Wavelets and functional magnetic resonance imaging of the human brain

Ed Bullmore, Jalal M. Fadili, Voichita Maxim, Levent Sendur, B. Whitcher, John Suckling, Michael Brammer, Michael Breakspear

Wavelets and functional magnetic resonance imaging of the human brain

Ed Bullmore,[a],* Jalal Fadili,[b] Voichita Maxim,[a] Levent Şendur,[a] Brandon Whitcher,[c] John Suckling,[a] Michael Brammer,[d] and Michael Breakspear[e]

[a]Brain Mapping Unit and Wolfson Brain Imaging Centre, University of Cambridge, Addenbrooke's Hospital, Cambridge, UK
[b]GREYC CNRS UMR 6072, 14050 Caen Cedex, France
[c]Translational Medicine and Technology, GlaxoSmithKline, Greenford, UK
[d]Department of Biostatistics and Computing, Institute of Psychiatry (King's College), London, UK
[e]Brain Dynamics Centre (Westmead Hospital) and School of Physics, University of Sydney, Sydney, Australia.

The discrete wavelet transform (DWT) is widely used for multiresolution analysis and decorrelation or "whitening" of nonstationary time series and spatial processes. Wavelets are naturally appropriate for analysis of biological data, such as functional magnetic resonance images of the human brain, which often demonstrate scale invariant or fractal properties. We provide a brief formal introduction to key properties of the DWT and review the growing literature on its application to fMRI. We focus on three applications in particular: (i) wavelet coefficient resampling or "wavestrapping" of 1-D time series, 2- to 3-D spatial maps and 4-D spatiotemporal processes; (ii) wavelet-based estimators for signal and noise parameters of time series regression models assuming the errors are fractional Gaussian noise (fGn); and (iii) wavelet shrinkage in frequentist and Bayesian frameworks to support multiresolution hypothesis testing on spatially extended statistic maps. We conclude that the wavelet domain is a rich source of new concepts and techniques to enhance the power of statistical analysis of human fMRI data.

Keywords: Scaling; 1/f; Long memory; Permutation; Nonparametric; False discovery rate

Introduction

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 (Fig. 1). Over the last 15 years or so, wavelets have emerged as powerful new mathematical tools for the analysis of complex data sets.[1] 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 (see Fig. 1).

This is evidently a multiresolution 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

* Corresponding author. Department of Psychiatry, University of Cambridge, Brain Mapping Unit, Addenbrooke's Hospital, Cambridge CB2 2QQ, UK. Fax: +44 1223 336581.
E-mail address: etb23@cam.ac.uk (E. Bullmore).
URL: http://www-bmu.psychiatry.cam.ac.uk.

[1] History of wavelets: 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 Mallat (1998) and the construction by Ingrid Daubechies (1992) 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. (2001).
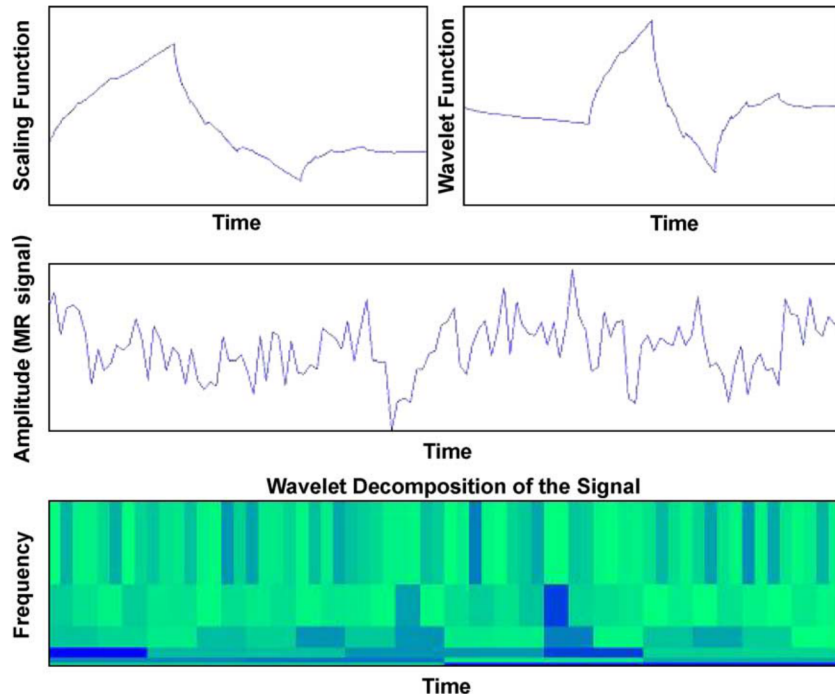
Fig. 1. Wavelets and time-scale decomposition of a representative fMRI time series. Top row: Daubechies scaling (father) and wavelet (mother) functions with regularity $R = 4$. Middle row: A "raw" functional MRI time series measured in the left superior temporal gyrus of a single individual. Bottom row: A time-scale plane tiling or scalogram illustrating the atomization or decomposition of the time series by the discrete wavelet transform using the Daubechies wavelets. The plane is covered by a set of rectangles of identical area, each of which is colored to indicate the sign and magnitude of the corresponding wavelet coefficient; large coefficients are coded dark blue or purple. Note that detail and approximation coefficients for the coarsest scale of the decomposition have the best resolution of scale (smallest height on y-axis) but the worst resolution of time (greatest width on x-axis) and vice versa for the detail coefficients for the finest scale of the decomposition.

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 Mallat (1998): "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 multiresolution nature and its adaptivity to nonstationary or local features in data—are sufficient to indicate that it will be of interest in the analysis of functional magnetic resonance imaging (fMRI) data, which we 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. Second, the wavelet transform has proven to be a useful basis for nonparametric regression, denoising or compression of large imaging data sets 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. Third, 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.

Wavelets, fractals, and the brain

In addition to these general technical advantages of statistical analysis in the wavelet domain, there is a related, more substantive argument favoring the use of wavelet methods specifically in analysis of brain imaging data that is based on the expectation that the brain may often demonstrate broadly fractal properties. The word fractal was originally coined by Mandelbrot (1977) to define a class of objects with the characteristic property of self-similarity (or self-affinity),[2] mean-

---

[2] Self-similarity: 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 $s^H B(t)$.
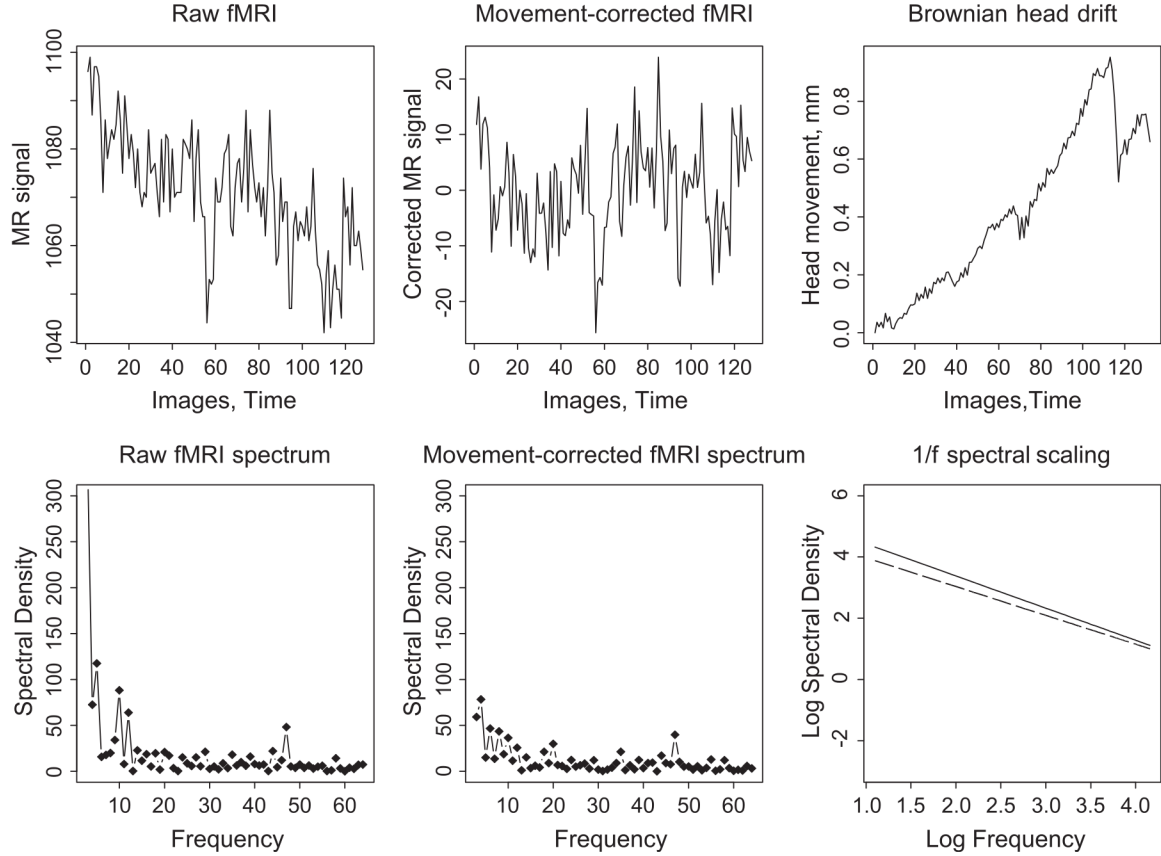
Fig. 2. Long memory or 1/f-like characteristics of fMRI time series. Top row: "Raw" fMRI time series (left) and the same time series after movement correction (middle) by regression on the time series of estimated head movements during scanning. A common mode of head movement is low-frequency drift illustrated by the time series of translations of the image center of gravity in the y-plane during scanning (right). Bottom row: Spectrum of raw time series (left) and spectrum of movement-corrected time series (middle). It is clear that both time series have disproportionate power at low frequencies and that correction for head movement relatively attenuates low-frequency power. The slope of a straight line fitted to a plot of log spectral density versus log frequency (right) is an estimator of the spectral exponent, γ, for raw time series (solid line; $\gamma = -1.05$, SE = 0.21) and movement-corrected time series (broken line; $\gamma = -0.94$, SE = 0.19).

ing 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. In other words, the structure of the process is approximately scale invariant or scale free.

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 which have $1 < D < 2$, like the human electrocardiogram (ECG) or a "raw" fMRI time series (Fig. 2), typically have long-range autocorrelations (long memory) in time and spectral density $S(f)$ related to a power law function of frequency:

$$S(|f|) \sim |f|^{\gamma} \text{ or } \log S(|f|) = c + \gamma \log |f|. \tag{1}$$

Fractal or scaling time series often have $\gamma \sim -1$ and are therefore sometimes also referred to as 1/f or 1/f-like processes.[3]

The fractal or scaling properties of the brain have a bearing on the suitability of wavelets for brain mapping because it has been recognized that wavelets are particularly apt for analysis and synthesis of fractal processes (Vidakovic, 1999; Wornell, 1993,

---

[3] Relations among γ, D, and H: 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 and fractional Gaussian noise (Schroeder, 1991; Staszewski and Worden, 1999):

$$D = T + 1 - H.$$

The Hurst exponent is in turn simply related to the spectral exponent. If $\gamma < -1$, $\gamma = 2H + 1$; if $\gamma > -1$, $\gamma = 2H - 1$.

1996). The key feature of wavelets that makes them a natural basis for analysis of self-similar or scale-invariant data is that each level of the (discrete) wavelet decomposition is a scaled (by factor 2) version of the next smallest scale: a family of wavelets is a fractal.

Three ways in which wavelets are technically attractive for statistical analysis of fractal processes are as follows: (i) wavelets effect a multiresolution decomposition that is advantageous for analysis of scale-invariant 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 (Fan, 2003) 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, the Hurst exponent and other measures of complexity (Baraniuk et al., 2002). We will return to these advantageous aspects of wavelet analysis in greater detail below; first, we briefly review some of the prior literature on fractal properties of natural and specifically neurobiological data.

Following Mandelbrot (1977), it has become increasingly clear that fractal, scale-invariant or scaling properties are shared by a wide variety of natural and social phenomena—ranging from Internet traffic (Abry et al., 2002) and econometric time series (Mandelbrot, 1997) through DNA base sequences (Peng et al., 1995) to collaborative and affiliative social networks (Strogatz, 2001) and ecosystems (Brown et al., 2002). The pathophysiology of the heart is arguably the human biological 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 fiber trees are self-similar and have fractional 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 improve diagnosis of coronary artery disease and prognosis of otherwise sudden arrhythmias (Ivanov et al., 1996). 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 (West et al., 1999).

There have been some comparable applications of fractal analysis to neurobiological data. Dendritic branching patterns of single neurons have been quantified in terms of fractal dimensions (Caserta et al., 1995). Fractal dimensions and $1/f$ spectral properties have been measured in electroencephalographic (EEG) signals (Bullmore et al., 1994a; Linkenkaer-Hansen et al., 2001; Senhadji et al., 1995). The fractal properties of anatomical surfaces and boundaries segmented in human MRI data have been measured (Blanton et al., 2001; Bullmore et al., 1994b; Free et al., 1996; Kiselev et al., 2003; Thompson et al., 1996). Fractal methods have been applied to analysis of radioligand SPET and PET images (Kuikka and Tiihonen, 1998), and imaging-orientated models for cerebral blood flow have been proposed on the basis of the probably fractal geometry of cerebrovascular architecture (Turner, 2001). There have also been some preliminary investigations of $1/f$ spectral properties in fMRI time series (Fadili and Bullmore, 2002; Fadili et al., 2001; Shimizu et al., 2004; Zarahn et al., 1997).

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 (Guan et al., 2002); to display the flow of information between scales in nonequilibrium fluid flows (Nakao et al., 2001); and to predict the behavior of spatially extended nonlinear dynamical equations (Parlitz and Meyer-Kress, 1995). 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.

In short, scale-invariant processes are abundant 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. There is a case for considering wavelets as *more than* "just another basis," one among many possible and equally plausible mathematical domains for statistical analysis of fMRI data.

Overview of wavelet-based methods for fMRI data analysis

Previous general reviews of wavelets in biomedical image processing, including some early work on fMRI, are provided by Aldroubi and Unser (1996), Laine (2000) and Bullmore et al. (2003). Statistical issues in wavelet analysis of time series are addressed comprehensively by Percival and Walden (2000). Wornell (1993, 1996) makes a detailed case for the general optimality of wavelets for analysis of fractal signals. Bruce and Gao (1996) describe the 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 (Abu-Rezq et al., 1999; Alexander et al., 2000; Angelidis, 1994; Iyriboz et al., 1999; Maldjian et al., 1997; Weaver et al., 1991; Wink and Roerdink, 2004; Wood and Johnson, 1999; Zaroubi and Goelman, 2000). Multiresolution hypothesis testing of spatial maps of fMRI time series statistics has been explored by Brammer (1998), Desco et al. (2001), and Ruttimann et al. (1998). Linear model estimation in the wavelet domain has been described by Fadili and Bullmore (2002), Meyer (2003) and Müller et al. (2003). Resampling of fMRI data in the wavelet domain has been developed by Breakspear et al. (2004), Bullmore et al. (2001a), Hossien-Zadeh et al. (2003) and Laird et al. (2004). Additionally, there have been applications of wavelets to the image processing problems of registering individual fMRI data sets in a standard anatomical space (Dinov et al., 2002) and correcting unidirectional geometric distortions in echoplanar imaging data (Kybic et al., 2000).

There have also been a number of interesting applications of wavelets to analysis of human brain mapping data in other modalities. Turkheimer et al. (2000, 2003) developed methods for multiresolution analysis and linear modeling (ANOVA, etc.) of multisubject positron emission tomography (PET) studies in the wavelet domain; Cselenyi et al. (2002) explored two and three-dimensional wavelet transforms as spatial filters of radioligand binding potential maps measured using PET; Raz et al. (1999) used wavelet packet analysis to decompose auditory electrophysiological potentials into component waveforms; and Barra and Boire (2000) reported a technique for brain tissue classification or segmentation of structural MRI based on fuzzy clustering of wavelet coefficients.
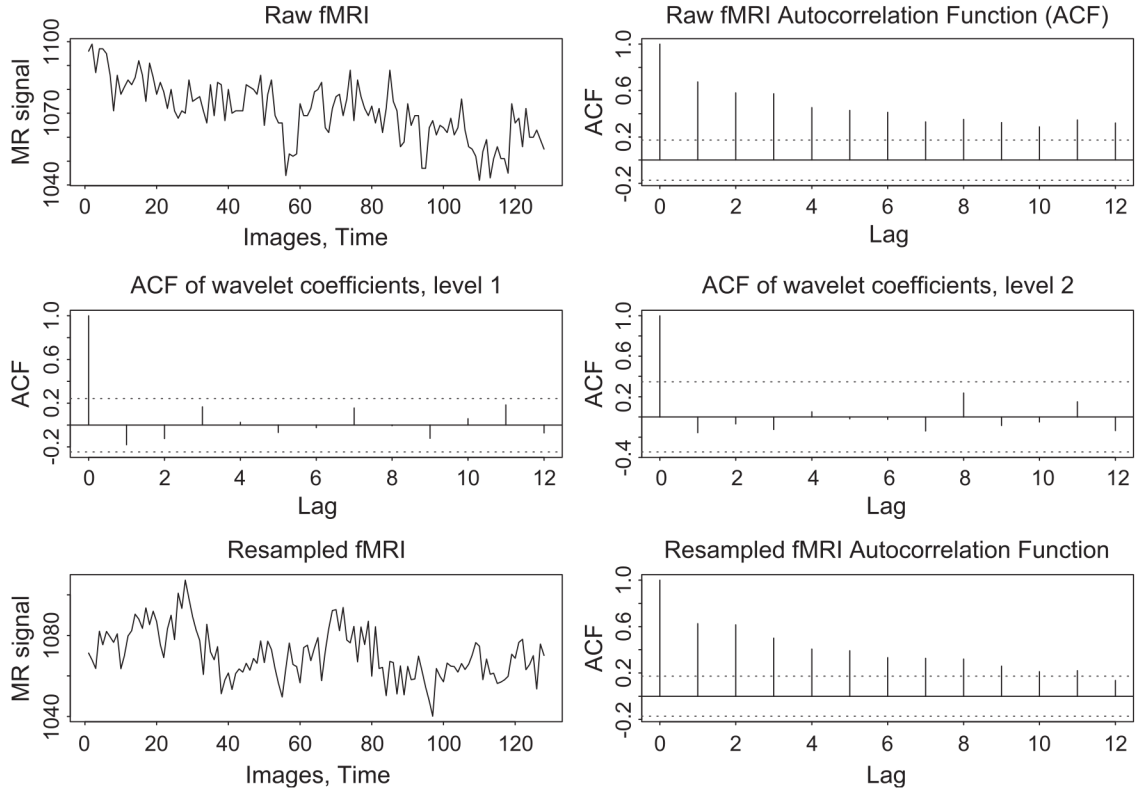
Fig. 3. Wavelet resampling in the time domain or "wavestrapping" of simulated $1/f$ noise. Top row: Raw fMRI time series (left) and its autocorrelation function (right) demonstrating long memory or persistent structure. Middle row: The wavelet coefficients of the fMRI series at the two finest scales of the decomposition, level 1 (left) and level 2 (right), are decorrelated or whitened, that is, all the autocorrelation coefficients fall within dashed lines indicating Bartlett's 95% confidence interval for zero, $0 \pm 2/\sqrt{n}$. Bottom row: The resampled time series (left) and its autocorrelation function (right) generated by taking the inverse wavelet transform after randomly reshuffling the order of decorrelated wavelet coefficients within each level of the decomposition. The key point is that although the original time series is strongly autocorrelated, its wavelet coefficients are exchangeable, and their random reordering therefore generates a resampled time series with an autocorrelation function very similar to the original data.

In what follows, we first provide a brief formal introduction to some key properties of the discrete wavelet transform (DWT)[4] and then discuss in more detail its application to three aspects of statistical analysis of fMRI data: (i) resampling of fMRI time series in time and space; (ii) time series modeling in the context of fractional Gaussian noise (fGn); and (iii) wavelet shrinkage or multiple hypothesis testing methods to control false discovery rate and to threshold Bayesian posterior probabilities.

The discrete wavelet transform (DWT)

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 dilating and translating a compactly supported "mother" wavelet ψ with zero integral over time $\int \psi(t)dt = 0$:

$$\Psi_{j,k}(t) = \frac{1}{\sqrt{2^j}} \Psi\left(\frac{t - 2^j k}{2^j}\right); \tag{2}$$

and by dilating and translating a "father" wavelet or scaling function ϕ with unit integral over time $\int \phi(t)dt = 1$:

$$\phi_{j,k}(t) = \frac{1}{\sqrt{2^j}} \phi\left(\frac{t - 2^j k}{2^j}\right); \tag{3}$$

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

Wavelets can be distinguished by their smoothness (also called regularity) and by their number of vanishing moments. The number of vanishing moments of a wavelet ψ is the largest integer R that satisfies $\int t^r \psi(t)dt = 0$, for all $r = 0, 1, \ldots, R - 1$.

Multiresolution analysis

The discrete wavelet transform of a spatial process or time series is a multiresolution analysis that distributes the total variance

---

[4] Existence of nonorthogonal wavelet transforms: Our focus has been on dyadic, orthonormal wavelets in the knowledge that related methods for time-invariant or undecimated wavelet transformation, which effect a redundant or nonorthogonal multiresolution decomposition, are also available and may have some advantages for aspects of time series analysis (for details, see Percival and Walden, 2000).

or energy of the data over a hierarchy of scales $\{S_j\}$. At each scale, the data are 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 perfectly recovered by adding the approximation at the coarsest scale and the details at this and all finer scales:

$$y = \sum_k a_{J,k}\phi_{J,k} + \sum_{j \leq J}\sum_k d_{j,k}\Psi_{j,k} \qquad (4)$$

Whitening properties of the discrete wavelet transform

The correlation between wavelet coefficients—both within and between scales of the decomposition—will generally be small even if the data are highly autocorrelated (Fig. 3). This is sometimes called the whitening or decorrelating property of the discrete wavelet transform and it was first defined theoretically for the class of self-affine signals known as fractional Brownian motion (Dijkerman and Mazumdar, 1994; Flandrin, 1992; Tewfik and Kim, 1992). Wornell (1993, 1996), and more recently Fan (2003), 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 $d_{j',k'}$

$$\langle d_{j,k}, d_{j',k'}\rangle \sim |\,2^j k - 2^{j'}k'\,|^{2(H-R)} \qquad (5)$$

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

$$\langle d_{j,k}, d_{j,k'}\rangle \sim |k - k'|^{2(H-R)} \qquad (6)$$

where R is the number of vanishing moments 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.

Wavelet estimators of fractal noise parameters

The Hurst exponent (and thereby the fractal dimension and spectral exponent)[3] 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 jth scale

$$\mathrm{Var}\{d_{.,j}\} = \frac{1}{N-1}\sum_k \left(d_{k,j} - \bar{d}_{.,j}\right)^2, \qquad (7)$$

is related to the scale by a power law with exponent $2H + 1$ (Flandrin, 1992)

$$\mathrm{Var}\{d_{.,j}\} \approx \left(2^j\right)^{2H+1}. \qquad (8)$$

From this expression, various estimators of H can be derived (Ninness, 1998; Wornell, 1993), of which the simplest is a least squares fit of the linear model (Flandrin, 1992)

$$\log_2\left(\mathrm{Var}\{d_{.,j}\}\right) = c + (2H + 1)j + \varepsilon_j. \qquad (9)$$

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 unrealistic assumptions about spatial auto-covariance and other distributional aspects of the data. Perhaps for these reasons, an appropriate nonparametric test can have superior sensitivity compared to a parametric alternative (Bullmore et al., 2001b; Hayasaka and Nichols, 2003; Nichols and Holmes, 2002). 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. In contrast, almost any statistic of interest may be accessible to inference on the basis of an appropriate resampling scheme (for examples, respectively, of resampling spatial and multivariate statistics in brain mapping, see Bullmore et al., 1999a; Welchew et al., 2002).

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

$$y = \mathbf{X}\beta + \varepsilon \qquad (10)$$

where y is the fMRI time series observed at a single voxel, $\mathbf{X}$ 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 residuals Σ will not generally be diagonal—there may be appreciable autocorrelation or serial dependency of the residuals in the time domain (Bullmore et al., 1996) (see Fig. 3).

The causes of residual or endogenous autocorrelation in fMRI time series are not yet certainly known: they probably include instrumental noise, head movement in the scanner, and aliased cardiac or respiratory cycle pulsation. In addition to these "nuisance" sources of colored noise, there may also be contributions from substantively more interesting neurophysiological processes. For example, Linkenkaer-Hansen et al. (2001) used detrended fluctuation analysis (a nonorthogonal multi-resolution technique) to demonstrate long-range autocorrelations in no-task or "resting" EEG and MEG data. Specifically, correlations ("dynamical memory") in electrical brain activity were observed over time-scales from milliseconds to tens of minutes, easily encompassing the time-scale of neuronal activity-related BOLD response.

The problem for resampling is that it would 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.[5] Various methods have been proposed to circumvent this problem (and are more completely reviewed by Davison and Hinkley, 1997). 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 correlation structure of the data may be preserved under resampling (Carlstein et al., 1998). 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

$$\varepsilon_t = \eta \varepsilon_{t-1} + \rho_t, \rho \sim N(0, \sigma^2 \mathbf{I}) \tag{11}$$

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

$$
\begin{aligned}
y_t^* &= y_t - \hat{\eta} y_{t-1} \\
\mathbf{X}_t^* &= \mathbf{X}_t - \hat{\eta} \mathbf{X}_{t-1}
\end{aligned}
\tag{12}
$$

then the residuals ε^* of the transformed model $y^* = \mathbf{X}^* \boldsymbol{\beta} + \varepsilon^*$ will be white, provided the AR(1) model is adequate to account for the autocorrelation in the observed time series under the null hypothesis (Bullmore et al., 1996). 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 (Laird et al., 2004; Theiler et al., 1992). 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 stochastic properties (Davison and Hinkley, 1997).

Wavestrapping in time

The discrete wavelet transform provides an alternative device for this strategy of resampling a time series after orthogonal transformation. In its simplest form, 1-D "wavestrapping" involves the following:

1. Computing the discrete wavelet transform of the observed series;

2. Randomly permuting the decorrelated detail and approximation coefficients at each scale of the decomposition; and
3. Computing the inverse wavelet transform to recover the resampled series in the time domain.

As illustrated in Fig. 3, this method can faithfully preserve the autocorrelation structure of a $1/f$-like time series because its wavelet coefficients will be whitened and therefore exchangeable in the wavelet domain. This method is more fully described by Bullmore et al. (2001a) for sampling the null distribution of the linear model parameter vector $\boldsymbol{\beta}$ by fitting the time series regression model to the fMRI data at each voxel after repeated permutation in the wavelet domain. Laird et al. (2004) compared this algorithm to a Fourier-based resampling scheme for activation mapping of fMRI data and concluded that wavelet resampling was superior, that is, associated with a greater area under the receiver operating characteristic (ROC) curve, indicating greater sensitivity to detect true signals.

One caveat is that fMRI time series will invariably be finite and are typically rather short ($n \sim 50$–500 time points). This potentially creates a problem for wavestrapping because estimation of wavelet coefficients at the boundaries of finite time series introduces artifactual correlations between coefficients. This is notably an issue with the widely used periodic assumption for boundary correction, although so-called interval wavelets may be less problematic in this respect (Cohen et al., 1993).

Breakspear et al. (2003) described two refinements to this resampling scheme, which more exactly preserved the phase and amplitude distributions 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 (see also Sabatini, 1999), and cyclic rotation of the wavelet coefficients within each level by a random shift. Breakspear et al. (2003) also demonstrated that cross correlations between time series in the same volume of data could be preserved under resampling simply by decomposing each of the time series in the wavelet domain and then applying exactly the same permutation to the wavelet coefficients of each time series. Thus, multiple one-dimensional resampling of fMRI time series volumes can preserve (approximately) both the autocorrelation and cross-correlation structure of the original data.

Wavestrapping in space and time

In addition to wavelet resampling of each voxel's one-dimensional time series (temporal wavestrapping), it is also possible to resample the data in the spatial domain (Breakspear et al., 2004). To do this, each two-dimensional slice of data is subject to a 2-D discrete wavelet transform. This yields, at each scale of the decomposition, three sets of detail coefficients, capturing the variance in the horizontal, diagonal, and vertical directions, respectively. A permutation scheme (random shuffle, block resampling, or cyclic rotation) is then applied to each of these sets. Bootstrapped data are retrieved in the spatial domain by the two-dimensional inverse wavelet transform of the permuted coefficients. The algorithm is illustrated by spatial resampling of a test image in Fig. 4, and the preservation of the original image's spatial spectrum under blocked 2-D wavelet resampling is demonstrated in Fig. 5 (Breakspear et al., 2004).

Because of the nonstationary or spatially localized nature of wavelets, this permutation scheme can be constrained so that only
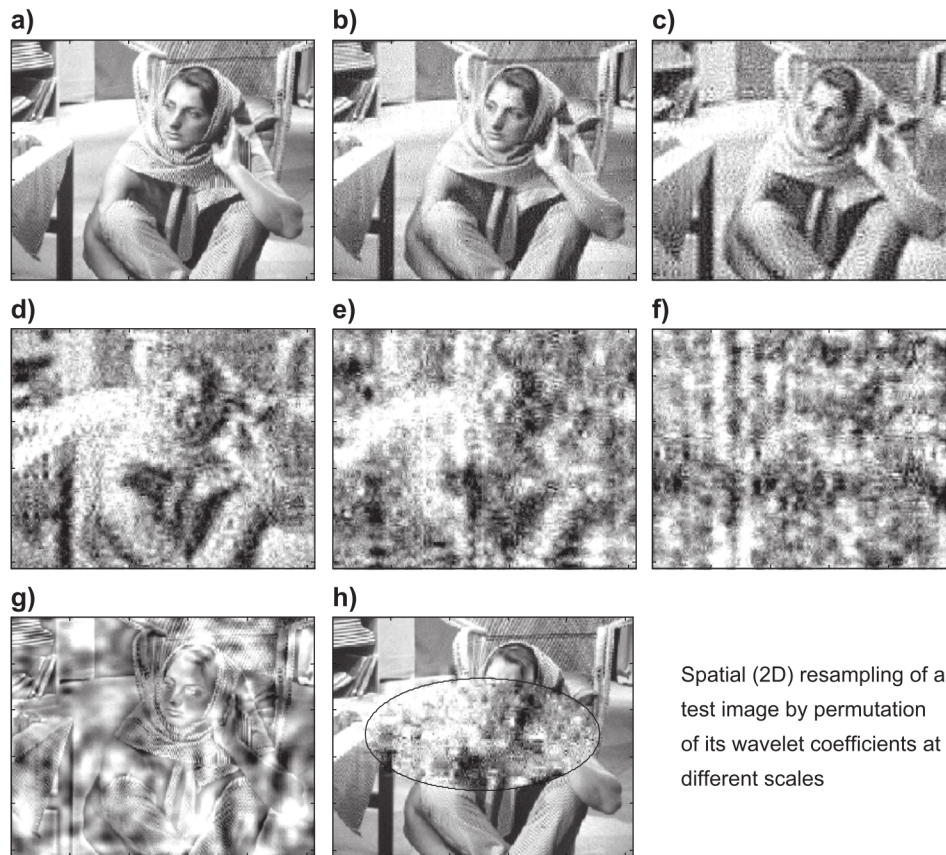

Fig. 4. Wavelet resampling in the spatial domain. The test image, shown in panel a, is a standard IEEE image and is chosen to demonstrate spatial wavelet resampling because certain features of the technique can be easily visualized in this context. In panel b, only the finest details of the wavelet decomposition have been resampled. In comparison to a, it is evident that the fine details of the scarf and the chair lattice have been moved. However, this structure has not been smoothed out of the picture but is evident diffusely through the background of the image. The number of scales on which the resampling operates has been increased to 2 in panel c, 3 (panel d), 4 (panel e), and all scales (panel f). Varying the scales at which the permutation acts makes evident the manner in which the structure at each scale contributes to the information in the image. For example, in panel e, only the coarsest shadows of the face remain in their original location. Structure at all smaller scales is now present diffusely throughout the image. In panel g, the structure represented by the three smallest scales has not been resampled. Hence, the detailed facial features and scarf pattern remain in their original spatial location. However, the shadows—which give the image its depth—are now present diffusely throughout the image. In this image, specific information at various scales conveys certain types of information (detail, depth, texture, etc.). In neuroimaging data, it is expected that fluctuations in neural activity at different scales are related to different types of neural interactions and information processing. Reprinted with permission from Breakspear et al. (2004).

those coefficients representing intracerebral tissue in an MR image are subject to permutation. Extracerebral coefficients are left unpermuted. In Fig. 4h, the effect of constraining the resampling of coefficients to an elliptical subdomain of the image is illustrated. It can be seen that the spatial correlations within this ellipse have been randomized, whereas those outside it have been unchanged. Such a spatially localized permutation scheme is unique to resampling in the wavelet domain. In the Fourier domain, the spatial localization of data is described by the phase of a complex two-dimensional Fourier transform and it is precisely the phases that must be randomized to produce resampled data. Hence, it is not possible to construct spatially localized 3- or 4-D bootstrap data using a Fourier-based technique.

The application of spatiotemporal wavestrapping to a full (4-D) fMRI data set requires a further important constraint: Resampling of each slice, at each time point, must be done as discussed above for multiple temporal resampling. That is, exactly the same permutation must be applied to the wavelet decomposition of each slice and at each time point in the fMRI experiment. In this way, cross correlations between slices can be (approximately) preserved.

A multivariate temporal resampling step can be added before or after spatial resampling. Such a process has been shown to produce reasonable type I error control in no-task or "resting" fMRI data sets while allowing identification of functional correlations in a visual stimulus experiment (Breakspear et al., 2004).

Why bother with spatiotemporal wavestrapping? An emerging body of research is currently being directed towards the study of functional connectivity—large scale network correlations—in neuroimaging data sets (Horwitz, 2003). A variety of novel connectivity measures have been developed for this purpose. Resampling of functional MRI data in the spatial and temporal domains represents a potentially powerful nonparametric approach to hypothesis testing for any such connectivity measure or experiment.

However, a crucial issue in the design and validation of wavestrapping schemes is the extent to which the wavelet coefficients are indeed decorrelated. For $1/f$-like time series, there is useful theory predicting decay of the correlation between coefficients within scale as an exponential function of the distance between them (Eqs. (5) and (6); Fan, 2003). There is as yet only
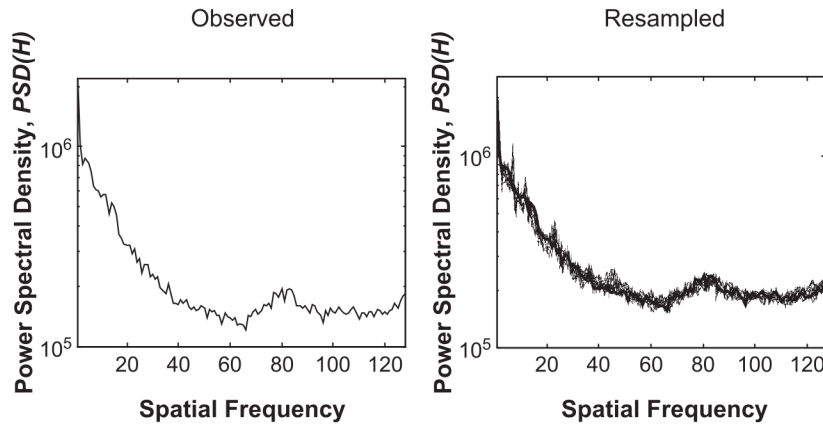
Fig. 5. Wavelet resampling in the spatial domain preserves the spatial spectrum of the original data. Left panel: Horizontal spatial spectrum of the test image in Fig. 4a. Right panel: An ensemble of 19 surrogate spectra generated by 2-D spatial blocked wavestrapping of the test image (dotted lines). It can be seen that the observed spectrum falls within the permutation distribution, indicating that spatial covariance of the image has been well preserved under resampling and that the key assumption of exchangeability of the resampled units (blocks of spatially adjacent wavelet coefficients) is reasonably well justified in this case. Violation of exchangeability would be associated with marked "whitening" of the surrogate spatial spectra.

preliminary evidence in support of the assumption that fMRI time series generally have $1/f$-like spectral densities (Fig. 2). This may not be such an important issue since approximate wavelet decorrelation has already been extended to a much wider class of time series models (Whitcher, 2001, 2004a), but the theoretical constraints on decay of correlation between higher dimensional (2- or 3-D) wavelet coefficients are not so well known. Initial theoretical work for 2-D fractional Brownian motion has shown that the autocorrelation functions for the detail coefficients decay at an exponential rate depending on the squared distance between coefficients (Wu and Su, 1998).

Because of the complicated nature of noise structure in the spatial domain of fMRI data, Breakspear et al. (2004) suggest an empirical trial and error approach toward finding a suitable spatial wavestrapping scheme. Whitcher (2004b, in review) has recently linked wavelet resampling of spatial patterns to a general class of spatial processes, specifically, the Matérn class of spectral density functions (Matérn, 1986). The rigid filtering sequence of the discrete wavelet transform was also generalized to the more flexible discrete wavelet packet transform (Mallat, 1998). The discrete wavelet packet transform provides a more adaptive partition (tessellation) of the time-frequency plane when compared to the standard discrete wavelet transform. This allows the wavelet basis set to adapt more easily to unknown, and potentially complex, spatial autocorrelation. In order to select an adequate basis set for spatial wavestrapping, a recursive hypothesis testing procedure is performed—from coarse to fine scales—that produces a valid orthonormal basis for the two-dimensional discrete wavelet packet transform. More work is required to match the spatial structure of fMRI data sets to meaningful neural models and stochastic processes in order to provide a more rigorous theoretical underpinning to spatial wavestrapping.

Time series modeling in the wavelet domain

The existence of serially dependent 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 (Eq. (11)); 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 (Bullmore et al., 1996; Friston et al., 2000; Locascio et al., 1997; Marchini and Smith, 2003; Purdon and Weisskoff, 1998;
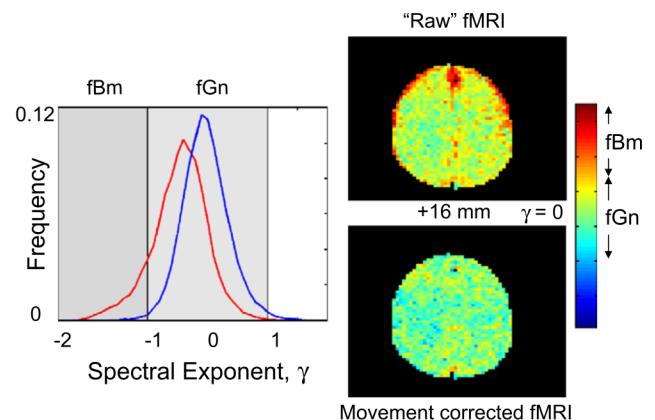

Fig. 6. Spectral exponents of functional MRI data acquired under no-task, "resting" conditions. Left panel: Frequency distributions of spectral exponent γ estimated in a slice of "raw" fMRI data (red line) and movement-corrected fMRI data (blue line). In the raw data, the distribution extends to large negative values of γ that are characteristic of fractional Brownian motion (fBm); the effect of movement correction is to shift the distribution to the left so that it is approximately centered on zero and falls largely within the typical range for fractional Gaussian noise (fGn). Right panel: The spatial distribution of estimated spectral exponents is shown for a single axial slice of data ($z = +16$ mm in Talairach space). The largest negative values—indicating the most persistent, longest memory noise—occur on the edge of frontal cortex and in medial posterior parietal cortex.

Woolrich et al., 2001; Worsley et al., 2002) but are susceptible to failure, with consequent loss of type I 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. Alternative approaches are possible in the wavelet domain.

Colors of fMRI noise

The power spectral exponent γ of a time series (see Eq. (1)) can be used to locate it on a continuum of colored noises. If $\gamma = 0$, all frequencies are uniformly represented in the spectrum and the noise is said to be white. When $\gamma < 0$, the time series will have positively autocorrelated, long memory or persistent structure; when $\gamma > 0$, the time series will have negatively autocorrelated or antipersistent structure. If $\gamma = -1$, low frequencies predominate in the spectrum and the noise may be called $1/f$ or pink; if $\gamma = -2$, low frequencies are yet more predominant in the spectrum and the noise may be called brown or Brownian (Voss, 1988). Two instances of long-memory noise are fractional Brownian motion (fBm; $-3 < \gamma < -1$) and its increment process, fractional Gaussian noise (fgN; $-1 \leq \gamma \leq 1$). Both fGn and fBm are Gaussian models and subsume classical Gaussian noise ($\gamma = 0$) and Brownian motion ($\gamma = -2$).

If we estimate γ at each voxel of an fMRI data set acquired under no-task or "resting" conditions, we can see that it is distributed mainly in the range corresponding to persistent fGn although there are an appreciable number of voxels with spectral exponents characteristic of fBm and some with spectral exponents characteristic of antipersistent fGn (Fig. 6). This distribution is, not surprisingly, conditioned by the extent to which the data have been corrected for effects of head movement during scanning. Of course the effects of head movement correction will differ considerably depending on the type and extent of movement that occurred during a particular scanning session. However, a very common mode of movement is a slow, almost linear drift as the subject's head relaxes into its most stable position (Fig. 2). If such a data set is uncorrected for movement, or corrected only by registration or realignment of each volume in the series with the first image or average image, then this will be reflected by persistent trends in the realigned time series. Secondary movement correction procedures, such as regression of the realigned time series on estimated voxel displacements over time (trigonometrically calculated for each voxel based on its distance from the image center of gravity and the rotations and translations over time of the image center of gravity; Bullmore et al., 1999b), can attenuate low-frequency trends and other movement-related features. The distribution of γ shows a corresponding shift towards zero, with fewer voxels falling in the range of fBm, and the large majority falling in the range of fGn (Fig. 6). These preliminary results suggest that fGn will generally provide an adequate and parsimonious model for the wide range of persistent and antipersistent noises encountered in fMRI: further empirical work to test this claim is currently ongoing.

Fractional Gaussian noise

Fractional Gaussian noise is more formally defined as a stationary process with zero mean and auto-covariance function

$$s_\tau = \frac{\sigma^2}{2}\left(|\tau + 1|^{2H} - 2|\tau|^{2H} + |\tau - 1|^{2H}\right), \quad (13)$$

where τ denotes the lag between time points and $s_\tau = s_{-\tau}$ for $\tau < 0$. It is evident that fGn is completely parameterized by its Hurst exponent H and its variance σ^2.

The spectral density function of fGn is (Beran, 1994; Percival and Walden, 2000):

$$S(f) = 4\sigma^2 C_H \sin^2(\pi f) \sum_{j=-\infty}^{j=\infty} \frac{1}{|f + j|^{2H+1}}, \quad -\frac{1}{2} \leq f \leq \frac{1}{2}. \quad (14)$$

When the frequency is close to zero, the spectral density function is approximately

$$S(f) \approx \frac{\sigma^2 C_H (2\pi)^2}{|f|^{2H-1}}, \quad (15)$$

with

$$C_H = \Gamma(2H + 1)\sin(\pi H)/(2\pi)^{2H+1}. \quad (16)$$

Wavelet estimators of time series model parameters in the context of fGn

Fadili and Bullmore (2002) specified the fMRI regression problem as

$$y = \mathbf{X}\beta + \varepsilon, \quad \varepsilon \sim N\big(0, \mathbf{\Sigma}\{H, \sigma^2\}\big) \quad (17)$$

that is, ε is a fractional Gaussian noise parameterized by H and σ^2. Taking the discrete wavelet transform of this model gives

$$y_w = \mathbf{X}_w\beta + \varepsilon_w, \quad \varepsilon_w \sim N\big(O, \sum{}_w\big), \quad (18)$$

where $\mathbf{X}_w$ is the result of applying the DWT separately to each column of the design matrix, and y_w and ε_w are the wavelet transforms of the data and the error process, respectively. As noted earlier, the orthonormal DWT is approximately a Karhunen–Loève
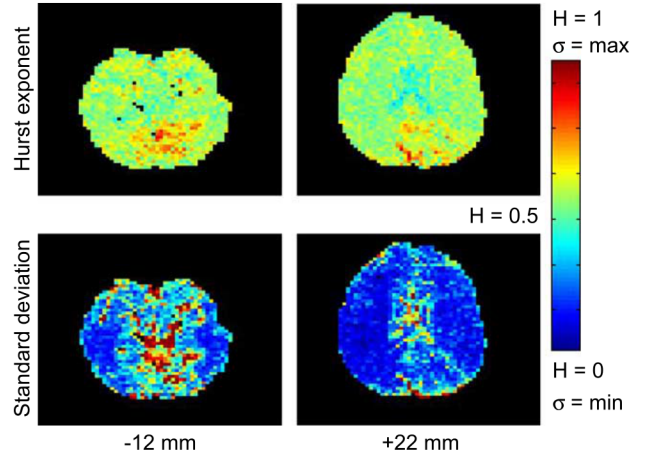

Fig. 7. Maps of the Hurst exponent ($0 < H < 1$) and error standard deviation σ estimated by wavelet-generalized least squares at each voxel of a single movement-corrected image acquired under no-task, "resting" conditions. Top row: Maps of the Hurst exponent in two slices (left and right located at $z = -12$ and $+22$ mm in Talairach space, respectively). Values of $H > 0.5$, indicative of persistent fractional Gaussian noise (fGn), are located in occipital and medial posterior parietal cortex; values of $H < 0.5$, indicative of antipersistent fGn, are located in the vicinity of the lateral ventricles. Bottom row: Maps of the noise standard deviation in the same two slices, indicating large (antipersistent) noise variance in the vicinity of CSF spaces.

expansion of $1/f$-like processes. To a good approximation, the variance–covariance matrix of fractional Gaussian 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; Fadili and Bullmore, 2002), which incorporated the approximate expression for spectral density of fGn (Eq. (15)). The WLS estimator was shown to be asymptotically the BLU estimator of regression model parameters β in the context of fGn; it also provided an ML estimate of the Hurst exponent and the variance σ^2 of the noise at each voxel (see Fig. 7). 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.

Nonstationary noise models and semiparametric regression

Two further extensions of time series modeling in the wavelet domain have been described in relation to fMRI data analysis. Fadili et al. (2001) 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 dynamic changes in fractal noise properties. Shimizu et al. (2004) described a comparable approach, using the continuous wavelet transform to estimate the multifractal spectrum (distribution of Hölder exponents) in fMRI time series: they showed that the distribution of Hölder exponents was wider, and its maximum value was greater, in "smoother" activated gray matter voxels than in unactivated white matter voxels.

Fadili and Bullmore (2002) also noted the possibility of estimating partial linear or semiparametric models in the wavelet domain. Such models can be written

$$y = \mathbf{X}\beta + g(t) + \varepsilon \qquad (19)$$

where $g(t)$ denotes a nonparametric function of time. Meyer (2003) describes the theory and application of semiparametric models to fMRI data in greater detail.

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 (Donoho and Johnstone, 1995). If we write a model of the data as the sum of an unknown function of time $g(t)$ plus error

$$y = g(t) + \varepsilon \qquad (20)$$

then we can optimally recover an estimate of the denoised signal $\hat{g}(t)$ under certain broad assumptions about the form of the signal.[6]

Wavelet shrinkage algorithms generically consist of three steps:

1. Calculate the wavelet transform of the noisy signal;

2. Modify the noisy wavelet coefficients according to some threshold value and thresholding rule (shrinkage step); and

3. Compute the inverse transform using the modified coefficients.

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. normal, then there is an argument for using the universal threshold proposed by Donoho and Johnstone (1995):

$$\lambda_U = \sigma\sqrt{2\log(n)} \qquad (21)$$

where n is the number of data points in the time series and σ is the standard deviation of the noise, robustly estimated by the median absolute deviation (MAD) of the wavelet coefficients at the finest scale of the decomposition: $\hat{\sigma} = \text{MAD}/0.6745$. However, subband adaptive algorithms, by which coefficients are compared to a level-dependent threshold, often have superior performance compared to rules that apply a universal threshold; an example is SureShrink (Donoho and Johnstone, 1995), which uses level-dependent thresholds to minimize Stein's unbiased risk estimate (SURE).

The most widely used thresholding rule is probably soft thresholding. The main idea of soft thresholding is to subtract the threshold value λ from all coefficients $\{d\}$ larger than λ and to set all other coefficients to zero. It can be defined as

$$\text{soft}(d, \lambda) = \text{sign}(d)(|d| - \lambda)_+ \qquad (22)$$

where $(s)_+$ is defined as

$$(s)_+ = \begin{cases} 0 & \text{if } s < 0 \\ s & \text{otherwise} \end{cases} \qquad (23)$$

Hard thresholding simply sets to zero any wavelet coefficient, which has absolute value less than the threshold and does not change coefficients which exceed the threshold.

Relatively speaking, hard thresholding is a low bias–high variance solution to the shrinkage problem; soft thresholding is a high bias–low variance solution. Numerous other thresholding rules and estimators of λ have been proposed in the signal processing literature and several have been explored in fMRI.

Alexander et al. (2000) described the application of wavelet shrinkage to complex- and real-valued fMRI data and compared 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 (Johnstone and Silverman, 1997). LaConte et al. (2000), Ngan et al. (2000) and colleagues 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 wavelet. Von Tscharner and Thulborn (2001) 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. Long et al. (2004) have shown that allowing the wavelet threshold rule to be spatially variant—and using the local temporal variance structure as the constraint on this rule—is associated with lower

---

[6] 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 (for details, see Mallat, 1998, and Vidakovic, 1999).

mean squared error and better localization in both simulated and experimental fMRI activation maps.

It is perhaps surprising that wavelet shrinkage has not yet been explored more extensively for 2- or 3-D denoising of fMRI statistic maps, prior to hypothesis testing in the spatial domain, although Gaussian smoothing is very widely applied for this purpose. Gaussian smoothing customarily entails application of a single smoothing kernel (for a multiresolution approach to PET data in Gaussian scale-space, see Worsley et al., 1996) 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 features of varying extent is not constrained by the arbitrary choice of a single kernel size. Wink and Roerdink (2004) compared Gaussian smoothing with wavelet denoising methods using several criteria to cross validate these alternatives: signal-to-noise ratio, the shape of activated regions, and quality of the resulting activation maps. They found that for medium to high levels of noise, the less-smoothing wavelet methods improved the signal-to-noise ratio, better preserved the shape of activated regions, and produced less overall errors (though more false negatives) than Gaussian smoothing. The relationship between Gaussian smoothing and wavelet shrinkage has also been explored theoretically and empirically by Fadili and Bullmore (2004).

Multiresolution hypothesis testing

Some work on wavelet-based denoising has taken a more probabilistic approach to defining threshold values and rules (Hyvarinen, 1999). It is easy to see that, by this slight change of emphasis towards a more statistical perspective, algorithms for wavelet shrinkage applied to fMRI can be reformulated as methods for massively univariate hypothesis testing and control of type I error in the context of multiple comparisons.

Ruttimann et al. (1998) exploited the multiresolution 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 $\varepsilon \sim N(0, \sigma^2)$, so the wavelet coefficients were also i.i.d. Gaussian[7] under the null hypothesis $d \sim N(0, \sigma^2)$, and therefore the sum of squared, standardized wavelet coefficients at each scale and orientation of the 2-D wavelet transform was distributed approximately as chi squared, that is,

$$\sum_{k=1}^{K^j} \left(\frac{d_{j,k}}{\sigma} \right)^2 \sim \chi^2 \tag{24}$$

on K^j d.f., where $K^j = n/2^j$ is the number of coefficients in the jth level of the decomposition. Levels of the decomposition for which the null hypothesis could not be refuted by this "omnibus" test were not examined further. Levels at which the null hypothesis was refuted were investigated further by individually testing each coefficient 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-motivated 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 innovative approach had 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 seems unlikely to be realistic in general.

Müller et al. (2003) 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 monoresolution Gaussian smoothing and robustness of the method to choice of wavelet. In agreement with studies of simulated data (Desco et al., 2001), they reported that lower order wavelets achieved better performance.

Hilton et al. (1996) and Brammer (1998) both 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, Brammer (1998) estimated the "Brownian bridge process"

$$B^Z(i/K) = \frac{1}{\sigma\sqrt{2K^j}} \sum_{k=1}^{k \leq K^j} \left(d_{j,k}^2 - \bar{d}_{.,k}^2 \right) \tag{25}$$

and tested the maximum value of this cumulative sum process against critical values of the Kolmogorov–Smirnoff 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 recursively or iteratively identifying wavelet coefficients that depart, both in terms of their absolute magnitude and their relation to neighboring coefficients, from the white Gaussian behavior predicted under the null hypothesis that the errors in the spatial map are i.i.d. Gaussian.

False discovery rate control in the wavelet domain

Some more recent statistical approaches to multiple hypothesis testing on maps or images in the wavelet domain have included efforts to control type I error in terms of the false discovery rate (FDR), to estimate and test statistics that are informed by the spatial relations between large coefficients within and between levels of the decomposition (Sendur and Selesnick, 2002a; Shen et al., 2002), and to incorporate prior distributions for wavelet coefficients under the alternative hypothesis in Bayesian rules (Vidakovic, 1999).

A wavelet-based algorithm for controlling the FDR was described by Fadili and Bullmore (2004). The FDR of empirical wavelet coefficients was conventionally defined as the expected false positive fraction E (FPF), that is, the proportion of all positive tests that are falsely positive. Following Abramovich et al. (Abramovich and Benjamini, 1996; Abramovich and Sapatinas, 1996), the maximum number of observed wavelet coefficients was retained subject to the constraint E (FPF) < α, using the following algorithm to calculate a global threshold for elimination/retention:

---

[7] The discrete wavelet transform of Gaussian white noise is Gaussian white noise.

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:

$$p_{j,k} = 2\left(1 - \Phi\left(\frac{|d_{j,k}|}{\hat{\sigma}}\right)\right) \tag{26}$$

where Φ is the cumulative standard normal distribution and $\hat{\sigma}$ is the robust estimate of the standard deviation of the noise (MAD/0.6745).

2. Sort the $p_{j,k}$ in an ascending order, $p_1 \leq p_2 \leq p_3 \ldots \leq p_n$.
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:

$$\lambda_{FDR} = \hat{\sigma}\Phi^{-1}\left(1 - \frac{p i_{FDR}}{2}\right) \tag{27}$$

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.

A more powerful variant of the FDR control algorithm, called enhanced false discovery rate (EFDR) control, has recently been proposed by Shen et al. (2002) and applied to fMRI by Pavlicova et al. (2003). The basic idea of EFDR is to reduce the number of hypothesis tests before applying a standard FDR procedure to the reduced set of test statistics in the wavelet domain. It has been shown that EFDR inherits the desirable type I error control properties of FDR, that is, it also controls the family-wise error, but is more powerful by virtue of testing a smaller number of hypotheses. Shen et al. (2002) used a test statistic that was informed by a local neighborhood of wavelet coefficients to take advantage of anticipated dependencies between neighboring wavelet coefficients and defined the reduced set of hypotheses to go forward to testing by the FDR algorithm using the generalized degrees of freedom (Shen et al., 2002).
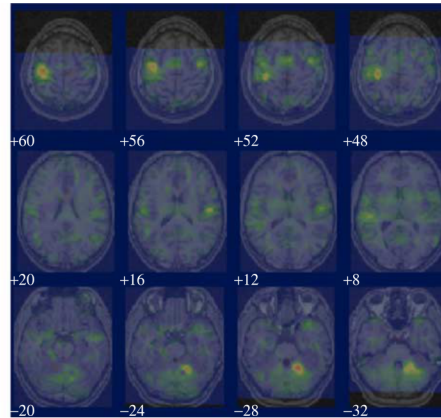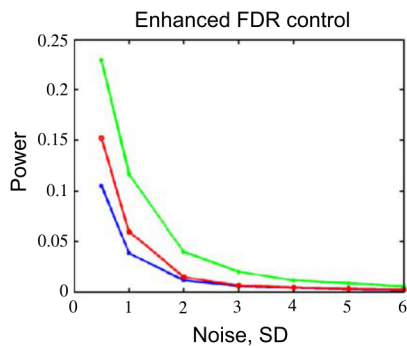
This logic can be used to develop algorithms that use different test statistics or different methods for restricting the number of tests in the wavelet domain. This more general wavelet-based EFDR algorithm can be summarized as follows:

1. Calculate the wavelet coefficients of the image.
2. Use a nonparametric wavelet shrinkage algorithm to eliminate the majority of noisy coefficients, retaining a reduced number of coefficients n^*.
3. Apply a standard FDR algorithm, as already defined, to test the reduced set of coefficients.
4. Estimate the signal by the inverse wavelet transform using only those coefficients which survive step 3.

As shown in Fig. 8, the EFDR algorithm provides considerably greater power than Bonferroni and classical FDR algorithms at the same level of type I error control.

Bayesian hypothesis testing in the wavelet domain

Bayesian methods for wavelet thresholding were described in detail by Vidakovic (1999) and seem likely to offer improved sensitivity in many applications. The Bayesian approach here means putting a prior distribution on the variability of the sparse wavelet coefficients of the true (unknown) image. The question arises: What is a good model for the prior distribution of the wavelet coefficients? Most often, the coefficients have been assumed to be independently distributed under Gaussian, Laplacian, generalized Gaussian, or other distributions (Fadili and Bullmore, 2004). For example, the classical soft thresholding operator can be obtained by a Laplacian assumption. Bayesian methods for image denoising using other priors have also been proposed (Gao, 1998).

Fig. 8. Multiresolution hypothesis testing in the wavelet domain. Left panel: Wavelet-based methods for control of the false discovery rate (FDR) show decreasing power to detect simulated signals as a function of increasing spatial noise. At any level of noise, Bonferroni correction of multiple tests on wavelet coefficients (blue line) is less powerful than the FDR control procedure of Abramovich and Benjamini (1995) (red line). An enhanced FDR control algorithm (green line; Shen et al., 2002), incorporating a bivariate wavelet shrinkage algorithm (Sendur and Selesnick, 2002a) to define the reduced set of hypotheses to be tested, is more powerful than conventional FDR. Right panel: Experimental fMRI data acquired during an event-related right-hand finger-tapping task were mapped by a Bayesian multiresolution hypothesis testing algorithm in the wavelet domain (posterior probability threshold $\alpha = 0.01$), incorporating a Gaussian prior for the sparse wavelet coefficients of the true image. Left brain is shown in the left side of each slice. There is intense, focal activation of contralateral motor cortex and ipsilateral cerebellum and more diffuse activation of supplementary motor area and ipsilateral motor cortex.

Sendur and Selesnick (2002a,b) proposed a non-Gaussian bivariate probability distribution function to model the bivariate statistics of wavelet coefficients of natural images (Field, 1987; Simoncelli and Olshausen, 2001). The model captures the dependence between a wavelet coefficient and its parent, that is, the coefficient at the same location but an immediately coarser scale. Using Bayesian estimation theory a simple nonlinear shrinkage function for wavelet denoising was derived from this model, which generalized the classical soft thresholding approach. The new shrinkage function, which depends on both the coefficient and its parent, yields improved results for wavelet-based image denoising.

Let $\delta_1 = \delta_{j+1,k}$ represent the "parent" of $\delta_2 = \delta_{j,k}$, both being wavelet coefficients of the true image. We can write $d_1 = \delta_1 + \varepsilon_1$ and $d_2 = \delta_2 + \varepsilon_2$, where d_1 and d_2 are noisy observations of δ_1 and δ_2; ε_1 and ε_2 are noise. The proposed non-Gaussian bivariate pdf for the coefficient and its parent can be written as

$$P_d(\delta_1, \delta_2) = \frac{3}{2\pi\sigma^2} \exp\left(\frac{\sqrt{3}}{\sigma} \sqrt{\delta_1^2 + \delta_2^2} \right). \qquad (28)$$

The maximum *a posteriori* (MAP) estimator of δ_1 is derived to be:

$$\hat{\delta}_1 = \frac{\left(\sqrt{d_1^2 + d_2^2} - \frac{\sqrt{3}\sigma_n^2}{\sigma} \right)_+}{\sqrt{d_1^2 + d_2^2}} d_1 \qquad (29)$$

which can be interpreted as a Bayesian bivariate shrinkage function.

An alternative approach to Bayesian thresholding was developed for fMRI by Fadili and Bullmore (2004) who used a Gaussian prior. For this model, Abramovich and Sapatinas (1996) have proposed a closed form expression for the ratio test (RT) statistic as the Bayes thresholding rule and an EM algorithm to estimate the hyperparameters including the level-dependent variance of the Gaussian pdf. In Fadili and Bullmore (2004), only wavelet coefficients with posterior probabilities greater than an arbitrary critical threshold were retained in the reconstruction by hard thresholding (for an illustrative activation map produced by this technique, see Fig. 8).

Conclusions

These are early days in the application of wavelets to the particular challenges of fMRI data analysis and much remains to be tried and tested. However, it already seems clear that the wavelet domain is a rich source of relatively new concepts and techniques to enhance the power of statistical analysis of scale-invariant time series and spatial processes. We have highlighted applications to the problems of resampling, time series modeling, and multiple hypothesis testing in fMRI, but there are many wavelet methods that have not yet been properly explored in relation to fMRI and there are many apparently suitable problems in fMRI—for example, the scaling properties of functional connectivity—that have not yet been addressed in the wavelet domain. It seems likely that interesting further developments lie ahead.

Acknowledgments

References

Abramovich, F., Benjamini, Y., 1995. Thresholding of wavelet coefficients as multiple hypotheses testing procedure. In: Antoniadis, A., Oppenheim, G. (Eds.), Wavelets and Statistics. Springer, New York, pp. 5–14.

Abramovich, F., Benjamini, Y., 1996. Adaptive thresholding of wavelet coefficients. Comput. Stat. Data Anal. 22, 351–361.

Abramovich, F., Sapatinas, T., 1996. Bayesian approach to wavelet decomposition and shrinkage. In: Muller, P., Vidakovic, B. (Eds.), Bayesian Inference in Wavelet Based Models. Springer, New York, pp. 33–50.

Abry, P., Baraniuk, R., Flandrin, P., Riedi, R., Veitch, D., 2002. Multiscale nature of network traffic. IEEE Signal Process. Mag. 19, 28–46.

Abu-Rezq, A.N., Tolba, A.S., Khuwaja, G.A., Foda, S.G., 1999. Best parameters selection for wavelet packet-based compression of magnetic resonance images. Comput. Biomed. Res. 12, 173–184.

Aldroubi, A., Unser, M., 1996. Wavelets in Biology and Medicine. CRC Press, Boca Raton.

Alexander, M.E., Baumgartner, R., Windischberger, C., Moser, E., Somorjai, R.L., 2000. Wavelet domain de-noising of time-courses in MR image sequences. Magn. Reson. Imaging 18, 1129–1134.

Angelidis, P.A., 1994. MR image compression using a wavelet transform coding algorithm. Magn. Reson. Imaging 12, 1111–1120.

Baraniuk, R.G., Flandrin, P., Janssen, A.J.E.M., Michel, O.J.J., 2002. Measuring time-frequency information content using the Renyi entropies. IEEE Trans. Inf. Theory 47, 1391–1409.

Barra, V., Boire, J.Y., 2000. Tissue segmentation on MR images of the brain by possibilistic clustering on a 3D wavelet representation. J. Magn. Reson. Imaging 11, 267–278.

Beran, J., 1994. Statistics for Long-Memory Processes. Chapman and Hall, Boca Raton.

Blanton, R.E., Levitt, J.G., Thompson, P.M., Narr, K.L., Captillo-Cunliffe, L., Nobel, A., Singerman, J.D., McCracken, J.T., Toga, A.W., 2001. Mapping cortical asymmetry and complexity patterns in normal children. Psychiatry Res. Neuroimaging 107, 29–43.

Brammer, M., 1998. Multidimensional wavelet analysis of functional magnetic resonance images. Hum. Brain Mapp. 6, 378–382.

Breakspear, M., Brammer, M.J., Robinson, P.A., 2003. Construction of multivariate surrogate sets from nonlinear data using the wavelet transform. Physica D 182, 1–22.

Breakspear, M., Brammer, M., Bullmore, E., Das, P., Williams, L.M., 2004. Spatio-temporal wavelet resampling for functional neuroimaging data. Hum. Brain Mapp. 23, 1–25.

Brown, J.H., Gupta, V.K., Li, B.L., Milne, B.T., Restrepo, C., West, G.B., 2002. The fractal nature of nature: power laws, ecological complexity and biodiversity. Philos. Trans. R. Soc. B 357, 619–626.

Bruce, A., Gao, H.-Y., 1996. Applied Wavelet Analysis with S-PLUS. Springer, New York, NY.

Bullmore, E.T., Brammer, M.J., Harvey, I., Persaud, R., Murray, R.M., Ron, M.A., 1994a. Fractal analysis of the boundary between white matter and cerebral cortex in magnetic resonance images: a controlled study of schizophrenic and manic-depressive patients. Psychol. Med. 24, 771–781.

Bullmore, E.T., Brammer, M.J., Bourlon, P., Alarcon, G., Polkey, C.E., Elwes, R., Binnie, C.D., 1994b. Fractal analysis of electroencephalographic signals intracerebrally recorded during 35 epileptic seizures: evaluation of a new method for synoptic visualization of ictal events. Electroencephalogr. Clin. Neurophysiol. 91, 337–345.

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

Bullmore, E., Suckling, J., Overmeyer, S., Rabe-Hesketh, S., Taylor, E., Brammer, M.J., 1999a. 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. Imaging 18, 32–42.

Bullmore, E., Brammer, M.J., Rabe-Hesketh, S., Curtis, V.A., Morris, R.G., Williams, S.C.R., Sharma, T., McGuire, P.K., 1999b. Methods for diagnosis and treatment of stimulus-correlated motion in generic brain activation studies using fMRI. Hum. Brain Mapp. 7, 38–48.

Bullmore, E., Long, C., Suckling, J., Fadili, M.J., Calvert, G., Zelaya, F., Carpenter, T.A., Brammer, M.J., 2001a. Colored noise and computational inference in neurophysiological fMRI time series analysis: resampling methods in time and wavelet domains. Hum. Brain Mapp. 12, 61–78.

Bullmore, E., Suckling, J., Brammer, M.J., 2001b. In praise of tedious permutation. In: Moore, M. (Ed.), Spatial Statistics: Methodological Aspects and Some Applications. Springer, New York, pp. 183–200.

Bullmore, E.T., Fadili, J., Breakspear, M., Salvador, R., Suckling, J., Brammer, M.J., 2003. Wavelets and statistical analysis of functional magnetic resonance images of the human brain. Stat. Methods Med. Res. 12, 375–399.

Carlstein, E., Do, K.-A., Hall, P., Hesterberg, T., Kunsch, H.R., 1998. Matched block bootstrap for dependent data. Bernouilli 4, 305–328.

Caserta, F., Eldred, W.D., Fernandez, E., Hausman, R.E., Stanford, L.R., Bulderev, S.V., Schwarzer, S., Stanley, H.E., 1995. Determination of fractal dimension of physiologically characterised neurons in 2-dimensions and 3-dimensions. J. Neurosci. Methods 56, 133–144.

Cohen, A., Daubechies, I., Vial, P., 1993. Wavelet bases on the interval and fast algorithms. J. Appl. Comput. Harm. Anal. 1, 54–81.

Cselenyi, Z., Olsson, H., Farde, L., Gulyas, B., 2002. Wavelet-aided parametric mapping of cerebral dopamine D2 receptors using the high affinity PET radioligand [^{11}C]FLB 457. NeuroImage 17, 47–60.

Daubechies, I., 1992. Ten Lectures on Wavelets. Society for Industrial and Applied Mathematics, Phildelphia, PA.

Davison, A.C., Hinkley, D.V., 1997. Bootstrap Methods and Their Application. Cambridge Univ. Press, Cambridge.

Desco, M., Hernandez, J.A., Santos, A., Brammer, M.J., 2001. Multi-resolution analysis in fMRI: sensitivity and specificity in the detection of brain activation. Hum. Brain Mapp. 14, 16–27.

Dijkerman, R.W., Mazumdar, R.R., 1994. On the correlation structure of the wavelet coefficients of fractional Brownian motion. IEEE Trans. Inf. Theory 40, 1609–1612.

Dinov, I.D., Mega, M.S., Thompson, P.M., Woods, R.P., Sumners de, W.L., Sowell, E.L., Toga, A.W., 2002. Quantitative comparison and analysis of brain image registration using frequency-adaptive wavelet shrinkage. IEEE Trans. Inf. Technol. Biomed. 6, 73–85.

Donoho, D.L., Johnstone, I.M., 1995. Adapting to unknown smoothness via wavelet shrinkage. J. Am. Stat. Assoc. 90, 1200–1224.

Fadili, M.J., Bullmore, E.T., 2002. Wavelet-generalised least squares: a new BLU estimator of linear regression models with $1/f$ errors. NeuroImage 15, 217–232.

Fadili, M.J., Bullmore, E.T., 2004. A comparative evaluation of wavelet-based methods for multiple hypothesis testing of brain activation maps. NeuroImage (in press).

Fadili, M.J., Bullmore, E.T., Brett, M., 2001. Wavelet methods for characterising mono-and poly-fractal noise structures in shortish time series: an application to functional MRI. Proc. IEEE Int. Conf. Image Process. 2001, 225–228.

Fan, Y., 2003. On the approximate decorrelation property of the discrete wavelet transform for fractionally differenced processes. IEEE Trans. Inf. Theory 49, 516–521.

Field, D., 1987. Relations between the statistics of natural images and the response properties of cortical cells. J. Opt. Soc. Am. A 4, 2379–2394.

Flandrin, P., 1992. Wavelet analysis and synthesis of fractional Brownian motion. IEEE Trans. Inf. Theory 38, 910–917.

Free, S.L., Sisodiya, S.M., Cook, M.J., Fish, D.R., Shorvon, S.D., 1996. Three-dimensional fractal analysis of the white matter surface from magnetic resonance images of the human brain. Cereb. Cortex 6, 830–836.

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

Gao, H., 1998. Wavelet shrinkage denoising using the non-negative garrote. J. Comput. Graph. Stat. 7, 469–488.

Guan, S., Lai, C.H., Wei, G.W., 2002. A wavelet method for the characterization of spatiotemporal patterns. Physica D 163, 49–79.

Hayasaka, S., Nichols, T.E., 2003. Validating cluster size inference: random field and permutation methods. NeuroImage 20, 2343–2356.

Hilton, M., Ogden, T., Hattery, D., Edenm, G., Jawerth, B., 1996. Wavelet processing of functional MRI data. In: Aldroubi, A., Unser, M. (Eds.), Wavelets in Medicine and Biology. CRC Press, Boca Raton, pp. 93–114.

Horwitz, B., 2003. The elusive concept of brain connectivity. NeuroImage 19, 466–470.

Hossien-Zadeh, G.A., Soltanian-Zadeh, H., Ardekani, B.A., 2003. Multi-resolution fMRI activation detection using translation invariant wavelet transform and statistical analysis based on resampling. IEEE Trans. Med. Imaging 22, 795–805.

Hyvarinen, A., 1999. Sparse code shrinkage: denoising of non-Gaussian data by maximum likelihood estimation. Neural Comput. 11, 1739–1768.

Ivanov, P.C., Rosenblum, M.G., Peng, C.K., Mietus, J., Havlin, S., Stanley, H.E., Goldberger, A.L., 1996. Scaling behavior of heartbeat intervals obtained by wavelet-based time-series analysis. Nature 383, 323–327.

Iyriboz, T.A., Zukoski, M.J., Hopper, K.D., Stagg, P.L., 1999. A comparison of wavelet and JPEG lossy compression methods applied to medical images. J. Digit. Imaging 12, 14–17.

Jaffard, S., Meyer, Y., Ryan, R.D., 2001. Wavelets: Tools for Science and Technology. Society for Industrial and Applied Mathematics, Phildelphia, PA.

Johnstone, I.M., Silverman, B.W., 1997. Wavelet threshold estimators for data with correlated noise. J. R. Stat. Soc., Ser. B Stat. Methodol. 59, 319–351.

Kiselev, V.G., Hahn, K.R., Auer, D.P., 2003. Is the brain cortex a fractal? NeuroImage 20, 1765–1774.

Kuikka, J., Tiihonen, J., 1998. Fractal analysis—A new approach in brain receptor imaging. Ann. Med. 30, 242–248.

Kybic, J., Thevenaz, P., Nirkko, A., Unser, M., 2000. Unwarping of unidirectionally distorted EPI images. IEEE Trans. Med. Imaging 19, 97–119.

LaConte, S.M., Ngan, S.C., Hu, X., 2000. Wavelet transform based Wiener filtering of event-related fMRI data. Magn. Reson. Med. 44, 746–757.

Laine, A., 2000. Wavelets in temporal and spatial processing of biomedical images. Ann. Rev. Biomed. Eng. 2, 511–550.

Laird, A.R., Rogers, B.P., Meyerand, M.E., 2004. Comparison of Fourier and wavelet resampling methods. Magn. Reson. Med. 51, 418–422.

Lindley, D.V., Novick, M.R., 1981. The role of exchangeability in inference. Ann. Stat. 9, 45–58.

Linkenkaer-Hansen, K., Nikouline, V.V., Palva, J.M., Ilmoniemi, R.J., 2001. Long-range temporal correlations and scaling behavior in human brain oscillations. J. Neurosci. 21, 1370–1377.

Locascio, J.L., Jennings, P.J., Moore, C.I., Corkin, S., 1997. Time series analysis in the time domain and resampling methods for studies of functional Magn. Reson. Imaging. Hum. Brain Mapp. 5, 168–193.

Long, C.J., Brown, E.N., Manoach, D.S., Solo, V., 2004. Spatiotemporal wavelet analysis for functional MRI. NeuroImage (in press).

Maldjian, J.A., Liu, W.C., Hirschorn, D., Murthy, R., Semanczuk, W., 1997. Wavelet transform-based image compression for transmission of MR data. Am. J. Roentgenol. 169, 23–26.

Mallat, S., 1998. A Wavelet Tour of Signal Processing. Academic Press, San Diego.

Mandelbrot, B.B., 1977. The Fractal Geometry of Nature. WH Freeman, New York.

Mandelbrot, B.B., 1997. Fractals and Scaling in Finance: Discontinuity, Concentration, Risk. Springer, New York.

Marchini, J., Smith, S., 2003. On bias in the estimation of autocorrelations for fMRI voxel time series analysis. NeuroImage 18, 83–90.

Matérn, B., 1986. Spatial Variation. Springer Verlag, Berlin.

Meyer, F.G., 2003. Wavelet based estimation of a semi-parametric generalized linear model of fMRI time series. IEEE Trans. Med. Imaging 22, 315–322.

Müller, K., Lohmann, G., Zysset, S., von Cramon, D.Y., 2003. Wavelet statistics of functional MRI data and the general linear model. J. Magn. Reson. Imaging 17, 20–30.

Nakao, H., Mishiro, T., Yamada, M., 2001. Visualization of correlation cascade in spatiotemporal chaos using wavelets. Int. J. Bifurc. Chaos 11, 1483–1493.

Ngan, S.C., LaConte, S.M., Xu, H., 2000. Temporal filtering of event-related fMRI data using cross-validation. NeuroImage 11, 797–804.

Nichols, T.E., Holmes, A.P., 2002. Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum. Brain Mapp. 15, 1–25.

Ninness, B., 1998. Estimation of $1/f$ noise. IEEE Trans. Inf. Theory 44, 32–46.

Parlitz, U., Meyer-Kress, G., 1995. Predicting low-dimensional spatiotemporal dynamics using discrete wavelet transforms. Phys. Rev., E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics 51, 2709–2711.

Pavlicova, M., Cressie, N., Santner, T.J., Algaze, A., 2003. Using enhanced FDR to find activation in fMRI images. NeuroImage 19, S910.

Peng, C.K., Buldryev, S.V., Goldberger, A.L., Havlin, S., Mategna, R.N., Simons, M., Stanley, H.E., 1995. Statistical properties of DNA sequences. Physica A 221, 180–192.

Percival, D.B., Walden, A.T., 2000. Wavelet Methods for Time Series Analysis. Cambridge Univ. Press, Cambridge.

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

Raz, J., Dickerson, L., Turetsky, B., 1999. A wavelet packet model of evoked potentials. Brain Lang. 66, 61–88.

Ruttimann, U., Unser, M., Rawlings, R., Rio, D., Ramsey, N., Mattay, V., Hommer, D., Frank, J., Weinberger, D., 1998. Statistical analysis of functional MRI data in the wavelet domain. IEEE Trans. Med. Imaging 17, 142–154.

Sabatini, A.M., 1999. Wavelet-based estimation of $1/f$-type signal parameters: confidence intervals using the bootstrap. IEEE Trans. Signal Process. 47, 3406–3409.

Schroeder, M., 1991. Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise. WH Freeman, New York.

Sendur, L., Selesnick, I.W., 2002a. Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency. IEEE Trans. Signal Process. 50, 2744–2756.

Sendur, L., Selesnick, I.W., 2002b. Bivariate shrinkage with local variance estimation. IEEE Signal Process. Lett 9, 438–441.

Senhadji, L., Dillenseger, J.L., Wendling, F., Rocha, C., Kinie, A., 1995. Wavelet analysis of EEG for three-dimensional mapping of epileptic events. Ann. Biomed. Eng. 32, 738–748.

Shen, X., Huang, H.-C., Cressie, N., 2002. Nonparametric hypothesis testing for a spatial signal. J. Am. Stat. Assoc. 97, 1122–1140.

Shimizu, Y., Barth, M., Windischberger, C., Moser, E., Thurner, S., 2004. Wavelet-based multifractal analysis of fMRI time series. NeuroImage 22, 1195–1202.

Simoncelli, E., Olshausen, B., 2001. Natural image statistics and neural representation. Annu. Rev. Neurosci. 24, 1193–1216.

Staszewski, W.J., Worden, K., 1999. Wavelet analysis of time series: coherent structures, chaos and noise. Int. J. Bifurc. Chaos 9, 455–471.

Strogatz, S.V., 2001. Exploring complex networks. Nature 410, 268–276.

Tewfik, A.H., Kim, M., 1992. Correlation structure of the discrete wavelet coefficients of fractional Brownian motion. IEEE Trans. Inf. Theory 38, 904–909.

Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., Farmer, J., 1992. Testing for nonlinearity: the method of surrogate data. Physica D 58, 77–94.

Thompson, P.M., Schwartz, C., Lin, R.T., Khan, A.A., Toga, A.W., 1996. Three-dimensional statistical analysis of sulcal variability in the human brain. J. Neurosci. 16, 4261–4274.

Turkheimer, F.E., Brett, M., Aston, J.A., Leff, A.P., Sargent, P.A., Wise, R.J., Grasby, P.M., Cunningham, V.J., 2000. Statistical modeling of positron emission tomography images in wavelet space. J. Cereb. Blood Flow Metab. 20, 1610–1618.

Turkheimer, F.E., Aston, J.A.D., Banati, R.B., Riddell, C., Cunningham, V.J., 2003. A linear wavelet filter for parametric imaging with dynamic PET. IEEE Trans. Med. Imaging 22, 289–301.

Turner, R., 2001. BOLD localization: the implications of vascular architecture. NeuroImage 13, S1011.

Vidakovic, B., 1999. Statistical Modeling by Wavelets. John Wiley and Sons, New York.

Von Tscharner, V., Thulborn, K.R., 2001. Specified-resolution wavelet analysis of activation patterns from BOLD contrast fMRI. IEEE Trans. Med. Imaging 20, 704–714.

Voss, R.F., 1988. Fractals in nature: from characterization to simulation. In: Barnsley, M.F., Devaney, R.L., Mandelbrot, B.B., Peitgen, H.-O., Saupe, D., Voss, R.F. (Eds.), The Science of Fractal Images. Springer, New York, pp. 21–70.

Weaver, J.B., Xu, Y.S., Healy, D.M., Cromwell, L.D., 1991. Filtering noise from images with wavelet transforms. Magn. Reson. Med. 21, 288–295.

Welchew, D.E., Honey, G.D., Sharma, T., Robbins, T.W., Bullmore, E.T., 2002. Multidimensional scaling of integrated neurocognitive function and schizophrenia as a disconnexion disorder. NeuroImage 17, 1227–1239.

West, G.B., Brown, J.H., Enquist, B.J., 1999. The fourth dimension of life: fractal geometry and allometric scaling of organisms. Science 284, 1677–1679.

Whitcher, B., 2001. Simulating Gaussian stationary processes with unbounded spectra. J. Comput. Graph. Stat. 10, 112–134.

Whitcher, B., 2004a. Wavelet-based estimation for seasonal long-memory processes. Technometrics 46, 225–238.

Whitcher, B., 2004b. Wavelet-based bootstrapping of spatial patterns on a finite lattice. Comput. Stat. Data Anal. (in review).

Wink, A.M., Roerdink, J.B.T.M., 2004. Denoising functional MR images: a comparison of wavelet denoising and Gaussian smoothing. IEEE Trans. Med. Imaging 23, 374–387.

Wood, J.C., Johnson, K.M., 1999. Wavelet packet denoising of magnetic resonance images: importance of Rician noise at low SNR. Magn. Reson. Med. 41, 631–635.

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

Wornell, G.W., 1993. Wavelet-based representations for the $1/f$ family of fractal processes. Proc. IEEE 81, 1428–1450.

Wornell, G.W., 1996. Signal Processing with Fractals: A Wavelet-Based Approach. Prentice Hall, Upper Saddle River, NJ.

Worsley, K.J., Marrett, S., Neelin, P., Evans, A.C., 1996. Searching scale-space for activation in PET images. Hum. Brain Mapp. 4, 74–90.

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

Wu, B.-F., Su, Y.-L., 1998. On stationarizability for nonstationary 2D random fields using discrete wavelet transform. IEEE Trans. Image Process. 7, 1359–1366.

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

Zaroubi, S., Goelman, G., 2000. Complex denoising of MR data via wavelet analysis: application for functional MRI. Magn. Reson. Imaging 18, 59–68.