Temporal Autocorrelation in Univariate Linear Modelling of FMRI Data

FMRIB Technical Report TR01MW1
(A related paper has been accepted for publication in NeuroImage)

Mark W. Woolrich[1,2], Brian D. Ripley[3], Michael Brady[2] and Stephen M. Smith[1]

1: Oxford Centre for Functional Magnetic Resonance Imaging of the Brain (FMRIB),
Department of Clinical Neurology, University of Oxford, John Radcliffe Hospital,
Headley Way, Headington, Oxford, UK
2: Medical Vision Laboratory, Department of Engineering Science,
University of Oxford, Oxford, UK
3: Department of Statistics, University of Oxford, Oxford, UK
Corresponding author — Mark Woolrich: woolrich@fmrib.ox.ac.uk

Abstract

In functional magnetic resonance imaging (FMRI) statistical analysis there are problems with accounting for temporal autocorrelations when assessing change within voxels. Techniques to date have utilised temporal filtering strategies to either shape these autocorrelations, or remove them. Shaping, or "colouring", attempts to negate the effects of not accurately knowing the intrinsic autocorrelations by imposing known autocorrelation via temporal filtering. Removing the autocorrelation, or "prewhitening" gives the best linear unbiased estimator, assuming that the autocorrelation is accurately known. For single-event designs, the efficiency of the estimator is considerably higher for prewhitening when compared with colouring. However, it has been suggested that sufficiently accurate estimates of the autocorrelation are currently not available to give prewhitening acceptable bias. To overcome this, we consider different ways to estimate the autocorrelation for use in prewhitening. Having performed high-pass filtering, a Tukey taper (set to smooth the spectral density more than would normally be used in spectral density estimation) performs best. Importantly, estimation is further improved by using nonlinear spatial filtering to smooth the estimated autocorrelation, but only within tissue type. Using this approach when prewhitening reduced bias to to close to zero at probability levels as low as 1×10^{-5}.

1 Introduction

In this paper we focus on the issues surrounding temporal autocorrelations in FMRI time series. These include understanding the nature of the autocorrelation, the effects of temporal filtering, the effect of different experimental designs and ways of performing efficient and accurate statistical tests. In particular, the aim is to deduce an autocorrelation estimation technique which gives acceptably low bias when prewhitening.

We start with a brief overview of previous work in the area. We then set up a familiar GLM framework to define the different strategies for dealing with autocorrelations in FMRI. Four different approaches to temporal autocorrelation estimation are considered and a qualitative data analysis is used to examine the way in which the estimated autocorrelation varies spatially, the effects of temporal filtering, and the effect of different design types. We then introduce nonlinear spatial smoothing of the autocorrelation as a means to improving the estimation further. Finally, quantitative assessment of the bias (calibration) for the different autocorrelation estimation techniques is performed, by computing null distributions from null/rest data and comparing them with the expected theoretical distributions.

2 Overview of Previous Work

Friston et al. (2000) suggested that current techniques for estimating the autocorrelation (Autoregressive(AR) and $1/f$ models where f is the frequency) are not accurate enough to give prewhitening acceptable bias. Therefore, estimation is made more robust to inaccurate autocorrelation estimates by swamping any intrinsic autocorrelation with the known autocorrelation introduced by band pass filtering, an approach often referred to as "colouring" (Friston et al., 1995), (Worsley and Friston, 1995). For this they use a Gaussian (or similar) low-pass filter matched to the haemodynamic response function (HRF) and a linear high-pass filter which aims to remove the majority of the autocorrelation due to low frequency components. Having shaped the autocorrelation, prewhitening is then not applicable, and the autocorrelation estimate is instead used to correct the variance of univariate linear model parameter estimates and the degrees of freedom used in the GLM.

Although colouring is unbiased, given an accurate autocorrelation estimate, Bullmore et al. (1996) noted the need for serially independent (whitened) residuals to obtain the Best Linear Unbiased Estimates (BLUE) of the GLM parameters. The parameter estimates are "best" in the sense that they are the unbiased estimates with the lowest variance. This was achieved using Pseudo-Generalised Least Squares (PGLS) - also known as the Cochrane-Orcutt transformation. The autocorrelation is estimated for the residuals from a first linear model and is then used to "prewhiten" the data and the design matrix, for use in a second linear model. The residuals of the second linear model should be close to white noise. Further iterations of this process are possible.

For inference, they ascertain the null distributions using randomisation; that is they randomly reorganise the order of signal intensity values in each observed time series and estimate the test statistic for each. It is important to note that such randomisation of the time series is only valid in the absence of autocorrelations, hence an accurate prewhitening step is necessary. Any randomisation approach on non-white data needs to randomise the data in such a way that the effective structure of the autocorrelation is maintained, for the null distribution to be valid.

To model the autocorrelation, Bullmore proposed an AR model of order 1 (AR(1)), which was shown to model the autocorrelations satisfactorily for the data used in their paper. Purdon and Weisskoff (1998) also suggested using an AR(1) model to do prewhitening, but they also included a white noise component. However, the main focus of their paper was to explore the effect, on the false positive rate, of not taking into account the temporal autocorrelation. For a desired false positive rate of $\alpha = 0.05$ they find false positive rates as high as $\alpha = 0.16$ in the uncorrected data. At $\alpha = 0.02$ the situation worsens further, with uncorrected data giving $\alpha = 0.095$. This is because any inaccuracies in the distribution compared with the assumed theoretical distribution are more prominent further down the tail of the distribution.

Locascio et al. (1997) used an Autoregressive Moving Average (ARMA) model (see also Chatfield (1996)) and incorporated it into an overall Contrast Autoregressive and Moving Average (CARMA) model. As well as the ARMA and modelled experimental responses, the CARMA model contains baseline, linear and quadratic terms for the removal of low frequency drift. They fit separate MA and AR models of up to order 3.

Locascio et al. (1997) also suggests that the existence of positive autocorrelation is due to carry over from one time point to the next, stemming from time intervals that are smaller than the actual temporal changes. They describe autocorrelation as the persistence of neuronal activation, cyclical events (presumably they are referring to aliased cardiac and respiratory cycles), or possibly characteristics or artefacts of the measurement process.

Zarahn et al. (1997) and Aguirre et al. (1997) observed $1/f$ noise profiles in FMRI data, and as a result attempted to use a $1/f$ noise model with three parameters to account for temporal autocorrelations. They also carried out a number of water phantom studies to establish how much, if any, of the $1/f$ noise is attributable to physiological processes. They concluded that the same $1/f$ noise was apparent in the phantoms and that therefore the noise was not of physiological origin.

They use their $1/f$ model of the intrinsic autocorrelation together with Worsley and Friston (1995)'s approach of colouring the data. Zarahn proposed to refine this by incorporating the $1/f$ model's representation of the intrinsic autocorrelation into the autocorrelation due to temporal filtering, in order to give a better estimate of the autocorrelation post-temporal filtering. Unfortunately, their attempts fail because they fit the $1/f$ model over the entire brain volume, therefore ignoring the potential for spatial non-stationarity of the noise profile.

Hu et al. (1995) concentrate on the components of the coloured noise in FMRI data that are due to physiological fluctuations. By recording respiration and cardiac cycle data at the same time as the FMRI data is acquired, some of these effects can be removed. This could help to reduce the autocorrelation and potentially improve any autocorrelation estimation subsequently employed. However, these are not the only causes of coloured noise in the data and hence whether or not respiration and cardiac cycle data is available, robust strategies for dealing with autocorrelation are still required.

3 Methods

3.1 GLM framework

In the basic GLM, $\mathbf{Y} = \mathbf{XB} + \mathbf{e}$, $\mathbf{Y}$ is the observed data, $\mathbf{X}$ is the matrix of "regressors" (often referred to as the design matrix) and $\mathbf{B}$ are the parameters to be estimated. The errors $\mathbf{e}$ are assumed to have a Normal distribution $N(0, \sigma^2\mathbf{V})$, where $\mathbf{V}$ is the autocorrelation matrix for the time series. There exists (Seber, 1977) a square, nonsingular matrix $\mathbf{K}$ such that $\mathbf{V} = \mathbf{KK^T}$, and that $\mathbf{e} = \mathbf{K}\epsilon$ where ϵ are $N(0, \sigma^2\mathbf{I})$.

Now consider a GLM which incorporates temporal filtering of the data, where $\mathbf{S}$ is the square matrix that performs the temporal filtering via matrix multiplication. $\mathbf{S}$ is a Toeplitz matrix produced from the impulse response; this is directly equivalent to convolving with the impulse response using zero padding. The design matrix is also temporally filtered using $\mathbf{S}$ to reflect the known change in the observed data. We now have:

$$\mathbf{SY} = \mathbf{SXB} + \eta \tag{1}$$

where η is $N(0, \sigma^2\mathbf{SVS^T})$. We use an ordinary least squares (OLS) estimate of $\mathbf{B}$, given by:

$$\hat{\mathbf{B}} = (\mathbf{SX})^+\mathbf{SY} \tag{2}$$

where $(\mathbf{SX})^+$ is the pseudo-inverse of $(\mathbf{SX})$ given by $(\mathbf{SX})^+ = ((SX)^T SX)^{-1}(SX)^T$. The variance of a contrast $\mathbf{c}$, of these parameter estimates, $\hat{\mathbf{B}}$, is given by:

$$\begin{aligned} &Var\{\mathbf{c^T\hat{B}}\} = k_{eff}\sigma^2 \\ &k_{eff} = \mathbf{c^T}(\mathbf{SX})^+\mathbf{SVS^T}((\mathbf{SX})^+)^\mathbf{T}\mathbf{c} \end{aligned} \tag{3}$$

Note that k_{eff} is a scalar that scales σ^2 by an amount that depends upon the design matrix $\mathbf{X}$, the temporal autocorrelation $\mathbf{V}$ and the contrast $\mathbf{c}$ to give the variance of the contrast of parameter estimates. For an estimate of σ^2 we use (Worsley and Friston, 1995), (Seber, 1977):

$$\hat{\sigma}^2 = \eta^\mathbf{T}\eta/\mathbf{trace}(\mathbf{RSVS^T}) \tag{4}$$

where $\mathbf{R} = \mathbf{I} - \mathbf{SX}(\mathbf{SX})^+$, the residual forming matrix, which can be used to obtain the residuals of the model fit:

$$\mathbf{r} = \mathbf{RSY} \tag{5}$$

3.2 Strategies for Dealing with Autocorrelation

For the moment we assume a known autocorrelation matrix $\mathbf{V} = \mathbf{KK^T}$. We investigate three approaches to dealing with the autocorrelation in FMRI; these are:

- Colouring (e.g. Friston et al. (1995)), with $\mathbf{S} = \mathbf{A}$ where $\mathbf{A}$ is a low-pass filter, giving:

$$k_{eff} = \mathbf{c^T}(\mathbf{AX})^+\mathbf{AVA^T}((\mathbf{AX})^+)^\mathbf{T}\mathbf{c} \tag{6}$$

- Variance Correction (where one corrects the statistics for autocorrelation in the data, but neither colours the data nor goes as far as prewhitening), with $\mathbf{S} = \mathbf{I}$ giving:

$$k_{eff} = \mathbf{c^T}(\mathbf{X})^+\mathbf{V}((\mathbf{X})^+)^\mathbf{T}\mathbf{c} \tag{7}$$

- Prewhitening (e.g. Bullmore et al. (1996)) gives the optimal BLUE, and is obtained by setting $\mathbf{S} = \mathbf{K^{-1}}$ giving:

$$k_{eff} = \mathbf{c^T}(\mathbf{X^T V^{-1} X})^{-1}\mathbf{c} \tag{8}$$

For all of these approaches we need an accurate estimate of the autocorrelation matrix $\mathbf{SVS^T}$. Techniques for calculating such an estimate are discussed in the next section.

3.3 High-pass filtering

The data to be considered is the FMRI time series at each voxel following motion correction. The raw motion-corrected time series have a considerably coloured noise structure, the majority of which occurs at low frequency. Therefore, in this paper our approach is to perform high-pass filtering to remove the worst of the low frequency components. This is also beneficial since it is the low frequency deterministic trends in the time series which contribute most to violating an assumption of second-order stationarity.

High-pass filtering can be performed by incorporating such things as a discrete cosine transform set (DCT) into the design matrix $\mathbf{X}$ or into the matrix $\mathbf{S}$ (Friston et al., 2000). However, such techniques produce large end-effects and so we prefer to use a non-linear filter as proposed by Marchini and Ripley (2000). This approach fits and removes Gaussian-weighted running-lines of fixed width using a least squares fit and was found to be a reliable method of trend removal in Marchini and Ripley (2000). As in Marchini and Ripley (2000), the width of the Gaussian is chosen to be twice the cycle length when using boxcar or single event with fixed inter-stimulus interval(ISI)(Bandettini and Cox, 2000) designs. However, for randomised ISI single event designs (Burock et al. (1998), Dale and Buckner (1997) and Dale (1999)) the situation is not as clear. This is because the signal contains power at virtually all frequencies (see figure 10(b)). Hence, a compromise is used by setting the full-width half-maximum (FWHM) to 45 scans. This removes the worst of the low frequency trends, allowing sensible autocorrelation modelling, whilst removing negligible power from the signal. Such nonlinear high-pass filtering is performed as a preprocessing step on all data sets subsequently used in this paper.

3.4 Autocorrelation Estimation

An estimate of the autocorrelation matrix $\mathbf{SVS^T}$ of the error η is required. We could estimate $\mathbf{B}$ using equation 2 to obtain the residuals $\mathbf{r}$ and then estimate the autocorrelation matrix of the residuals. However, it can be shown that:

$$\mathbf{r} = \mathbf{RSY} = \mathbf{R}\eta \tag{9}$$

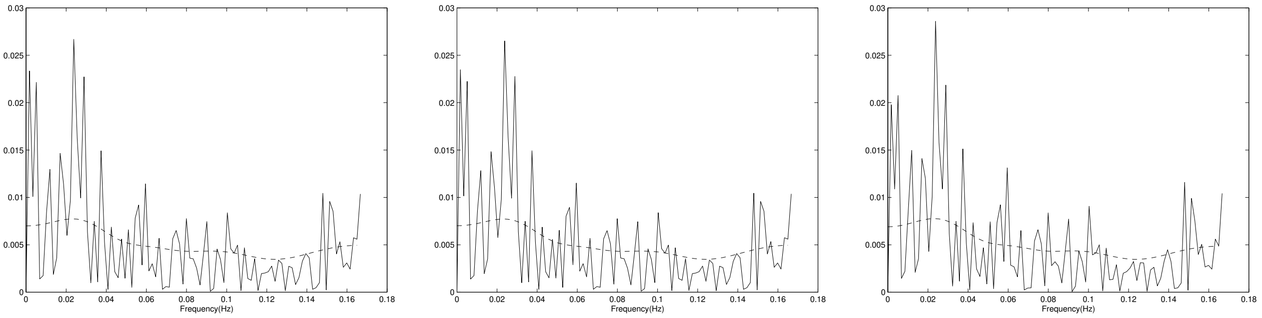
and then it follows that the autocovariance of the residuals is given by:

$$Cov\{\mathbf{r}\} = \sigma^2 \mathbf{RSVS^T R^T} \tag{10}$$

which is *not* $\mathbf{SVS^T}$. The difference between the autocorrelation of the error and the autocorrelation of the residuals is clearly due to regression onto the design matrix and it is impossible to unravel $\mathbf{SVS^T}$ from $\mathbf{RSVS^T R^T}$ since $\mathbf{R}$ is noninvertible. This turns out to be not too much of a problem, as can be illustrated using an example time series generated artificially using the noise process $N(0, \sigma_t V_t)$, where V_t is the autocorrelation matrix for the typical grey matter autocorrelation in figure 5(a). Figure 1(a) shows the spectral density of the time series, figure 1(b) shows the spectral density of the residuals $\mathbf{r}$ for the same time series after regression onto the HRF convolved boxcar (shown in figure 7), and figure 1(c) shows the same after regression onto the randomised ISI design (shown in figure 10). Despite some subtle differences in the raw spectral density estimates the Tukey estimates are remarkably similar. This is perhaps surprising, particularly for the randomised ISI design, which covers a large frequency range (see figure10(b)). The primary reason for this is that the spectral density will only be affected at the frequency *and* phase of the regressors. Secondly, when the regressor has high power at a particular frequency but not at its neighbouring frequencies (this is less true for the randomised ISI design but still has some effect), then spectral density estimation techniques which heavily smooth the spectral density will help rectify this problem further. All techniques considered in this paper do effectively smooth the spectral density.

For variance correction or colouring, an estimate of $\mathbf{SVS^T}$ can be calculated from the residuals after equation 2 is used to obtain the parameter estimates. This estimate of $\mathbf{SVS^T}$ is used in equation 3 to give the variance of the parameter estimates.

However, prewhitening requires an estimate of $\mathbf{SVS^T}$ *before* the BLUE can be computed and equation 3 used. To get round this an iterative procedure is used (Bullmore et al., 1996). Firstly, we obtain the residuals $\mathbf{r}$ using a GLM with $\mathbf{S} = \mathbf{I}$. The autocorrelation $\mathbf{V}$ is then estimated for these residuals. Given an estimate of $\mathbf{V}$, $\mathbf{V^{-1}}$ and hence $\mathbf{K^{-1}}$ can be obtained by inverting in the spectral domain (some autocorrelation models, e.g. autoregressive, have simple parametrised forms for $\mathbf{K^{-1}}$, and hence inversion in the spectral domain is not necessary). Next, we use a second linear model with $\mathbf{S} = \mathbf{K^{-1}}$, and the process can then be repeated to obtain new residuals from which $\mathbf{V}$ can be re-estimated and so on. We use just one iteration and find that it performs sufficiently well in practice. Further iterations either give no further benefit or cause over-fitting, depending upon the autocorrelation



(a) (b) (c)

Figure 1: (a) Shows the spectral density of an artificially generated time series using the noise process $N(0, \sigma_t V_t)$, (b) and (c) show the spectral densities of the residuals $\mathbf{r}$ for the same time series after regression onto the HRF convolved boxcar shown in figure 7 and the randomised ISI shown in figure 10 respectively. Raw spectral density estimates are shown by the solid line and estimation using Tukey windowing (with $M = 15$) are shown by the broken line. There is no visible difference between the Tukey estimated spectral densities.

estimation technique used. Autoregressive model fitting procedures which determine the order would do the former, nonparametric approaches(Tukey, multitapering etc.) the latter.

Whether for use in prewhitening, or for correcting the variance and degrees of freedom of the test statistic, an accurate, robust estimate of the autocorrelation is necessary. This estimation could be carried out in either the spectral or temporal domain - they are interchangeable. Raw estimates (equation 11) can not be used since they are very noisy and introduce an unacceptably large bias. Hence some means of improving the estimate is required.

All approaches considered assume second-order stationarity - an assumption whose validity is helped by the use of the non-linear high pass filtering mentioned in the previous section. We consider standard windowing or tapering spectral analysis approaches, multitapering, parametric ARMA and a nonparametric technique which uses some simple constraints. The results of these different techniques applied to a typical grey-matter voxel in a rest/null data are shown for comparison in figure 2.

3.4.1 Single Tapers

A standard approach to taking raw estimates of the autocorrelation or equivalently of the spectral density is to window the raw time series prior to taking a Fourier Transform. This down-weights points at either end of the time series, reducing leakage due to end effects (Bracewell, 1978). Equivalently the raw autocorrelation estimate can be tapered such that it is downweighted at high lags. Intuitively, this seems reasonable since the precision of the raw autocorrelation estimates clearly decrease with high lag.

With either windowing of the time series or of the raw autocorrelation estimate, the shape and size of the window needs to be decided upon. Here, we prefer to use windowing of the raw autocorrelation estimate. This is because of considerations of spatial regularisation which we will come to later. For a time series $x(t)$ for $t = 1, \ldots, N$ the raw autocorrelation estimate at lag τ is given by:

$$r_{xx}(\tau) = \frac{1}{\hat{\sigma}^2} \sum_{t=1}^{N-\tau} x(t)x(t+\tau)/(N-\tau) \tag{11}$$

The two favoured windows in the time series literature are the Tukey and Parzen windows, which appear to perform equally well (Chatfield, 1996). Hence, we arbitrarily concentrate on the Tukey window which is defined as:

$$\hat{\rho}_{xx}(\tau) = \begin{cases} \frac{1}{2}\left(1 + cos\left(\frac{\pi\tau}{M}\right)\right) r_{xx}(\tau) & \text{if } \tau < M \\ 0 & \text{if } \tau \geq M \end{cases} \tag{12}$$

where M is the truncation point such that for $\tau > M$, $\hat{\rho}_{xx} = 0$. This window smooths the spectral density by an amount determined by M.

The choice of the value for M is a balance between reducing the variance whilst minimising the distortion of the autocorrelation/spectral density estimate. The variance in the estimation of the spectral density is given by (Chatfield, 1996):

$$Var[\hat{\rho}_{xx}(\tau)/\rho_{xx}(\tau)] = \frac{3M}{4N} \tag{13}$$

Large M corresponds to less smoothing in the spectral domain. A rough guide in the literature is to set M to be about $2\sqrt{N}$ (Chatfield, 1996). For $N = 200$ this gives $M = 28$.

3.4.2 Non-parametric Estimation

Instead of presetting M to what is considered to be a reasonable value, we instead apply some constraints. The first assumption is that $\rho_{xx}(\tau) > 0$ for all τ. The second assumption is that the autocorrelation is monotonically decreasing. This means that low frequency components are favoured, which are widely accepted in the literature as being the most important to account for.

The autocorrelation is estimated using a standard unbiased estimator (equation 11) and then the best least squares fit that satisfies the constraint of monotonicity can be obtained using techniques from the literature of isotonic regression. The particular algorithm that we use is the Pool Adjacent Violators Algorithm (PAVA) (T. Robertson and Dykstra, 1988), which provides a unique, least squares fit under the constraint. Before using the algorithm, we set $\hat{\rho}_{xx}(\tau) = r_{xx}(\tau)$ for $\tau = 1, \ldots, N/3$ and $\hat{\rho}_{xx}(\tau) = 0$ for $\tau > N/3$. This is done partly because it reduces the amount of data the algorithm iterates over, and also because the raw autocorrelation estimate is very noisy for $\tau \geq N/3$ (there is less data available to compute autocorrelations at high lags), and we do not expect significant autocorrelation at such high lags. Furthermore, the value of zero will propagate, eventually stopping at the lag which gives M. For the purpose of the algorithm it is also necessary to define a weighting function $w(\tau) = 1$ for $\tau = 1, \ldots, N$. The algorithm then proceeds as follows:

1. If $\hat{\rho}_{xx}(\tau)$ is not isotonic there must exist a violator at k such that $\hat{\rho}_{xx}(k-1) > \hat{\rho}_{xx}(k)$.
2. Pool these two values, by replacing them both with their weighted average:
$$[\hat{\rho}_{xx}(k-1)w(k-1) + \hat{\rho}_{xx}(k)w(k)]/[w(k-1) + w(k)] \tag{14}$$
3. Replace $w(k-1)$ and $w(k)$ by $w(k-1) + w(k)$
4. Repeat until no more violators.

The algorithm was tested on artificial data which consisted of white noise of length $N = 200$, and had been low-pass filtered with a Gaussian of varying standard deviation. This highlighted a slight bias for white noise data, which was easily remedied by setting $\hat{\rho}_{xx}(1) = 0$ if $\hat{\rho}_{xx}(1) < 0.1$.

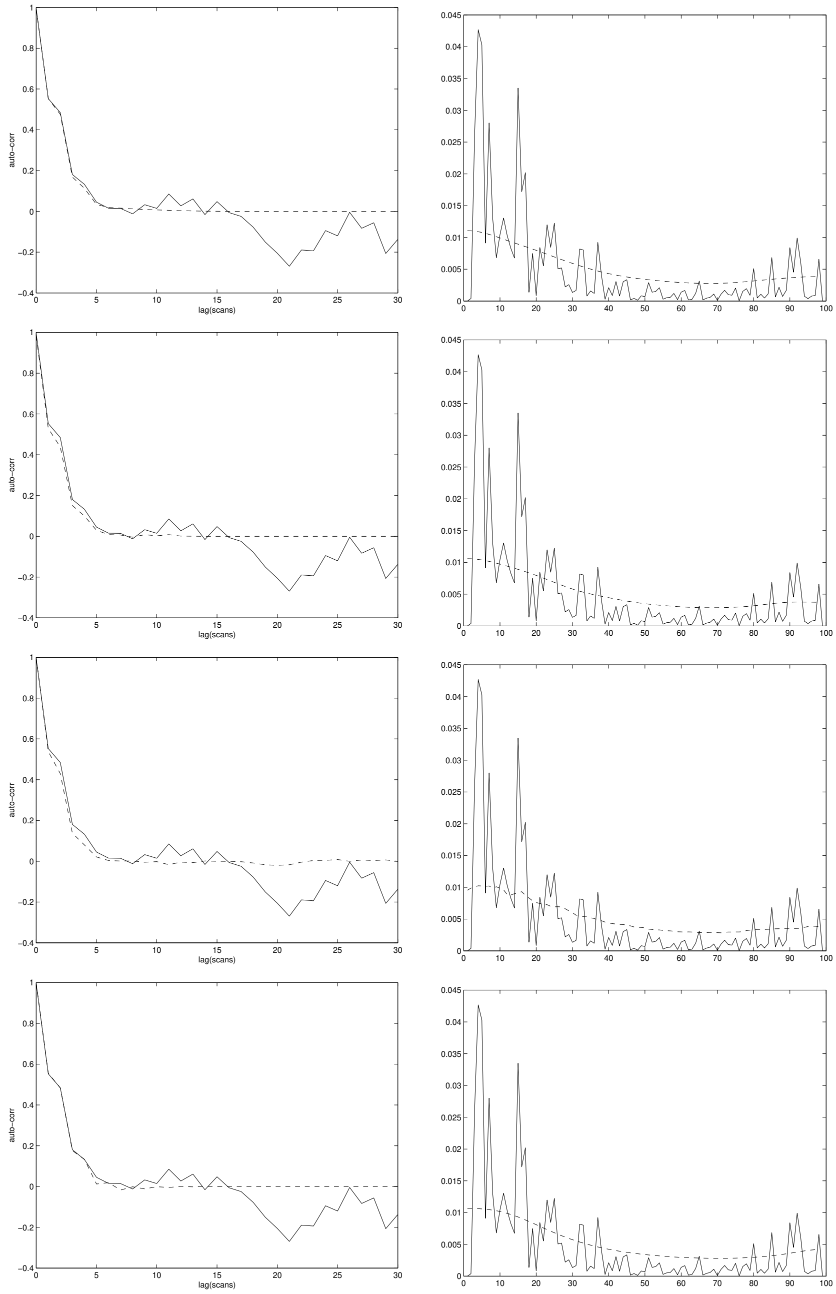
3.4.3 Multitapering

Multitapering is an extension of single taper approaches and consists of dividing the data into overlapping subsets that are each individually tapered, and then Fourier transformed. The individual spectral coefficients of each subset are averaged to reduce the variance. The way in which the data is to be subdivided is defined by a set of tapers indexed by $l = 1 \ldots L$, the estimated spectral density at frequency bin f is then given by:

$$S(\mathbf{f}) = \frac{\sum_{l=0}^{L-1} \lambda_l S_l(\mathbf{f})}{\sum_{l=0}^{L-1} \lambda_l} \tag{15}$$

where $S_l(\mathbf{f})$ is the estimated spectral density using taper l and λ_l are weights for each tapered spectral density estimate.

As with the single-window approaches the spectral density is effectively smoothed, but without losing information at the end of the time series. The windows are chosen so that they are orthogonal and reduce leakage as much as possible. Under these requirements the optimal choice is the Discrete Prolate Spheroidal Sequences or Slepian



(a) (b)

Figure 2: (a) Autocorrelation, and (b) Spectral density, for a typical grey-matter voxel in a rest/null data set. Raw estimates are shown by the solid line, and (from top to bottom) estimates from Tukey windowing with $M = 15$, nonparametric PAVA, multitapering with $NW = 13$, and a general order AR model are shown by broken lines.

sequences (Percival and Walden, 1993). The Slepian sequence used is determined by the length of the time series N, and by a parameter W which corresponds to the half-bandwidth (i.e. $w = 2\pi W$ is the half-bandwidth in radians). When using the Slepian sequences to give $S_l(\mathbf{f})$, the weights λ_l used in equation 15 could simply be unity or something more complex. In this paper we use the eigenvalues of the Toeplitz matrix associated with the Fourier transformation as the weights (Percival and Walden, 1993). However, even these weights aren't optimal - more elaborate weighting such as Thomsen's non-linear approach Percival and Walden (1993), which adapt to the local variations in the spectral density could be used instead.

As with the single taper approaches the parameter W needs to be chosen to balance the desired reduction in variance with minimising the distortion of the spectral estimate. The variance in the estimation of the spectral density is given by (Percival and Walden, 1993): For Slepian sequence multi-tapering the variance in the estimate is:

$$Var[\hat{\rho}_{xx}(\tau)/\rho_{xx}(\tau)] = 2NW \tag{16}$$

Hence comparing equation 16 with equation 13 a Tukey tapering approach can be compared with multitapering by setting the variances the same. For example, for $N = 200$ the recommended Tukey parameter was $M = 28$. To give the same variance the requires a multitaper parameter of $NW \approx 7$.

3.4.4 Autoregressive Parametric Model Estimation

Stationary stochastic time series can be modelled using an autoregressive process of sufficiently high order p (AR(p)):

$$x(t) = \phi_1 x(t-1) + \phi_2 x(t-2) + ... + \phi_p x(t-p) + e(t) \tag{17}$$

where $e(t)$ is a white noise process and $\phi_1, \phi_2, \ldots, \phi_p$ are the autoregressive model parameters. There is the option to fit more complex ARMA models (autoregressive process forced by a moving average process). However, AR models are often used on their own due to their relative simplicity in fitting. It may then require more parameters to fit the process with no significant loss of accuracy. This is the approach considered here.

The time series literature, including Chatfield (1996), describes various techniques for determining the order p and parameters of AR models. Here, we use the partial autocorrelation function (PACF) to find p and ordinary least squares to fit the parameters. When fitting an AR(p) model, the last partial coefficient α_p measures the excess correlation at lag p not accounted for by an AR($p-1$) model; α_p plotted for all p is the PACF. The lowest value of p for which α_p in the PACF is not significantly different to zero (using the 95% confidence limits of approximately $\pm 2/\sqrt{N}$ (Chatfield, 1996)), is the order used.

4 Qualitative Data Analysis

4.1 Methods

The intention here is to explore the effects of temporal filtering and the spatial variation of the autocorrelations in real FMRI data. We could attempt to examine the autocorrelation, or equivalently the power spectral density, itself. However, this would give N (number of scans or time points) data points for each voxel. Instead, we use:

$$S_\rho = N/[1 + 2\sum_{\tau=1}^{N-1} \rho_{xx}(\tau)] \tag{18}$$

This produces a single value whose variation can then be easily visualised. The value S_ρ corresponds approximately to $1/k_{\mathit{eff}}$ in equation 7 when performing a t-test ($c = [1]$ and $X = [1, ..., 1]^T$) with large N.

An $S_\rho = n$ indicates white noise and $0 < S_\rho < n$ indicates a time series with positive autocorrelation. We examined one rest/null dataset from a normal volunteer. Two hundred echo planar images (EPI) were acquired using a 3 Tesla system with time to echo (TE) = 30ms, TR=3 secs, in-plane resolution 4mm and slice thickness 7mm. The first 4 scans were discarded to leave $N = 196$ scans and the data was motion corrected using AIR (Woods et al., 1993). To calculate S_ρ at each voxel we arbitrarily used the nonparametric PAVA autocorrelation approach to estimate the autocorrelation.

4.2 Results

Figure 3 shows a set of histograms of S_ρ for the entire brain volume with the skull and background removed. With no temporal filtering, the histogram has a peak at the low values of $S_\rho \approx 15$ representing tissue with high autocorrelation, and a peak at $S_\rho = 196$ corresponding to white noise.

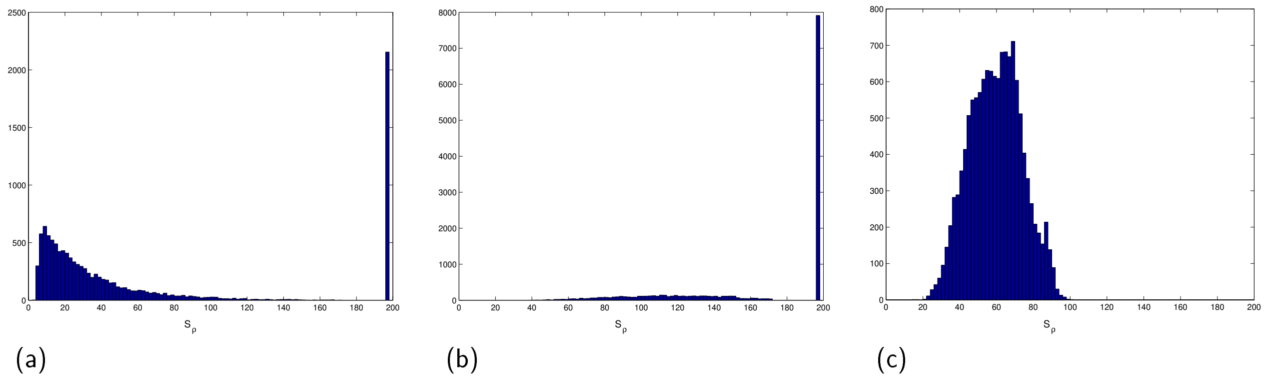


Figure 3: Histograms of S_ρ for a null FMRI dataset with (a) no temporal filtering (b) non-linear high-pass filtering (c) Low pass filtering matched to a Gaussian HRF followed by non-linear high-pass filtering. The only preprocessing is motion correction. Similar results were obtained for five other null datasets from three different subjects on the same scanner. The two apparent populations in (b) are due largely to white and grey-matter; this can be seen in the spatial map of S_ρ in figure 4.

4.2.1 High-pass Filtering

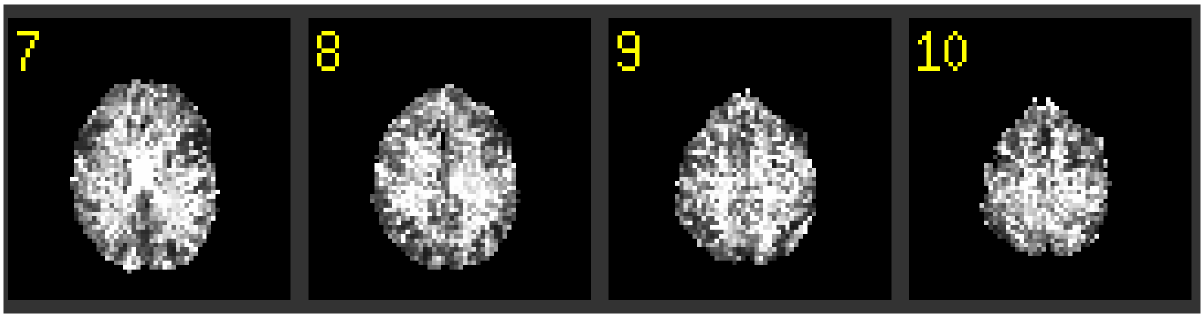
High pass filtering is used to remove the worst of the low frequency noise in the FMRI time series. Here we are using the non-linear high-pass filtering discussed in the previous section. In figure 3(b) a FWHM of 40 scans was used and has the effect of removing the lower peak and pushing the whole histogram to higher values. This shift to higher S_ρ values represents a decrease in the positive autocorrelation in the data; this corresponds to a decrease in the parameter variances, as desired. The use of high-pass filtering reduces the low frequency noise and non-stationarity of the time series, making the estimation of the autocorrelation more robust and valid. However, having used a non-linear high-pass filter, the power spectral density of the time series has been shaped in such a way that it is not easily modelled by a low order parametric AR model.

4.2.2 Low-pass filtering

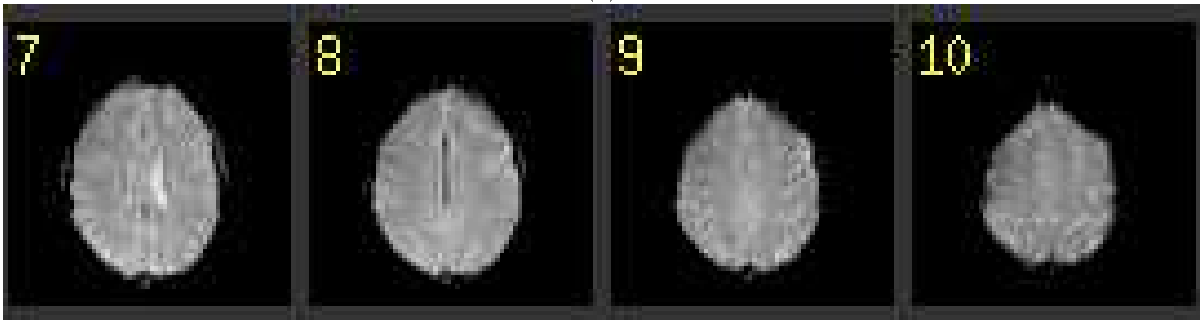
The low-pass filter used here is a Gaussian filter matched to a commonly assumed HRF with parameters $\sigma = 2.98secs$ and $\mu = 3secs$ (Friston et al. (1995)). The histogram for low-pass filtering demonstrates the idea of colouring, in that the whole histogram is focused to a peak centered on a S_ρ close to that entirely due to the low-pass filtering. However, the standard deviation around this peak (standard deviation=17.6) is greater than the standard deviation of the estimator of S_ρ (determined empirically on artificial data as standard deviation=12.9). Since the data is showing a greater variability in S_ρ than there is in just estimating S_ρ, this suggests the requirement for local estimation of autocorrelation even when low pass filtering (colouring) is performed. Although the autocorrelation estimate is made more robust, the autocorrelation imposed by the colouring does not completely smother the intrinsic autocorrelation.

4.2.3 Spatial Variation

Figure 4 shows four slices of the S_ρ values in the brain volume after non-linear high-pass filtering has been performed, along with the original functional image, for comparison. Figure 4 shows considerable spatial variation and structure, with lower S_ρ corresponding to increased autocorrelation in the grey matter compared to the white-matter and CSF. Exactly the same characteristic could be easily observed in five other null data sets.



(a)



(b)

Figure 4: (a) Spatial maps of S_ρ in the brain volume after the non-linear high-pass filtering has been performed. This corresponds to the histogram in figure 3(b), with high S_ρ displayed as lighter grey and low S_ρ as darker grey. (b) EPI for the same slices. Exactly the same characteristic could be easily observed in five other null data sets.

These findings appear to contradict Zarahn's and Lund's conclusions that the autocorrelation is not at all physiological in origin. It may be that the greater smoothness in the grey-matter is due to a larger number of edges in grey-matter compared to white matter. Edges within voxels may be subject to motion of any type (inaccurate motion correction, physiological pulsations), and this motion along with a partial volume effect may produce increased low frequency noise. Further analysis is required to understand such sources of the autocorrelations that could exhibit these characteristics. This will be a topic for future research.

5 Effect of Different Regressors

5.1 Methods

It turns out that the regressor used in the design matrix considerably affects the relative efficiency between colouring and prewhitening. To illustrate this we use a typical autocorrelation estimated for a grey-matter voxel (shown in figure 5) to give a typical $\mathbf{V}$ matrix, and use for $\mathbf{X}$ one of four different types of regressors of particular interest:

(a) a boxcar design with period 60 secs

(b) a single-event (SE) design with fixed inter-stimulus interval (ISI) of 15 secs and stimulus duration of 0.1 secs (Bandettini and Cox, 2000)

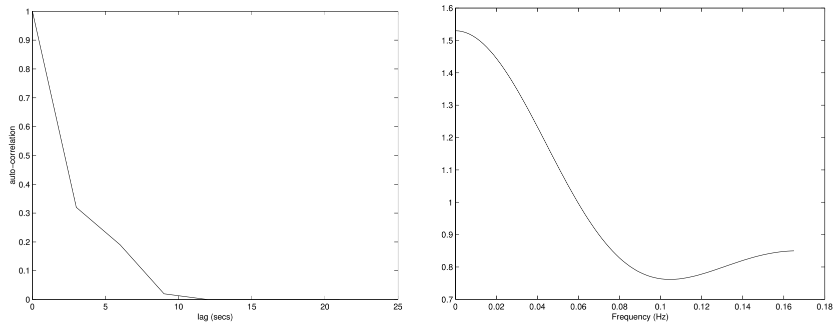
(c) a single-event design with stimulus duration of 0.1 secs with jittering such that the ISIs are drawn from a uniform distribution U(13.5 secs, 16.5 secs)(Josephs et al., 1997)

(d) a single-event design with randomized ISI taken from a normal distribution with mean $6secs$ and standard deviation $2secs$ with no ISI less than $2secs$ (Burock et al. (1998), Dale and Buckner (1997) and Dale (1999)) and stimulus duration of 0.1 secs.

All four designs are convolved with the same gamma HRF:

$$f_G(t; a, b) = \frac{b^a}{\Gamma(a)} t^{a-1} e^{-bt} \tag{19}$$

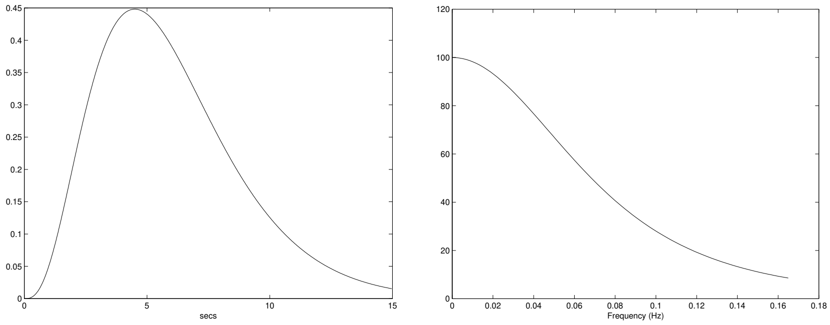
where the Gamma parameters a, b are set according to mean $a/b = 6secs$ and variance $a/b^2 = 9secs^2$. The Gamma HRF for these parameters is shown in figure 6. More complicated HRF models could be used. However, any reasonable HRF model would be expected to have a similar spectral density and therefore behave in a similar way in this context. For all regressors TR is taken as 3 secs and all regressors have their means removed.



(a) (b)

Figure 5: (a) Autocorrelation for a typical grey-matter voxel in a rest/null data set, and (b) the power spectral density for the same autocorrelation.

The variance of the parameter estimates, $k_{eff}\sigma^2$, is inversely proportional to the efficiency and can give us a measure of the relative efficiency of the different temporal filtering strategies. Although the estimation of σ does



(a) (b)

Figure 6: (a) Gamma HRF, $f_G(t; a, b)$ has parameters set according to mean $a/b = 6secs$ and variance $a/b^2 = 9secs^2$, and (b) the spectral density for the HRF.

depend upon the temporal filtering strategy used (equation 4 depends on $\mathbf{S}$), this effect is negligible. Subsequently, we define a measure of efficiency, E, relative to the maximally efficient prewhitening estimator for the regressor as:

$$E = \frac{k_{eff} \text{ for prewhitening}}{k_{eff}} \tag{20}$$

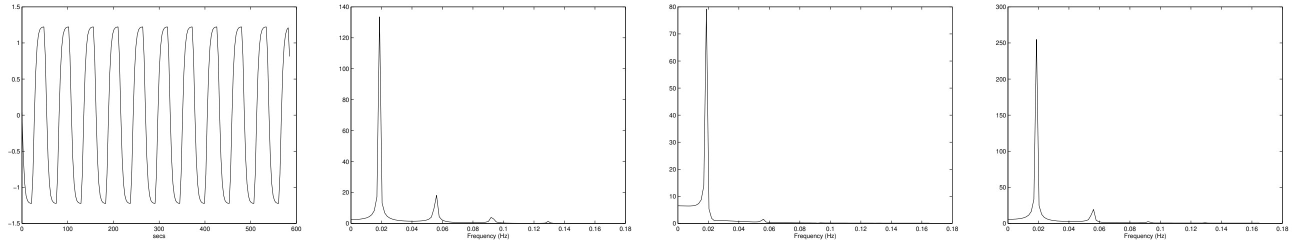
This was computed for each regressor using each of the three different temporal filtering strategies using equations 6, 7, 8 accordingly (there is only one regressor in each case so we use $c = [1]$). The low pass filter used for the colouring was matched to the HRF (figure 6). The values of E for the randomised ISI and jittered ISI designs were averaged over 100 randomly generated designs.

5.2 Results

The results are shown in table 1. It can be seen that even for the boxcar design, colouring is not as efficient as variance correction or prewhitening. For the randomized ISI design, the contrast is even more apparent, with prewhitening being more efficient than variance correction, which in turn is much more efficient than colouring. This concurs with the theory and work by Friston et al. (2000).

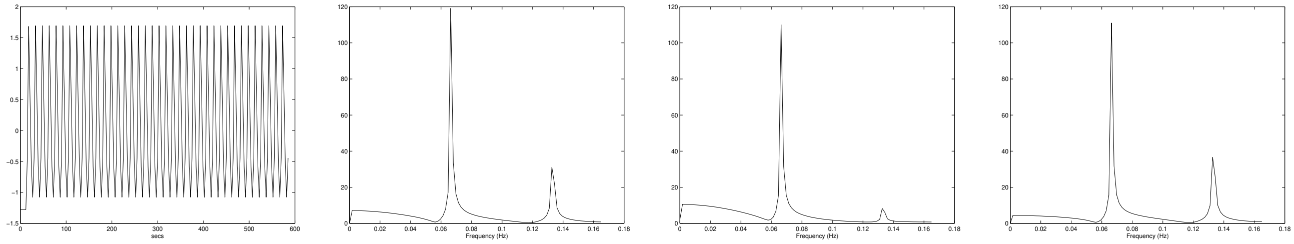
Examination of the spectral density for the randomized ISI design before and after it has been low-pass filtered (when colouring) illustrates the reason for the loss in efficiency. These spectra are shown in figures 10(b) and (c) respectively. It might have been expected that the frequency response in figure 10(b) would have relatively little high frequency content due to convolution with the smooth HRF whose spectral density is shown in figure 6(b). However, there are clearly strong high frequency components in the regressor. These are introduced when the high temporal resolution version of the regressor is sampled to a lower temporal resolution. Figure 10(c) clearly shows a reduction in these high frequency components due to the low pass filtering, resulting in a loss of efficiency. In contrast, comparing figure 7(b) with figure 7(c) reveals little difference particularly with regard to most of the power being in the fundamental frequency. Hence for the boxcar, colouring has similar efficiency to variance correction and prewhitening.

However, this does not explain why variance correction is less efficient than prewhitening for the randomized ISI design. This loss in efficiency is instead due to using a more inefficient estimator when using variance correction, compared with the best linear unbiased estimator (BLUE) of prewhitening. Here, the prewhitening down-weights the low frequencies compared to the high frequencies (inverse of figure 5) to give the BLUE. This is of particular benefit when the regressor has substantial power across a larger range of frequencies being weighted, such as is the case for the randomized ISI design (compare figure 10(b) with 10(d))



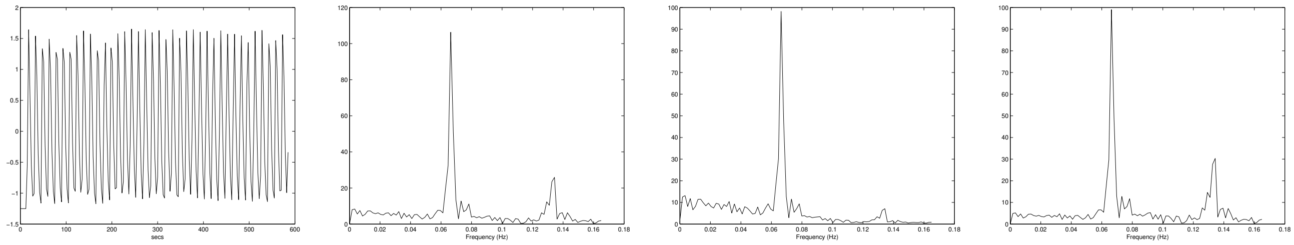
(a) (b) (c) (d)

Figure 7: Plots of (a) the regressor $\mathbf{X}$ (b) the spectral density of $\mathbf{X}$ (c) the spectral density of $\mathbf{SX}$ where $\mathbf{S}$ is the colouring low-pass filtering Toeplitz matrix (d) the spectral density of $\mathbf{K^{-1}X}$ where $\mathbf{K^{-1}}$ is the prewhitening matrix from $\mathbf{V} = \mathbf{KK^T}$. Where $\mathbf{X}$ is a boxcar.



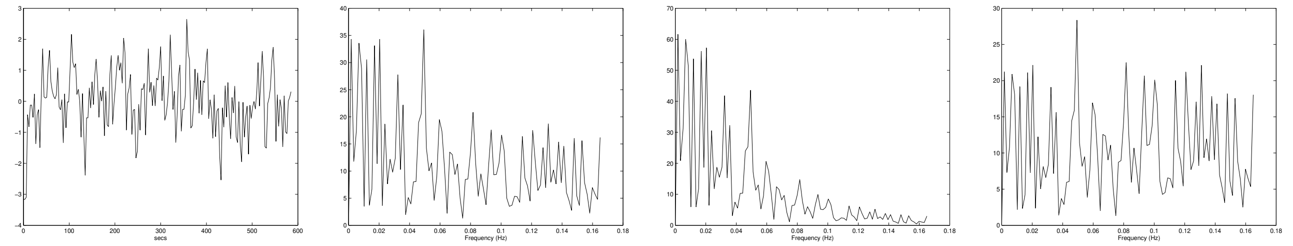
(a) (b) (c) (d)

Figure 8: As for figure 7, but $\mathbf{X}$ is a single-event design with fixed ISI.



(a) (b) (c) (d)

Figure 9: As for figure 7, but $\mathbf{X}$ is a single-event design with jittered ISI.



(a) (b) (c) (d)

Figure 10: As for figure 7, but $\mathbf{X}$ is a single-event design with randomized ISI.

	boxcar	SE/fixed ISI	SE/jittered ISI	SE/random ISI
Colouring(eq.6)	0.96	0.70	0.68	0.21
Variance correction(eq.7)	0.99	0.98	0.98	0.80
Prewhitening(eq.8)	1.00	1.00	1.00	1.00

Table 1: Table showing relative efficiency E (equation 20) calculated using the three different strategies for dealing with the autocorrelation and for four different types of regressor.

6 Nonlinear Spatial Smoothing

Given the similarity of temporal autocorrelation in the local neighbourhood of each voxel, we attempt to improve the robustness of the autocorrelation estimates using a small amount of local spatial smoothing. However, our qualitative data analysis indicated clearly that the autocorrelation differs between tissue types. Isotropic spatial smoothing would blur the autocorrelation estimates across tissue boundaries, resulting in biased estimates of the autocorrelation near tissue boundaries. An alternative approach is to use some form of nonlinear spatial smoothing that does not smooth across such boundaries.

Accurate segmentation of white-matter, grey-matter and CSF would provide the necessary information to avoid blurring across tissue types. However, segmentation of EPI images is hard due to poor tissue type contrast, bias field effects, low resolution and the partial volume effect.

Instead we use the "Smoothing over Univalue Segment Assimilating Nucleus" (SUSAN) noise reduction filter (Smith and Brady, 1997), which is a nonlinear filter designed to preserve image structure by only smoothing over those neighbours which form part of what is believed to be the "same region", or USAN, as the central voxel under consideration. This concept is illustrated in 1-D in Figure 11.

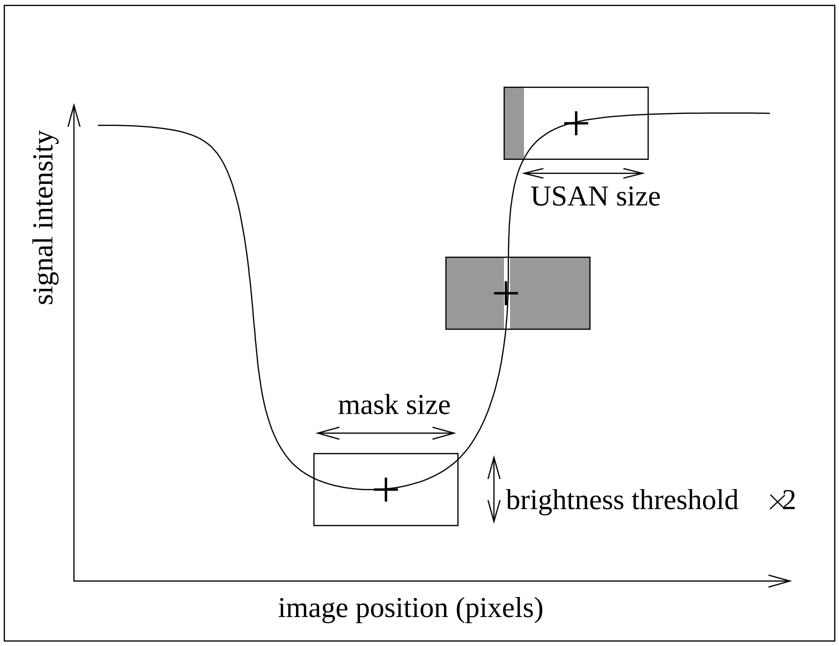


Figure 11: The SUSAN smoothing method smooths only similar pixels within a local region known as the Univalue Segment Assimilating Nucleus (USAN). Here the USAN is shown as the white region within the mask.

The filter averages over all the pixels in the locality which lie in the USAN using the weighting,

$$w(\vec{r}, \vec{r_0}) = e^{-\frac{(I(\vec{r})-I(\vec{r_0}))^2}{2t^2}} \tag{21}$$

where t is the brightness difference threshold, $I(\vec{r})$ is the brightness at any pixel $\vec{r}$ and w is the output of the comparison. In 1-D the full univariate equation is:

$$J(x) = \frac{\sum_{i\neq 0} I(x+i) e^{-\frac{i^2}{2\sigma_s^2} - \frac{(I(x+i)-I(x))^2}{2t^2}}}{\sum_{i\neq 0} e^{-\frac{i^2}{2\sigma_s^2} - \frac{(I(x+i)-I(x))^2}{2t^2}}} \tag{22}$$

where σ_s controls the scale of spatial smoothing, i.e. the mask size.

We can use a 3D version of this filter to spatially smooth the raw autocorrelation estimate at each lag or to spatially smooth the spectral density estimate at each frequency. Whether we smooth the autocorrelation or the spectral density estimates will depend upon the autocorrelation/spectral density estimation technique then used on the spatially regularised data.

Normally, the USAN is estimated from the same data, or image, as that which is being smoothed. However, here we will be using one of the EPI volumes to generate the USAN.

We would expect a σ_s of approximately 1 or 2 voxels to be optimal, since the spatial autocorrelation of S_ρ suggests that smoothness is in the immediate neighbourhood only. Simple histogram techniques are used to assess the approximate standard deviation of grey matter in the EPI. The brightness threshold t is then set to 1/3 of the approximate standard deviation of grey matter. This gives a very conservative brightness threshold t, one which allows a small amount of smoothing within grey matter whilst allowing negligible smoothing between matter types.

7 Calibration

7.1 Methods

We now want to ascertain the difference in the the bias of the resulting statistical distributions that exists for the different approaches for estimating the autocorrelation. This is determined experimentally on real rest (null) FMRI data by computing the t-statistic at each voxel for a dummy design paradigm The t-statistic is given by $t = \mathbf{c}\hat{\mathbf{B}}/\sqrt{Var\{\mathbf{c}\hat{\mathbf{B}}\}}$ where $\hat{\mathbf{B}}$ and $Var\{\mathbf{c}\hat{\mathbf{B}}\}$ are given by equations 2 and 3 respectively. The t-statistics are then probability transformed to z-statistics. The probability transform involves converting the t-statistic into its corresponding probability (by integrating the t-distribution from the t-statistic's value to infinity) and then calculating the z-statistic that corresponds to the same probability (by integrating the normal distribution from the z-statistic's value to infinity).

These z-statistics form what we refer to as the null distribution. A technique with low bias should give a null distribution that closely approximates the theoretical z-distribution (or Normal distribution).

For the theoretical, Normal probability density function, $f(z)$, we can obtain the z-statistic, z_p, for a chosen probability p such that $p = \int_{z_p}^{\infty} f(z)dz$. This can then be compared to $p_{null} = \frac{1}{2}Prob(|z| > z_p)$ for the empirically obtained null distribution, $d(z)$. This is given by:

$$p_{null} = \frac{\sum_{|z|>z_p} d(z)}{2\sum_z d(z)} \tag{23}$$

Since for purposes of inference the tail is the most important part of the distribution, we examine p_{null} as far into the tail as the sample size will allow. This is aided by using both tails of the empirically obtained null distribution.

We intend to study data taken at TR=3 and 1.5 secs. Six different rest/null datasets (3 normal volunteers, 2 datasets per volunteer) were obtained using TR=3 secs and 9 null datasets (3 normal volunteers, 3 datasets per volunteer) were obtained using TR=1.5 secs. For each dataset 204 echo planar images (EPI) were acquired using a 3 Tesla system with time to echo (TE) = 30ms, in-plane resolution 4mm and slice thickness 7mm. The first 8 scans were discarded to leave $N = 196$ scans and the data was motion corrected, intensity normalised by subtracting the global mean time series from each voxel's time series, and non-linear high-pass filtered. We computed an empirical distribution based on either all of the TR=3 secs data or on all of the TR=1.5 secs data. The z-statistics for all of the brain voxels in the six or nine null datasets are all pooled together to give one empirical null distribution. The resulting distributions consisted of z-statistics from approximately 80000 voxels. This allowed for examination into the tail to probabilities as low as $1e-5$. It is important that we examine this far into the tail of the distribution as this is approximately where inference needs to take place when multiple comparison corrections are taken into account (Worsley et al., 1992).

We will consider two different paradigms - the simple boxcar HRF convolved paradigm (on the TR=3 secs data) and the single-event with randomized ISI design (on the TR=1.5 and 3 secs data) as described earlier. Various autocorrelation estimation techniques will be compared on the calibration plots when performing prewhitening.

7.2 Results

7.2.1 Tukey single taper

Recall that a suggested value for M is $2\sqrt{N}$, which for $N = 196$ gives $M \approx 28$. This value is compared along with a range of others ($M = 5, 10, 15, 28$) when using Tukey tapering with no spatial smoothing of the autocorrelation estimate. The results are shown in figure 13 (TR=3 secs) and figure 17(a) (TR=1.5 secs).

The first thing to note from figure 13 is that when no autocorrelation estimation is made and the residuals are assumed to be white (i.e. $S = I$), then the boxcar design deviates far more from the theoretical distribution than the single-event design. This is because the single-event design has power at frequencies across the full range and is therefore less effected by not correcting for the coloured noise in the data which is concentrated at low frequency, whereas the boxcar design's power is mostly at its fundamental frequency, which is within the range of low frequency noise. This makes the obvious point that designs which concentrate their power as far away as possible from the low frequency end will suffer less from the low frequency noise in the data. However, the amount of bias for a randomised ISI when no autocorrelation estimation is made is still quite considerable, and so there is still the requirement for the estimation and correction of the autocorrelation.

The values $M = 5$ and $M = 10$ perform about the same in figure 13(a) and 17(a) and $M = 5$ performs slightly better than $M = 10$ in figure 13(b). The key point is that lower values of M, i.e. those that smooth the spectral density more than is normally recommended, perform better.

7.2.2 Multitapering

Having established that lower values of M perform well when using Tukey tapering, we can use equations 16 and 13 with $N = 196$, to give a multitaper with approximately the same spectral density estimate variance. For $M = 10$ this gives $NW = 13$. We compare this with $NW = 4$, a commonly used value in the literature and corresponds to $M = 32$. The results are shown in figure 14 for the TR=3 secs data. $NW = 13$ performs much better and is similar to the Tukey estimator with $M = 10$, which is not surprising since they correspond to the same variance and therefore to similar amounts of spectral density smoothing. Again, increased spectral density smoothing compared with that which is normally recommended is better.

7.2.3 Autoregressive model and Nonparametric PAVA

When using the autoregressive model of general order it was found to require orders of up to 6; figure 12 shows the histogram of different AR orders required, pooled over all six of the null datasets taken with TR=3 secs. The results of using the autoregressive model and the nonparametric PAVA with the TR=3 secs data are shown in figure 15, along with Tukey with $M = 10$ and multitapering with $NW = 13$. Particularly for the boxcar design the single taper Tukey with $M = 10$ performs the best.

7.2.4 Autocorrelation Spatial Smoothing

We have established that without any spatial regularisation of the autocorrelation estimate, the single taper Tukey with $M = 5, 10$ perform best. We now want to explore the additional benefits, if any, of using the SUSAN spatial smoothing of the raw autocorrelation estimate before the Tukey tapering is applied and also to establish how much smoothing is of benefit. Spatial autocorrelation of the S_ρ map suggests that the autocorrelation is only correlated over a short range. The voxel dimensions for the 6 datasets are $4 \times 4 \times 7mm$ and hence we consider SUSAN filtering with $\sigma_s = 4, 8, 12mm$.

Although the single taper Tukey performed better with $M = 10$, because we are now regularising spatially, it turns out to be better to allow more flexibility (i.e. less smoothing) of the spectral density by choosing a Tukey taper with $M = 15$. Figures 16 (TR=3 secs) and 17(b) (TR=1.5 secs) show the results of the different amounts of

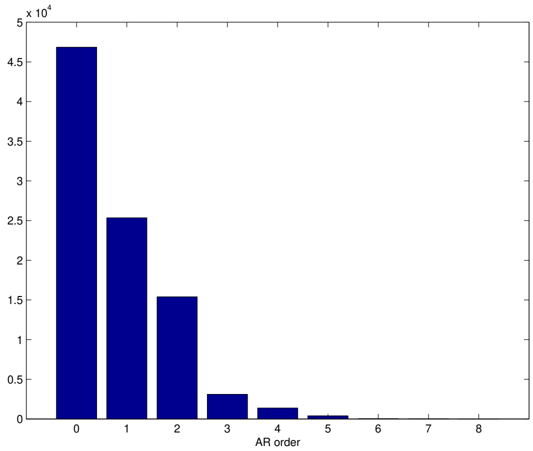
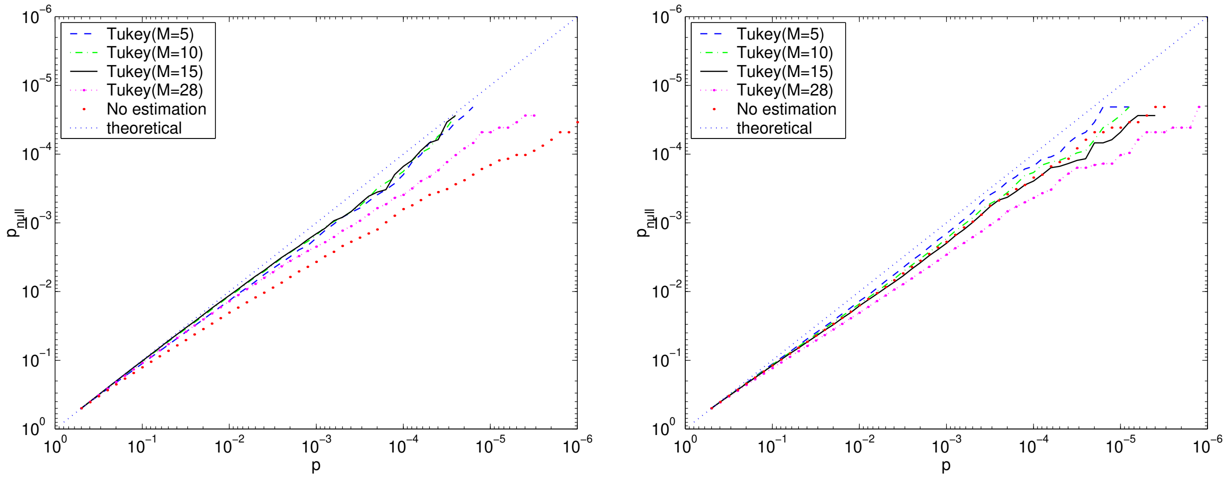


Figure 12: Histogram showing the orders of the AR models required to estimate the autocorrelation for all voxels in all six null datasets (TR=3 secs) after applying the randomized ISI single-event design

spatial smoothing. A σ_s of $8mm$ performs best and shows improvement over performing no spatial smoothing for the TR=3 secs data and performs similarly for the TR=1.5 secs data.

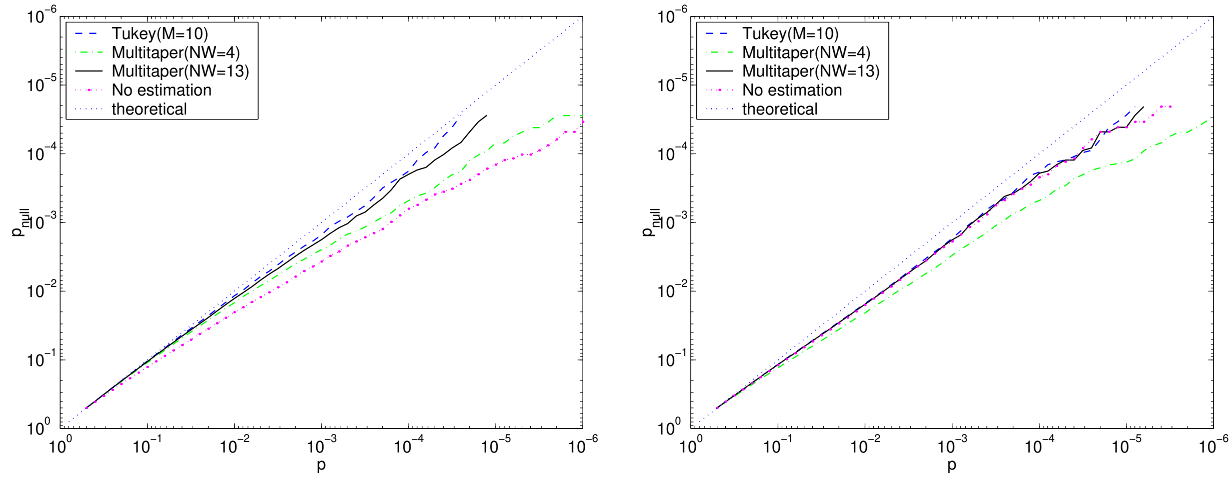


(a) (b)

Figure 13: Comparison of Tukey autocorrelation estimation for different values of M via log probability plots comparing theoretical p against null distribution p_{null} obtained from six different null datasets using TR=3 secs for (a) a boxcar design convolved with a gamma HRF, and (b) a stochastic single-event design convolved with a gamma HRF. All are calculated using prewhitening. The straight dotted line shows the result for what would be a perfect match between theoretical and null distribution.

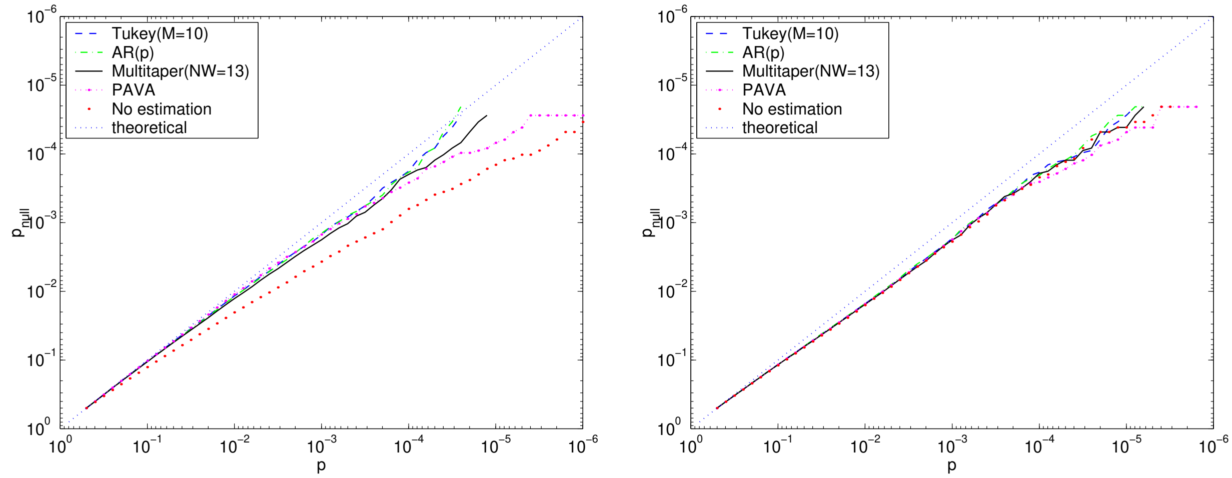
8 Discussion

Prewhitening requires a robust estimator of the autocorrelation to maintain low bias and Friston et al. (2000) suggests that current techniques for estimating the autocorrelation are not accurate enough to give prewhitening acceptable bias. However, in Friston et al. (2000) efforts were focussed on using global estimates of autocorrelation for reasons of computational efficiency. In this paper local estimation techniques were considered and were found



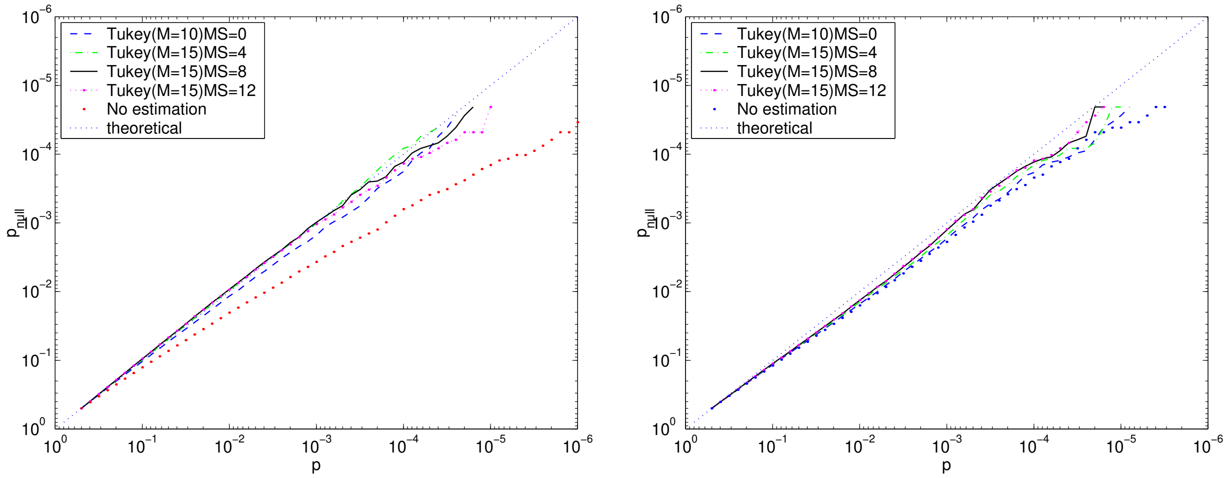
(a) (b)

Figure 14: Comparison of Multitapering autocorrelation estimation for different values of NW via log probability plots comparing theoretical p against null distribution p_{null} obtained from six different null datasets using TR=3 secs for (a) a boxcar design convolved with a gamma HRF, and (b) a stochastic single-event design convolved with a gamma HRF. All are calculated using prewhitening. The straight dotted line shows the result for what would be a perfect match between theoretical and null distribution.



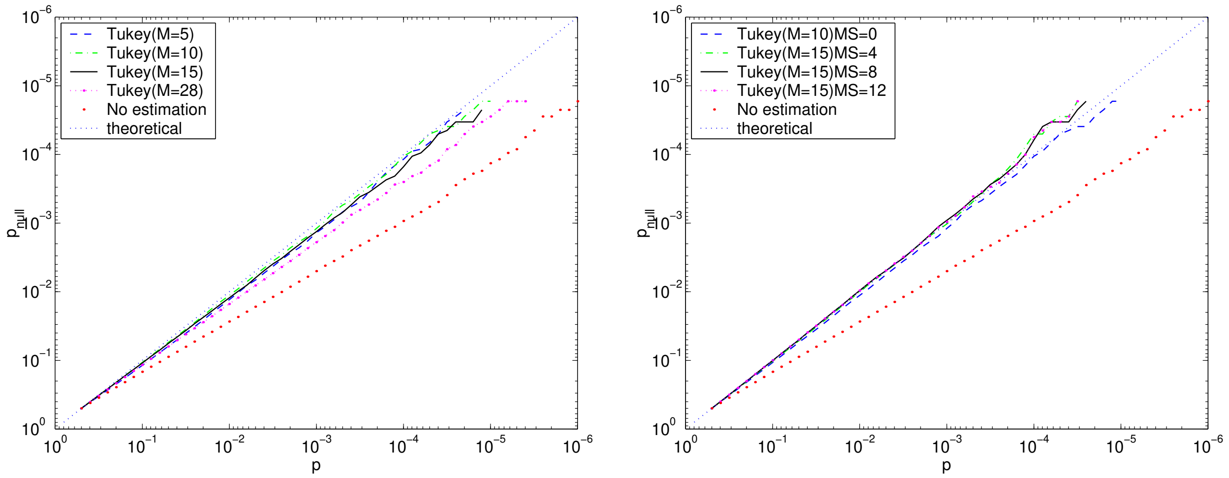
(a) (b)

Figure 15: Comparison of the different autocorrelation estimation techniques via log probability plots comparing theoretical p against null distribution p_{null} obtained from six different null datasets using TR=3 secs for (a) a boxcar design convolved with a gamma HRF, and (b) a stochastic single-event design convolved with a gamma HRF. All are calculated using prewhitening. The straight dotted line shows the result for what would be a perfect match between theoretical and null distribution.



(a) (b)

Figure 16: Comparison of different amounts of spatial smoothing of the raw autocorrelation estimate prior to using Tukey tapering with $M = 15$, via log probability plots comparing theoretical p against null distribution p_{null} obtained from six different null datasets using TR=3 secs for (a) a boxcar design convolved with a gamma HRF, and (b) a stochastic single-event design convolved with a gamma HRF. All are calculated using prewhitening. The straight dotted line shows the result for what would be a perfect match between theoretical and null distribution. MS is the mask-size used in the SUSAN smoothing and corresponds to σ_s.



(a) (b)

Figure 17: (a) Comparison of Tukey autocorrelation estimation for different values of M, and (b) Comparison of different amounts of spatial smoothing of the raw autocorrelation estimate prior to using Tukey tapering with $M = 15$, for data taken with TR=1.5 secs. The comparison is made via log probability plots comparing theoretical p against null distribution p_{null} obtained from nine different null datasets with a stochastic single-event design convolved with a gamma HRF. All are calculated using prewhitening. The straight dotted line shows the result for what would be a perfect match between theoretical and null distribution. MS is the mask-size used in the SUSAN smoothing and corresponds to σ_s.

to perform at acceptable speeds (less than 5 minutes for the null datasets used in this paper).

One interesting characteristic of the calibration/bias plots in figures 13-16 is that the empirically obtained probabilities are predominantly less than the expected theoretical probabilities. It is not clear why this should be the case. One possibility is that all of the autocorrelation techniques are overestimating the noise at low frequency. This could be a symptom of the trade-off between being sufficiently flexible to model the low frequency components and avoiding over-fitting at higher frequencies. One solution could be to use a nonparametric model fitted in the spectral domain which allows more flexibility at low frequencies (Marchini and Ripley, 2000). In this paper nonlinear spatial smoothing is used to regularise spatially and this allows the use of a more flexible $M = 15$ Tukey window, whilst at the same time avoiding over-fitting. This reduces bias to close to zero.

In similar work by Burock and Dale (2000) a first order Autoregressive model with an extra white noise component is used when performing prewhitening on randomised ISI designs. They also demonstrate the efficiency gained through prewhitening and show that their estimates are unbiased. However, they do not appear to examine the bias as far into the tail as in this paper. Perhaps more importantly, the plots used to examine the bias are on a linear scale and this makes assessment of bias at low probabilities very difficult to assess. In this paper we use a log-log scale and findings suggest that bias is evident in the tail for general order AR models.

It would also be interesting to know more about the source of temporal autocorrelation, particularly with regards to its spatial nature. In particular, it would be interesting to understand the source of the increased autocorrelation in the grey-matter, whether or not it is physiological in origin and how it varies within the grey-matter itself.

9 Conclusions

As in Friston et al. (2000) we have demonstrated that when using designs such as a box car convolved with a gamma HRF, colouring can be used with minimal loss of efficiency. However, for single-event designs with randomized ISIs, jittering or just very short ISIs, colouring is much less efficient and hence prewhitening is desirable.

Prewhitening requires a robust estimator of the autocorrelation to maintain low bias. To estimate the autocorrelation or equivalently the spectral density for use in prewhitening, different techniques were considered. These were single tapering Tukey, multitapering, autoregressive model of general order and a nonparametric approach that assumes monotonicity in the autocorrelation.

Crucially, nonlinear high-pass filtering is performed as a preprocessing step to remove the worst of the non-stationary components and low frequency noise. A Tukey taper, with much greater smoothing of the spectral density than is normally recommended in the literature, performed the best when prewhitening.

Importantly, a small amount of spatial smoothing of the autocorrelation estimates was also found to be necessary to reduce bias to close to zero at low probability levels. The autocorrelation was found to vary considerably between matter types, with higher positive autocorrelation (low frequency noise) in the grey matter when compared with the white matter. Therefore, non-linear spatial smoothing of the autocorrelation was used, which only smoothed within matter types. Using a Tukey taper ($M = 15$) along with the non-linear spatial smoothing we were able to reduce bias to close to zero at probability levels as low as 1×10^{-5}.

References

Aguirre, G., Zarahn, E., and 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.

Bandettini, P. and Cox, W. (2000). Event-related fMRI contrast when using constant interstimulus interval: Theory and experiment. *Magnetic Resonance in Medicine*, 43:540–548.

Bracewell, R. N. (1978). *The Fourier Transform and its Application.* McGraw-Hill Kogakusha, 2nd edition.

Bullmore, E., Brammer, M., Williams, S., Rabe-Hesketh, S., Janot, N., David, A., Mellers, J., Howard, R., and Sham, P. (1996). Statistical methods of estimation and inference for functional MR image analysis. *Magnetic Resonance in Medicine*, 35(2):261–277.

Burock, M. A., Buckner, R. L., Woldorff, M. G., Rosen, B. R., and Dale, A. M. (1998). Randomized event-related experimental designs allow for extremely rapid presentation rates using functional MRI. *NeuroReport*, 9(16):3735–9.

Burock, M. A. and Dale, A. M. (2000). Estimation and detection of event-related fMRI signals with temporally correlated noise: A statistically efficient and unbiased approach. *Human Brain Mapping*, 11:249–260.

Chatfield, C. (1996). *The Analysis of Time Series*. Chapman and Hall.

Dale, A. (1999). Optimal experimental design for event-related fMRI. *Human Brain Mapping*, 8(2-3):109–114.

Dale, A. and Buckner, R. (1997). Selective averaging of rapidly presented individual trials using fMRI. *Human Brain Mapping*, 5:329–340.

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

Friston, K., Josephs, O., Zarahn, E., Holmes, A., Rouquette, S., and Poline, J.-B. (2000). To smooth or not to smooth? *NeuroImage*, 12:196–208.

Hu, X., Le, T., Parrish, T., and Erhard, P. (1995). Retrospective estimation and correction of physiological fluctutaion in functional MRI. *Magnetic Resonance in Medicine*, 34(2):201–212.

Josephs, O., Turner, R., and Friston, K. (1997). Event-related fMRI. *Human Brain Mapping*, 5:1–7.

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

Marchini, J. and Ripley, B. (2000). A new statistical approach to detecting significant activation in functional MRI. *NeuroImage*, 12(4):366–380.

Percival, D. B. and Walden, A. T. (1993). *Spectral Analysis for Physical Applications, Multitaper and Conventional Univariate Techniques*. Cambridge University Press.

Purdon, P. and Weisskoff, R. (1998). Effect of temporal autocorrelation due to physiological noise and stimulus paradigm on voxel-level false-positive rates in fMRI. *Human Brain Mapping*, 6:239–249.

Seber, G. (1977). *Linear Regression Analysis*. Wiley.

Smith, S. and Brady, J. (1997). SUSAN - a new approach to low level image processing. *International Journal of Computer Vision*, 23(1):45–78.

T. Robertson, F. W. and Dykstra, R. (1988). *Order Restricted Statistical Inference*. Wiley.

Woods, R., Mazziotta, J., and Cherry, S. (1993). MRI–PET registration with automated algorithm. *Journal of Computer Assisted Tomography*, 17(4):536–546.

Worsley, K., Evans, A., Marrett, S., and Neelin, P. (1992). A three-dimensional statistical analysis for CBF activation studies in human brain. *Journal of Cerebral Blood Flow and Metabolism*, 12:900–918.

Worsley, K. and Friston, K. (1995). Analysis of fMRI time series revisited - again. *NeuroImage*, 2:173–181.

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

10 Acknowledgements

The authors would like to thank Dr Mark Jenkinson (FMRIB Centre) and Jonathan Marchini (Statistics Dept., University of Oxford) for their help and input into this work, and acknowledge support from the UK MRC and EPSRC.