Research Article

Statistics in Medicine

Received 4 March 2014, Accepted 14 May 2015 Published online 26 August 2015 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.6592

Estimating the population local wavelet spectrum with application to non-stationary functional magnetic resonance imaging time series

Aimee N. Gott,[a] Idris A. Eckley[a*†] and John A. D. Aston[b]

Functional magnetic resonance imaging (fMRI) is a dynamic four-dimensional imaging modality. However, in almost all fMRI analyses, the time series elements of this data are assumed to be second-order stationary. In this paper, we examine, using time series spectral methods, whether such stationary assumptions can be made and whether estimates of non-stationarity can be used to gain understanding into fMRI experiments. A non-stationary version of replicated stationary time series analysis is proposed that takes into account the replicated time series that are available from nearby voxels in a region of interest (ROI). These are used to investigate non-stationarities in both the ROI itself and the variations within the ROI. The proposed techniques are applied to simulated data and to an anxiety-inducing fMRI experiment. Copyright © 2015 John Wiley & Sons, Ltd.

Keywords: locally stationary; replicate; random effects; wavelet processes; fMRI;

1. Introduction

Functional magnetic resonance imaging (fMRI) is a neuroimaging methodology that has revolutionized the study of the brain, allowing researchers to understand neurological function and processes in the living brain. fMRI is a dynamic imaging technique, in that it generates not only spatial information about activity in brain locations but does this rapidly (approximately every 2 s) over time. This yields a large number of time series of data associated with the experimental response. Typically, these time series are complex to model and can contain non-stationarities. However, one advantage of fMRI is that the spatial resolution is high, and therefore, many voxels (volume elements, typically small regions of the order of 2–3 mm in each direction of the brain) are associated with a single anatomical region of interest (ROI) within the brain, leading to effective replication of the time series of interest. In this paper, we propose a methodology for analyzing replicated time series, which may or may not be non-stationary. Specifically, we consider the challenge of identifying voxel and ROI specific features of the time series through a spectral time series approach. In this setting, the notion of replicates pertains to the time series recorded at different voxels in space. In particular, we consider the mean trend-removed time series from each voxel and seek to identify the common structure across all voxels. Whilst this is a non-standard application of the term 'replicate time series', it can be both a useful expression of the concepts and framework for carrying out the analyses.

Spectral analysis is a widely used technique in the analysis of time series. Estimating the sample spectrum for a single time series has been considered greatly for both stationary and non-stationary data; see, for example, [1] and [2] for stationary data and [3] and [4] for examples of non-stationary spectral analysis. Indeed, it has been considered previously for biomedical and neuroimaging time series [5]. However, there has been limited consideration given to spectral analysis in the situation where replicates of time series from multiple sources (in this case voxels) exist. More widely considered in the literature is the

[a] *Department of Mathematics and Statistics, Lancaster University, Lancaster, U.K.*
[b] *Statistical Laboratory, University of Cambridge, Cambridge, U.K.*
[*] *Correspondence to: Idris Eckley, Department of Mathematics and Statistics, Lancaster University, Lancaster, LA1 4YF, U.K.*
[†] *E-mail: i.eckley@lancaster.ac.uk*

problem of estimating the population trend, which would in this case be associated with an ROI, from replicated time series. An overview of the field is given by [6].

For the case of spectral analysis, we wish to consider mean zero (or detrended) time series where the interest lies in the form of the underlying second-order structure of the series. As we are dealing with time series from multiple voxels in the same ROI, we assume that they are all formed from the same underlying process with each series having additional variation from this population effect. We are then interested in estimating both the population effect and the variability between voxels. This will allow understanding of how the dynamics of fMRI time series operate on a whole and also the types of variability that might be seen across an ROI.

We are by no means the first to consider this problem for general time series data. In their work, based in a second-order stationary setting, [7] assume some population spectral structure common to all time series. Each series is then assumed to have some specific random effect that leads to a specific spectral structure from which the observed time series is assumed to be a realisation. They propose fitting a parametric model to estimate the population spectrum and the additional between series variability. This approach was then reconsidered by [8] who proposed fitting a semiparametric model, making model fitting more efficient. A non-parametric approach was considered by [9]. They made use of tree-structured wavelet methods to estimate both the population spectrum and the between series variability, again in a stationary setting.

A common requirement of these methods is that the observed time series are assumed to be stationary. In all three cases, the spectrum being estimated is the second-order stationary (Fourier) spectrum. Our objective in this paper is to extend the random effects model to estimating the spectral structure of locally stationary time series. Locally stationary replicated time series have been considered by [10]. In their work, they considered factor analysis of locally stationary replicated data, whereas our interest is in the *population spectral structure* of such data. The approach we take assumes that our series fit the locally stationary wavelet (LSW) process model of [4], and as such, we consider estimating the population evolutionary wavelet spectrum. Wavelets have been extensively used in the analysis of fMRI time series, see, for example, [11], and they naturally allow for non-stationarity to be included in the method.

Throughout, we assume that we have an LSW process $\{X^m_{t,T}, t = 1, \dots, T\}$ for each voxel m, where $m = 1, \dots, M$. Following the existing methods, we assume that voxels are independent but share some true underlying spectral structure, $S_j(\nu)$, for $j = 1, \dots, J$. While this is likely to be a deviation from the truth, given the spatial blurring present in fMRI data, it is a good starting point, and common to all analysis where ROIs are typically just averaged over the individual voxel time series within the region. Our aim is to estimate this structure from the estimates of the voxel-specific spectra, $S^m_j(\nu)$ for $m = 1, \dots, M$ and $j = 1, \dots, J$, along with estimates of the between voxel variability.

This paper is organised as follows: in Section 2, we provide background to the LSW framework and extend the definitions to the replicated data setting. Section 3 outlines the random effects model for our setting. The approach we take for estimating the population spectrum is given in Section 4. In Section 5, we give the results of a simulation study and in Section 6, we consider the method for the analysis of an anxiety-inducing fMRI experiment.

2. Locally stationary wavelet processes and extension to replicated data

The approach we take for estimating the spectral structure of a replicated non-stationary time series is based upon the spectral structure of LSW processes as defined by [4]. Here, we give an overview of the LSW process, extending it to the case of replicated data. For an introduction to wavelets and wavelet transforms, see [12] or [13].

Before we define an LSW process, we first recap discrete wavelets upon which an LSW process is built. We begin by recalling the low and high-pass quadrature mirror filters, $\{h_k\}_{k\in\mathbb{Z}}$ and $\{g_k\}_{k\in\mathbb{Z}}$, used in the construction of Daubechies compactly supported wavelets [14]. Then, we define discrete wavelets, $\psi_j = (\psi_{j0}, \dots, \psi_{j(N_j-1)})$, following [4] such that

$$\psi_{1n} = \sum_k g_{n-2k}\delta_{0k} = g_n,$$
$$\psi_{(j-1)n} = \sum_k h_{n-2k}\psi_{jk},$$

for $n = 0, \ldots, N_{j-1} - 1$, and where δ_{0k} is the Kronecker delta function. From this definition, we also have $N_J = (2^j - 1)(N_h - 1) + 1$, where N_h is the number of non-zero elements of $\{h_k\}$.

With the aforementioned notation in place, we may consider the definition of an LSW process. LSW processes are time series process representations built upon discrete non-decimated wavelets, $\psi_{j,k}(t)$, as defined earlier. More formally following [4], we define a voxel-specific LSW process as follows:

Definition 1
Assume we have $m = 1, \ldots, M$ voxels. For $T = 2^J$, the voxel-specific LSW process $\{X^m_{t,T}, t = 1, \ldots, T\}$ has the following representation :

$$X^m_{t,T} = \sum_{j,k} w^m_{j,k;T} \psi_{j,k}(t) \xi^m_{j,k}.$$

Here, $\xi^m_{j,k}$ is a voxel-specific orthonormal random increment sequence, $\psi_{j,k}(t)$ is a discrete non-decimated wavelet and $w^m_{j,k;T}$ is a voxel-specific amplitude. Henceforth, we will omit the indexing of the process by T, although it naturally will still be assumed.

We assume that the LSW process model given by equation (1) comes from some underlying two-stage process, an example of which is illustrated in Figure 1. In the first stage of this process, the voxel-specific amplitude function, $w^m_{j,k}$, is defined according to the random effect applied to the population amplitude. Figure 1(a) and (b) shows amplitude functions for each time and scale of a process. In Figure 1(a), we have a population amplitude function which, as can be seen in Figure 1(b), has then had a random effect applied to give a voxel-specific amplitude function. In the second stage of this model forming process,

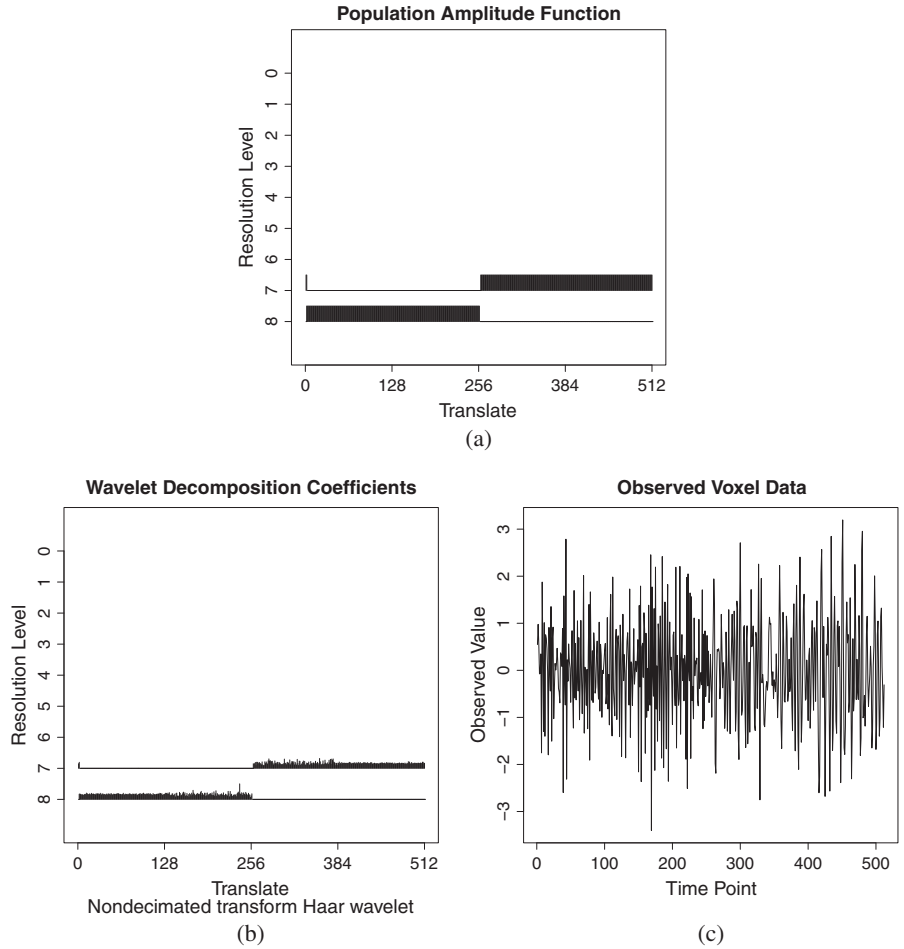

Figure 1. Illustration of the two-stage process allowing the independence assumption. The process starts with the population amplitude function (a) from which the voxel-specific amplitude (b) is obtained. Finally, observed sequence (c), determined by the stochastic within voxel variation.

the voxel-specific LSW process is obtained from the voxel-specific amplitude with the random increment sequence. Figure 1(c) shows an LSW process simulated from the voxel-specific spectral structure in Figure 1(b). By construction, these two stages are independent, and therefore, the two random processes can be treated as independent.

It is the independence of the two random processes that means that replicate data of this kind may be considered in the LSW framework and that the properties required of an LSW process are satisfied. That is, as defined by [4], the aforementioned LSW process satisfies the following properties:

(a) $\mathbb{E}(\xi_{jk}^m) = 0$; therefore, due to the independence of the random processes, $\mathbb{E}(X_{jk}^m) = 0$.
(b) $\mathrm{cov}(\xi_{jk}^m, \xi_{ln}^m) = \delta_{jl}\delta_{kn}$
(c) For each scale j there exists a Lipschitz continuous function $W_j^m(k/T)$ such that

$$\sup_k \left| w_{j,k}^m - W_j^m\left(\frac{k}{T}\right)\right| \leqslant C_j/T,$$

where the $\{C_j\}$ are a sequence of constants.

Note that the random increment sequence is considered to be the stochastic variation within a voxel, that is, the variation that would be expected by repeating, for example, the same experiment at the same voxel. The random effect, which we will denote $z_{j,k}^m$, is therefore the between voxel variation, or the variability we would expect from the true population effect relating to a specific voxel. This may be considered as being conditionally fixed as we assume that this effect would be the same for a voxel if the experiment were to be repeated, whereas the within voxel variation would change with each repetition.

2.1. *Estimating the voxel-specific evolutionary wavelet spectrum*

As in [4], a measure of the local contribution to power of an LSW process can be obtained from the evolutionary wavelet spectrum.

Definition 2
For a voxel, $m = 1, \ldots, M$, the evolutionary wavelet spectrum is defined as

$$S_j^m(v) = lim_{T\to\infty}\left(|w_{j,vT}^m|^2\right).$$

The voxel-specific evolutionary wavelet spectrum (EWS) is given by the local wavelet periodogram (LWP) as defined by [4]. Consequently, for an LSW process, X_t^m, for $t = 0, \ldots, T-1$ and for voxels $m = 1, \ldots, M$ this can be estimated as follows. First, we define the empirical wavelet coefficients of this process as

$$d_{j,k}^m = \sum_t X_t^m \psi_{j,k}(t).$$

The LWP, $I_{j,k}^m$, which forms an estimate of the voxel-specific EWS, is then obtained by taking the square of the empirical wavelet coefficients. So, we have

$$I_{j,k}^m = |d_{j,k}^m|^2. \tag{1}$$

Nason *et al.* [4] showed that this is in fact a biased and inconsistent estimator of the EWS. Specifically they established that

$$\mathbb{E}(I_{j,k}^m) = \sum_\ell A_{j,\ell} S_\ell^m(k/T) + O(T^{-1}), \tag{2}$$

and

$$\mathrm{Var}(I_{j,k}^m) = 2\left\{\sum_\ell A_{j,\ell} S_\ell^m(k/T)\right\}^2 + O(2^{-j}/T), \tag{3}$$

where we define an inner product matrix of autocorrelation wavelets

$$A_{j,\ell} = \langle \Psi_j \Psi_\ell \rangle = \sum_\tau \Psi_j(\tau)\Psi_\ell(\tau). \tag{4}$$

Fortunately, an (asymptotically) unbiased estimator is easily obtained by multiplication of the raw wavelet periodogram with the inverse of the inner product matrix. So, letting $\mathbf{I}_k^m$ be a vector of all scales of the periodogram at location k, that is, $\mathbf{I}_k^m = \left(I_{1,k}^m, I_{2,k}^m, \ldots, I_{J,k}^m\right)$, then we define the corrected estimate of the EWS to be

$$\mathbf{L}_k^m = A_J^{-1}\mathbf{I}_k^m. \tag{5}$$

As in the classical (stationary) setting, the periodogram needs to be smoothed to obtain consistency (see [4] for details).

3. A random effects model for the wavelet spectrum

In this section, we develop a random effects model to identify both the voxels-specific and population spectra. This is somewhat related to the random effects model for spectra given in [15], but here applied to non-stationary spectra.

In order to be able to separate and estimate both the local spectral structure and the between voxel variability, we need to consider the log-periodogram. To this end, let $\mathbf{Y}_k^m = \log \mathbf{L}_k^m$ so that we may write our mixed effects model in the additive form as follows:

$$\mathbf{Y}_k^m = \log\left(\mathbf{S}(\nu)\right) + \mathbf{Z}^m(\nu) + \epsilon^m. \tag{6}$$

We now have a voxel-specific random effect $\mathbf{Z}^m(\nu)$, such that $\mathbb{E}[\mathbf{Z}^m(\nu)] = 0$ and following [9], we express the variance of this random effect as $V(\nu) = \mathrm{Var}[Z^m(\nu)]$. We also have $\epsilon^m = \log E^m$.

With our mixed effects model in the form given by equation (6), the question of interest is whether we are able to use the voxel-specific spectra to estimate the population spectrum and the between voxel variability, in the form of the variance function, $V(\nu)$.

As we shall observe, in practice, the voxel-specific spectrum often contains zero, or close to zero, values. This causes problems given the need to use a log transformation to separate the effects of population and between voxel variability. Such problems also occur with the methods proposed by [7] and [9]. In the next section, we describe one approach to dealing with this important issue.

4. Estimating the population spectral structure

4.1. Methodology

We now describe the stages of the method we adopt for estimating the population wavelet spectrum in the case of fMRI data.

Step 1: Estimate the voxel-specific spectra. Assume that we have data of the form, $\{X_t^m, t = 1, \ldots, T\}$ for $T = 2^J$, $J \in \mathbb{Z}$ and for voxels $m = 1, \ldots, M$. From this data, we, first of all, wish to estimate the corrected voxel-specific periodogram, $\mathbf{L}_k^m$, as outlined in Section 2.1.
Step 2: Estimate the voxel-specific log-spectra. In Section 3, we showed that we are required to work with the log-periodogram in order to separate the fixed and random effects. As such, we must estimate the log-spectrum for each voxel. When trying to estimate the log, we have to consider the subtle, but important point, of how to deal with scale and location pairs where the spectrum may take the value zero. We outline in Section 4.2 an approach to dealing with the issue of spectral zeros. However, in the case where spectral zeros are present, the log-spectrum will be given the value *NA* at this point.
Step 3: Estimate the population spectrum. With estimates of the voxel-specific spectra, we are now able to obtain population estimates by averaging the voxel-specific log-spectra. This is possible in this case, because of the balanced design of the fMRI experiment. That is, we have

$$\tilde{S}_j(\nu) = \log\left\{\hat{S}_j(\nu)\right\} = \frac{1}{M}\sum_{m=1}^{M} \tilde{L}_{j,k}^m$$

For locations where all voxels have been set to *NA*, we also define the average log-spectrum to be *NA*. That is, for j, k such that $\tilde{L}^m_{j,k} = NA$ for all $m = 1, \ldots, M$ then $\tilde{S}_j(\nu) = NA$. The estimate of the spectrum $\hat{S}_j(\nu)$ is found by inverting the logarithm operation, giving the value zero to coefficients defined as *NA*. That is

$$\hat{S}_j(\nu) = \begin{cases} \exp\left\{\tilde{S}_j(\nu)\right\} & \text{if } \tilde{S}_j(\nu) \neq NA \\ 0 & \text{if } \tilde{S}_j(\nu) = NA \end{cases}$$

Step 4: Estimate the variance functions. Finally, we may estimate the variance functions, $V_j(\nu)$, from the voxel-specific log-spectra. Here, we simply estimate the variance across the log-spectra of all voxels. So, we have

$$\hat{V}_j(\nu) = \text{Var}\left(\tilde{L}^1_{j,k}, \ldots, \tilde{L}^M_{j,k}\right)$$

Note that due to the form of the mixed effects model (equation (6)), we do not need to invert the logarithmic transform to estimate the variance function. Additionally, if required, voxel-specific random effect estimates can also be computed at this point.

It should also be recalled that for locations where the population spectrum is equal to zero, the variance function will also take the value zero. This is because of the fact that it is not possible to estimate the between voxel variability at these locations as there is no power in the voxel spectra. So, if $\tilde{L}^1_{j,k} = \tilde{L}^2_{j,k} = \ldots = \tilde{L}^m_{j,k} = NA$ then $\hat{V}_j(\nu) = 0$.

4.2. Modelling considerations: spectral zeros

This is an issue that commonly occurs when dealing with log spectra, but has, so far, received little attention in the literature. Suppose that we have the case where the true value of the population spectrum is equal to zero.

As an illustration, consider Figure 2. If, at a particular scale, the population spectrum were to be of the form shown on the left of Figure 2, then the voxel-specific spectra would take the form shown on the right. At the locations where the population spectrum is equal to zero, the voxel-specific spectra are equal to zero for all voxels. It is clear that there is no detectable variability between voxels at these locations.

Clearly attempting to take logarithms of these values would not be informative. We, therefore, leave these values as being equal to zero and introduce a threshold, λ, above which we estimate the log-spectrum, below which we treat as being zero. In practice, to avoid confusion with the case that $L^m_{j,k} = 1$ (and therefore $\log L^m_{j,k} = 0$), we define values below the threshold for particular scale and location pairs

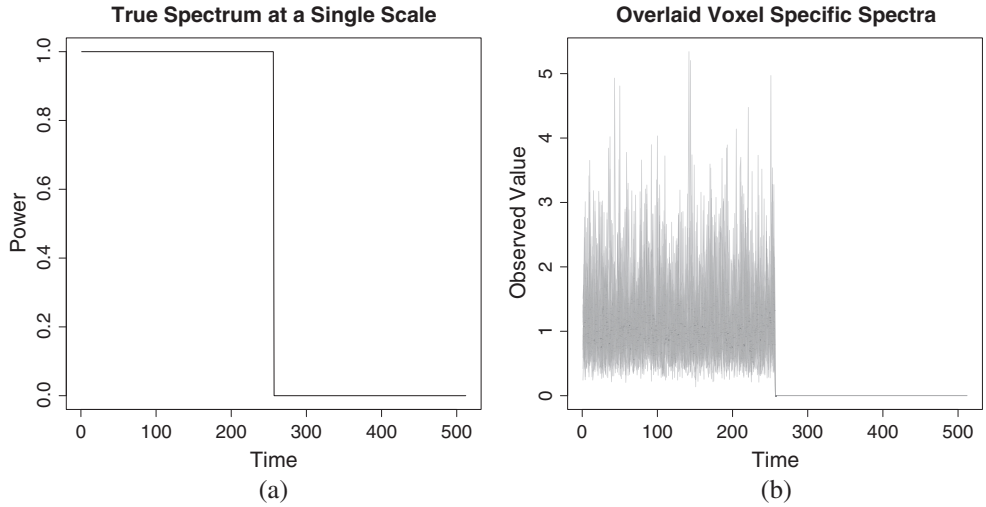

Figure 2. Example of the effect of a zero population spectrum. On the left, we have the true population spectrum for a single scale that is equal to zero where $t = 257, \ldots, 512$. On the right, the voxel-specific spectra, all of which are zero at the same locations that the population spectrum is zero.

as having no defined value, that is, *NA* to use programming parlance. More formally, we estimate the voxel-specific log-spectrum as

$$\tilde{L}^m_{j,k} = \begin{cases} \log L^m_{j,k} & \text{if } L^m_{j,k} > \lambda \\ NA & \text{if } L^m_{j,k} \leqslant \lambda \end{cases}$$

where $\tilde{L}^m_{j,k}$ is an estimate of the log-spectrum $Y^m_{j,k}$ for j, k such that $\tilde{L}^m_{j,k} \neq 0$.

The choice of threshold, λ, is clearly important. We consider the case where the innovation process underlying the LSW process is Gaussian. In this case, it is well-known that the wavelet periodogram has a (correlated) χ^2 structure. Hence, we draw on the work of [16, 17] and [4] who consider thresholding data with an χ^2 structure and adopt a cut-off value of the form $\lambda = \sigma \log(T)$. In all cases considered to date, this yields good results for both the population spectrum and the variance function.

Whilst this appears to solve the problem, it is also possible that the voxel-specific process, $Z^m_{j,k}$, from which $z^m_{j,k}$ is a realisation, leads to a voxel-specific amplitude function $w^m_{j,k} < \sqrt{\lambda}$. In this case, the process described earlier would lead to the voxel-specific spectrum at this scale and location, j, k, being set to zero, or as defined earlier in practice *NA*. However, in reality, the spectrum at this location and scale is not equal to zero. If this were to happen for a number of voxels at a specific time point, then we may question whether the population spectrum were in fact zero. Therefore, we introduce an additional parameter, β, that relates to the probability that the value of the true population spectrum is zero. If the proportion of voxels whose spectrum at this location is greater than β, we set all voxels to be zero at that location, which should be the case when the population is zero as described earlier. If the proportion is less than β, then we assume that the zeros are simply due to the voxel-specific effect.

More formally, we have

$$\tilde{L}^m_{j,k} = \begin{cases} \tilde{L}^m_{j,k} & \text{if } \frac{1}{M}\sum_{m=1}^{M} \mathbb{I}(\tilde{L}^m_{j,k} = NA) < \beta \\ NA & \text{if } \frac{1}{M}\sum_{m=1}^{M} \mathbb{I}(\tilde{L}^m_{j,k} = NA) \geqslant \beta \end{cases}$$

We suggest that values for β in the range 0.9 to 0.95 would be a sensible choice, akin to a 10 or 5% critical value in a hypothesis testing framework. Of course, this can be adjusted for knowledge of a specific application.

5. Application to simulated data

We now consider simulated examples to determine how well the method performs in a variety of settings. We report the results for a number of different population spectra and variance functions and assess the performance based upon the squared bias. Specifically, we consider the performance of the method for both stationary and non-stationary data as well as for varying complexity of structure of the population spectrum and variance function. For all of the simulation results given, we simulated 100 sets of replicated data with various combinations of numbers of voxels, M, and lengths of series, T. For all examples, we set $\beta = 0.95$.

Although the primary aim of this method is to provide a means for estimating the spectral structure of non-stationary data, our first example (Simulation 1) aims to determine how well our approach deals with stationary data. The next two cases simulate a population spectrum in the form of the Haar MA example of [4] with smaller (Simulation 2) and larger (Simulation 3) amounts of between voxel variability. We then consider a case where the amount of between voxel variability changes across a single scale in the spectral structure (Simulation 4) and a case where power exists in the population spectrum at multiple scales at a given location (Simulation 5). Finally, we consider an example where the population spectrum takes a more complex form (Simulation 6). For the exact form of the population spectrum and variance function for each simulation, see the Appendix.

Values for the squared bias and variance when estimating the population spectrum for each of the simulation cases are shown in Table I. In general, as the number of voxels increases, the bias decreases, although the variance is greater for larger numbers of voxels.

Simulation 1, in general, has a much higher bias than any of the other simulations, with the exception of Simulation 6. This would suggest that whilst the method is able to deal with stationary data, it is more suited to non-stationary data, although struggles as the complexity of the spectral structure is increased.

Table I. Bias (variance) for population spectrum. Bias increases as T increases due to the number of data points compared, but at a slower rate than the increase in T.

		$T=256$			$T=512$			$T=1024$		
	Level	$M=10$	$M=20$	$M=100$	$M=10$	$M=20$	$M=100$	$M=10$	$M=20$	$M=100$
Simulation 1	1	235 (24.5)	181 (70.1)	129 (80.7)	486 (27.6)	406 (96.3)	368 (151)	992 (32.7)	876 (146)	885 (179)
Simulation 2	1	47.3 (26)	28.8 (25.6)	12.6 (11.6)	99.5 (43)	59.3 (55.1)	12.9 (7.03)	208 (41.9)	127 (75.4)	24.2 (13.7)
	2	73.5 (102)	85.4 (94.8)	95.4 (34.3)	133 (134)	136 (144)	133 (121)	253 (189)	229 (328)	181 (141)
	3	82 (177)	95.2 (125)	105 (46.8)	150 (323)	161 (340)	177 (86.4)	285 (530)	281 (479)	288 (158)
	4	49.1 (65.5)	35.5 (83.6)	27.7 (45.7)	96.6 (72.4)	61 (115)	30.8 (47.4)	196 (153)	111 (222)	38.3 (34.8)
Simulation 3	1	47.3 (23.7)	29.8 (26.8)	13.1 (12.3)	99.1 (26.1)	60.7 (43.1)	14.2 (10.2)	210 (73.3)	128 (97.6)	25.4 (17.7)
	2	73 (86.1)	86.5 (89.1)	96.5 (42.4)	134 (130)	140 (170)	137 (92.9)	256 (301)	235 (223)	187 (173)
	3	82.7 (174)	96.2 (107)	106 (41.7)	154 (356)	163 (248)	176 (86.6)	289 (556)	289 (536)	295 (166)
	4	48.3 (74.9)	36 (81.8)	27.7 (35.9)	98.4 (128)	62.3 (142)	32.9 (49.8)	196 (183)	110 (187)	39.3 (32)
Simulation 4	1	116 (14.1)	85.1 (43.8)	47.1 (35.6)	240 (19.5)	194 (63.8)	154 (71.5)	491 (20.8)	422 (110)	399 (120)
	2	111 (19.2)	67.9 (43.7)	18 (11.3)	229 (31.3)	154 (66.2)	64.6 (57.2)	471 (43.5)	346 (148)	209 (110)
Simulation 5	1	147 (10.6)	110 (43.9)	77.3 (43.8)	302 (20.4)	244 (66.9)	206 (57.5)	616 (24.2)	528 (123)	486 (80.4)
	2	142 (21.1)	91.7 (29.1)	40.2 (23.1)	293 (24.7)	206 (79)	122 (74.3)	601 (45.5)	456 (176)	333 (136)
Simulation 6	2	561 (2010)	354 (814)	275 (71.5)	1210 (4320)	784 (1260)	586 (199)	2580 (10100)	1720 (3040)	1310 (558)
	4	577 (5910)	398 (2020)	319 (220)	1190 (10200)	794 (5150)	626 (631)	2480 (26400)	1650 (11800)	1330 (1200)

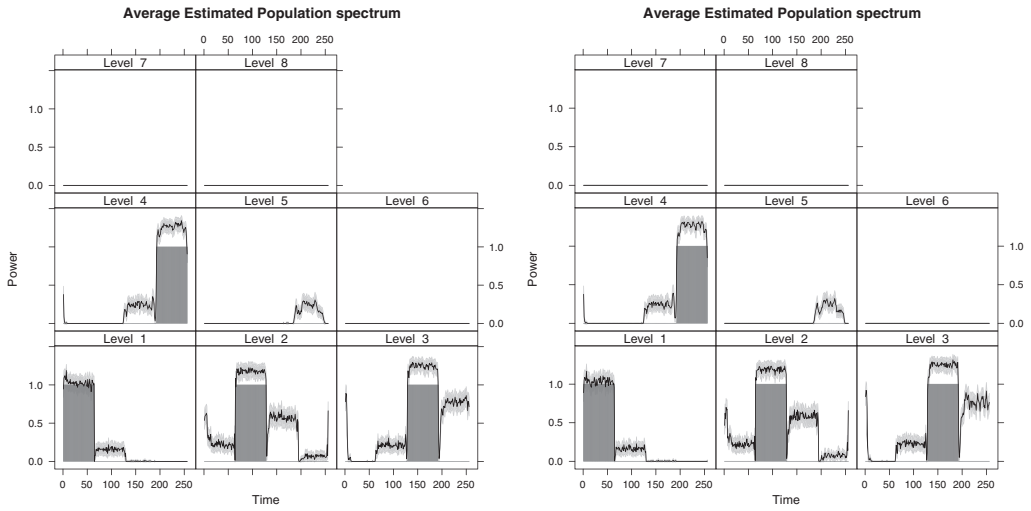

Figure 3. Plots showing the true value of the population spectrum (grey) with the mean estimated value (black) and approximate 95% confidence intervals (light grey) overlaid. Plot (a) shows the estimates for Simulation 2 and plot (b) shows the estimates for Simulation 3.

Some interesting results can be seen in Simulations 2 and 3 where in the middle levels (2 and 3) the bias increases by a considerable amount. Figure 3 shows, for Simulations 2 and 3, plots of the true population spectral structure with the mean estimated spectrum and 95% confidence intervals overlaid. In both cases, the estimates are quite reasonable at the points where power exists. It is also clear that there is some leakage of power across scales, which is most pronounced at levels 2 and 3. This may explain the reason for the much larger bias in these levels for Simulations 2 and 3. As power is leaking across scales and there is power in the levels at either side of levels 2 and 3, the leakage is most pronounced here. It is also interesting to note here that the bias changes very little despite the increase in between voxel variability from Simulation 2 to 3.

Clearly, the greater complexity in structure in Simulations 4 and 5 has had some impact with larger bias than the simpler structures.

From all of the results, it seems that the method is able to estimate the population spectrum well, although struggles more with the more complex structures as would be expected and performs less well for stationary data.

Table II shows the squared bias and variance for estimating the variance functions associated with each of the simulations discussed earlier. As with the population spectrum, the bias, in general, decreases as the number of voxels is increased. In this case, the variance is also lower for larger numbers of voxels.

Clearly, the stationary spectral structure has impacted the estimation of the variance, with Simulation 1 giving some of the larger bias values. It is interesting to note that, in fact, the largest values for bias are found to come from Simulation 4, where the between voxel variability had a more complex form, changing in value across a single scale. In this case, the method has not estimated the spectrum as well as in other cases.

Again, all of the simulations show that the method performs well and is able to obtain reasonable estimates of the variance function for both the simple and more complex cases. The method is able to estimate the variance function when the population structure is both stationary and non-stationary.

6. Functional magnetic resonance imaging data from an anxiety-inducing experiment

In this application, we consider data obtained from an experiment to induce anxiety. Functional magnetic resonance imaging relies on the observation of relative changes in blood oxygen levels in the brain. This is a result of the properties of deoxy-hemoglobin (hemoglobin not carrying oxygen) and oxy-hemoglobin (hemoglobin carrying oxygen). Deoxy-hemoglobin has the ability to suppress a magnetic signal whilst oxy-hemoglobin does not, due to their paramagnetic and dimagnetic properties, respectively. Thus, changes in oxygenation are observable over time from a sequence of brain images obtained

Table II. Bias (variance) for estimation of the variance function, that is, $\mathrm{Var}(z_j(t))$. Bias increases as T increases due to the number of data points compared, but at a slower rate than the increase in T.

		$T=256$			$T=512$			$T=1024$		
	Level	$M=10$	$M=20$	$M=100$	$M=10$	$M=20$	$M=100$	$M=10$	$M=20$	$M=100$
Simulation 1	1	15.5 (0.154)	14.4 (0.437)	11 (0.161)	31.3 (0.096)	29.7 (0.349)	27.1 (0.228)	63.1 (0.081)	61.1 (0.3)	59.8 (0.23)
Simulation 2	1	1.93 (1.26)	3.04 (1.66)	1.57 (0.051)	2.51 (1.73)	4.15 (2.47)	2.08 (0.066)	3.17 (2.42)	5.69 (4.23)	2.99 (0.1)
	2	4.49 (5.56)	6.87 (5.55)	4.07 (0.178)	5.56 (6.09)	9.01 (5.88)	5.25 (0.255)	7.6 (6.47)	12.2 (7.71)	7.17 (0.25)
	3	5.51 (8.74)	8.08 (7.48)	4.88 (0.347)	7.56 (8.51)	11.5 (10.6)	7.08 (0.537)	10.9 (14.2)	16.8 (10.8)	10.1 (0.61)
	4	4 (9.72)	5.48 (8.35)	3.04 (0.262)	6.07 (17.4)	8.33 (14.5)	4.42 (0.525)	9.18 (16.6)	13 (18.3)	6.84 (0.75)
Simulation 3	1	2.63 (1.97)	2.54 (1.21)	0.495 (0.02)	4.09 (1.26)	4.05 (1.14)	0.755 (0.01)	7.2 (1.15)	6.95 (1.21)	1.72 (0.03)
	2	4.52 (4.18)	5.74 (4.88)	2.44 (0.136)	6.14 (3.47)	7.35 (4.23)	2.77 (0.129)	10 (4.05)	11.8 (7.03)	3.46 (0.13)
	3	5.27 (7.48)	6.72 (7.63)	3.18 (0.278)	8.07 (7.66)	9.06 (7.45)	4.03 (0.275)	11.8 (8.31)	14.2 (7.86)	5.72 (0.47)
	4	4.01 (9.82)	3.38 (2.84)	1.07 (0.118)	6.4 (8.51)	5.67 (6.1)	1.28 (0.097)	10.6 (13.2)	9.99 (12.5)	1.76 (0.12)
Simulation 4	1	16.6 (0.124)	16.2 (0.659)	14.1 (0.148)	33.3 (0.16)	32.5 (0.37)	30.4 (0.184)	66.7 (0.091)	65.6 (0.29)	63.7 (0.23)
	2	16.8 (0.677)	16.4 (1.39)	12.9 (0.114)	33.2 (0.48)	32.3 (0.89)	27.3 (0.314)	66.5 (0.736)	64.7 (1.1)	57.6 (0.58)
Simulation 5	1	6.63 (0.215)	6.71 (0.565)	4.92 (0.024)	13.2 (0.41)	13 (0.553)	10.9 (0.029)	26.1 (0.106)	25.7 (0.25)	23.4 (0.03)
	2	7.08 (0.721)	7.23 (1.52)	4.17 (0.031)	13.5 (0.77)	13.5 (1.29)	9.19 (0.039)	26.6 (0.711)	26.2 (1.77)	20.3 (0.07)
Simulation 6	2	11.4 (13.3)	9.89 (6)	3.95 (0.096)	19 (9.67)	17.2 (5.87)	7.64 (0.104)	34.9 (12.5)	32.1 (12.3)	16.1 (0.16)
	4	10.1 (19.6)	8.55 (5.91)	4.15 (0.178)	18.4 (18.4)	17 (11.2)	8.82 (0.247)	34.7 (19.4)	32.1 (18.8)	18.4 (0.39)

by measuring fluctuations in magnetic properties in a strong magnetic field. More details of fMRI data and statistical analysis of such data can be found in [18].

Typically an fMRI analysis proceeds on a voxel-by-voxel basis, with correction at the end for multiple comparison [19]. However, there is a growing literature of using within and between ROI time series as models for the analysis of fMRI [5, 20, 21]. These works also include analysis of the within and between ROI dependency structures. It could be seen that our approach is using a mixture of these approaches along with notions of random effect models akin to those in [15], where a mixed effect view of spectral analysis is presented, although we do not specifically account for the spatial dependency in our analysis.

The data set which we consider here was previously analysed by [22] and concerns an anxiety-inducing experiment. Subjects firstly viewed a fixation cross for 2 min to obtain a resting baseline. They then viewed an instruction slide giving the topic of a speech they had been previously told they would have 2 min to prepare. Prior to scanning, they had been told they would have 2 min to prepare a 7-min speech that they would have to give to a panel of expert judges, although there was a small chance they would not have to give the speech. After those 2 min, a further slide was shown informing them they did not have to give the speech. This was followed by a further 2 min resting baseline. This induced both anxiety and relaxation states within the subjects. For full details, see [22] or [23] who demonstrate evidence of temporal nonstationarity for this specific data set, thus further highlighting the need for statistical methods, which can account for temporal non-stationarity.

Data was collected every 2 s for a total of 215 observations. More specifically, $T2^*$ fMRI images were acquired in a $3.12 \times 3.12\,\text{mm} \times 3\text{mm}$ array, with associated anatomical image for spatial normalisation to a template. Clustering was performed based on functional localisation using k-means clustering [22] and then associated with anatomical regions. The data we consider here comes from two regions of the brain of interest in this experiment. The first is from the rostral medial prefrontal cortex (RMPFC), which is known to be associated with anxiety. This area contains 11 voxels. As each voxel is observed in the same area of the brain, they may be assumed to have the same underlying population spectral structure. The second area is the visual cortex (VC), which is associated with task-related instructions. This cluster contains 25 voxels. As a first approximation, the voxels were assumed to be spatially independent. Spatial dependence in fMRI is present, but it is fairly weak so this is unlikely to be a strong assumption for spectral estimation (as opposed to hypothesis testing, for example). We proceed here with the analysis of a single subject's data.

It is common to assume that the data arising from fMRI experiments are (second-order) stationary with analysis following from this assumption. We wish to consider here whether this is a reasonable assumption based on the population spectral structure. Prior to our analysis, we detrended the data [19] to satisfy the required condition of our process having a zero mean and added zeros at each end to obtain the required 2^J length. As the start and end of these series are resting states, this padding is reasonable and will have limited impact on our results.

In Figure 4, we can see the results of the application of our method to the RMPFC data. In Figure 4(a), we have the population spectrum and in Figure 4(b), the estimate of the variance function by level. Recall that the variance function will only be estimated at locations where the population spectrum is non-zero, something which may occur naturally here due to the preprocessing of the data to remove effects such as cardiac and respiratory cycles and so on.

In Figure 4(a), we can see that we have power in the spectrum in the first five levels. Of most interest is the level five data. We can see here that the spectrum takes the value zero until approximately time 50 to 60 where power appears for a short time before taking the value zero again. Power reappears at this scale again at around time 120, also for a short time before returning to zero. This corresponds roughly to the time point at which the first instruction, to prepare a speech, was given and the second slide was shown telling them they did not have to give the speech, which were given at 2 and 4 min 15, respectively. These results would suggest that in this case, the second-order structure of the population effect is not stationary. This could imply change points (which is likely given the previous analysis by [22]), as well as other non-stationary variance effects, are present, as even mean changes will affect the second-order structure if not properly accounted for. This adds additional information to that gleaned from the data by [22], where only mean stationarity was investigated. If the mean were of interest, then our analysis could proceed on a two-stage basis, mean analysis first, then second-order analysis.

Figure 5 shows the results after applying our approach to the VC data. Again, the population spectrum is given in Figure 5(a) and the variance function by level is given in Figure 5(b).

First of all, consider Figure 5(a). As with the RMPFC data, we notice that the power at level five first appears at around the time of the first visual instruction and appears again around the time of the second

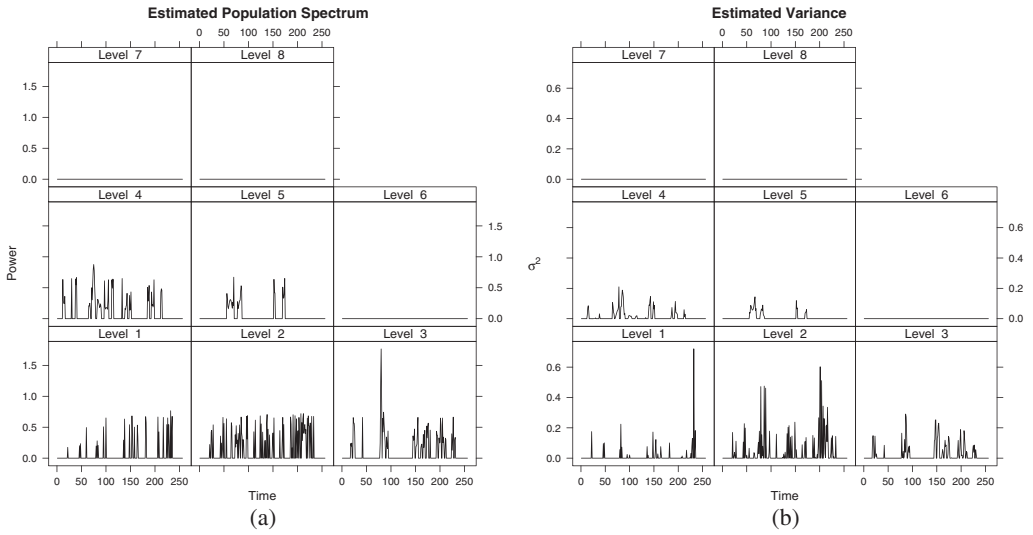

Figure 4. Results for rostral medial prefrontal cortex data. The estimate of the population spectrum is shown in (a) and the estimate of the variance function, by level, is shown in (b).

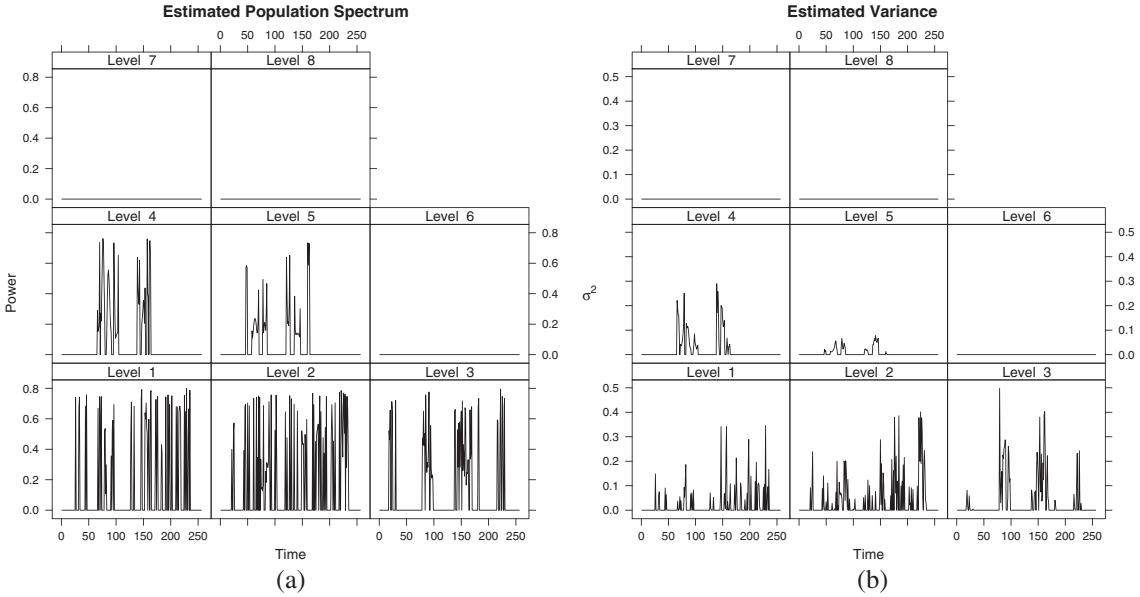

Figure 5. Results for visual cortex data. The estimate of the population spectrum is shown in (a) and the estimate of the variance function, by level, is shown in (b).

visual instruction. This effect is also visible for VC data at level four. We also notice an interesting effect at level three with, in addition to peaks at around times 60 and 120, peaks appearing at the start and the end of the time points.

As with RMPFC, these peaks appearing in the spectrum would suggest that the VC data is not second-order stationary and that possible change points are present.

7. Concluding remarks

This paper has shown how a random effects model for the spectral structure of replicated time series, such as that proposed by [7], may be extended to the locally stationary setting, which has particular relevance to fMRI. We have shown that it is possible to estimate the population LSW spectrum for replicated time series as well as being able to obtain a measure of the between voxel variability at each scale of

the spectrum. Unlike the existing methods, we focus on an approach that allows us to account for local changes in the spectral structure, as non-stationarities are known to be present in fMRI.

Our approach has been applied to simulated examples as well as fMRI data. The simulation results obtained show that the approach performs well in a variety of situations, although performs less accurately as the complexity of the population spectral structure increases, although this would be expected. When applied to fMRI data, we can see that data that may have been treated as stationary in previous analysis would seem to exhibit features suggesting that it is in fact non-stationary. Indeed, given the nature of the anxiety experiment analysed, it is highly improbable that the data would be stationary, and furthermore, the positions of change in the spectral powers appear to coincide with the experimental setup. The analysis was predicated on the choice of ROIs in the fMRI study. One interesting possible extension of the methodology would be to use determination of non-stationarities to attempt to verify whether the ROI definition is consistent with the assumption of homogeneity.

It should be noted that our methodology here has been predicated on independence between the replicates, which is a major simplification for fMRI data. However, some preliminary simulation results (data not shown) seem to indicate that for spatial correlation levels approximately similar to that of unsmoothed fMRI data, variances of the estimates are not substantially affected but biases are increased with increasing correlation. The data in the experiment was subjected to a small smoothing kernel, much smaller than the heavy smoothing often seen in fMRI analysis, and so it is likely that equivalent results hold here. Incorporating spatial dependencies into spectral estimates would be a useful and challenging future research topic both from a theoretical and methodological point of view. In addition, our approach at present is a single subject approach – this could be extended by adding another layer of hierarchy to the random effects model. However, non-stationarities are likely to be specific to an individual so a single subject approach is possibly most suitable here.

Finally, we turn to a philosophical point. Within this article, we have focussed on the development of a non-stationary wavelet framework; however, one could equally seek to identify the *population spectral structure* within a locally stationary Fourier framework. Naturally, this would give a frequency-specific analysis, as opposed to the time-frequency band decomposition afforded by the proposed wavelet framework. The development of this methodology, together with a comparison of the two approaches, is left as an interesting avenue for future research.

Appendix A

Simulation 1: Stationary data. We show here an example of the performance of our method when the data is stationary. In this case, we have a spectrum such that $w_{1,k} = 1$ for all locations k and $w_{j,k} = 0$ for all $j \neq 1$ and all k. The variance function is defined so that $V_j(\nu) = (0.4)^2$ for all scales and locations where $w_{j,k} \neq 0$.

Simulation 2: Haar MA process. For this case, we simulate a population spectrum to take the form of the Haar MA example of [4]. To this end, we define the amplitude function such that

$$w_{j,k} = \begin{cases} 1 \text{ if } j = 1 \text{ and } k \in \{1, \dots, n/4\} \\ 1 \text{ if } j = 2 \text{ and } k \in \{n/4 + 1, \dots, n/2\} \\ 1 \text{ if } j = 3 \text{ and } k \in \{n/2 + 1, \dots, 3n/4\} \\ 1 \text{ if } j = 4 \text{ and } k \in \{3n/4 + 1, \dots, n\} \\ 0 \text{ otherwise} \end{cases} \tag{7}$$

Additionally, we have, $V_j(\nu) = (0.2)^2$ for all j, k such that $w_{j,k} \neq 0$, where $\nu = k/T$.
Simulation 3: Haar MA process with greater between voxel variability. Here, we retain the same population spectrum as in the previous simulation, as defined by [7] earlier. However, we increase the amount of between voxel variability, to determine the performance of the method when the variability is larger.

So, for this simulation, we define the variance function to take the form, $V_j(\nu) = (0.4)^2$ for all j, k such that $w_{j,k} \neq 0$.

Simulation 4: Changing between voxel variability. Here, we consider our approach performance when the amount of between voxel variability is location dependent, increasing at some locations across a

given scale. We define the population spectrum to take a reasonably simple structure, with power only at two scales:

$$w_{j,k} = \begin{cases} 1 \text{ if } j = 1 \text{ and } k \in \{1, \dots, n/2\} \\ 1 \text{ if } j = 2 \text{ and } k \in \{n/2+1, \dots, n\} \\ 0 \text{ otherwise.} \end{cases} \tag{8}$$

We then define our variance function as follows:

$$V_{1,k} = \begin{cases} (0.2)^2 & \text{if } k \in \{1, \dots, n/4\} \\ (0.6)^2 & \text{if } k \in \{n/4+1, \dots, n/2\} \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

$$V_{2,k} = \begin{cases} (0.6)^2 & \text{if } k \in \{n/2+1, \dots, 3n/4\} \\ (0.2)^2 & \text{if } k \in \{3n/4+1, \dots, n\} \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

For all other scale, the variance will not be estimated as the population spectrum is zero, so for simplicity, we set $V_{j,k} = 0$ for $j \neq 1, 2$.

Simulation 5: Overlapping population power. For this simulation, we consider how the method performs when there is power in the population spectrum at more than one scale, j, at the same locations k.

For this example, we define the population amplitude function as

$$w_{j,k} = \begin{cases} 1 \text{ if } j = 1 \text{ and } k \in \{1, \dots, 5n/8\} \\ 1 \text{ if } j = 2 \text{ and } k \in \{3n/8+1, \dots, n\} \\ 0 \text{ otherwise.} \end{cases} \tag{11}$$

We define the variance function so that $V_j(v) = (0.4)^2$ for all scales and locations where $w_{j,k} \neq 0$.

Simulation 6: Non-blocky spectral structure. Our final non-stationary example considers the performance in a more complex structure where the power changes in value across the same level.

We define the population amplitude function as follows:

$$w_{j,k} = \begin{cases} \frac{1}{2} \exp\left(\frac{3n_k}{256}\right) & \text{if } j = 2, \\ \frac{1}{2} \exp\left(\frac{3(n_k-64)}{256}\right) & \text{if } j = 4, \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

where $n_k \equiv k \mod 128$. As for Simulation 4, we set the variance function to be $V_j(v) = (0.4)^2$ for all scales and locations where $w_{j,k} \neq 0$.

Acknowledgements

The authors wish to thank Professor Martin Lindquist for providing the fMRI data. Gott gratefully acknowledges EPSRC for her studentship.

References

1. Priestley M. Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society Series B* 1965; **27**:204–237.
2. Chatfield C. *The Analysis of Time Series*. Chapman & Hall: Boca Raton, 2004.
3. Dahlhaus R. Fitting time series models to nonstationary processes. *Annals of Statistics* 1997; **25**:1–37.
4. Nason GP, von Sachs R, Kroisandt G. Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *Journal of the Royal Statistical Society Series B* 2000; **62**:271–292.

5. Kang H, Ombao H, Linkletter C, Long N, Badre D. Spatio-spectral mixed-effects model for functional magnetic resonance imaging data. *Journal of the American Statistical Association* 2012; **107**(498):568–577.
6. Diggle PJ, Lang KY, Zeger SL. *Analysis of Longitudinal Data*. Oxford University Press: Oxford, 1994.
7. Diggle P, Al Wasel I. Spectral analysis of replicated biomedical time series. *Applied Statistics* 1997; **46**(1):31–71.
8. Iannaccone R, Coles S. Semiparametric models and inference for biomedical time series with extra variation. *Biostatistics* 2001; **2**(3):261–276.
9. Freyermuth J, Ombao H, von Sachs R. Tree-structured wavelet estimation in a mixed effects model for spectra of replicated time series. *Journal of the American Statistical Association* 2010; **105**(490):634–646.
10. Motta G, Ombao H. Evolutionary factor analysis of replicated time series. *Biometrics* 2012; **68**(3):825–836.
11. Bullmore E, Fadili J, Breakspear M, Salvador R, Suckling J, Brammer M. Wavelets and statistical analysis of functional magnetic resonance images of the human brain. *Statistical Methods in Medical Research* 2003; **12**:375–399.
12. Vidakovic B. *Statistical Modeling by Wavelets*. Wiley: New York, 1999.
13. Abramovich F, Bailey T, Sapatinas T. Wavelet analysis and its statistical applications. *The Statistician* 2000; **49**:1–29.
14. Daubechies I. *Ten Lectures on Wavelets*. SIAM: Philadelphia, 1992.
15. Krafty RT, Hall M, Guo W. Functional mixed effects spectral analysis. *Biometrika* 2011; **98**(3):583–598.
16. Gao HY. Wavelet estimation of spectral densities in time series analysis. *Ph.D. Thesis*, University of California, Berkeley, 1993.
17. von Sachs R, Schneider K. Wavelet smoothing of evolutionary spectra by nonlinear thresholding. *Applied and Computational Harmonic Analysis* 1996; **3**:268–282.
18. Lindquist M. The statistical analysis of fMRI data. *Statistical Science* 2008; **23**(4):439–464.
19. Worsley KJ, Liao C, Aston JAD, Petre V, Duncan G, Evans AC. A general statistical analysis for fMRI data. *Neuroimage* 2002; **15**(1):1–15.
20. Bowman FD. Spatiotemporal models for region of interest analyses of functional neuroimaging data. *Journal of the American Statistical Association* 2007; **102**(478):442–453.
21. Derado G, Bowman FD, Kilts CD. Modeling the spatial and temporal dependence in fMRI data. *Biometrics* 2010; **66**(3):949–957.
22. Lindquist M, Waugh C, Wagner T. Modeling state-related fMRI activity using change-point theory. *NeuroImage* 2007; **35**(3):1125–1141.
23. Cribben I, Haraldsdottir R, Atlas L, Wager T, Lindquist M. Dynamic connectivity regression: determining state-related changes in brain connectivity. *Neuroimage* 2012; **61**:907–920.