◆ Human Brain Mapping 12:61–78(2001) ◆

Colored Noise and Computational Inference in Neurophysiological (fMRI) Time Series Analysis: Resampling Methods in Time and Wavelet Domains

Ed Bullmore,[1,4,*] Chris Long,[4] John Suckling,[4,5] Jalal Fadili,[3] Gemma Calvert,[6] Fernando Zelaya,[4] T. Adrian Carpenter,[2] and Mick Brammer[4]

[1]*Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom*
[2]*Wolfson Brain Imaging Centre, University of Cambridge, Cambridge, United Kingdom*
[3]*Department of Experimental Psychology, University of Cambridge, Cambridge, United Kingdom*
[4]*Institute of Psychiatry, King's College, London, United Kingdom*
[5]*Guy's, King's and St Thomas's Medical School, London, United Kingdom*
[6]*FMRIB Centre, University of Oxford, Oxford, United Kingdom*

Edited by: Karl Friston, Associate Editor

Abstract: Even in the absence of an experimental effect, functional magnetic resonance imaging (fMRI) time series generally demonstrate serial dependence. This colored noise or endogenous autocorrelation typically has disproportionate spectral power at low frequencies, i.e., its spectrum is $\frac{1}{f}$-like. Various pre-whitening and pre-coloring strategies have been proposed to make valid inference on standardised test statistics estimated by time series regression in this context of residually autocorrelated errors. Here we introduce a new method based on random permutation after orthogonal transformation of the observed time series to the wavelet domain. This scheme exploits the general whitening or decorrelating property of the discrete wavelet transform and is implemented using a Daubechies wavelet with four vanishing moments to ensure exchangeability of wavelet coefficients within each scale of decomposition. For $\frac{1}{f}$-like or fractal noises, e.g., realisations of fractional Brownian motion (fBm) parameterised by Hurst exponent $0 < H < 1$, this resampling algorithm exactly preserves wavelet-based estimates of the second order stochastic properties of the (possibly nonstationary) time series. Performance of the method is assessed empirically using $\frac{1}{f}$-like noise simulated by multiple physical relaxation processes, and experimental fMRI data. Nominal type 1 error control in brain activation mapping is demonstrated by analysis of 13 images acquired under null or resting conditions. Compared to autoregressive pre-whitening methods for computational inference, a key advantage of wavelet resampling seems to be its robustness in activation mapping of experimental fMRI data acquired at 3 Tesla field strength. We conclude that wavelet resampling may be a generally useful method for inference on naturally complex time series. *Hum. Brain Mapping 12:61–78, 2001.*

Key words: fractional noise; neuroimaging; brain mapping; nonparametric; high field MRI

INTRODUCTION

Colored noise

Neurophysiological time series obtained by functional magnetic resonance imaging (fMRI) are commonly tested for an experimental effect in the time domain, by fitting a linear regression model to the time series $\mathbf{y}$ or $\{Y_t\}$ $t = 1, 2, 3, \ldots, N$ at each voxel:

$$\mathbf{y} = \mathbf{X}\beta + \rho. \quad (1)$$

This model partitions each time series into a component determined by the experimental design $\mathbf{X}\beta$, where $\mathbf{X}$ is the design matrix and β is the vector of regression model parameters, and a residual component $\mathbf{y} - \mathbf{X}\beta = \rho$ or $\{\rho_t\}$ with variance σ^2 and covari-

*Correspondence to: Ed Bullmore, Brain Mapping Unit, Dept Psychiatry (Box 255), Addenbrooke's Hospital, Cambridge CB2 2QQ, UK. E-mail: etb23@cam.ac.uk

Received for publication 26 May 2000; accepted 14 September 2000

ance matrix Ω. If the model residuals are serially independent or white noise then the off-diagonal elements of Ω are zero, i.e., $\Omega = \sigma^2 \mathsf{I}$, where I is the identity matrix, and the Gauss-Markov theorem guarantees that ordinary least squares estimates will be unbiased and will have minimum variance over all linear estimates. As a rule, however, the residuals generated by fitting a linear model to an fMRI time series will *not* be white noise; more usually, they will be colored noise demonstrating some degree of serial dependency or temporal autocorrelation, i.e., the off-diagonal elements of Ω will not be zero.

In the frequency domain, this is equivalent to stating that the power spectral density of the noise $S_f(\rho)$ is not uniform over all frequencies f, i.e., $S_f(\rho) \not\approx f^0$. A careful analysis of the spectral characteristics of fMRI noise, examined in datasets acquired under resting or "null" conditions, has instead suggested that the noise may generally have $\frac{1}{f}$-like properties with disproportionate power at lower frequencies, i.e., $S_f(\rho) \approx f^{-\alpha}$ with spectral exponent $\alpha > 0$ [Zarahn et al., 1997; Aguirre et al., 1997].[1]

Several sources of color in fMRI have been suggested, including hemodynamically-convolved neuronal or instrumental (white) noise, aliased cardiorespiratory pulsation, uncorrected head movement, and experimentally-induced variance that is not well-modelled by the design matrix. It is clear from studies of cadavers and phantoms that colored noise may arise in the absence of physiological processes and must therefore be due, at least partly, to physical effects [Zarahn et al., 1997; Smith et al., 1999]. It is also known that the magnitude and form of residual autocorrelation may vary considerably from voxel to voxel within the brain [Bullmore et al., 1996; Locascio et al., 1997] and may be influenced by experimental parameters including the repetition time (TR) between consecutive time points in the series, and the period length in blocked periodic experimental designs [Purdon and Weisskoff, 1998]. The characterisation of fMRI noise as a $\frac{1}{f}$-like process does not by itself discriminate between physical and physiological sources. $\frac{1}{f}$-like processes, which are generally self-affine and have non-integer fractal dimensions in the time domain, are naturally ubiquitous [Mandelbrot, 1977; Voss, 1988]; see Appendix. Physical systems in which many particles are relaxing from excited states at different rates are well known generators of $\frac{1}{f}$-like noise [Schroeder, 1991], which seems potentially relevant to fMRI. And physiological time series recorded by electrocardiography [Goldberger et al., 1990] and electroencephalography [Bullmore et al., 1994a] have demonstrated $\frac{1}{f}$-like or fractal properties, indicating that this may be a common mode of dynamic behavior for human biological processes also.

[1]A notational note: The term $\frac{1}{f}$-like is used here in preference to the simpler $\frac{1}{f}$ partly because the spectral exponent $1 \leq \alpha \leq 3$ may not always be exactly equal to 1 and partly because many such "pink" or fractal or fractional noises are not stationary and therefore cannot strictly be described in terms of a power spectrum; see Appendix.

Pre-whitening and pre-coloring

But regardless of the source(s) of colored noise in fMRI, it represents a problem for inference based on time series regression parameters. Ordinary least squares estimates of parameter standard errors, given by $\sigma^2(\mathsf{X}^T\mathsf{X})^{-1}$, will then be under-estimates, meaning that standardised test statistics, constructed as the ratio of a parameter to its standard error, will be over-estimates, and type 1 error will be inflated above its nominal rate. This problem has been addressed in the time domain in one of two main ways: either by estimating the off-diagonal elements of Ω by modelling the residual series ρ [Bullmore et al., 1996; Locascio et al., 1997; Purdon and Weisskoff, 1998; Dale, 1999]; or by predicting the off-diagonal elements of Ω by convolving the data and design matrix with a known smoothing kernel [Friston et al., 1995; Worsley and Friston, 1995; Friston et al., 2000]. The first method is often called *pre-whitening* because it attempts to remove endogenous autocorrelation from the residuals of an iterative least squares fit; the second method has been called *pre-coloring* because it attempts to replace endogenous autocorrelation of unknown form by imposing exogenous auto-correlation of a known form. Fundamentally, the two approaches are very similar.

To pre-whiten the data, we find the whitening or deconvolution matrix K^{-1}, where $\mathsf{K}\mathsf{K}^T = \Omega$ is the factorisation or Cholesky decomposition of the covariance matrix Ω, and pre-multiply both the data and the design matrix by this whitening matrix before fitting the transformed model,

$$\mathsf{K}^{-1}\mathbf{y} = \mathsf{K}^{-1}\mathsf{X}\beta + \epsilon, \tag{2}$$

by least squares to yield a serially independent or white noise residual series $\{\epsilon_t\}$. The estimated parameters and their standard errors will then have the desirable properties of being unbiased with minimum variance.

To pre-color the data, we specify a coloring or convolution matrix K, and pre-multiply both the data and

the design matrix by this coloring matrix before fitting the transformed model,

$$\mathbf{K}\mathbf{y} = \mathbf{K}\mathbf{X}\beta + \mathbf{K}\epsilon, \quad (3)$$

by least squares to yield a serially dependent or colored noise series $\mathbf{K}\epsilon$ which has covariance matrix $\Omega = \sigma^2\mathbf{K}\mathbf{K}^T$. Appropriately adjusted estimates of parameter standard errors are then given by the theory of generalised least squares [Worsley and Friston, 1995].

These two approaches have some complementary strengths and weaknesses: pre-whitening is theoretically more efficient but may be biased if the unknown form of the endogenous autocorrelation is not correctly modelled; pre-coloring is less efficient but less prone to bias provided the coloring matrix is robust enough to impose its predicted form on the residual autocorrelation [Friston et al., 2000]. Pre-coloring will also incur a risk of attenuating high frequency features in the data that may be experimentally determined [Marchini and Ripley, 2000], perhaps especially if the experimental design involves rapidly sequential presentation of discrete trials. But, at present, there is probably insufficient data to judge the adequacy of either approach in dealing with the greater challenges of treating colored noise in event-related experiments or experiments conducted at high field strength, i.e., $\geq$ 3 Tesla.

Both pre-coloring and pre-whitening share strengths and weaknesses in common compared to alternative methods in the frequency domain [Lange and Zeger, 1997; Marchini and Ripley, 2000]. The main relative strength of Fourier-based methods, for periodically designed experiments, is that the Fourier transform is asymptotically independent at different frequencies and this simplifies inference. For example, a test statistic constructed as the ratio of the log periodogram to the smoothed spectral density of the noise at a given frequency will be unbiased by endogenous color contributing power at other frequencies [Marchini and Ripley, 2000]. But time domain methods seem relatively stronger in two related ways: they are adaptable to a wider range of experimental designs; and so may be conceptually more accessible to psychologically-minded consumers, rather than statistically-minded producers, of brain mapping methods. To illustrate this point, consider an experiment in which blocks of four contrasting conditions are presented in a regularly repeated sequence ABCDABCD.... Periodogram-based methods will not be able to distinguish the contrast between conditions A and C from the contrast between conditions B and D, since both have the same frequency; whereas this distinction is trivial in the time domain. Non-periodic or event-related designs will also be less naturally treated in the frequency domain than in the time domain.

Finally, it is worth noting that both time and frequency domain methods share an important limitation: they assume that fMRI noise is stationary. This assumption may simplify the mathematical treatment of noise but it is unlikely to be realistic. More generally, we should allow that the stochastic properties of fMRI noise may change over time and/or that the noise may include transients and singularities.

Theoretical and computational inference

Returning to the choice between pre-whitening and pre-coloring methods of estimation, one important criterion to consider is the method by which statistical inferences will be made on these estimates. If the probability distribution of the test statistic under the null hypothesis will be theoretically derived then it is important above all that estimation is unbiased and some loss of efficiency may be tolerated to that end; in other words, a pre-coloring strategy might seem attractive. However, if the null distribution of the test statistic will be ascertained computationally, i.e., by repeated random resampling of the observed data [Efron and Tibshirani, 1993; Edgington, 1995; Davison and Hinkley, 1997], then any bias in parameters estimated on the observed data may equally apply to parameters estimated on resampled data. In which case, a pre-whitening strategy will seem preferable.

However, in order for a computationally ascertained null distribution "automatically" to correct the probabilities of observed test statistics for bias in their estimation, the data must be *exchangeable* or, approximately, independent in whatever form they are resampled [Lehmann, 1986]. If this is so, then the autocorrelational structure of the data will be preserved under resampling and biased estimation will be irrelevant to valid inference.

It is easy enough to prescribe a resampling scheme for fMRI which *fails* to meet these conditions. We can estimate the test statistic by ordinary least squares before and after naïvely, repeatedly reshuffling the time points at random. This will fail because consecutive time points in the raw time series are not independent or exchangeable under the null hypothesis; but they will be rendered independent by reshuffling. The observed and resampled statistics will thus be estimated on the basis of time series with different second order properties and will be differentially biased as a result. However, if we have adequately

pre-whitened the raw data, so that given $\beta = 0$ we have $\mathbf{K}^{-1}\mathbf{y} = \boldsymbol{\epsilon}$ (see Eq. 2), then we *can* legitimately resample the time series by permuting $\{Y_t\}$ or equivalently $\{\epsilon_t\}$. This kind of model-based resampling in the time domain has previously been used for computational inference on fMRI data [Bullmore et al., 1996; Locascio et al., 1997] and examples of this approach are described in greater detail below. Clearly the validity of such methods depends critically on the adequacy of the pre-whitening matrix $\mathbf{K}^{-1}$ to render the residuals $\{\epsilon_t\}$ independent. If the residuals are not in fact whitened by $\mathbf{K}^{-1}$ then the pre-whitened data will not be exchangeable, there will be differential bias in estimation of parameters on observed and resampled data, and uncontrolled type 1 error.

A rather different approach to resampling in the time domain is to make the resampling unit a block of consecutive time points rather than a single time point. Provided the number of points per block is large enough to encompass all long-range dependencies in the data, we might assume that the blocks are exchangeable and that the unknown stochastic properties of the time series under the null hypothesis will then be preserved after resampling. There is a large literature on block resampling for time series analysis and sophisticated refinements have been proposed to the basic scheme of randomly permuting non-overlapping blocks of arbitrary length [Carlstein et al., 1998]. But, in practice, block resampling can fail "catastrophically" to preserve serial dependencies in the data, so that the resampled series are whiter than the observed data, leading again to uncontrolled type 1 error [Davison and Hinkley, 1997].

Another possibility is to resample the observed time series after orthogonal transformation to another domain. An established 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 Fourier transform of the "phase scrambled" data to obtain a resampled time series which should preserve the spectral properties of the observed data. This method, like theoretical methods of inference in the frequency domain [Lange and Zeger, 1997; Marchini and Ripley, 2000], is justified 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 non-linear processes or long-range dependencies then this phase scrambling algorithm also may fail adequately to preserve the second order stochastic properties of the observed data [Davison and Hinkley, 1997; but see also Christofferson, 1997].

The rest of this paper

Here we introduce a new method for inference on endogenously colored fMRI time series. Briefly, this involves taking the discrete wavelet transform of the time series, randomly permuting the wavelet coefficients, then inverting the transform and thereby reconstituting the resampled data in the time domain. This algorithm exploits the whitening property of the discrete wavelet transform, which has previously been exploited for theoretical inference on positron emission tomography (PET) data [Ruttimann et al., 1996]; and it exactly preserves wavelet-estimated second order stochastic properties of $\frac{1}{f}$-like noises, such as fractional Brownian motion. In the next section, we describe experimental data and statistical methods; this is followed in turn by some results of applying the method to simulated time series and experimental fMRI data, and then by a discussion. Some supplementary material on fractals and fractional Brownian motion is consigned to an Appendix.

METHODS AND MATERIALS

Experimental designs and data acquisition

Gradient-echo echoplanar imaging data were acquired as follows:

Null (1.5T)

Thirteen normal volunteers were studied while they lay quietly in the scanner with their eyes closed for 5 min. One hundred T_2^*-weighted images were acquired at each of 14 non-contiguous slices of data in an oblique axial plane using the 1.5 Tesla (T) GE Signa system (General Electric, Milwaukee WI) at the Maudsley Hospital, London, UK: time to echo (TE) = 40 ms, TR = 3 sec, in-plane resolution = 3 mm, slice thickness = 7.7 mm, number of excitations = 1.

Bimodal sensory stimulation (1.5T AB)

One normal volunteer was studied during blocked periodic stimulation in both visual and auditory modalities. Visual stimulation consisted of 30 sec epochs of 8 Hz photic flash alternating with 30 sec epochs of darkness. Auditory stimulation consisted of 39 sec epochs of a spoken narrative alternating with 39 sec epochs of silence. There were five cycles of alternation

between activation (A) and baseline (B) conditions in the visual modality and 3.8 cycles of AB alternation in the auditory modality. These data were acquired using the 1.5T system at the Maudsley Hospital with the same acquisition parameters as for the null data. This experiment was designed to activate (at different frequencies) primary visual and auditory cortices.

Spatial working memory (3.0T AB)

One normal volunteer was studied using a blocked periodic design which alternated between 32 sec epochs of a spatial working memory task (A) and 32 sec epochs of visual fixation on a central crosshair (B). During each epoch of the spatial working memory condition, two or four elements of a 5×5 square grid were sequentially highlighted, each element being highlighted for 1,000 ms. Then the empty grid was shown for 2,000 or 4,000 ms before a single element of the grid was differently highlighted. The subject's task was to indicate by right-handed button press if the (blue) element highlighted after the delay was or was not the same as one of the series of (black) elements highlighted before the delay. A different series of elements was highlighted before the delay period in each trial. 256 T_2^*-weighted images were acquired at each of 17 non-contiguous slices in an oblique axial plane using the 3T Bruker system (Bruker Medical, Etlinger, Germany) at the Wolfson Brain Imaging Centre (WBIC), Cambridge, UK: TE = 40 ms, TR = 2 sec, in-plane resolution = 1.95×3.9 mm, slice thickness = 5.5 mm, number of excitations = 1. This experiment was designed to activate lateral prefrontal, medial premotor, and bilateral posterior parietal regions.

Sign language (3.0T ABAC)

One normal volunteer was studied using a blocked ABACABAC . . . design, each epoch lasting for 21 sec. During the A condition, the subject saw a static image of a person, "the signer;" during the B condition, the signer moved his hands to express a sentence using British Sign Language (BSL); during the C condition, the signer moved his hands to express single words in BSL. The subject's task was to look at these displays and understand the signs they were shown. 140 T_2^*-weighted images were acquired at each of 16 non-contiguous slices in an oblique axial plane using the 3T Varian system (Varian NMR Instruments, Palo Alto CA) at the Functional Magnetic Resonance Imaging of the Brain (FMRIB) Centre, Oxford, UK: TE = 30 ms, TR = 3 sec, in-plane resolution = 3 mm, slice thickness = 8 mm, number of excitations = 1. This experiment was designed to activate visual areas in the vicinity of the occipito-temporal junction that have previously been activated in studies of lip-reading [Calvert et al., 1997], which also demands extraction of semantic information from visual stimuli.

Motor attention and reaction (3.0T ER)

One normal volunteer was studied using a discrete trial or event-related (ER) design. Images were acquired continuously for 10 min, with the same acquisition parameters as for the sign language experiment, and the subject was asked to wait for a simple visual cue (a white circle on a black background) before immediately opposing the thumb and index finger of his right hand. Only five cues were presented in the course of the experiment (mean ISI = 2 min). This experiment was designed to activate systems for motor response selection, motor action, and visual cue perception.

Time domain models for the experimental effect

For all of these experimental data the same general approach to regression model specification was adopted [Friston et al., 1995]. An N-length vector was constructed to indicate which images were acquired during presentation of an activation condition (1) and which were acquired during presentation of a baseline condition (0). This vector was separately convolved with two Poisson kernels, parameterised by $\lambda = 4$ s or $\lambda = 8$ s. These two Poisson-convolved input functions were combined to form the $(N \times 2)$ design matrix $\mathbf{X}$. The design matrix was fit after motion correction [Bullmore et al., 1999b], linear detrending, and mean-zeroing of each observed fMRI time series.

Autoregressive pre-whitening

To implement model-based resampling in the time domain, we must inevitably assume some model for the endogenous autocorrelation. We chose to specify autoregressive AR(p) models of orders $p = 1$ and 3 for the colored component in the least squares residuals,

$$\rho = \mathbf{y} - \mathbf{X}\beta,$$

$$\rho_t = \sum_{i=1}^{p} \alpha_i \rho_{t-i} + \epsilon_t, \quad \epsilon \sim \mathcal{N}(0, \sigma^2 \mathbf{I}). \tag{4}$$

We have previously used an AR(1) model for autoregressive pre-whitening and shown that this was suf-

ficient to whiten periodically designed data acquired at 1.5T [Bullmore et al., 1996]. We expected that the AR(3) model might have some additional capacity to whiten any residual series not satisfactorily treated by the simpler form. It is predictable that even higher order AR processes might have even greater whitening properties but this will be at the cost of degrees of freedom in the prewhitened time series and the specification of very high order AR models in this context is not recommended [Marchini and Ripley, 2000].

The estimated autoregressive coefficients $\{\hat{\alpha}_i\}$ were used to pre-whiten the observed data and design matrix by the Cochrane-Orcutt procedure [Cochrane and Orcutt, 1949; Venables and Ripley, 1999]. This algorithm is computationally quicker than explicitly constructing the whitening matrix $\mathbf{K}^{-1}$ but achieves almost exactly the same result:

1. Fit the model $\mathbf{y} = \mathbf{X}\beta + \rho$ by least squares
2. Fit the AR(p) model, Eq 4, to the residuals ρ
3. Transform the data and each row of the design matrix $\mathbf{X}_t$ by the estimated AR coefficients

$$Y_t^* = Y_t - \sum_{i=1}^{p} \hat{\alpha}_i Y_{t-i}$$

$$\mathbf{X}_t^* = \mathbf{X}_t - \sum_{i=1}^{p} \hat{\alpha}_i \mathbf{X}_{t-i} \quad (5)$$

4. Fit the transformed model to the transformed data $\mathbf{y}^* = \mathbf{X}^*\beta^* + \epsilon$ to obtain a new set of regression parameter estimates $\hat{\beta}^*$ and their standard errors $\sigma_\epsilon^2(\mathbf{X}^{*T}\mathbf{X}^*)^{-1}$.
5. Iterate the AR modelling process by multiplying the original design matrix by the new parameter vector to generate a new set of residuals $\rho^* = \mathbf{y} - \mathbf{X}\beta^*$, which are returned to the algorithm at step 2. This loop is iterated, usually two to three times, until the sum of squared differences between consecutive estimates of the AR parameter vector is less than 10^{-5}.

For activation mapping, our test statistic S for an experimental effect is the sum of squared standardised regression coefficients β or $\{\beta_1, \beta_2\}$:

$$S = \left(\frac{\hat{\beta}_1}{\mathrm{SE}(\hat{\beta}_1)}\right)^2 + \left(\frac{\hat{\beta}_2}{\mathrm{SE}(\hat{\beta}_2)}\right)^2, \quad (6)$$

where the estimated parameters and their standard errors (SE) are those returned by the Cochrane-Orcutt algorithm at convergence (asterisks are omitted for notational convenience).

To sample the permutation distribution of S under the null hypothesis, we randomly resample the observed time series $\mathbf{y}$ without replacement to obtain a permuted time series $\tilde{\mathbf{y}}$ and estimate the same test statistic in the permuted data $\tilde{S}$ by exactly the same procedures as those used to estimate S in the observed data. Resampling is repeated an arbitrary number of times at each voxel and the resulting test statistics may be pooled over all intracerebral voxels in the image to sample the permutation distribution expeditiously.

However, as noted earlier, this resampling scheme is only valid if the residuals of the pre-whitened data $\{\epsilon_t\}$ are serially independent. We therefore estimated the Box-Pierce statistic B_Q in each set of residuals [Box and Pierce, 1970]:

$$B_Q = N_\epsilon \times \sum_{i=1}^{Q} \gamma_i^2. \quad (7)$$

where $\{\gamma_i\}$ $i = 1, 2, 3, \ldots, Q$ are the autocorrelation coefficients of the residual process ϵ up to an arbitrary maximum lag $Q = 10$. Under the null hypothesis that the residual process is white noise, $B_Q \sim \chi^2$ on $Q - p$ degrees of freedom. So if this null hypothesis was refuted with $P \leq 0.05$ at a given voxel then the observed time series was *not* resampled to contribute to the pooled permutation distribution and the observed statistic S was not tested; in short, that voxel was excluded from inferential consideration.

Resampling in the wavelet domain

Wavelets are families of orthogonal basis functions defined by dilation and translation of a mother wavelet ψ [Mallat, 1989],

$$\psi_{j,k}(t) = \frac{1}{\sqrt{2^j}} \psi\left(\frac{t - 2^j k}{2^j}\right), \quad (8)$$

and by dilation and translation of a father wavelet ϕ or scaling function,

$$\phi_{J,k}(t) = \frac{1}{\sqrt{2^J}} \phi\left(\frac{t - 2^J k}{2^J}\right), \quad (9)$$

where $j = 1, 2, 3, \ldots, J$ indexes the scale $S_j = 2^j = 2, 4, \ldots, 2^J$ or, approximately, the inverse frequency to which the mother or father wavelet has been dilated;

and $k = 1, \ldots, N/2^j$ indexes the location in time to which it has been translated. Note that the father wavelet is only dilated at the largest scale, meaning it represents the smoothest, lowest frequency components of the signal.

Assuming that the number of time points in the series is (or can be zero-padded to) a power of 2, the discrete wavelet transform of $\mathbf{y}$ will yield a vector of N wavelet coefficients $\mathbf{w}$, i.e., DWT($\mathbf{y}$) = $\mathbf{w}$. These coefficients can be understood as the weightings to be applied to the N locally supported basis elements or wavelets $\{\psi_{j,k}\}$, $\{\phi_{J,k}\}$ in recovering the signal by the inverse wavelet transform, i.e., IWT($\mathbf{w}$) = $\mathbf{y}$. Individual wavelet coefficients are often denoted $w_{j,k}$ where j and k index scale and time, as before. Wavelet decomposition therefore projects the signal onto a set of basis functions or a family of wavelets each of which is localised simultaneously in time and frequency. The Heisenberg uncertainty principle means that time and frequency cannot both be resolved with arbitrary precision, i.e., the product of temporal resolution Δt and frequency resolution Δf is lower-bounded, $\Delta t \Delta f \geq 1/4\pi$. Wavelets at the highest level of detail $j = 1$, $S_j = 2$ have the finest temporal resolution but the poorest scale resolution, whereas wavelets at the lowest level of detail $j = J$ have poor resolution in time, but good resolution of scale.

Wavelets have been widely used in various signal and image processing contexts since their mathematical development in the late 1980s, including many prior applications to image compression, non-parametric regression, and problems in brain mapping [Ruttimann et al., 1996, 1998; Brammer, 1998; Turkheimer et al., 1999; Raz et al., 1999; Samar et al., 1999]. There are also general introductory texts by Bruce and Gao [1996] and Mallat [1999].

From the present point of view, however, the single most important property of the discrete wavelet transform is that generally the correlation between the wavelet coefficients of a signal will be small even if the signal itself is highly autocorrelated in time. This is sometimes called the *whitening* or *decorrelating* property of the wavelet transform and it was first understood theoretically for the class of signals known as fractional Brownian motion [Flandrin, 1992; Tewfik and Kim, 1992; Dijkerman and Mazumdar, 1994]. Wornell [1993, 1996] later showed that wavelet decomposition also has optimally decorrelating or Karhunen-Loève properties for the wider class of $\frac{1}{f}$-like signals.

For fractional Brownian motion or other $\frac{1}{f}$-like noise we can thus more specifically say that the expected correlation between any two wavelet coefficients $w_{j,k}$ and $w_{j',k'}$

$$\langle w_{j,k}, w_{j',k'} \rangle \sim \mathbb{O}(|2^j k - 2^{j'} k'|^{2(H-R)}), \qquad (10)$$

or that the expected correlation between any two wavelet coefficients at the same scale $w_{j,k}$ and $w_{j,k'}$

$$\langle w_{j,k}, w_{j,k'} \rangle \sim \mathbb{O}(|k - k'|^{2(H-R)}), \qquad (11)$$

where $\langle \cdot \rangle$ denotes expectation, R is the number of vanishing moments of the mother wavelet and H is the Hurst exponent of the process; see Appendix. These results imply, as can be shown more explicitly [Flandrin, 1992], that the second order stochastic properties of wavelet coefficients are stationary within and between levels of detail (even if the data are non-stationary in the time domain). Remembering that the Hurst exponent for fBm cannot be ≥ 1, we can see that, provided the number of vanishing moments R is greater than 2, the correlation between any pair of wavelets will decay rapidly as an inverse power of their separation within and between levels of detail. More precisely, Dijkerman and Mazumdar [1994] showed that provided $R > 2H + 1$ the intercoefficient correlations will decay hyperbolically fast within levels and exponentially fast between levels. This suggests that the minimum number of vanishing moments required to decorrelate fractal noises with $H < 1$ will be 4.

It does not follow, however, that increasing the number of vanishing moments above the lower limit for effective decorrelation necessarily achieves greater decorrelation. This is because the number of vanishing moments of a wavelet is related to its local support in time, which will be at least of size $2R - 1$ [Daubechies, 1988]; and as the local support of a wavelet becomes less compact, the risk of artefactual intercoefficient correlations due to periodic boundary correction in computing the DWT of a finite time series becomes greater [Tewfik and Kim, 1992]. For this reason, it is recommended that the maximum scale of the decomposition 2^J should be set such that $N/2^{J-1} \geq 2R$ [Tewfik and Kim, 1992].

Bearing in mind that an fMRI time series is typically rather short ($N \sim 128$), and its intercoefficient correlational structure may therefore be especially sensitive to boundary correction artefacts, it seemed sensible to adopt the most compactly supported wavelet with four vanishing moments, which is the fourth-order Daubechies wavelet [Daubechies, 1988] supported over eight time points; see Figure 1. (Interestingly,

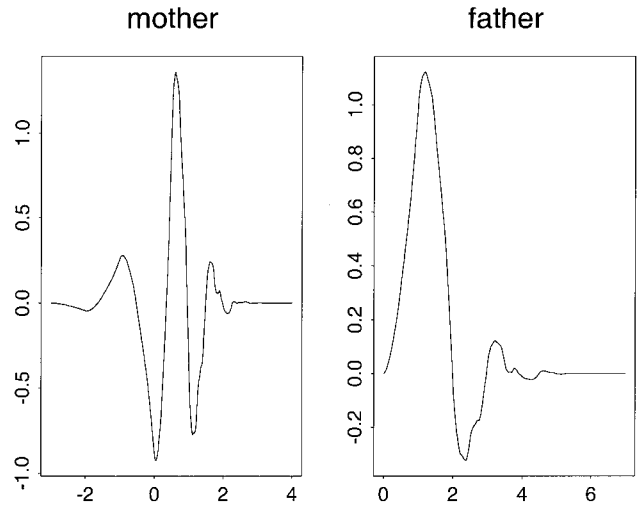

Figure 1.
Daubechies mother ψ and father ϕ wavelets with four vanishing moments and local support over eight time points. This is the basis adopted for resampling of time series in the wavelet domain because this number of vanishing moments is sufficient to decorrelate the wavelet coefficients of fractal noises with Hurst exponent $H < 1$; Daubechies wavelets have minimum local support for any given number of vanishing moments; and compactness of support minimises artefactual intercoefficient correlations due to boundary correction in finite time series.

Ruttimann et al. [1996] likewise found empirically that a wavelet with $R = 4$ vanishing moments was sufficient to decorrelate the wavelet transform of statistic maps derived from PET data [without explicitly assuming a fractal form for the noise]; although they chose to use the third order [cubic spline] Battle-Lemarié wavelet rather than the more compactly supported Daubechies wavelet.)

The algorithm

We apply these ideas to resampling of time series quite directly by the following algorithm:

1. Select the fourth order Daubechies wavelet, $R = 4$
2. Set $J = 5$, so that $N/2^{J-1} = 8 = 2R$ (assuming $N = 128$)
3. Take the discrete wavelet transform of the zero mean data, $\mathrm{DWT}(\mathbf{y}) = \mathbf{w}$
4. Randomly resample without replacement the wavelet coefficients $\mathbf{w} = \{w_{i,j}\}$ at each level of detail j to obtain a permuted coefficient vector $\tilde{\mathbf{w}}$
5. Take the inverse wavelet transform of the permuted coefficients $\mathrm{IWT}(\tilde{\mathbf{w}}) = \tilde{\mathbf{y}}$ to reconstruct the resampled data in the time domain.

For activation mapping, we identically computed the test statistic S by ordinary least squares in observed and resampled data alike. We resampled each time series in the image 10 times and pooled the resulting estimates of $\tilde{S}$ over all voxels to sample its permutation distribution.

Computational issues

Results on simulated data were obtained using S-PLUS [Venables and Ripley, 1999]. For activation mapping, the wavelet resampling algorithm was implemented in C [Press et al., 1992]. A pragmatic advantage of resampling in the wavelet domain is that the discrete wavelet transform is famously fast to compute by Mallat's pyramidal algorithm, which has complexity $\mathcal{O}(N)$, i.e., it is faster than fast Fourier transforms with complexity $\mathcal{O}(N \log N)$ [Mallat, 1999].

RESULTS

Simulated noises

Fractal or $\frac{1}{f}$-like noises have been extensively used in the movie industry and elsewhere to simulate natural landscapes and several algorithms for synthesis of so-called "fractal forgeries" [Voss, 1988] are described in the computer graphics and statistics literatures [Wornell, 1993; Stoksik et al., 1994; Abry and Sellan, 1996]. Here we use two methods of synthesis—one based on the statistical model of fBm and the other based on a physical process known to generate fractional noise.

Fractional Brownian motion

To synthesise a realisation of fBm with arbitrary H, a Gaussian white noise process ϵ is multiplied by the coloring matrix $\mathsf{K_B}$ given by factorisation (Cholesky decomposition) of the theoretically predicted covariance matrix $\Omega_B = \mathsf{K_B K_B^T}$ [Krueger et al., 1996]. The $\langle t, u \rangle$th element of Ω_B is defined by substituting t, u, and the variance of the Gaussian white noise process σ^2 in Eq. 17. The process is then realised (in vector notation) by

$$\mathbf{b} = \epsilon^T \mathsf{K}_B. \tag{12}$$

The expected covariance matrices Ω_B for $H = 0.5$ and $H = 0.25$ are shown in Figure 2, together with realisations of the corresponding fractional Brownian processes. The Hurst exponents empirically estimated

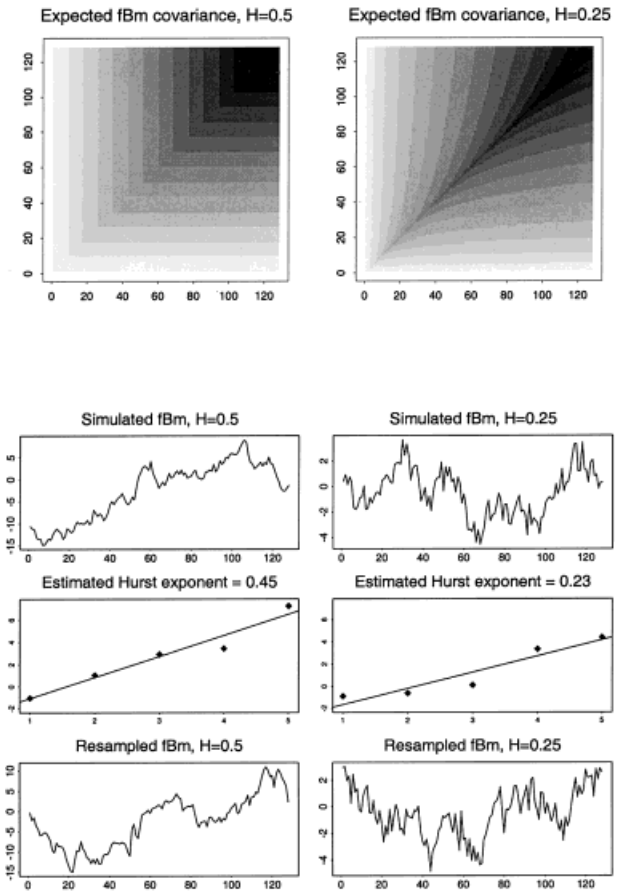

Figure 2.

Synthesis, analysis, and resampling of fractional Brownian motion (fBm). Top row: Expected autocovariance matrices Ω_B for fBm with Hurst exponents $H = 0.5$ and $H = 0.25$. Second row: Fractional Brownian processes ($N = 128$) realised by coloring a normal white process ϵ with the Cholesky factor of Ω_B, see Eq 12. Third row: Empirical estimation of Hurst exponents by least squares fit of Eq 21. The slope of the straight line drawn through the five points in each plot of level j (on the x-axis) vs. log variance of wavelet coefficients (on the y-axis) equals $2\hat{H} + 1$. Fourth row: fBm resampled by random permutation of the wavelet coefficients of the simulated series. The Hurst exponents estimated by Eq 21, $\hat{H} = 0.45$ and 0.23, are identical for the simulated and resampled series.

in these simulated data by fitting Eq. 21 were $\hat{H} = 0.45$ and 0.23, respectively. Also shown in Figure 2 are resampled processes obtained by random permutation of the wavelet coefficients of the simulated processes. The estimated Hurst exponents for the resampled series are exactly the same as for the simulated series. This is not surprising because this estimator of H depends only on the variance of the wavelet coefficients at a given scale, and this quantity is not changed by random permutation of wavelet coefficients separately within each scale. Therefore we can say that wavelet-estimated second order stochastic properties of fractional Brownian noise are perfectly preserved by this resampling scheme.

Parallel hyperbolic relaxation

Not all fractional or $\frac{1}{f}$-like noises represent fractional Brownian motion; indeed, in some respects, fractional Brownian motion is an awkward model for naturally fractal noises [Wornell, 1996]. To assess the performance of wavelet resampling in the broader class of $\frac{1}{f}$-like noises, we also simulated fractional noise by a physical model [Schroeder, 1991]. Superposition of many independent, parallel relaxation processes with hyperbolically distributed relaxation times $\{\tau\}$ can be shown theoretically to generate $\frac{1}{f}$-like behavior over several decades of frequency, and empirically this behavior has been demonstrated in many physical systems. It is simply simulated by generating a few, say three, relaxation processes $\{r_t\}$ $t = 1, 2, 4, \ldots, N$ and summing them. The discrete elements of each process are defined

$$r_{t+1} = \kappa r_t + \sqrt{1 - \kappa^2}\epsilon_t \qquad (13)$$

where $r_0 = 0$, ϵ_t is a random variable independently drawn from a standard normal distribution, and $\kappa = \exp(-1/\tau)$. Schroeder [1990] suggests using a set of three relaxation times that increase by a factor of 10, e.g., $\tau = 1, 10, 100$, giving correlations $\kappa = 0.37, 0.90, 0.97$.

An example of $\frac{1}{f}$-like noise simulated by this method is shown in Figure 3. Also shown in Figure 3 are the autocorrelation function and discrete wavelet transform of the simulated series, and the autocorrelation functions of the wavelet coefficients at each of the top three levels of detail. The whitening property of the DWT is evident. (The coefficients at lower levels of detail were also not significantly autocorrelated but these plots are not shown.) Random resampling of wavelet coefficients within levels, followed by the inverse wavelet transform, generates the resampled time series shown in the bottom row together with its autocorrelation function.

These plots suggest that wavelet resampling preserves the autocorrelational properties of physically simulated $\frac{1}{f}$ noise. To test this impression more thoroughly, we generated 50 realisations of this noise and calculated the autocorrelation function for each real-

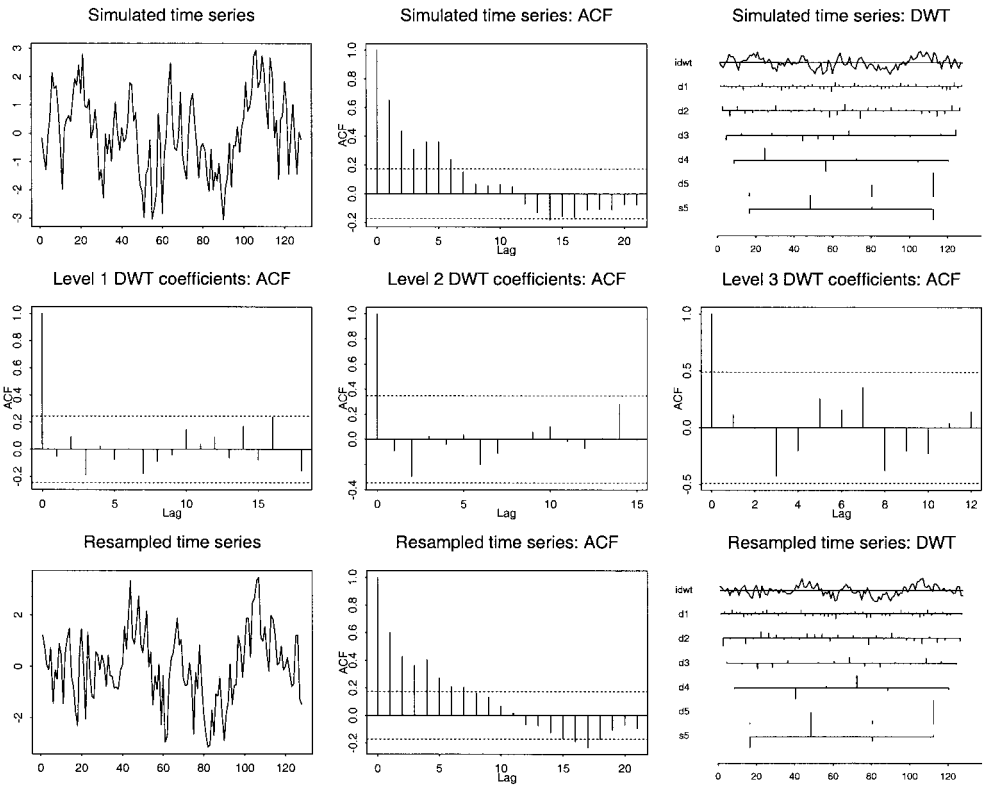

Figure 3.

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

isation to estimate the mean autocorrelation coefficient $\bar{\gamma}_i$ $i = 1, 2, 3, \ldots, 20$ at each of 20 lags and its standard deviation. This defined a 95% confidence envelope for the autocorrelation function of the *simulated* noise. We compared this to the mean and 95% confidence envelope of the autocorrelation function of the *resampled* noise. This required generating a single new realisation of the noise then resampling it 50 times in the wavelet domain and computing the autocorrelation function for each resampled series. The results are shown in Figure 4. It is clear that the mean and variability of the autocorrelation function over multiple realisations are very similar to the mean and variability of the autocorrelation function over multiple resamples of the same realisation.

Functional MRI: null data

Time series

A single fMRI time series measured during null or resting conditions at 1.5 Tesla is shown in Figure 5, together with its autocorrelation function, and the mean and 95% confidence envelope for the autocorrelation function estimated over 50 resamples of the observed series. It is clear that this fMRI series is less strongly autocorrelated than any of these simulated series (although fMRI time series can be much more strongly autocorrelated than this particular example); however, the observed autocorrelation function lies

within the 95% confidence envelope of the resampled autocorrelation function.

Activation mapping

Each of the 13 null datasets was analysed assuming a periodic input function with period length = 24 images and each of the V intracerebral voxels in the resulting statistic images was tested with a probability of type 1 error P corresponding to an expected number of false positive tests over the whole image $E = PV$, $1 \leq E \leq 200$; see Figure 6 and Table I.

For a valid test, the number of positive voxels observed when the null hypothesis (of zero periodic trend) is true, as presumably it is in these data, should be less than or equal to the expected number of positive voxels. By this criterion, all three methods are valid on average. However, there is more variability between images in the number of positive tests after AR(1) prewhitening and the observed number exceeds the expected number of positive tests in three images. As expected, AR(3) pre-whitening is a more conservative procedure and is valid for all images. The wavelet resampling scheme performs about as well as AR(3) pre-whitening and is valid for all images. Note that both AR(3) and wavelet resampling schemes are evidently somewhat over-conservative *on average;* but a scheme which was less conservative on average would likely allow uncontrolled error in at least some individual images, as indeed is the case with the AR(1) results shown here.

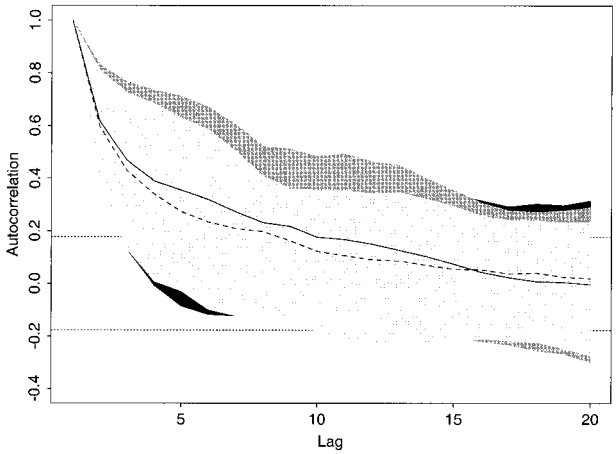

Figure 4.
Wavelet resampling of fractal noise simulated by multiple relaxation processes. The mean (dashed line) and 95% confidence envelope (black shading) for the autocorrelation function estimated over 50 realisations of $\frac{1}{f}$-like noise (N = 128) can be compared to the mean (solid line) and 95% confidence envelope (dark grey shading) for the autocorrelation function estimated over 50 resamples of a single simulated series. The area of the ACF envelopes common to both simulated and resampled series is shaded pale grey. Evidently the resampling scheme mimics closely the mean and variability of autocorrelational structure in the multiple simulated series.

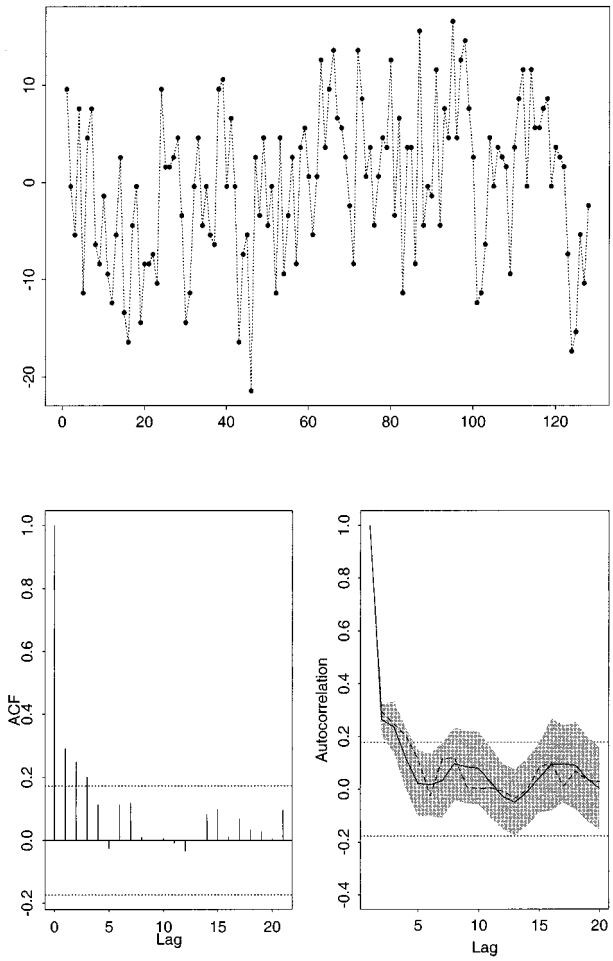

Figure 5.
Wavelet resampling of functional MRI noise. Top: A functional MRI time series observed under resting or null conditions at 1.5 Tesla (N = 128). Bottom row, left: Autocorrelation function of observed fMRI series. Bottom row, right: The observed (dashed line) autocorrelation function can be compared to the mean (solid line) and 95% confidence envelope (grey shading) for the autocorrelation function estimated over 50 resamples of the observed fMRI series. The observed autocorrelation function lies within the 95% confidence envelope estimated by (apparently unbiased) wavelet resampling.

Functional MRI: experimental data

Selected slices of the activation maps computed from the experimental data (acquired while subjects were asked to perform some sensory or cognitive task)

are shown in Figure 7. It is immediately clear that both autoregressive schemes entail some loss of inferential scope, since there are a number of voxels (colored blue) where the regression residuals have not been whitened by the Box-Pierce test. As expected, this problem is generally more severe for the AR(1) pre-whitening scheme than for the AR(3) pre-whitening scheme; and also relatively more severe for both schemes at 3.0T compared to 1.5T. In some slices, clumps of blue voxels can be seen to encroach on areas of activated (red) voxels, suggesting that one cause of color in regression residuals may be unmodelled experimental variance. However, in other slices of these maps of 3.0T data we have seen blue voxels concentrated around ventricular spaces and in the vicinity of susceptibility artefacts, suggesting that there may be many sources of potentially troublesome color at high field strength. It is also clear by inspection of the maps created from the ER and periodic ABAC studies at 3.0T that the problem of colored noise may be quite variable from image to image, even when they have been acquired using the same MR system.

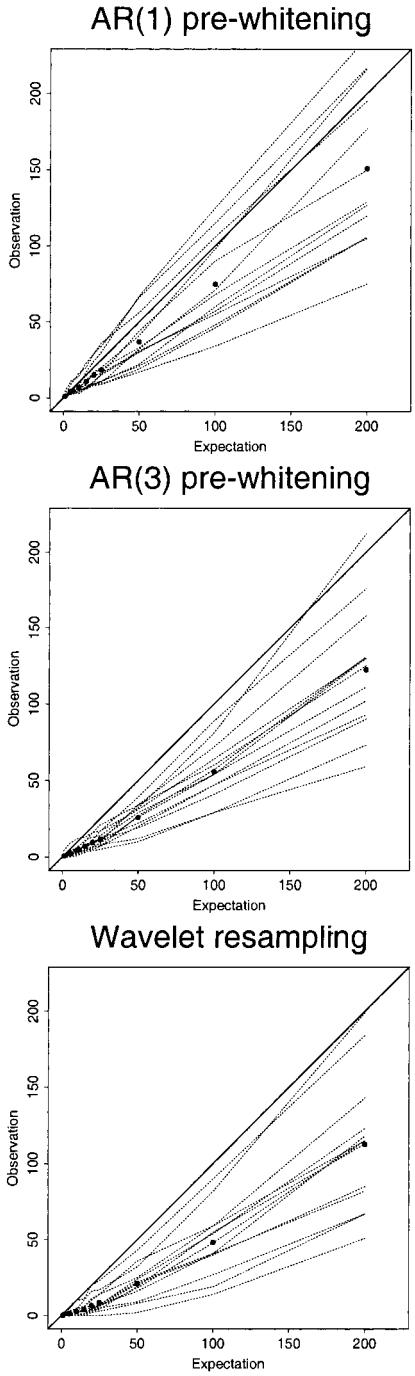

Figure 6.

TABLE I. Expected (E) and mean observed numbers of false positive tests for autoregressive pre-whitening and wavelet resampling schemes applied to activation mapping of 13 null fMRI datasets*

E	AR(1)	AR(3)	Wavelet
1	0.85 (1.21)	0.54 (1.13)	0.15 (0.55)
5	4.23 (2.71)	2.31 (2.32)	1.15 (1.28)
10	6.93 (3.86)	4.85 (2.73)	2.62 (2.84)
15	10.69 (5.75)	7.15 (2.91)	3.85 (3.72)
20	14.92 (7.77)	9.62 (3.78)	6.69 (5.91)
25	18.15 (9.44)	11.54 (4.67)	8.69 (6.43)
50	36.85 (17.07)	25.77 (8.76)	21.15 (11.43)
100	74.85 (28.98)	55.64 (18.03)	48.38 (27.07)
200	150.77 (52.1)	122.31 (42.12)	112.46 (44.05)

* Standard deviations are in parentheses.

The validity of inference by wavelet resampling is not conditional on the whiteness of the residuals; so all intracerebral voxels can be tested for an effect. The areas of activation demonstrated by the wavelet-based method are very similar to those demonstrated by the pre-whitening methods. There is no obvious evidence for differential sensitivity of the various methods to

Figure 6.

Type I error calibration curves for time domain resampling schemes based on AR(1) and AR(3) pre-whitening by the Cochrane-Orcutt procedure, and for wavelet domain resampling. In each plot of the observed number of positive tests vs. the expected number of positive tests, each of the dotted lines represents the results for one of 13 images acquired under null or resting conditions; the points represent the observed number of positive tests averaged over images; and the solid line $y = x$ indicates perfect agreement between observation and expectation. For a valid test, the observed number of positive tests under the null hypothesis must be less than or equal to the expected number. By this criterion all three tests are valid on average; but AR(3) pre-whitening and wavelet resampling are also valid for each of the individual images.

detect activation based on these results. However, any differential sensitivity of these and other methods of activation mapping is likely dependent on properties of both the noise and the design matrix. We are currently working on a more formal and comprehensive investigation of the relative sensitivities of these and other methods under diverse experimental conditions.

DISCUSSION

The wavelet resampling algorithm at the heart of this paper is simple. We take the discrete wavelet transform of a time series; randomly permute the wavelet coefficients within each level of detail; then reconstitute the data in the time domain by taking the inverse wavelet transform of the shuffled coefficients. To the best of our knowledge, this scheme has not previously been proposed as a basis for inference on time series although the whitening or decorrelating property of the wavelet transform, which underpins the validity of the resampling, has been best-established theoretically for fractal time series with $\frac{1}{f}$-like spectral properties.

We have comparatively evaluated the performance of this novel resampling scheme in functional MRI data analysis. First, we have calibrated the number of positive tests observed in analysis of 13 null datasets against the expected number. Wavelet resampling, like the more established alternative procedures based on autoregressive pre-whitening, was shown to support valid hypothesis testing in the context of residually autocorrelated regression errors. Such a demonstration of nominal type 1 error control is a necessary step in the validation of any candidate algorithm for hypothesis testing; but by this criterion alone there was little to choose between wavelet resampling and a Cochrane-Orcutt procedure modelling the noise as a third order autoregressive process. Nor was there a great deal of difference between these methods in analysis of a periodically designed 1.5T experiment. However, we also compared their performance in the apparently more exacting context of experimental fMRI data acquired at 3T. At the higher field strength, both pre-whitening procedures often failed adequately to pre-whiten the residuals. Since standardised statistics estimated by least squares in the context of colored residuals will be over-estimated, leading to a greater risk of type 1 error at truly unactivated voxels, we must exclude unwhitened voxels from inferential consideration. This meant that we were unable to make any decision about activation at very large numbers of voxels in some of the 3T images. Although this problem was predictably more severe for the simpler AR(1) pre-whitening scheme it also afflicted the maps supposedly pre-whitened by fitting an AR(3) model to the regression residuals. In contrast, the wavelet resampling scheme, which does not depend for its validity on the whiteness of the model residuals, could be used to infer activation at every voxel in all images. These results suggest a decisive advantage for wavelet resampling in terms of its robustness to cope with fMRI data acquired under various experimental conditions from two different high field MR systems (at Cambridge and Oxford). They also illustrate the value of continually re-evaluating statistical methods, such as AR(1) pre-whitening, as imaging technology evolves: what worked at 1.5T may not always be sufficient at 3T.

Although we have considered several methods for treating the problem of colored noise in fMRI, we have not investigated its sources or spatial distribution in any detail. We have suggested, on the basis of limited data, that colored noise may be especially problematic at higher field strength but this assertion needs to be tested and explained more carefully. If it is true that high field strength generally predisposes to problematically colored noise then what accounts for the variable severity of this problem, as illustrated by the three high-field datasets considered here? It would also be interesting to characterise resting and residual fMRI noise in greater detail, especially at high field strength, and in this respect we note an additional advantage of wavelet decomposition: it can be used conveniently to estimate the parameters of $\frac{1}{f}$-like noise. For example, one could use Eq. 21 or a more elaborate maximum likelihood estimator [Wornell, 1996] to produce a map of the Hurst exponent, spectral exponent or fractal dimension of the noise at each voxel of brain or phantom images. This analysis would likely shed light on the sources of colored noise; more fundamentally, it would also test our key assumption, based on the natural ubiquity of fractal processes and, more specifically, on prior work by Zarahn and colleagues [1997], that fMRI noise is $\frac{1}{f}$-like. Similarly, it would be useful to clarify to what extent fMRI noise empirically demonstrates nonstationary features and, therefore, to what extent wavelet resampling can be theoretically expected to cope better than time or frequency domain resampling schemes which assume stationarity of noise.

It will also be interesting in the future to compare the performance of wavelet resampling to pre-coloring strategies in the time domain as, for example, implemented in SPM (http://www.fil.ion.ucl.ac.uk/spm). One advantage in principle of wavelet resampling is that it can cope with nonstationary noise

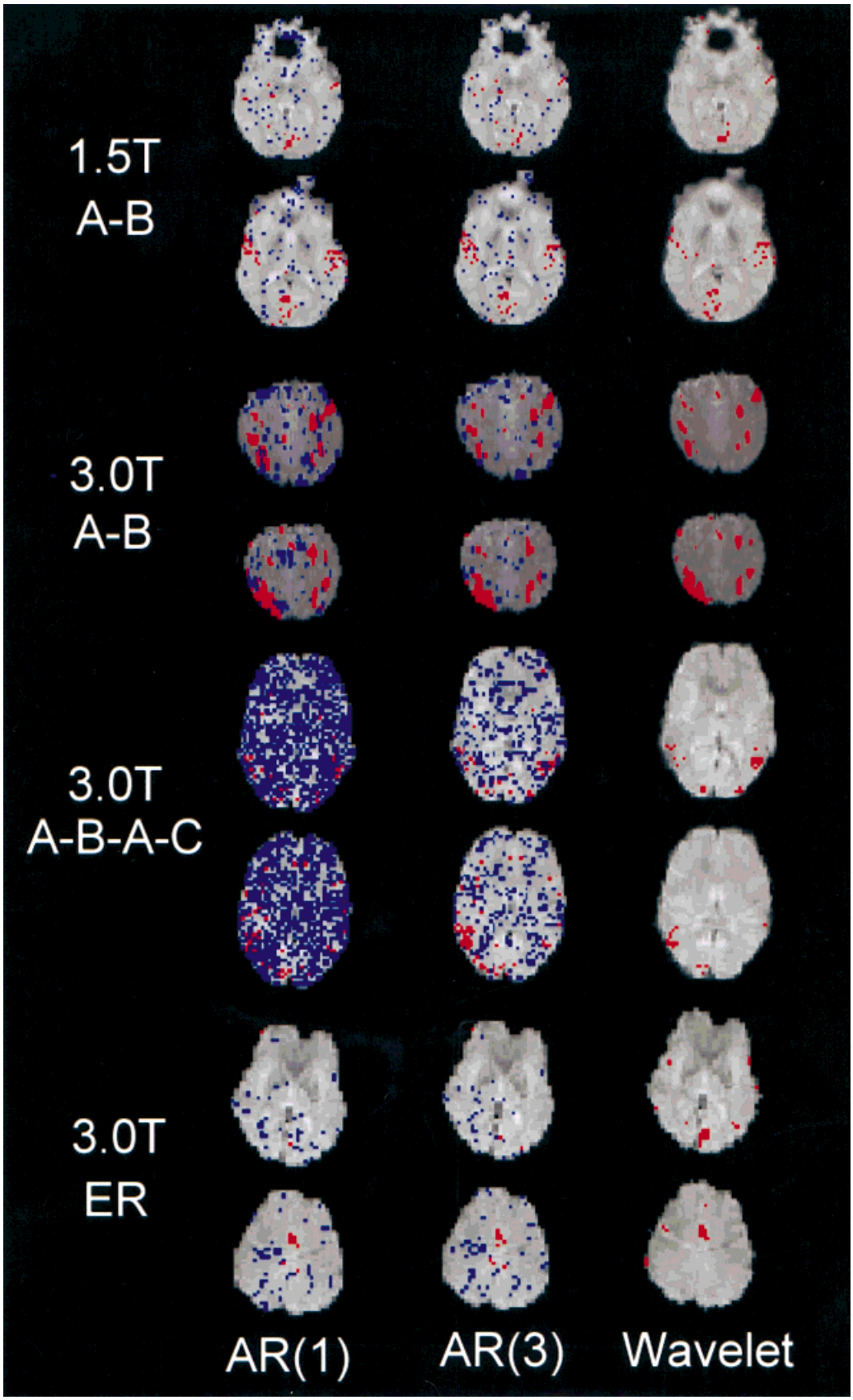

Figure 7.

structure; whereas pre-coloring (or pre-whitening) is limited to predicting (or estimating) a stationary or linear time invariant noise structure. Another potential advantage of wavelet resampling is that, unlike pre-coloring, it does not risk attenuating sensitivity to detect experimentally designed variance by temporal smoothing of the data [Hopfinger et al., 2000]. However, these preliminary observations are enough to suggest that the relative merits of various estimators and inferential procedures will be conditional on both the properties of the noise and the form of the design matrix. We shouldn't expect that one or other method will be unconditionally superior under all conceivable experimental conditions. It would clearly be very useful to have a more comprehensive understanding of the experimental conditions under which various methods perform well. But it is difficult to see how this collective understanding could arise without provision of an internationally shared library of fMRI datasets acquired under diverse experimental conditions and freely accessible over the internet to all. It must surely be a high priority for the brain mapping community to organise such a facility for itself [Koslow, 2000].

Every method has its limitations, of course, and wavelet resampling is no exception. First, the method is computationally intensive and therefore more time-consuming than referring test statistics to a theoretically-derived distribution. Traditionally, the tediousness of permutation testing has counted heavily against it [Fisher, 1936]; but the rising tide of cheap and powerful microprocessors has much eroded this objection [see Bullmore et al., 2001 for a brief historical review of resampling methods in brain mapping]. To put it in a more local and contemporary context, the central processing time costs of permutation testing are generally small, in our experience, compared to the costs of other widely-adopted image analysis processes, such as image realignment using sinc interpolation. And in this implementation specifically, the time costs of the procedure are minimised by using a fast pyramidal algorithm for the discrete wavelet transform and by pooling test statistics estimated on 10 permutations of each series over tens of thousands of voxels to sample their permutation distribution. Second, because the regression model parameters are estimated in the context of colored noise, the parameters will not be estimated as efficiently as they would be by an adequately pre-whitening scheme. Since we have shown that autoregressive pre-whitening is *not* always adequate to deal with the autocorrelational structure of fMRI noise this may seem rather a moot point. But there are other ways in which the data might be pre-whitened, including some wavelet-based methods for diagonalising the residual covariance matrix Ω and, if these were as robust as the wavelet resampling scheme we have described, their greater efficiency would be relatively advantageous. Third, there is the technically detailed question of minimising intercoefficient correlations. This requires that the wavelet basis has a minimum number of vanishing moments (four) and the most compact possible support, and that the time series is as long as possible to reduce the influence of artefactual intercoefficient correlations caused by boundary correction procedures. Here we have adopted the commonly-used periodic boundary correction algorithm but the risk of artefactual intercoefficient correlations might be reduced further by methods using non-convolutional filters at the edges of each level of detail [Cohen et al., 1993]. Fourth, it would be desirable to incorporate this particular resampling scheme in general algorithms for strong type 1 error control in the context of multiple comparisons made by computational inference [Westfall and Young, 1993; Holmes et al., 1996]. Finally, there remains the challenge of extending this method to incorporate spatial information in the neighborhood of each voxel. Spatially-informed test statistics are generally more sensitive than those informed solely by data at a single voxel [Poline et al., 1997] and spatial statistics can be readily tested by computational inference [Bullmore et al., 1999a]. However, to preserve the spatial correlational structure of the data under resampling in the time domain requires that the same set of permutations is applied to all time series in the image. It is not immediately clear what is the analogous method for preserving the spatial autocorrelations of the data under resampling in the wavelet domain.

Figure 7.

Brain activation mapping of experimental fMRI data based on different treatments of residual autocorrelation: AR(1) pre-whitening, middle row AR(3) pre-whitening, bottom row wavelet resampling. *1.5T AB* two-condition periodically designed data acquired at 1.5T; *3.0T AB* two-condition periodically designed data acquired at 3.0T (Cambridge); *3.0T ABAC* three-condition periodically designed data acquired at 3.0T (Oxford); *3.0T ER* event-related data acquired at 3.0T (Oxford). In all maps, red voxels indicate significant activation with voxel-wise probability of type I error $P < 1 \times 10^{-3}$; blue voxels indicate where the "pre-whitened" regression model residuals were not in fact serially independent by the Box-Pierce test. The key point to note is that wavelet resampling is considerably more robust than the pre-whitening schemes in dealing with some of the higher field datasets, but of approximately equivalent sensitivity.

CONCLUSIONS

We conclude that fMRI time series resampling in the wavelet domain provides a robust and valid alternative to methods of inferential brain mapping based on autoregressive pre-whitening in the time domain; and that the advantages of wavelet resampling were most salient in analysis of data acquired at 3T field strength. More generally, we note that this novel wavelet resampling scheme exactly preserves estimated second order properties of fractal noises and might therefore be useful for inference on a wide variety of other naturally complex time series.

ACKNOWLEDGMENTS

We thank Professors Steve Williams and John Pickard and Drs Emma Williams and Peter Jezzard and other colleagues in Cambridge, London and Oxford who assisted in collection of these imaging data. E.B. was supported by the Wellcome Trust. Null data collection at the Institute of Psychiatry, London, was partly supported by an experimental medicine research grant from SmithKline Beecham plc to E.B., F.Z., J.S., and others.

REFERENCES

Abry P, Sellan F (1996): The wavelet-based synthesis for fractional Brownian motion proposed by F. Sellan and Y. Meyer: Remarks and fast implementation. Appl Comput Harmonic Anal 3:377–383.

Aguirre GK, Zarahn E, D'Esposito M (1997): Empirical analyses of BOLD fMRI statistics II. Spatially smoothed data collected under null hypothesis and experimental conditions. NeuroImage 5:199–212.

Box GEP, Pierce DA (1970): Distribution of residual autocorrelations in autoregressive integrated moving average time series models. J Am Stat Assoc 65:1509–1526.

Brammer MJ (1998): Multidimensional wavelet analysis of functional magnetic resonance images. Hum Brain Mapp 6:378–382.

Bruce A, Gao H-Y (1996): Applied wavelet analysis with S-PLUS. New York, NY: Springer.

Bullmore ET, Brammer MJ, Bourlon P, Alarcon G, Polkey CE, Elwes R, Binnie CD (1994a): Fractal analysis of electroencephalographic signals intracerebrally recorded during 35 epileptic seizures: Evaluation of a new method for synoptic visualisation of ictal events. Electroencephalography Clin Neurophysiol 91:337–345.

Bullmore ET, Brammer MJ, Harvey I, Persaud R, Murray RM, Ron MA (1994b): 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 ET, Brammer MJ, Williams SCR, Rabe-Hesketh S, Janot N, David AS, Mellers JDC, Howard R, Sham P (1996): Statistical methods of estimation and inference for functional MR image analysis. Magn Reson Med 35:261–277.

Bullmore ET, Suckling J, Overmeyer S, Rabe-Hesketh S, Taylor E, Brammer MJ (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 Imag 18:32–42.

Bullmore ET, Brammer MJ, Rabe-Hesketh S, Curtis V, Williams SCR, Sharma T, McGuire PK (1999b): Methods for diagnosis and treatment of stimulus correlated motion in generic brain activation studies using fMRI. Hum Brain Mapp 7:38–48.

Bullmore ET, Suckling J, Brammer MJ (2001): In praise of tedious permutation. In: Spatial statistics: methodological aspects and some applications. Moore, M, editor. Lecture notes in statistics. New York: Springer, p 183–200.

Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SCR, McGuire PK, Woodruff PWR, Iversen SD, David AS (1977): Activation of auditory cortex during silent lipreading. Science 276:593–596.

Carlstein E, Do K-A, Hall P, Hesterberg T, Künsch HR (1998): Matched-block bootstrap for dependent data. Bernouilli 4:305–328.

Christofferson J (1997): A resampling method for regression models with serially correlated errors. Comput Stat Data Anal 25:43–53.

Cochrane D, Orcutt GH (1949): Application of least squares regression to relationships containing autocorrelated error terms. J Am Stat Assoc 44:32–61.

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

Dale AM (1999): Optimal experimental design for event-related fMRI. Hum Brain Mapp 8:109–114.

Daubechies I (1988): Orthonormal bases of compactly supported wavelets. Comm Pure Appl Math 41:909–996.

Davison AC, Hinkley DV (1997): Bootstrap Methods and their Application. Cambridge, UK: Cambridge University Press.

Dijkerman RW, Mazumdar RR (1994): On the correlation structure of the wavelet coefficients of fractional Brownian motion. IEEE Trans Info Theory 40:1609–1612.

Edgington ES (1995): Randomisation tests. New York: Marcel Dekker.

Efron B, Tibshirani RJ (1993): An introduction to the bootstrap. New York, NY: Chapman & Hall.

Fisher RA (1936): The coefficient of racial likeness and the future of craniometry. J Royal Anthropol Soc 66:57–63.

Flandrin P (1992): Wavelet analysis and synthesis of fractional Brownian motion. IEEE Trans Info Theory 38:910–917.

Friston KJ, Holmes AP, Poline J-B, Grasby PM, Williams SCR, Frackowiak RSJ, Turner R (1995): Analysis of fMRI time series revisited. NeuroImage 2:45–53.

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

Goldberger AL, Rigney DR, West BJ (1990): Chaos and fractals in human physiology. Sci Am 46:42–49.

Holmes AP, Blair RC, Watson JDG, Ford I (1996): Nonparametric analysis of statistic images from functional mapping experiments. J Cereb Blood Flow Metab 16:7–22.

Hopfinger JB, Büchel C, Holmes AP, Friston KJ (2000): A study of analysis parameters that influence the sensitivity of event-related fMRI analyses. NeuroImage 11:326–333.

Koslow, SH (2000): Should the neuroscience community make a paradigm shift to sharing primary data? Nature Neurosci 3:863–865.

Krueger WM, Jost SD, Rossi K, Axen U (1996): On synthesizing discrete fractional Brownian motion with applications to image processing. Graphical Models Image Proc 58:334–344.

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

Lehmann EL (1986): Testing statistical hypotheses. New York: John Wiley & Sons.

Locascio JL, Jennings PJ, Moore CI, Corkin S (1997): Time series analysis in the time domain and resampling methods for studies of functional magnetic resonance brain imaging. Hum Brain Mapp 5:168–193.

Mallat S (1989): A theory for multiresolution signal decomposition: The wavelet representation. IEEE Trans Pattern Anal Mach Intelligence 11:674–693.

Mallat S (1999): A wavelet tour of signal processing. San Diego, CA: Academic Press.

Mandelbrot BB (1977): The fractal geometry of nature. New York: W.H. Freeman and Co.

Mandelbrot BB, van Ness JW (1968): Fractional Brownian motions, fractional noises and applications. SIAM Rev 10:422–437.

Marchini JL, Ripley BD (2000): A new statistical approach to detecting significant activation in functional MRI. NeuroImage 12:366–380.

Ninness B (1998): Estimation of $\frac{1}{f}$ noise. IEEE Trans Info Theory 44:32–46.

Poline JB, Worsley KJ, Evans AC, Friston KJ (1997): Combining spatial extent and peak intensity to test for activations in functional imaging. NeuroImage 5:83–96.

Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1992): Numerical recipes in C: The art of scientific computing. Cambridge: Cambridge University Press.

Purdon PL, Weisskoff RM (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 UE, Unser M, Thevenaz P, Lee C, Rio D, Hommer DW (1996): Statistical analysis of image differences by wavelet decomposition. In: Aldroubi A, Unser M, editors. Wavelets in medicine and biology. Boca Raton, FL: CRC Press. p 115–144.

Ruttimann UE, Unser M, Rawlings RR, Rio D, Ramsey NF, Mattay VS, Hommer DW, Frank JA, Weinberger DR (1998): Statistical analysis of functional MRI data in the wavelet domain. IEEE Trans Med Imaging 17:142–154.

Samar VJ, Bopardikar A, Rao R, Swartz K (1999): Wavelet analysis of neuroelectric waveforms: A conceptual tutorial. Brain Lang 66:7–60.

Schroeder M (1991): Fractals, chaos, power laws: Minutes from an infinite paradise. New York: W.H. Freeman and Co.

Smith AM, Lewis BK, Ruttimann UE, Ye FQ, Sinnwell TM, Yang Y, Duyn JH, Frank JA (1999): Investigation of low frequency drift in fMRI signal. NeuroImage 9:526–533.

Stoksik MA, Lane RG, Nguyen DT (1994): Accurate synthesis of fractional Brownian motion using wavelets. Electron Lett 30: 383–384.

Tewfik AH, Kim M (1992): Correlation structure of the discrete wavelet coefficients of fractional Brownian motion. IEEE Trans Info Theory 38:904–909.

Turkheimer FE, Brett M, Visvikis D, Cunningham VJ (1999): Multiresolution analysis of emission tomography images in the wavelet domain. J Cereb Blood Flow Metabol 19:1189–1208.

Venables WD, Ripley BD (1999): Modern applied statistics with S-PLUS. New York: Springer Verlag.

Voss RF (1988): Fractals in nature: From characterisation to simulation. In: Peitgen H-O, Saupe D, editors. The science of fractal images. New York, NY: Springer-Verlag. p 21–70.

Westfall PH, Young SS (1993): Resampling-based multiple testing. New York: John Wiley & Sons.

Wornell GW (1993): Wavelet-based representations for the $\frac{1}{f}$ family of fractal processes. Proc IEEE 81:1428–1450.

Wornell GW (1996): Signal processing with fractals: A wavelet-based approach. Upper Saddle River, NJ: Prentice Hall.

Worsley KJ, Friston KJ (1995): Analysis of fMRI time series revisited—again. NeuroImage 2:173–181.

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

APPENDIX

Fractals and fractional Brownian motion

Fractal is a word coined by Mandelbrot [1977] from the Latin adjective *fractus* meaning broken or fractured. Fractals are formally irregular, rough, non-differentiable, and often have non-integer (fractal) dimensions. A key property of fractals is that they should look more or less the same under magnification, i.e., they are at least approximately self-similar. A familiar example from nature is that magnifying one of the branches of a tree can make it look approximately the same as the whole tree, and one of its twigs may look approximately the same after magnification as the branch. Historically, some mathematically-generated fractals have been considered monstrous or pathological because of their inaccessibility to treatment by classical methods such as calculus and Euclidean geometry. More recently, it has been recognised that fractals are naturally far more common than the canonical forms they offend; indeed, Mandelbrot [1977] has claimed that fractal geometry is the geometry of nature. In the human brain, fractal properties have been demonstrated for both anatomical structures and physiological processes [Bullmore et al., 1994a,b].

Fractional Brownian motions (fBm) or processes are a class of fractals described by Mandelbrot and van Ness [1968] as a generalisation of the special case of ordinary Brownian motion (as defined by Kolmogorov in the 1940s). A fractional Brownian motion $B(t)$ is a zero-mean, nonstationary, and nondifferentiable function of time such that the mean square difference in value of the function between two time points is proportional to the time difference Δ raised to the power of two times the Hurst exponent H which has value $0 < H < 1$; i.e.,

$$\langle |B(t) - B(t - \Delta)|^2 \rangle \approx |\Delta|^{2H}. \qquad (14)$$

From this definition several properties of fBm can be deduced. Setting $\Delta = 1$, we can see that the mean square or variance of the fractional Gaussian noise or increments process $I(t) = B(t) - B(t-1)$ is stationary,

$$\langle I(t)^2 \rangle \approx 1^{2H}. \tag{15}$$

Setting $\Delta = t$, we confirm that the variance of the process itself is nonstationary,

$$\langle B(t)^2 \rangle \approx t^{2H}. \tag{16}$$

And setting $\Delta = t - u$, we can define the expected (non-stationary) covariance between the process at two times t and u,

$$\langle B(t), B(u) \rangle = \frac{\sigma^2}{2}(|t|^{2H} + |u|^{2H} - |t-u|^{2H}),$$

$$\frac{\sigma^2}{2} = \frac{\Gamma(2-2H)\cos(\pi H)}{2\pi H(1-2H)}. \tag{17}$$

Finally, it can be shown from Eq 17 that $B(t)$ is self-affine, i.e., rescaling the process in time by an arbitrary scalar $s > 0$ yields a series $B(st)$ with identical statistical properties to the process $s^H B(t)$ generated by rescaling its values on the original time scale by the scalar s^H. And from this self-affine property it follows [Voss, 1988] that the fractal dimension D of the process is simply related to its Hurst exponent:

$$D = T + 1 - H, \tag{18}$$

where T is the topological dimension of the data. For a time series, $T = 1$; therefore $D = 2 - H$.

Thus if the Hurst exponent $H \to 0$, the process is highly irregular or antipersistent, $D \to 2$; whereas if $H \to 1$, the fractal dimension also approaches unity and the process looks smoother or persistent. Ordinary Brownian motion is the special case of fBm occurring when the increments are normal, $H = \frac{1}{2}$ and $D = 1.5$.

Estimation of $\frac{1}{f}$-like noise parameters

Since fBm is non-stationary its second order properties are not properly described by a power spectrum or autocorrelation function; although Mandelbrot and van Ness [1968] introduced the notion of a "generalised" power spectrum for $B(t)$, $S_f(B(t)) \approx f^{-\alpha}$, $\alpha = 2H + 1$. And Wornell [1993, 1996] later showed that any $\frac{1}{f}$-like process F passed by an ideal bandpass filter was thereby rendered stationary with spectrum

$$S_f(F) = \frac{\sigma_F^2}{|f|^\alpha}. \tag{19}$$

Since any real, discrete time series *has* effectively been bandpass filtered between the limiting frequencies imposed by the sampling rate and duration of the finite series, this suggests that $\alpha = 2H + 1$ may appropriately be estimated from the Fourier transform of the series, e.g., by finding the gradient of a straight line fitted to log power vs. log frequency.

However, it is also possible to estimate H (and thereby α and D) from the wavelet transform of the time series. If a $\frac{1}{f}$-like noise is projected onto a wavelet basis with R vanishing moments then, provided $0 < (2H + 1) < 2R$, the variance of the wavelet coefficients at the jth scale, $\mathrm{Var}\{w_k\}_j$, is related to the scale by a power law with exponent $2H + 1$ [Flandrin, 1992], i.e.,

$$\mathrm{Var}\{w_k\}_j \approx (2^j)^{2H+1}. \tag{20}$$

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

$$\log_2(\mathrm{Var}\{w_k\}_j) = c + (2H + 1)j + \epsilon_j. \tag{21}$$