ‘wavestrapping’ of simulated $1/f$ noise. Top row, from left to right: A long memory time series simulated by a physical model of multiple relaxation processes ($N = 128$); its autocorrelation function (ACF), with dashed lines indicating Bartlett’s 95% confidence interval for zero $0 \pm 2/\sqrt{n}$, and its discrete wavelet transform (DWT). In the top right panel, the coefficients of the dilated and translated mother wavelets are shown for five levels of detail j , labelled d1–d5; and for the father wavelet, labelled s5. The top row of this panel shows the time series reconstructed by the inverse wavelet transform. Middle row, from left to right: The autocorrelation functions of the wavelet coefficients at levels d1–d3 are shown with dashed lines indicating 95% CI for zero, $0 \pm 2/\sqrt{n/2^j}$. Bottom row, right to left: The wavelet coefficients after random permutation within each level of detail; the autocorrelation function of the time series obtained by the inverse wavelet transform on the resampled coefficients; the resampled time series. The key point is that although the original time series is strongly autocorrelated, its wavelet coefficients are relatively whitened or decorrelated, and random permutation of these serially independent or exchangeable coefficients generates a resampled time series with an autocorrelation function very similar to the original. Figures produced by permission of Wiley.

possibly irregular subsets of an image, such as the domain of voxels representing intracerebral tissue in an MR image. Resampling of a locally defined subset of the data cannot so readily be achieved by Fourier techniques.

This method is more fully described by Bullmore *et al.*⁵⁰ for sampling the null distribution of the linear model parameter vector β by fitting the time series regression model to the fMRI data at each voxel after repeated permutation in the wavelet domain. Breakspear *et al.*⁵² describe two variations on this resampling scheme that more exactly preserve the phase and amplitude distribution of the original data: block resampling in the wavelet domain, which involves permutation of blocks of adjacent wavelet coefficients within each level of the decomposition; and cyclic rotation of the wavelet coefficients within each level by a random shift. Breakspear *et al.*⁵² use these methods for generation of surrogate data ensembles to test the null hypothesis of zero nonlinear interaction between coupled attractors in a dynamical systems analysis of electrophysiological data.

A crucial issue in the design and validation of wavestrapping schemes is the extent to which the wavelet coefficients are indeed decorrelated. For fractal time series there is useful theory predicting decay of the correlation between coefficients within scale as an exponential function of the distance between them; but the theoretical constraints on decay of the correlation between higher-dimensional, for example, two- or three-dimensional, wavelet coefficients are not so clearly specified, at least to our knowledge. Moreover, there is as yet only preliminary evidence in support of the assumption that fMRI time series generally have $1/f$ -like spectral densities or long memory autocorrelation functions. In any case, fMRI time series will invariably be finite and are typically rather short ($n \sim 50$ – 200 time points). This potentially creates a problem for wavestrapping because estimation of wavelet coefficients at the boundaries of finite time series can introduce artefactual correlations between coefficients; this is notably an issue with the widely used periodic convolutional filter for boundary correction, although nonconvolutional filters may be less problematic in this respect.⁷³

These caveats imply that development and optimization of resampling schemes in the wavelet domain remains an active focus for methodological research, but the potential advantages of this approach seem clear for resampling a wide range of test statistics on the null hypothesis of long memory autocorrelational structure in time.

4 Linear modeling in the wavelet domain

The existence of long memory noise in fMRI time series not only complicates resampling but also impacts on the efficiency of estimation of the linear model parameter vector β . It is well known that ordinary least squares (OLS) will be the best linear unbiased (BLU) estimator of β if the residual series ϵ is white. However, if ϵ is autocorrelated, OLS will be less than optimally efficient and will severely underestimate the standard error of β . One response to this problem is to formulate a linear time invariant (LTI) model for the serial dependency in ϵ , the simplest example of which is the AR(1) model already discussed; then use the estimated parameters of the LTI model to prewhiten the data and design matrix, or diagonalize the error covariance matrix σ . Variants of this approach – autoregressive least squares (ARLS) – have been

quite widely advocated or discussed in fMRI data analysis^{68,74–79} but are susceptible to failure, with consequent loss of type 1 error control due to overestimation of standardized test statistics, in the context of noise structures more complex than predicted, for example, by low-order AR models.

An alternative strategy to achieve diagonalization of the error covariance matrix, and thereby optimize efficiency in estimation of β , is to transform both data and design matrix into the wavelet domain prior to signal and noise parameter estimation.³⁰ This approach is predicated on the assumption that the residual process is fractal or has scale-invariant properties. More formally, we write the regression model as

$$y = X\beta + \epsilon, \quad \epsilon \sim N(0, \Sigma\{H\})$$

that is, we suppose that the error covariance matrix is not diagonal but has an off-diagonal structure parameterized by the Hurst exponent, H . Taking the discrete wavelet transform of this model gives

$$y_w = X_w\beta + \epsilon_w, \quad \epsilon_w \sim N(0, \Sigma_w)$$

where X_w is the result of applying the DWT separately to each column of the design matrix, and y_w and ϵ_w are, respectively, the wavelet transforms of the data and the error process. As noted earlier, the orthonormal DWT is a Karhunen–Loève expansion of $1/f$ processes. To a good approximation, the variance–covariance matrix of the noise is therefore *diagonalized* by the wavelet transform, that is, Σ_w is a diagonal matrix.

The L_2 regression problem can be solved in the wavelet domain using an iterative maximum likelihood (ML) estimator, called wavelet-generalized least squares (WLS).³⁰ The WLS estimator was shown to be the BLU estimator of regression model parameters

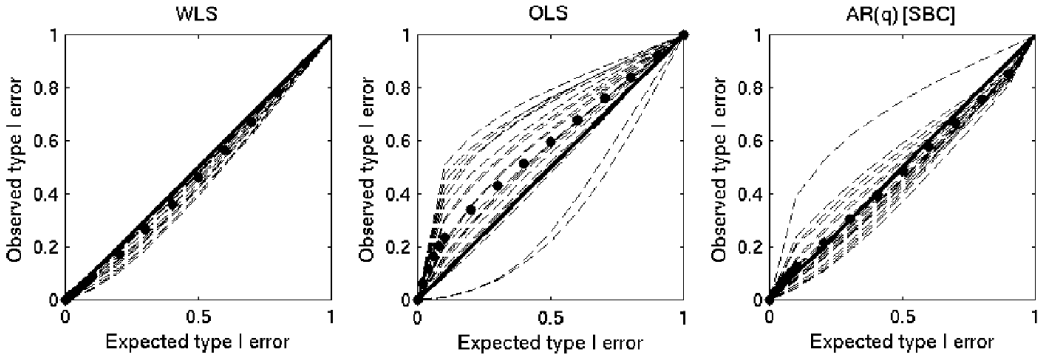


Figure 5 Type 1 error calibration curves for three linear model estimators applied to analysis of ‘null’ fMRI datasets. (a) WLS=wavelet-generalized least squares; (b) OLS=ordinary least squares; (c) AR(q)[SBC]=autoregressive least squares using AR models of order q specified by minimization of the Schwarz Bayesian criterion (SBC) at each voxel. Exact type 1 error control is indicated by correspondence between observed and expected numbers of positive tests. For both OLS and AR(q)[SBC] estimators, there is loss of type 1 error control (observed false positives > expected false positives) implying inadequate whitening of autocorrelated fMRI noise. The WLS estimator provides the basis for a valid, sometimes slightly conservative test. Figures produced by permission of Elsevier.

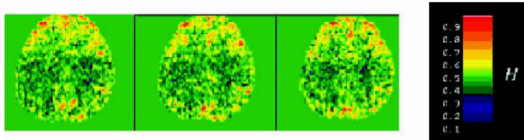


Figure 6 Maps of the Hurst exponent ($0 < H < 1$) estimated at each voxel of a single 'null' image by wavelet-generalized least squares. Selected slices of the parameter map are shown corresponding to Talairach z-coordinates +26mm (a), +30mm (b) and +34mm (c). H is colour-coded as shown in the bar to the right so that time series with long memory noise parameterized by $H \rightarrow 1$ are coloured orange/red. These voxels are concentrated symmetrically in the frontal cortex and in the medial posterior parietal cortex. Central areas of white matter are characterized by lower values of \hat{H} (dark green/blue voxels).

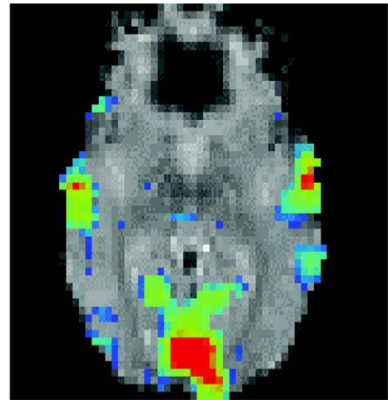


Figure 7 Multiple hypothesis testing in the wavelet domain. The method for multiple hypothesis testing proposed by Brammer⁴⁶ (1998) was here applied to detection of activation in a single slice of fMRI data acquired during bimodal (visual and auditory) stimulation at different frequencies. The Kolmogorov-Smirnov test statistic is colour-coded at each voxel. Red voxels indicate brain regions where the behaviour of the wavelet coefficients at scales corresponding to the frequencies of bimodal stimulation departed significantly from the expectation under the null hypothesis of an i.i.d. Gaussian process. Figure produced by permission of Wiley.

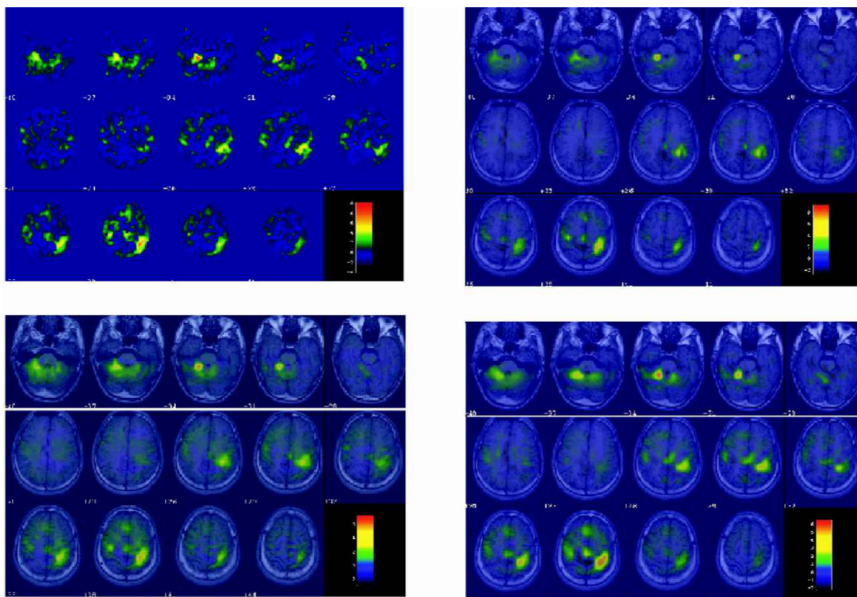


Figure 8 Multiple hypothesis testing in the wavelet domain: control of FDR and Bayesian approaches. (a) Spatial map of test statistics estimated by linear modelling (WLS) of the fMRI time series at each voxel of an individual image. (b) Thresholded map controlling false discovery rate at $P=0.01$. (c) Thresholded map using Ogden's recursive hypothesis testing algorithm with $P=0.01$. (d) Thresholded map using Vidakovic's Bayesian rule for wavelet thresholding. See Fadili and Bullmore⁸⁷ for further details. Figure produced by permission of SPIE.

in the context of long-memory errors; it also provides an ML estimate of the Hurst exponent and the error variance σ^2 of the noise at each voxel. Properties of this estimator, including its Cramèr–Rao bounds, were derived theoretically and compared satisfactorily to its empirical performance on a range of simulated data.

One useful empirical technique to evaluate the performance of different estimators and inferential procedures in fMRI data analysis is to apply them to analysis of ‘resting’ or ‘null’ datasets acquired while human subjects lie quietly in the scanner with their eyes closed. These data provide an experimental realization of the null hypothesis that brain activity is not determined by experimental design, since there is no explicit experimental input driving brain function in these data. If the linear model parameter β is then estimated at each voxel, for an arbitrary design matrix X , and tested against the null hypothesis $\beta = 0$, we expect αV (false) positive tests over the whole image, where $0 < \alpha < 1$ is the size of hypothesis test at each voxel and V is the search volume or number of voxels tested. As shown in Figure 5, parameter estimation by WLS gives almost exact type 1 error control over all sizes of test applied to null images acquired from seven individuals and analysed with three different design matrices each. As expected, OLS is associated with unacceptable loss of type 1 error control: the number of positive tests is much greater than αV because the standard error of β is substantially underestimated by OLS in the context of autocorrelated noise. The same is true, to a lesser extent, for an estimator using an arbitrary order of autoregressive model in an attempt to prewhiten the model residuals (the order of AR model was individually specified for each voxel to minimize the Schwarz Bayesian criterion). This suggests that the noise structure in fMRI time series may be more complex than reasonable choices of AR model order may be able to deal with robustly.

Another benefit of WLS is that it provides a map of the Hurst exponent estimated at each voxel. This confirms that some voxels have highly persistent autocorrelational structure, indicated by $\hat{H} \rightarrow 1$, and preliminary data suggest that the anatomical distribution of such long memory noise is not random. In the image represented in Figure 6, for example, long memory noise is concentrated symmetrically in areas of the frontal cortex and in the medial posterior parietal cortex; voxels representing central white matter are characterized by lower values of \hat{H} . Further experimental studies to clarify the source(s) of long memory noise in fMRI are possible on this basis, using WLS estimates of \hat{H} to map changes in fMRI noise properties related to systematic manipulation of relevant instrumental or physiological variables.

4.1 Nonstationary noise models and semiparametric regression

Two further extensions of linear modeling in the wavelet domain have been described in relation to fMRI data analysis. Fadili *et al.*²⁹ specified a *nonstationary fractal noise model* in which the Hurst exponent was allowed to vary in time. They found that voxels representing cortical regions of the brain were particularly likely to demonstrate nonstationarity of \hat{H} . Fadili and Bullmore³⁰ also noted the possibility of estimating partial linear or semiparametric models in the wavelet domain. Such models can be written

$$y = X\beta + g(t) + \epsilon$$

where $g(t)$ denotes a nonparametric function of time. Meyer⁴⁸ describes the theory and application of semiparametric models to fMRI data in greater detail.

4.2 Nonparametric regression

It is also possible to take an entirely nonparametric approach to signal detection or denoising in fMRI data analysis by adopting well-known techniques of *wavelet shrinkage*.⁸⁰ If we write a model of the data as the sum of an unknown function of time $g(t)$ plus error

$$y = g(t) + \epsilon$$

then we can optimally recover an estimate of the denoised signal $\hat{g}(t)$, under certain broad assumptions about the form of the signal by computing the DWT of the data; thresholding or shrinking the wavelet coefficients in each level according to some rule; and computing the inverse wavelet transform of the ‘shrunk’ coefficients. Vidakovic⁶ and Mallat¹ provide more detail about optimality of wavelet shrinkage. It can be proven that soft or hard thresholding with the universal threshold approximates by a log factor the minimax risk for estimation of a large class of functions, for example, Besov space, in Gaussian noise.

Within this general scheme, several choices of threshold λ and thresholding rule are available. If we are prepared to assume ϵ is i.i.d. Gaussian there is an argument for using the universal threshold proposed by Donoho and Johnstone⁸⁰

$$\lambda_U = \sigma\sqrt{2 \log(n)}$$

where n is the number of data points in the time series and σ is the standard deviation of the noise, which can be estimated by scaling the median absolute deviation (MAD) of the wavelet coefficients at the finest scale of the decomposition, $\hat{\sigma} = \text{MAD}/0.6745$. Hard thresholding sets to zero any wavelet coefficient which has absolute value less than the threshold and leaves untouched coefficients that exceed threshold. A widely used soft thresholding rule is defined as

$$T(d, \lambda) = \begin{cases} d - \lambda & \text{if } d \geq \lambda \\ d + \lambda & \text{if } d \leq -\lambda \\ 0 & \text{if } |d| \leq \lambda \end{cases}$$

where d is a wavelet coefficient (scale and location subscripts omitted for clarity); this assumes that all coefficients are contaminated by noise. Hard thresholding achieves a low bias–high variance solution to the shrinkage problem; soft thresholding is relatively a high bias–low variance solution. Numerous other thresholding rules and estimators of λ have been proposed in the general statistical literature but have not yet been explored systematically in fMRI. Examples of hard and soft thresholding applied to a representative fMRI time series are illustrated in Figure 4.

Alexander *et al.*⁴⁴ described the application of wavelet shrinkage to complex- and real-valued fMRI data and compare hard and soft thresholding rules to wavelet domain Wiener filtering. They advocated estimating the variance of the noise specifically for each level of the decomposition, rather than estimating σ from the finest scale wavelet

coefficients and assuming that it applies universally to the variance of the noise at all scales of the decomposition. Level-specific thresholds are appropriate to shrinkage of the more general class of data in which errors are correlated. LaConte, Ngan and colleagues^{81,82} described a time-varying filter based on the stationary or translation invariant wavelet transform and applied it to denoising of event-related fMRI time series. The threshold for shrinkage of wavelet coefficients was estimated by a ‘leave-one-out’ cross-validation procedure and the method was shown to be robust to reasonable choices of mother wavelet. Von Tscharner and Thulborn⁸³ used a wavelet tuned to the frequency of periodic alternation of the experimental input function to optimize time-frequency analysis of fMRI data acquired using a blocked periodic activation paradigm.

Although wavelet shrinkage of statistic maps overlaps substantively with the work on multiple hypothesis testing described below, it is perhaps surprising that wavelet shrinkage has not yet been explored extensively for two- or three-dimensional denoising of fMRI statistic maps, prior to statistical testing in the spatial domain, although Gaussian smoothing is very widely applied for this purpose. Gaussian smoothing customarily entails application of single smoothing kernel (though see Worsley *et al.*⁸⁴ for a multiresolutional approach to PET data in Gaussian scale-space) with the obvious risk of loss of power to detect spatial features incommensurate with smoothing kernel size. Smoothing by wavelet shrinkage has the relative merit of locally adaptive bandwidth so that the power to detect spatial feature of varying extent is not constrained by the arbitrary choice of a single kernel size.

5 Multiple hypothesis testing

A single fMRI dataset typically comprises $\sim 10^4$ voxels representing brain tissue. A common, massively univariate approach to analysis of these data is to estimate a test statistic by time series regression modeling at each voxel. This will result in a map of $\sim 10^4$ test statistics, each of which is to be tested under the null hypothesis of no experimental effect. This is evidently a multiple comparisons problem, which is complicated by the likelihood of some degree of spatial smoothness in the statistic maps under the null hypothesis, perhaps sufficient to render the Bonferroni correction overconservative (though see Ref. 66).

Ruttimann *et al.*⁴⁵ exploited the multiresolutional and decorrelating properties of the wavelet transform of spatial statistic maps in an attempt to mitigate the multiple comparisons problem. They assumed that the noise in the image was i.i.d. Gaussian $\epsilon \sim N(0, \sigma^2)$ and, therefore, that the wavelet coefficients were also i.i.d. Gaussian under the null hypothesis $d \sim N(0, \sigma^2)$. (The DWT of Gaussian white noise is Gaussian white noise.) They then proposed a hierarchical approach to hypothesis testing in which each level (and orientation) of the two-dimensional wavelet transform was subject to an ‘omnibus’ test of significance by comparing the sum of squared, standardized wavelet

coefficients it comprised against critical values of an appropriate chi-square distribution. In other words they assumed that under the null hypothesis

$$\sum_{k=1}^{K^j} \left(\frac{d_{j,k}}{\sigma} \right)^2 \sim \chi^2 K^j$$

where $K^j = n/2^j$ is the number of coefficients in the j th level of the decomposition. Levels of the decomposition for which the null hypothesis could not be refuted by this test were not examined further. Whereas each standardized coefficient in levels that passed this omnibus test for significance was then individually tested against the standard Normal distribution. The independence both between levels and between coefficients within levels that is implied by the assumption of i.i.d. Gaussian noise in the data encouraged control of the family-wise error by Bonferroni correction for both stages of the hypothesis testing algorithm. An activation map was finally constructed by the inverse wavelet transform using only those coefficients that had survived both tests for significance. This approach has the merit of reducing overall the number of tests to be conducted, but the validity of the χ^2 and Normal approximations, and of the Bonferroni correction, all depend on the assumption that the errors in the imaging data have an independent Normal distribution, which currently seems unlikely to be realistic in general.

Hilton *et al.*⁸⁵ and Brammer⁴⁶ all explored a related approach whereby the coefficients within each level of the wavelet decomposition were recursively tested against the null hypothesis that they resembled a white noise process. For example, Brammer⁴⁶ estimated the ‘Brownian bridge process’

$$B^Z\left(\frac{i}{K}\right) = \frac{1}{\sigma\sqrt{2K^j}} \sum_{k=1}^{i \leq K^j} (d_{j,k}^2 - \hat{d}_{j,k}^2)$$

and tested the maximum value of this cumulative sum process against critical values of the Kolmogorov–Smirnov statistic. The coefficients corresponding to maxima of $B^Z(i/K)$ were repeatedly removed and replaced by linear interpolation until the maximum nowhere exceeded its expected value under the null hypothesis. This process can be regarded as one way of iteratively identifying wavelet coefficients that depart, both in terms of their absolute magnitude and their relation to neighbouring coefficients, from the white Gaussian behaviour predicted under the null hypothesis that the errors in the spatial map are i.i.d. Gaussian. An activation map is constructed by taking the inverse wavelet transform using only those coefficients that were identified as significant (see Figure 7).

Müller *et al.*⁴⁹ extended Ruttimann’s method to hierarchical hypothesis testing on spatial maps of an arbitrary linear model parameter and applied this to analysis of event-related fMRI data. They demonstrated superior SNR compared to monoresolutional Gaussian smoothing and robustness of the method to type of wavelet. In agreement with studies of simulated data⁴⁷ they reported that lower-order wavelets achieved better performance.

More recent approaches to multiple hypotheses testing on maps or images in the wavelet domain have included efforts to define test statistics that are informed by the spatial relations between large coefficients within and between levels of the decomposition; methods for control of type 1 error in terms of the false discovery rate;⁸⁶ and incorporation of prior distributions for wavelet coefficients under the alternative hypothesis in Bayesian rules.⁶

In the work by Fadili and Bullmore,⁸⁷ we applied and compared methods for controlling the false discovery rate (FDR), and for Bayesian decision making, in the wavelet domain. The FDR of empirical wavelet coefficients can be defined as the expected false positive fraction $E(FPF)$, that is, the proportion of all positive tests that are falsely positive. Following Abramovich and colleagues,⁸⁸⁻⁹⁰ we aimed to retain the maximum number of observed wavelet coefficients subject to the constraint $E(FPF) < \alpha$, using the following algorithm to calculate a global threshold for elimination/retention:

- 1) For each of the n wavelet coefficients $d_{j,k}$ inside the brain at scale j , location k and each orientation, calculate the corresponding double-sided P -value, $p_{j,k}$ under H_0 :

$$p_{j,k} = 2 \left(1 - \Phi \left(\frac{|d_{j,k}|}{\hat{\sigma}} \right) \right)$$

where Φ is the cumulative standard normal distribution and $\hat{\sigma}$ is the robust MAD estimate of the standard deviation of the noise.

- 2) Sort the $p_{j,k}$ in an ascending order, $p_1 \leq p_2 \leq p_3 \cdots \leq p_n$.
- 3) Find the last index such that, $i_{FDR} = \max(i/p_i \leq (i/n)\alpha)$.
- 4) For this index, calculate the critical threshold corresponding to this double-sided p -value:

$$\lambda_{FDR} = \hat{\sigma} \Phi^{-1} \left(1 - \frac{p_{i_{FDR}}}{2} \right)$$

- 5) Use λ_{FDR} and apply classical hard thresholding or soft thresholding rules.
- 6) Apply the inverse DWT to obtain the thresholded image in the spatial domain.

However, this approach is based on a global threshold. To overcome this limitation, Ogden *et al.*⁹¹ developed a recursive hypothesis testing procedure that produces level-dependent thresholds. Rather than seeking to include as many wavelet coefficients as possible (subject to condition), the algorithm includes a wavelet coefficient only when there is strong evidence that it is needed in the reconstruction. A soft thresholding rule is applied with the level-dependent thresholds; see Ref. 87 for details of the algorithm.

A Bayesian method for wavelet thresholding was considered by Vidakovic.⁶ The Bayesian framework here imposes a prior which describes the variability of the wavelet coefficients of the true unknown image. This requires a prior distribution that has a point mass component at zero. Otherwise, the testing is impossible because any continuous prior density will give the prior (and hence the posterior) probability of zero to the precise hypothesis. In Ref. 87 we used a mixture of a central Gaussian and a point mass at zero. For this mixture model, Abramovitch *et al.*⁹⁰ proposed a closed-form expression for the ratio test (RT) statistic as the Bayes thresholding rule. They also proposed an EM algorithm to estimate the hyperparameters, namely the mixing

proportion and the level-dependent variance of the Gaussian pdf in the mixture model. In Fadili and Bullmore,⁸⁷ we derived an expression of the RT statistic under the null hypothesis. Only coefficients whose RTs exceed the α -level critical threshold are retained in the reconstruction by hard thresholding.

Global FDR, recursive and Bayesian approaches to multiple hypothesis testing are compared for an illustrative example of fMRI data in Figure 8. The activation task was simply repeated finger-thumb opposition with the right hand. Because of its soft thresholding rule, the recursive hypothesis testing procedure gives a smoother estimate than the FDR-based method. However, these two methods seem to have comparable power. The Bayesian approach, on the other hand, reveals a somewhat fuller characterization of the cerebral response in frontal areas important for movement control and planning.

6 Conclusions

These are early days in the extension of wavelet methods to the particular challenges of fMRI data analysis and much remains to be tried and tested. The noise properties of fMRI data both in time and space need to be further clarified in relation to the theoretical constraints on decorrelating and other properties of the DWT. There is a rich variety of wavelet methods for image and signal processing in general, which has yet to be fully evaluated in relation to fMRI. There are several areas of potential application – such as multiscale analysis of brain connectivity in the wavelet domain – that are almost entirely unexplored at present. In this article, we hope simply to have motivated the use of wavelets as an appropriate basis for analysis of nonstationary brain imaging data and to have highlighted a few methodological issues – resampling, model estimation and hypothesis testing – that can benefit specifically from the DWT as an optimal whitening filter of fractal processes.

Acknowledgements

This Human Brain Project/Neuroinformatics research is funded by the National Institute of Biomedical Imaging and Bioengineering and the National Institute of Mental Health. This work was also supported by the Wellcome Trust. The Wolfson Brain Imaging Centre is supported by an MRC Co-operative Group grant.

References

- 1 Mallat S. *A wavelet tour of signal processing*. San Diego: Academic Press, 1998.
- 2 Daubechies I. *Ten lectures on wavelets*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 1992.
- 3 Jaffard S, Meyer Y, Ryan RD. *Wavelets: Tools for science and technology*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2001.
- 4 Wornell GW. Wavelet-based representations for the $1/f$ family of fractal processes. *Proceedings IEEE* 1993; **81**: 1428–50.
- 5 Wornell GW. *Signal processing with fractals: A wavelet-based approach*. Upper Saddle River, NJ: Prentice Hall, 1996.
- 6 Vidakovic B. *Statistical modeling by wavelets*. New York: John Wiley & Sons, 1999.

- 7 Baraniuk RG, Flandrin P, Janssen AJEM, Michel OJJ. Measuring time-frequency information content using the Renyi entropies. *IEEE Transactions on Information Theory* 2002; **47**: 1391–1409.
- 8 Mandelbrot BB. *The fractal geometry of nature*. New York: WH Freeman, 1977.
- 9 Staszewski WJ, Worden K. Wavelet analysis of time series: coherent structures, chaos and noise. *International Journal of Bifurcation and Chaos* 1999; **9**: 455–71.
- 10 Schroeder M. *Fractals, chaos, power laws: Minutes from an infinite paradise*. New York: WH Freeman, 1991.
- 11 Abry P, Baraniuk R, Flandrin P, Riedi R, Veitch D. Multiscale nature of network traffic. *IEEE Signal Processing Magazine* 2002; **19**: 28–46.
- 12 Mandelbrot BB. *Fractals and scaling in finance: Discontinuity, concentration, risk*. New York: Springer, 1997.
- 13 Peng CK, Buldryev SV, Goldberger AL *et al.* Statistical properties of DNA sequences. *Physica A* 1995; **221**: 180–92.
- 14 Brown JH, Gupta VK, Li BL, Milne BT, Restrepo C, West GB. The fractal nature of nature: power laws, ecological complexity and biodiversity. *Philosophical Transactions of the Royal Society (B)* 2002; **357**: 619–26.
- 15 Ivanov PC, Rosenblum MG, Peng CK *et al.* Scaling behaviour of heartbeat intervals obtained by wavelet-based time-series analysis. *Nature* 1996; **383**: 323–27.
- 16 West GB, Brown JH, Enquist BJ. The fourth dimension of life: fractal geometry and allometric scaling of organisms. *Science* 1999; **284**: 1677–79.
- 17 Caserta F, Eldred WD, Fernandez E *et al.* Determination of fractal dimension of physiologically characterised neurons in 2-dimensions and 3-dimensions. *Journal of Neuroscience Methods* 1995; **56**: 133–44.
- 18 Bullmore ET, Brammer MJ, Bourlon P *et al.* Fractal analysis of electroencephalographic signals intracerebrally recorded during 35 epileptic seizures: Evaluation of a new method for synoptic visualization of ictal events. *Electroencephalography & Clinical Neurophysiology* 1994; **91**: 337–45.
- 19 Linkenkaer-Hansen K, Nikouline VV, Palva JM, Ilmoniemi. Long-range temporal correlations and scaling behavior in human brain oscillations. *Journal of Neuroscience* 2001; **21**: 1370–77.
- 20 Senhadji, Dillenseger JL, Wendling F, Rocha C, Kinie A. Wavelet analysis of EEG for three-dimensional mapping of epileptic events. *Annals of Biomedical Engineering* 1995; **32**: 738–48.
- 21 Bullmore ET, Brammer MJ, Harvey I, Persaud R, Murray RM, Ron MA. Fractal analysis of the boundary between white matter and cerebral cortex in magnetic resonance images: a controlled study of schizophrenic and manic-depressive patients. *Psychological Medicine* 1994; **24**: 771–81.
- 22 Blanton RE, Levitt JG, Thompson PM *et al.* Mapping cortical asymmetry and complexity patterns in normal children. *Psychiatry Research Neuroimaging* 2001; **107**: 29–43.
- 23 Free SL, Sisodiya SM, Cook MJ, Fish DR, Shorvon SD. Three-dimensional fractal analysis of the white matter surface from magnetic resonance images of the human brain. *Cerebral Cortex* 1996; **6**: 830–36.
- 24 Thompson PM, Schwartz C, Lin RT, Khan AA, Toga AW. Three-dimensional statistical analysis of sulcal variability in the human brain. *Journal of Neuroscience* 1996; **16**: 4261–74.
- 25 Thompson PM, Moussai J, Zohoori S *et al.* Cortical variability and asymmetry in normal aging and Alzheimer's disease. *Cerebral Cortex* 1998; **8**: 492–509.
- 26 Kuikka J, Tiihonen J. Fractal analysis – a new approach in brain receptor imaging. *Annals of Medicine* 1998; **30**: 242–48.
- 27 Turner R. BOLD localization: the implications of vascular architecture. *NeuroImage* 2001; **13**: S1011.
- 28 Zarahn E, Aguirre GK, D'Esposito M. Empirical analyses of BOLD fMRI statistics I. Spatially unsmoothed data collected under null hypothesis conditions. *NeuroImage* 1997; **5**: 179–97.
- 29 Fadili MJ, Bullmore ET, Brett M. Wavelet methods for characterising mono- and poly-fractal noise structures in shortish time series: An application to functional MRI. *Proceedings IEEE International Conference on Image Processing* 2001; 225–28.
- 30 Fadili MJ, Bullmore ET. Wavelet-generalised least squares: a new BLU estimator of linear regression models with 1/f errors. *NeuroImage* 2002; **15**: 217–32.
- 31 Guan S, Lai CH, Wei GW. A wavelet method for the characterization of spatiotemporal patterns. *Physica D* 2002; **163**: 49–79.

- 32 Nakao H, Mishiro T, Yamada M. Visualization of correlation cascade in spatiotemporal chaos using wavelets. *International Journal of Bifurcations & Chaos* 2001; 11: 1483–93.
- 33 Parlitz U, Meyer-Kress G. Predicting low-dimensional spatiotemporal dynamics using discrete wavelet transforms. *Physical Review E* 1995; 51: 2709–11.
- 34 Aldroubi A, Unser M, eds. *Wavelets in biology and medicine*. Boca Raton: CRC Press, 1996.
- 35 Laine A. Wavelets in temporal and spatial processing of biomedical images. *Annual Review of Biomedical Engineering* 2000; 2: 511–50.
- 36 Percival DB, Walden AT. *Wavelet methods for time series analysis*. Cambridge: Cambridge University Press, 2000.
- 37 Bruce A, Gao H-Y. *Applied wavelet analysis with S-PLUS*. New York, NY: Springer, 1996.
- 38 Weaver JB, Xu YS, Healy DM, Cromwell LD. Filtering noise from images with wavelet transforms. *Magnetic Resonance in Medicine* 1991; 21: 288–95.
- 39 Angelidis PA. MR image compression using a wavelet transform coding algorithm. *Magnetic Resonance Imaging* 1994; 12: 1111–20.
- 40 Maldjian JA, Liu WC, Hirschorn D, Murthy R, Semanczuk W. Wavelet transform-based image compression for transmission of MR data. *American Journal of Roentgenology* 1997; 169: 23–26.
- 41 Abu-Rezq AN, Tolba AS, Khuwaja GA, Foda SG. Best parameters selection for wavelet packet-based compression of magnetic resonance images. *Computing in Biomedical Research* 1999; 12: 173–84.
- 42 Iyriboz TA, Zukoski MJ, Hopper KD, Staggs PL. A comparison of wavelet and JPEG lossy compression methods applied to medical images. *Journal of Digital Imaging* 1999; 12: 14–17.
- 43 Zaroubi S, Goelman G. Complex denoising of MR data via wavelet analysis: application for functional MRI. *Magnetic Resonance Imaging* 2000; 18: 59–68.
- 44 Alexander ME, Baumgartner R, Windischberger C, Moser E, Somorjai RL. Wavelet domain de-noising of time-courses in MR image sequences. *Magnetic Resonance Imaging* 2000; 18: 1129–34.
- 45 Ruttimann U, Unser M, Rawlings R *et al*. Statistical analysis of functional MRI data in the wavelet domain. *IEEE Transactions on Medical Imaging* 1998; 17: 142–54.
- 46 Brammer M. Multidimensional wavelet analysis of functional magnetic resonance images. *Human Brain Mapping* 1998; 6: 378–82.
- 47 Desco M, Hernandez JA, Santos A, Brammer MJ. Multiresolution analysis in fMRI: sensitivity and specificity in the detection of brain activation. *Human Brain Mapping* 2001; 14: 16–27.
- 48 Meyer FG. Wavelet based estimation of a semi-parametric generalized linear model of fMRI time series. *IEEE Transactions on Medical Imaging* 2003; 22: 315–22.
- 49 Müller K, Lohmann G, Zysset S, von Cramon DY. Wavelet statistics of functional MRI data and the general linear model. *Journal of Magnetic Resonance Imaging* 2003; 17: 20–30.
- 50 Bullmore E, Long C, Suckling J *et al*. Colored noise and computational inference in neurophysiological fMRI time series analysis: Resampling methods in time and wavelet domains. *Human Brain Mapping* 2001; 12: 61–78.
- 51 Hossien-Zadeh GA, Soltanian-Zadeh H, Ardekani BA. Multi-resolution fMRI activation detection using translation invariant wavelet transform and statistical analysis based on resampling. *IEEE Transactions on Medical Imaging* 2003; 22: 795–805.
- 52 Breakspear M, Brammer MJ, Robinson PA. Construction of multivariate surrogate sets from nonlinear data using the wavelet transform. *Physica D* 2003; 182: 1–22.
- 53 Dinov ID, Mega MS, Thompson PM *et al*. Quantitative comparison and analysis of brain image registration using frequency-adaptive wavelet shrinkage. *IEEE Transactions on Information Technology in Biomedicine* 2002; 6: 73–85.
- 54 Kybic J, Thevenaz P, Nirkko A, Unser M. Unwarping of unidirectionally distorted EPI images. *IEEE Transactions on Medical Imaging* 2000; 19: 97–119.
- 55 Turkheimer FE, Brett M, Aston JA *et al*. Statistical modeling of positron emission tomography images in wavelet space. *Journal of Cerebral Blood Flow & Metabolism* 2000; 20: 1610–1618.
- 56 Turkheimer FE, Aston JAD, Banati RB, Riddell C, Cunningham VJ. A linear wavelet filter for parametric imaging with dynamic PET. *IEEE Transactions on Medical Imaging* 2003; 22: 289–301.

- 57 Cselenyi Z, Olsson H, Farde L, Gulyas B. Wavelet-aided parametric mapping of cerebral dopamine D2 receptors using the high affinity PET radioligand [^{11}C]FLB 457. *NeuroImage* 2002; 17: 47–60.
- 58 Raz J, Dickerson L, Turetsky B. A wavelet packet model of evoked potentials. *Brain and Language* 1999; 66: 61–88.
- 59 Barra V, Boire JY. Tissue segmentation on MR images of the brain by possibilistic clustering on a 3D wavelet representation. *Journal of Magnetic Resonance Imaging* 2000; 11: 267–78.
- 60 Flandrin P. Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Transactions on Information Theory* 1992; 38: 910–17.
- 61 Tewfik AH, Kim M. Correlation structure of the discrete wavelet coefficients of fractional Brownian motion. *IEEE Transactions on Information Theory* 1992; 38: 904–909.
- 62 Dijkerman RW, Mazumdar RR. On the correlation structure of the wavelet coefficients of fractional Brownian motion. *IEEE Transactions on Information Theory* 1994; 40: 1609–1612.
- 63 Ninness B. Estimation of 1/f noise. *IEEE Transactions on Information Theory* 1998; 44: 32–46.
- 64 Bullmore E, Suckling J, Brammer MJ. In praise of tedious permutation. In: Moore M, ed. *Spatial statistics: Methodological aspects and some applications*. New York: Springer, 2001, pp 183–200.
- 65 Nichols TE, Holmes AP. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human Brain Mapping* 2002; 15: 1–25.
- 66 Bullmore E, Suckling J, Overmeyer S, Rabe-Hesketh S, Taylor E, Brammer MJ. Global, voxel and cluster tests, by theory and permutation, for a difference between two groups of structural MR images of the brain. *IEEE Transactions on Medical Imaging* 1999; 18: 32–42.
- 67 Welchew DE, Honey GD, Sharma T, Robbins TW, Bullmore ET. Multidimensional scaling of integrated neurocognitive function and schizophrenia as a disconnection disorder. *NeuroImage* 2002; 17: 1227–39.
- 68 Bullmore E, Brammer MJ, Williams SCR *et al*. Statistical methods for estimation and inference for functional MR image analysis. *Magnetic Resonance in Medicine* 1996; 35: 261–77.
- 69 Lindley DV, Novick MR. The role of exchangeability in inference. *Annals of Statistics* 1981; 9: 45–58.
- 70 Davison AC, Hinkley DV. *Bootstrap methods and their application*. Cambridge: Cambridge University Press, 1998.
- 71 Carlstein E, Do K-A, Hall P, Hesterberg T, Kunsch HR. Matched block bootstrap for dependent data. *Bernoulli* 1998; 4: 305–28.
- 72 Theiler J, Eubank S, Longtin A, Galdrikian B, Farmer J. Testing for nonlinearity: the method of surrogate data. *Physica D* 1992; 58: 77–94.
- 73 Cohen A, Daubechies I, Vial P. Wavelet bases on the interval and fast algorithms. *Journal of Applied & Computational Harmonic Analysis* 1993; 1: 54–81.
- 74 Locascio JL, Jennings PJ, Moore CI, Corkin S. Time series analysis in the time domain and resampling methods for studies of functional magnetic resonance imaging. *Human Brain Mapping* 1997; 5: 168–93.
- 75 Friston KJ, Josephs O, Zarahn E, Holmes AP, Poline J-B. To smooth or not to smooth? Bias and efficiency in fMRI time series analysis. *NeuroImage* 2000; 12: 196–208.
- 76 Purdon PL, Weisskoff RM. Effect of temporal autocorrelation due to physiological noise and stimulus paradigm on voxel-level false-positive rates in fMRI. *Human Brain Mapping* 1998; 6: 239–49.
- 77 Woolrich MW, Ripley BD, Brady M, Smith SM. Temporal autocorrelation in univariate linear modeling of fMRI data. *NeuroImage* 2001; 14: 1370–86.
- 78 Worsley KJ, Liao CH, Aston J *et al*. A general statistical analysis for fMRI data. *NeuroImage* 2002; 15: 1–15.
- 79 Marchini J, Smith S. On bias in the estimation of autocorrelations for fMRI voxel time series analysis. *NeuroImage* 2003; 18: 83–90.
- 80 Donoho DL, Johnstone IM. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association* 1995; 90: 1200–24.
- 81 LaConte SM, Ngan SC, Hu X. Wavelet transform based Wiener filtering of event-related fMRI data. *Magnetic Resonance in Medicine* 2000; 44: 746–57.
- 82 Ngan SC, LaConte SM, Xu H. Temporal filtering of event-related fMRI data using cross-validation. *NeuroImage* 2000; 11: 797–804.

- 83 Von Tscharner V, Thulborn KR. Specified-resolution wavelet analysis of activation patterns from BOLD contrast fMRI. *IEEE Transactions on Medical Imaging* 2001; **20**: 704–14.
- 84 Worsley KJ, Marrett S, Neelin P, Evans AC. Searching scale-space for activation in PET images. *Human Brain Mapping* 1996; **4**: 74–90.
- 85 Hilton M, Ogden T, Hattery D, Edenm G, Jawerth B. Wavelet processing of functional MRI data. In: Aldroubi A, Unser M, eds. *Wavelets in medicine and biology*. Boca Raton: CRC Press, 1996, pp 93–114.
- 86 Shen X, Huang H-C, Cressie N. Nonparametric hypothesis testing for a spatial signal. *Journal of the American Statistical Association* 2002; **97**: 1122–40.
- 87 Fadili MJ, Bullmore ET. Wavelet-based multiple hypothesis testing approaches for activation map estimation. *Proceedings SPIE: Wavelets X* 2003, in press.
- 88 Abramovich F, Benjamini Y. Thresholding of wavelet coefficients as multiple hypotheses testing procedure. In: Antoniadis A, Oppenheim G, eds. *Wavelets and statistics*. New York: Springer, 1995, pp 5–14.
- 89 Abramovich F, Benjamini Y. Adaptive thresholding of wavelet coefficients. *Computational & Statistical Data Analysis* 1996; **22**: 351–61.
- 90 Abramovich F, Sapatinas T. Bayesian approach to wavelet decomposition and shrinkage. In: Muller P, Vidakovic B, eds. *Bayesian inference in wavelet based models*. New York: Springer, 1996, pp 33–50.
- 91 Ogden RT, Parzen E. Change-point approach to data analytic wavelet thresholding. *Statistics & Computing* 1996; **6**: 93–99.