The Uncertainty of Storm Season Changes: Quantifying the Uncertainty of Autocovariance Changepoints

Christopher F. H. Nam[1], John A. D. Aston[1], Idris A. Eckley[2], Rebecca Killick[2]

[1]Department of Statistics, University of Warwick, Coventry, CV4 7AL, UK

[2]Department of Mathematics and Statistics, University of Lancaster, Lancaster, LA1 4YF, UK

{c.f.h.nam | j.a.d.aston }@warwick.ac.uk

{i.eckley | r.killick }@lancaster.ac.uk

February 26, 2013

Abstract

In oceanography, there is interest in determining storm season changes for logistical reasons such as equipment maintenance scheduling. In particular, there is interest in capturing the uncertainty associated with these changes in terms of the number and location of them. Such changes are associated with autocovariance changes. This paper proposes a framework to quantify the uncertainty of autocovariance changepoints in time series motivated by this oceanographic application. More specifically, the framework considers time series under the Locally Stationary Wavelet framework, deriving a joint density for scale processes in the raw wavelet periodogram. By embedding this density within a Hidden Markov Model framework, we consider changepoint characteristics under this multiscale setting. Such a methodology allows us to model changepoints and their uncertainty for a wide range of models, including piecewise second-order stationary processes, for example piecewise Moving Average processes.

Keywords: Changepoints; Hidden Markov Models; Locally Stationary Wavelet processes; Oceanography; Sequential Monte Carlo.

1 Introduction

In oceanography, historic wave height data is often used to determine storm season changes. Identifying such changes in season provides a better understanding of the data for oceanographers which may help them in planning future maintenance work of equipment such as offshore oil rigs. Changes in autocovariance structure are associated with these storm season changes and thus autocovariance changepoint methods are employed in determining these changes. However, there is often uncertainty and ambiguity associated with these changes, such as their number and location, which traditional changepoint methods often fail to capture. This paper thus proposes a methodology in which changes in autocovariance structure are considered and the uncertainty associated with such changes is captured explicitly.

Work in changepoint detection and estimation has focused on detecting changes in mean, trend (regression), variance, and combinations there of; see Chen and Gupta (2000) and Eckley et al. (2011) for overviews. However, changes in autocovariance can also occur, although there is comparatively little changepoint literature dedicated to such changes. In addition, different changepoint methods often provide different changepoint estimates, for example the number and location of changepoints, and many fail to capture explicitly the uncertainty of these estimates. As Nam et al. (2012b) argue, there is consequently a need to assess the plausibility of estimates provided by different changepoint methods in general. Quantifying the uncertainty of changepoints provides one method of doing so for this autocovariance setting.

Methods for detecting and estimating changes in autocovariance have recently been proposed. Davis et al. (2006) propose the Automatic Piecewise Autoregressive Modelling (AutoPARM) procedure which models observed time series as piecewise AR processes with varying orders and AR coefficients. Changepoints are identified via optimisation of the minimum description length criteria which provides the best segmentation configuration. However the assumption of piecewise AR processes is a strong assumption and may not always be appropriate. Uncertainty is implicitly captured via asymptotic arguments in obtaining consistent estimates of the changepoint locations, conditional on the number of changepoints being known, and thus not reported explicitly.

Cho and Fryzlewicz (2012) consider identifying changepoints in periodograms of the time series. Periodograms describe the autocovariance structure of a time series in the frequency

domain. Specifically, Cho and Fryzlewicz (2012) consider modelling time series under the Locally Stationary Wavelet (LSW) framework. Under this framework, the Evolutionary Wavelet Spectrum (EWS) describes the autocovariance structure of a time series at different scales (frequency bands) and locations. Autocovariance changepoints in the time series thus correspond to changes in the scale processes of the EWS and vice versa. Changepoint analysis now focuses on identifying changepoints in these scale processes. Cho and Fryzlewicz (2012) (CF) analyse each scale process independently for changepoints via a non-parametric test statistic, and then combine changepoint results from each scale to obtain a single set of results for the observed time series. The non-parametric test statistic places less restriction on the time series considered although several tuning parameters are required under this approach and so care is required. In general, the wavelet-domain and the LSW framework potentially allow new types of models and data to be considered which are not feasible in the time-domain. This includes moving average processes.

Figure 1 successfully illustrates the phenomenon of obtaining different changepoint results when the aforementioned changepoint methods are applied to the wave height data. This paper consequently attempts to address this discrepancy in changepoint results by quantifying the uncertainty of autocovariance (second-order) changes.

[Figure 1 about here.]

Building upon the existing wavelet-based approach of Cho and Fryzlewicz (2012), we model the time series as a LSW process and perform our analysis using the wavelet periodogram, an estimate of the EWS. We derive a joint density for scale processes of the raw wavelet periodogram which can be embedded into a Hidden Markov Model (HMM) framework, a popular framework to model non-linear and non-stationary time series. This HMM framework allows a variety of existing changepoint methods to potentially be applied (for example changes in state in the Viterbi sequence (Viterbi, 1967)), with our focus being that of quantifying the uncertainty of changepoints as proposed in Nam et al. (2012b).

Time-domain HMMs are currently unable to model some time series with changing autocovariance structures without some approximation taking place. This includes piecewise moving average processes. Our work thus proposes a wavelet based HMM framework such that time series exhibiting piecewise autocovariance structures can be considered more actively. As such,

this framework allows us to quantify explicitly the uncertainty of second-order changepoints, an aspect which is not considered in existing changepoint methods.

The structure of this paper is as follows: Section 2 reviews the statistical background that the proposed methodology is built upon. Section 3 explains the proposed framework and modelling approach. Section 4 applies the proposed framework to a variety of simulated data and the oceanographic data as presented in Figure 1. Section 5 concludes the paper.

2 Statistical Background

Let $y_1, \ldots, y_n$ denote a potential non-stationary time series, observed at equally spaced discrete time points. We assume that the non-stationarity arises due to a varying second-order structure such that for any lag $v \geq 0$, there exists a τ such that

$$\text{Cov}(Y_1, Y_v) = \ldots = \text{Cov}(Y_{\tau-1}, Y_{\tau-v}) \neq \text{Cov}(Y_\tau, Y_{\tau-v+1}) = \ldots = \text{Cov}(Y_{n-v+1}, Y_n),$$

and that the mean remains constant. In situations where the mean is not constant, pre-processing of the data can be performed. We refer to τ as a changepoint. Changes in second-order structure can be constructed easily; for example by a piecewise autoregressive moving average (ARMA) process.

One approach in modelling time series exhibiting non-stationarity such as changes in mean and variance is via Hidden Markov Models (HMMs). For overviews of HMMs, we refer the reader to MacDonald and Zucchini (1997) and Cappé et al. (2005). The use of HMMs provides a sophisticated modelling framework for a variety of problems and applications including changepoint analysis (for example Chib (1998), Aston et al. (2011)) and thus forms one of the many building blocks in our proposed methodology. Within a HMM framework, we have the observation process $\{Y_t\}_{t\geq 1}$ which is dependent on an underlying latent finite state Markov chain (MC), $\{X_t\}_{t\geq 0} \in \Omega_X$ with $|\Omega_X| < \infty$. The states of the underlying MC can represent different data generating mechanisms, for example, "stormy" and "non-stormy" seasons in the oceanographic application we are interested in. Under the HMM framework, the observations, $Y_{1:n} \equiv (Y_1, \ldots, Y_n)$ are conditionally independent given the underlying state sequence $X_{0:n} \equiv (X_0, \ldots, X_n)$.

However as Frühwirth-Schantter (2005) observe, a time-domain based HMM framework is

not suitable for directly modelling time series data arising from piecewise ARMA models as the entire underlying state sequence needs to be recorded for inference. In order to apply the desired HMM framework for such data, an alternative approach is required. In this paper, we investigate the potential of transforming the problem to an alternative domain, namely the wavelet domain.

2.1 Locally Stationary Wavelet Processes

Wavelets are compactly supported oscillating functions which permit a time series or function to be equivalently represented at different scales (frequency bands) and locations. We refer interested readers to Vidakovic (1999), Percival and Walden (2007) and Nason (2008) for comprehensive overviews of wavelets in statistics and time series analysis. Due to the time localisation properties of wavelets, they are a natural method to use when considering change-points and discontinuities in time series. Hence, in recent years, they have been used to detect changes in means (Wang, 1995), changes in variance (Whitcher et al., 2000) and changes in autocorrelation structure (Choi et al., 2008).

The Locally Stationary Wavelet (LSW) framework is a popular wavelet based modelling framework for non-stationary time series arising from a varying second-order structure (Nason et al., 2000). Following Fryzlewicz and Nason (2006), we adopt the following definition of a LSW process,

Definition 1. *$\{Y_t\}_{t=1}^n$ for $n = 1, 2, \ldots$, is said to be a Locally Stationary Wavelet (LSW) process if the following mean-square representation exists,*

$$Y_t = \sum_{j=1}^{\infty} \sum_{k \in \mathbb{Z}} \psi_{j,k}(t) W_j \left(\frac{k}{n}\right) \xi_{j,k} \tag{1}$$

where $j \in \mathbb{N}$ and $k \in \mathbb{Z}$ denote the scale and location parameters respectively. $\psi_j = \{\psi_{j,k}\}_{k \in \mathbb{Z}}$ is a discrete, real-valued, compactly supported, non-decimated wavelet vector with support lengths $\mathcal{L}_j = \mathcal{O}(2^j)$ at each scale, with $\psi_{j,k}(t) = \psi_{j,k-t}$, the wavelet shifted by t positions. $\xi_{j,k}$ is a zero-mean, orthonormal, identically distributed incremental error process.

For each $j \geq 1$, $W_j(z) : [0, 1] \to \mathbb{R}$ is a real valued, piecewise constant function with a finite (but unknown) number of jumps. Let $\mathcal{N}_j$ denote the total magnitude of the jumps in $W_j^2(z)$, the variability of function $W_j^2(z)$ is controlled so that

- $\sum_{j=1}^{\infty} W_j(z) < \infty$ *uniformly in* z.
- $\sum_{j=1}^{J} 2^j \mathcal{N}_j = \mathcal{O}(\log n)$ *where* $J = \lfloor \log_2 n \rfloor$.

By definition, an LSW process assumes Y_t has mean zero for all t. The motivation for the LSW framework is that observed time series are often non-stationary over the entire observed time period (globally), but may be stationary if we analyse them in shorter time windows (locally). Analogous to classical Fourier time series analysis, the Evolutionary Wavelet Spectrum (EWS), $W_j^2(\frac{k}{n}), j = 1, 2, \ldots$, characterises the second-order structure of the LSW process, Y_t, up to the choice of wavelet basis. Note in particular that under this definition, the EWS is piecewise constant.

Given an observed time series $\{Y_t\}_{t=1}^n$, an estimate of the EWS can be obtained by considering the square of the empirical wavelet coefficients from a non-decimated wavelet transform (NDWT) of the series . That is,

$$W_j^2\left(\frac{k}{n}\right) \approx I_{j,k} = D_{j,k}^2 = \left(\sum_{t=1}^{n} \psi_{j,k}(t) Y_t\right)^2 \tag{2}$$

This is referred to as the raw wavelet periodogram in the literature. For a sequence of random variables, $Y_{1:n}$, we denote the corresponding raw wavelet periodogram as $\mathbf{I}_{1:n}$. This is a multivariate time series consisting of $J = \lfloor \log_2 n \rfloor$ components at each location, with each component denoting a different scale. $\mathbf{i}_{1:n} = \mathbf{d}_{1:n}^2$ denotes the empirical raw wavelet periodogram corresponding to observed time series $y_{1:n}$.

Both the HMM and LSW frameworks are powerful tools in modelling different types of non-stationary time series. A natural question to thus ask is whether it is possible to combine the two frameworks such that HMM-based changepoint methods can be applied within the LSW framework. The hybrid framework would thus allow us to consider changes in second-order structure through the LSW framework, whilst applying a multitude of existing HMM-based changepoint methods (for example Chib (1998), Aston et al. (2011) and Nam et al. (2012b)). Section 3 proposes a framework in which this can be achieved.

3 Methodology

As previously described, our goal is to quantify the uncertainty of autocovariance changepoints for a time series by considering its spectral structure. Quantities of interest include the change-

point probability $P(\tau = t|y_{1:n})$ (CPP, the probability of a changepoint at time t), and the distribution of number of changepoints within the observed time series $P(N_{CP} = n_{CP}|y_{1:n})$. Other changepoint characteristics such as joint or conditional changepoint distributions are also available using the proposed methodology.

The raw wavelet periodogram characterises how the second-order structure evolves over time. Thus, we perform analysis on the periodogram to quantify the uncertainty of second-order changepoints. This is achieved by modelling the periodogram via a HMM framework, and quantifying the changepoint uncertainty via an existing HMM approach (Nam et al., 2012b). In proposing the new methodology, several challenges need to be addressed.

Firstly, the multivariate joint density of $\mathbf{I}_k$ is unknown and needs to be derived. This density also captures the dependence structure introduced by the use of the NDWT in estimating the periodogram. The derivation of this joint density and its embedding in a HMM modelling framework is detailed in Section 3.1. As the model parameters, θ, associated with the HMM framework are unknown, these need to be estimated and we turn to Sequential Monte Carlo samplers (SMC, Del Moral et al. (2006)) in considering the posterior of the parameters. These model parameters can be shown to be directly associated with the EWS. An example SMC implementation is provided in Section 3.2. Section 3.3 details some aspects concerning the computation of the distribution of changepoint characteristics. Section 3.4 provides an outline of the overall proposed approach.

There are many advantages to considering the observed time series under the LSW framework. In particular, time series exhibiting piecewise second-order structure can be more readily analysed under this framework compared to a time-domain approach. For example, for a piecewise moving average processes, the associated EWS has a piecewise constant structure at each scale; a sparser representation where the discontinuities can be analysed with fewer issues potentially arising. This sparser representation is not possible in the time-domain and is thus one of the attractions of the LSW framework.

By combining the use of wavelets in conjunction with a HMM framework, we can systematically induce a dependence structure in the HMM framework compared to choosing an arbitrary dependence structure in the time-domain.

We assume in this paper that the error process in the LSW model is Gaussian, that is

$\xi_{j,k} \overset{\text{iid}}{\sim} \text{N}(0,1)$. This leads to Y_t being Gaussian itself and is commonly referred to as a Gaussian LSW process. Recall from Section 2.1 that our EWS is piecewise constant. That is,

$$W_j^2\left(\frac{k}{n}\right) = \sum_{r=1}^{H^*} w_{j,r}^2 \mathbf{1}_{\mathcal{W}_r}(k) \qquad j = 1, \ldots, J, \tag{3}$$

where $w_{j,r}^2$ are some unknown constants, and $\mathcal{W}_r, r = 1, \ldots, H^*$ is an unknown disjoint partitioning of $1, \ldots, n$ over all scales j simultaneously. Each $\mathcal{W}_r$ has a particular EWS power structure associated with it, such that consecutive $\mathcal{W}_r$ have changes in power in at least one scale. H^* denotes the unknown number of partitions there are in the EWS, and ultimately correspond to the segments in the data and in turn the number of changepoints.

We now propose the LSW-HMM modelling framework in quantifying the uncertainty of autocovariance changepoints under the assumptions outlined above.

3.1 LSW-HMM modelling framework

Recall that an estimate of the EWS is provided by the square of the empirical wavelet coefficients under a NDWT,

$$W_j^2\left(\frac{k}{n}\right) \approx I_{j,k} = D_{j,k}^2 = \left(\sum_{t=1}^{n} \psi_{j,k}(t) Y_t\right)^2. \tag{4}$$

We consider modelling the raw wavelet periodogram across location k over the different scales j. We adopt the convention that $j = 1$ is the finest scale, and $j = 2, \ldots, J$ as the subsequent coarser scales (where $J = \lfloor \log_2 n \rfloor$). Within-scale dependence induced by the NDWT can be accounted for by the HMM framework. We refer to the collection of J periodogram coefficients at a particular time point as $\mathbf{I}_k = \{I_{j,k}\}_{j=1,\ldots,J}$ (random variable) and $\mathbf{d}_k^2 = \{d_{j,k}^2\}_{j=1,\ldots,J}$ (observed, empirical) from here onwards. We next turn to deriving the joint density of $\mathbf{I}_k$.

3.1.1 Distribution of $\mathbf{I}_k$

Recall that since we have assumed an LSW model and Gaussian innovations, Y_t is Gaussian with mean zero. By performing a wavelet transform, the wavelet coefficients $D_{j,k}$ are Gaussian distributed themselves with mean zero. As is well documented in the literature (see Nason et al. (2000)), the use of NDWT induces a dependence structure between neighbouring $D_{j,k}$. We consider in particular, $\mathbf{D}_k = \{D_{j,k}\}_{j=1,\ldots,J}$, the coefficients across J scales considered at a

given location, k. Thus,

$$\mathbf{D}_k \sim \text{MVN}(\mathbf{0}, \Sigma_k^D) \qquad k = 1, \ldots, n,$$

where Σ_k^D specifies the covariance structure between the wavelet coefficients at location k across the J scales. Section 3.1.2 discusses how Σ_k^D can be computed from the spectrum $W_j^2(\frac{k}{n})$.

As $\mathbf{I}_k = \mathbf{D}_k^2 = (D_{1,k}^2, \ldots, D_{J,k}^2)$, the following result can be established.

Proposition 1. *The density of* $\mathbf{I}_k$ *is,*

$$\begin{aligned} g(\mathbf{d}_k^2|\Sigma_k^D) &= g(d_{1,k}^2, \ldots, d_{J,k}^2|\Sigma_k^D) \\ &= \frac{1}{2^J \prod_{j=1}^J |d_{j,k}|} \sum_{a_1,\ldots,a_J=\{+,-\}} f\left(a_1|d_{1,k}|, \ldots, a_J|d_{J,k}| \middle| \mathbf{0}, \Sigma_k^D\right), \end{aligned} \tag{5}$$

where $f(\cdot|\mathbf{0}, \Sigma_k^D)$ *is the joint density corresponding to* $\text{MVN}(\mathbf{0}, \Sigma_k^D)$.

Proof. This is based on a change of variables argument detailed further in Section A.1. □

We can thus use the joint density of wavelet coefficients, $\mathbf{D}_k$, to deduce the joint density for the squared wavelet coefficients $\mathbf{I}_k = \mathbf{D}_k^2$. A similar joint density can be computed if we consider each scale process of the periodogram, $\mathbf{I}^j = \{I_{j,k}\}_{k=1}^n$, although the order of computation increases exponentially.

3.1.2 Computing Σ_k^D

We next turn to the problem of accounting for the dependence between the coefficients, induced by a NDWT, which feeds into the joint densities of $\mathbf{D}_k$ and $\mathbf{I}_k$. Recall that the EWS characterises the autocovariance structure of the observation process for any orthonormal incremental process as follows (Nason et al., 2000):

$$\text{Cov}(Y_t, Y_{t-v}) = \sum_l \sum_m W_l^2\left(\frac{m}{n}\right) \psi_{l,m}(t)\psi_{l,m}(t-v).$$

It is possible to compute this autocovariance quantity without knowing the entire EWS due to the compact support of wavelets.

As the following proposition demonstrates, the autocovariance structure of the observations also feeds into the covariance structure of the wavelet coefficients.

Proposition 2. *For a LSW process, the covariance structure between specific wavelet coefficients of a NDWT, $D_{j,k}$ is of the following form:*

$$\text{Cov}(D_{j,k}, D_{j',k'}) = \sum_t \sum_v \psi_{j,k}(t)\psi_{j',k'}(t-v)\text{Cov}(Y_t, Y_{t-v}). \tag{6}$$

Proof. See Section A.2. □

We can thus deduce the covariance structure for the wavelet coefficients $\mathbf{D}_k$, Σ_k^D, from the EWS. Again, due to the compact support associated with wavelets, only a finite number of covariances in the summation are needed to evaluate this quantity. Consequently, the entire EWS does not need to be known to calculate the covariance between the wavelet coefficients.

One can show that to compute Σ_k^D, the covariance structure of the wavelet coefficients at location k, the power from locations $k - 2(\mathcal{L}_j - 1), \ldots, k$ for scale $j = 1, \ldots, J$ needs to be recorded where $\mathcal{L}_j$ denotes the number of non-zero filter elements in the wavelet at scale j (see Section A.3).

3.1.3 The HMM framework

Having derived a joint density for the wavelet periodogram, we now turn our attention to the question of how this can be incorporated appropriately within a HMM framework. The J multivariate scale processes from a raw wavelet periodogram can be modelled simultaneously via a single HMM framework with the derived multivariate emission density. That is, at location k, we consider $\mathbf{I}_k = \{I_{j,k}\}_{j=1,\ldots,J}$, and model it as being dependent on a single underlying, unobserved Markov chain (MC), X_k, which takes values from $\Omega_X = \{1, \ldots, H\}$ with $H = |\Omega_X| < \infty$,

$$p(x_k|x_{1:k-1}, \theta) = p(x_k|x_{k-1}, \theta) \qquad k = 1, \ldots, n \quad \text{(Transition)}$$

$$\mathbf{I}_k|\{X_{1:k-1}, \mathbf{I}_{1:k-1} = \mathbf{d}^2_{1:k-1}\} \sim g(\mathbf{I}_k = \mathbf{d}^2_k|x_{k-2(\mathcal{L}_J-1):k}, \theta) \qquad k = 1, \ldots, n \quad \text{(Emission)}$$

The HMM framework assumes that the emission density of $\mathbf{I}_k$ is determined by the latent process X_k, such that the process follows the Markov property and the $\mathbf{I}_{1:n}$ are conditionally independent given $X_{1:n}$. This latter remark allows us to account for the within-scale dependence induced by a NDWT. H denotes the number of underlying states the latent MC, X_k, can take and corresponds to different data generating mechanisms, for example "stormy" and "non-stormy" seasons in the motivating oceanographic application. Under our setup, this corresponds

to the number of unique power configurations over the disjoint partitioning $\mathcal{W}_1, \ldots, \mathcal{W}_{H^*}$. That is $H \leq H^*$ is the number of states that generate the H^* partitions, with some partitions possibly being generated by the same state. We assume in our analysis that H is known a priori, as we want to give a specific interpretation to the states in the application, that of "stormy" or "non-stormy"seasons. Typically, in more general time series, this is not the case, although Zhou et al. (2012) and Nam et al. (2012a) discuss how the number of states can be deduced via the existing use of SMC samplers. We assume that the underlying unobserved MC, X_k, is first order Markov, although extensions to an q-th order Markov Chain are permitted via the use of embedding arguments.

The state-dependent emission density, $g(\mathbf{I}_k|X_{k-2(\mathcal{L}_J-1):k})$, is that proposed in Equation 5, with the covariance structure Σ_k^D being dependent on $X_{k-2(\mathcal{L}_J-1):k}$. Rather than estimating entries of Σ_k^D directly, we instead estimate the powers, $w_{j,r}^2$ as in Equation 3, that feed directly into and populate Σ_k^D. More specifically, we estimate state-dependent powers $w_{j,r}^2$ in

$$W_{j,X_k}^2\left(\frac{k}{n}\right) = \sum_{r=1}^{H} w_{j,r}^2 \mathbf{1}_{[X_k=r]} \qquad j = 1, \ldots, J. \tag{7}$$

This state-dependent power structure is equivalent to the piecewise constant EWS as in Equation 3. As X_k is permitted to move freely between all states of Ω_X, we are able to reduce the summation limit in Equation 3 to H from H^*. Returning to previous power configurations in the EWS is therefore possible, with a change in state corresponding to a change in power in at least one scale. Σ_k^D is dependent on the underlying states of X_k from times $k-2(\mathcal{L}_J-1), \ldots, k$ (see Section A.3) and thus the order of the HMM is $2\mathcal{L}_J - 1$.

Here, θ denotes the model parameters that need to be estimated which consists of the transition matrix $\mathbf{P}$ and the aforementioned state-dependent power $W^2 = \{W_{\cdot,1}^2, \ldots, W_{\cdot,H}^2\}$, where $W_{\cdot,r}^2 = \{w_{j,r}^2\}_{j=1}^J$ for all $r \in \Omega_X$, is associated with the emission density. We can thus partition the model parameters into transition and emission parameters, $\theta = (\mathbf{P}, W^2)$. As θ is unknown, we turn to SMC samplers (Del Moral et al., 2006) for their estimation. Section 3.2 outlines an example implementation in approximating the posterior of θ, $p(\theta|\mathbf{d}_{1:n}^2)$.

3.2 SMC samplers implementation

This section outlines an example SMC implementation in approximating the parameter posterior, $p(\theta|\mathbf{d}_{1:n}^2, H)$ via a weighted cloud of N particles, $\{\theta^i, U^i|H\}_{i=1}^N$, since $\theta = (\mathbf{P}, W^2)$ is

unknown. SMC samplers provide an algorithm to sample from a sequence of connected distributions via importance sampling and resampling techniques (Del Moral et al., 2006). Defining $l(\mathbf{d}^2_{1:n}|\theta, H)$ as the likelihood, and $p(\theta|H)$ as the prior of the model parameters, we can define the following sequence of distributions,

$$\pi_b(\theta) \propto l(\mathbf{d}^2_{1:n}|\theta, H)^{\gamma_b} p(\theta|H) \qquad b = 1, \ldots, B, \tag{8}$$

where $\{\gamma_b\}_{b=1}^B$ is a non-decreasing tempering schedule such that $\gamma_1 = 0$ and $\gamma_B = 1$. The sequence of distributions thus introduces the effect of the likelihood gradually such that we eventually sample from the parameter posterior of interest. We sample initially from $\pi_1(\theta) = p(\theta|H)$ either directly or via importance sampling, and continually mutate and reweight existing samples from the current distribution to sample the next distribution in the sequence. Resampling occasionally occurs to maintain stability in the approximation. Under the HMM framework, this does not require sampling the underlying state sequence due to the exact computation of the likelihood via the Forward-Backward Equations (Baum et al., 1970), which leads to a reduction in Monte Carlo sampling error.

Section A.5 provides a detailed outline of the example SMC implementation used within our framework. The general points to note are that we consider the transition probability row vectors, $p_r, r = 1, \ldots, H$ constructing transition matrix $\mathbf{P}$ independently from the inverse state dependent powers $\frac{1}{w^2_{j,r}}, j = 1, \ldots, J, r = 1, \ldots, H$. The re-parametrisation of the state dependent powers to its inverse is analogous to the re-parametrisation of variance to precision (inverse variance) in typical time-domain models. In practice, the series we consider will all contain at least a small portion of variation, and as such issues regarding zero or infinite power for particular frequencies will not arise. We initialise by sampling from a Dirichlet and Gamma prior distribution respectively for transition probability vectors and state dependent powers, and mutate according to a Random Walk Metropolis Hastings Markov kernel on the appropriate domain for each component. There is a great deal of flexibility within the SMC samplers framework with regards to the type of mutation and sampling schemes from the prior. The example implementation presented is in no way the only implementation or optimal with respect to optimising mixing and acceptance rates. However, this design provides results which appear sensible without a great deal of manual tuning.

3.3 Exact changepoint distributions

Having formulated an appropriate HMM framework to model the periodogram $\mathbf{d}^2_{1:n}$, and accounting for unknown θ via SMC samplers, it is now possible to compute the changepoint distributions of interest. Conditioned on θ, exact changepoint distributions, such as $P(\tau^{(k_{\text{CP}})} = t|\mathbf{d}^2_{1:n}, \theta, H)$, can be computed via Finite Markov Chain Imbedding (FMCI) in a HMM framework (see Aston et al. (2011) and references therein). The exact nature refers to the fact that the distributions are not subject to sampling or approximation error, conditioned on θ. The FMCI framework uses a generalised changepoint definition such that sustained changes in state are permitted (k_{CP} and k'_{CP}, which corresponds to the sustained nature of seasons lasting at least a few time periods), and turns the changepoint problem into a waiting time distribution problem for runs in the underlying state sequence. The framework is flexible and efficient such that a variety of distributions regarding changepoint characteristics can be computed.

3.4 Outline of Approach

An outline of the final algorithm is as follows:

1. Perform a NDWT to time series $y_{1:n}, n = 2^J, J \in \mathbb{N}$ to obtain the wavelet periodogram.
2. Let $\mathbf{d}^2_{1:n}$ denote the corresponding J multivariate time series from the periodogram.
3. Assuming H underlying states, model $\mathbf{d}^2_{1:n}$ by a HMM framework with the corresponding joint emission density. This density accounts for the dependence structure between scale processes.
4. Account for the uncertainty of the unknown HMM model parameters, θ, via Sequential Monte Carlo samplers. This results in approximating the posterior, $p(\theta|\mathbf{d}^2_{1:n}, H)$, by a weighted cloud of N particles $\{\theta^i, U^i|H\}_{i=1}^N$.
5. To obtain the changepoint probability of interest, approximate as follows. Let k_{CP} denotes the sustained condition under a generalised changepoint definition (Nam et al., 2012b), and $P(\tau^{(k_{\text{CP}})} = t|\mathbf{d}^2_{1:n}, \theta^i, H)$ to be the exact changepoint distribution conditional on θ^i. Then the changepoint probability is,

$$P(\tau = t|y_{1:n}) \equiv P(\tau^{(k_{\text{CP}})} = t|\mathbf{d}^2_{1:n}, H) \approx \sum_{i=1}^{N} U^i P(\tau^{(k_{\text{CP}})} = t|\mathbf{d}^2_{1:n}, \theta^i, H). \tag{9}$$

That is, the weighted average of conditional exact changepoint distributions with respect to different model parameter configurations. $P(N_{CP} = n_{CP}|y_{1:n}) \equiv P(N_{CP}^{(k_{\text{CP}})} = n_{CP}|\mathbf{d}_{1:n}^2, H)$ follows analogously.

Computationally, it is not possible to consider all J scales of the periodogram as the order of the HMM increases exponentially (see Section A.4 for further details). Consequently, we approximate the by considering $J^* \leq J$ finer scales of the periodogram, a common approach in time series analysis (see for example Cho and Fryzlewicz (2012)). This restricts our attention to changes in autocovariance structure associated at higher frequencies which seem more apparent in the oceanographic data of interest. This should therefore not hinder our proposed methodology with regards to the motivating application.

We assume for that the choice of analysing wavelet used for the transform is known a priori, and is the same as the generating wavelet. However, this is often unknown and we note that wavelet choice is an area of ongoing interest with the effect between differing generating and analysing wavelets for EWS estimation investigated in Gott and Eckley (2013).

4 Results and Applications

We next consider the performance of our proposed methodology on both simulated and oceanographic data.

We first consider simulated white noise and MA processes with piecewise second-order structures. In particular, white noise processes are considered and compared to a time-domain HMM approach because this type of process can be modelled exactly in the time-domain. Hence our proposed wavelet method should thus compliment it. The potential benefit of the proposed wavelet approach is then demonstrated on piecewise MA processes in which an exact time-domain HMM is not possible.

We also return to the motivating oceanographic application. In addition to quantifying the uncertainty of the changepoints, we demonstrate concurrence with estimates provided by other autocovariance changepoint methods and those provided by expert oceanographers.

4.1 Simulated Data

We consider simulated processes of length 512 and with defined changepoints (red dotted lines at times 151, 301 and 451). We compare our proposed method to a time-domain Gaussian Markov Mixture model on the time series itself, regardless of how the data is actually generated.

In generating our results, the following SMC samplers settings have been used; $N = 500$ samples to approximate the defined sequence of $B = 100$ distributions. The hyperparameter for the r-th transition probability vector, α_r, is a H-long vector of ones with 10 in the r-th position which encourages the underlying MC to remain in the same state. The shape and scale hyperparameters for the inverse power parameters priors are $\alpha_\lambda = 1$ and $\beta_\lambda = 1$ respectively. A linear tempering schedule, that is $\gamma_b = \frac{b-1}{B-1}, b = 1, \ldots, B$, and a baseline line proposal variance of 10 which decreases linearly with respect to the iteration of the sampler, are utilised.

We consider processes arising from two possible generating mechanisms in the time-domain, and we thus assume $H = 2$ in our HMM framework, and $k_{\text{CP}} = 20, k'_{\text{CP}} = 10$ for the required sustained change in state under our changepoint definition. $J^* = 3$ scale processes of the periodogram under a Haar LSW framework are considered, a computationally efficient setting under the conditions presented.

For the SMC implementation regarding the Gaussian Markov Mixture model, the following priors were implemented: $\mu_r \overset{\text{iid}}{\sim} \text{N}(0, 10)$, $\frac{1}{\sigma_r^2} \overset{\text{iid}}{\sim} \text{Gamma}(\text{shape} = 1, \text{scale} = 1), r = 1, 2$.

4.1.1 Gaussian White Noise Processes with Switches in Variance

The following experiment concerns independent Gaussian data which exhibits a change in variance at defined time points. It is well known that the corresponding true EWS is $W_j^2(\frac{k}{n}) = \frac{\sigma_k^2}{2^j}, j = 1, \ldots, J$. A change in variance thus causes a change in power across all scales simultaneously. No approximation is required under a time-domain approach. A realisation of the data and corresponding changepoint analysis are displayed in Figure 2. The top panel is a plot of the simulated data analysed. The second and third panel display the changepoint probability plot (CPP, the probability of a changepoint occurring at each time) under the wavelet and time-domain approaches respectively. The fourth panel presents the distribution of the number of changepoints from both approaches.

We observe that our proposed methodology has peaked and centred CPP around the defined

changepoint locations and provides similar results to the time-domain approach. This type of CPP behaviour provides an indication of the changepoint location estimates. In some instances, the wavelet approach outperforms the time-domain approach, for example the changepoint associated with time point 301 is more certain. We note that there is some significant CPP assigned to the first few time points under the wavelet approach. This arises due to a label identifiability issue common with HMMs (see Scott (2002)). As such, an additional changepoint is often detected at the start of the data and this is reflected in the changepoint distribution. Disregarding this artefact, we observe that three changepoints occurring is almost certain under the wavelet approach. This is in accordance with the time-domain approach and truth.

[Figure 2 about here.]

The results demonstrate that there is potential in providing an alternative method when dealing with this type of data as the wavelet based method identifies changepoints near the defined locations. However some differences and discrepancies do exist between the proposed wavelet approach, the truth and time-domain approach. In particular, the CPP under the proposed approach is slightly offset from the truth. However, these estimates are still in line with what we might observe in the time series realisation and compares favourably to the time-domain approach.

4.1.2 Piecewise MA processes - Piecewise Haar MA processes

The following scenario considers piecewise MA processes with changing MA order. We consider in particular piecewise Haar MA processes where the coefficients of the MA process are the Haar wavelet coefficients with a piecewise constant power structure in the EWS being present. Such processes are the types of data that our proposed methodology should perform well on and for which time-domain HMM methods require some approximation involving high (but arbitrary and fixed) AR orders. In this case, we model the observed time series as a Gaussian Markov Mixture in the time-domain (AR order is zero). This incorrect modelling approach is also equally applicable when dealing with real data where the "true" model is unknown.

Stationary Haar MA processes have constant power structure in a single scale j' of the EWS, namely $W_j^2(\frac{k}{n}) = \mathbf{1}_{[j=j']}\sigma^2, \quad j' \in \{1, \ldots, J\}$, and a Haar generating wavelet, where σ^2 is the time-domain innovation variance of the process. The equivalent time-domain representation of

this model is a MA($2^{j'}-1$) process with innovation variance σ^2 and MA coefficients determined by the Haar wavelet at scale j'. Piecewise Haar MA processes can thus be constructed by considering piecewise constant EWS. Changes in power across scales correspond to changes in MA order and changes in power within-scales correspond to changes in variance of Y_t. Nason et al. (2000) remark that any MA process can be written as a linear combination of Haar MA processes, with the wavelet representation often being sparse.

Figure 3 considers a change in order from MA(1) $\leftrightarrow$ MA(7) and constant variance $\sigma^2 = 1$. These results show the real potential of the proposed method in that the CPP are centred and peaked around the defined changepoint locations, with additional changepoint potentially being present. The potential presence of additional changepoints is also reflected in the distribution of the number of changepoints with probability assigned to these number of changepoints. In contrast, the time-domain method is unable to identify these changepoints completely due to the highly correlated nature and change of autocovariance present in the the data. This thus demonstrates that there is an advantage in considering the changepoint problem in the wavelet-domain over the time-domain, in light of incorrect model specification.

[Figure 3 about here.]

Further piecewise MA simulations were performed with regards to changing variance, and both changing variance and MA order simultaneously (results not shown here). Under such scenarios, the proposed methodology outperformed or compared favourably to the approximating time-domain approach.

4.2 Oceanographic Application

We now return to consider the oceanographic data example introduced in Section 1. Clearly there is ambiguity as to when storm seasons start and the number that have occurred. Hence there is particular interest in quantifying the uncertainty of storm seasons. We therefore apply our proposed methodology to the data from a location in the North Sea.

The analysed data is plotted in the top panel of Figure 4 along with changepoint estimates from existing change in autocovariance methods namely, Cho and Fryzlewicz (2012) (CF, blue top ticks) and Davis et al. (2006) (AutoPARM, red bottom ticks). The data consists of dif-

ferenced wave heights measured at 12 hour intervals from March 1992 - December 1994 in a central North Sea location.

The following inputs have been used to achieve the presented changepoint results in Figure 4: $J^* = 2$ corresponding to higher frequency time series behaviour (where changes are expected), and $H = 2$ states have been assumed reflecting the belief that there are "stormy" and "non-stormy" seasons. The same SMC samplers settings utilised in the simulated data analysis have been used ($N = 500$ particles, $B = 100$ distributions, linear tempering schedule). Under a sustained changepoint definition, $k_{\text{CP}} = 40$ and $k'_{\text{CP}} = 30$, have been used to reflect the general sustained nature of seasons (seasons last for at least a few weeks).

Ocean engineers have indicated that it is typical to see two changes in storm season each year occurring in the Spring (March-April) and Autumn (September-October). The results displayed in Figure 4 concur with this statement; five and six storm season changes are most likely according to the number of changepoints distribution, and with the CPP being centred and peaked around these times. The uncertainty encapsulated by the number of changepoint distribution demonstrates that there are potentially more or fewer storm seasons than five or six, although these are less certain, along with the corresponding locations.

Results also concur with changepoint estimates from the other two methods, with our method highlighting another possible configuration. A few discrepancies exist, for example the changepoint estimated in the middle of 1993 according to CF and AutoPARM. These potential changes in state do not seem sufficiently sustained for a change in season to have occurred and thus our methodology has not identified them. Lowering the associated values of k_{CP} and k'_{CP}, does begin to identify these as changepoint instances, in addition to others. Changes identified in the middle of 1992 and end of 1994 by CF and AutoPARM are suspected to be due to an insufficient number of states to account for these more subtle changes. HMM model selection methods may therefore be worth implementing, although the current two state assumption corresponds directly to "stormy" and "non-stormy" seasons, allowing the model to be easily interpreted by ocean engineers.

[Figure 4 about here.]

5 Discussion

This paper has proposed a methodology for quantifying the uncertainty of autocovariance changepoints in time series. This is achieved by considering the estimate of the Evolutionary Wavelet Spectrum which fully characterises the potentially varying second-order structure of a time series. By appropriately modelling this estimate as a multivariate time series under a Hidden Markov Model framework and deriving the corresponding multivariate emission density which accounts for the dependence structure between processes, we can quantify the uncertainty of changepoints. The uncertainty of autocovariance changepoints has not explicitly been considered by existing methods in the literature.

This methodology has been motivated by oceanographic data exhibiting changes in second-order structure (corresponding to changes in storm season) where there is interest in the uncertainty of storm season changes due to their inherent ambiguity. Our method has showed accordance with various existing changepoint methods including expert ocean engineers. Our methodology allows us to assess the plausibility and performance of changepoint estimates and provide further information in planning future operations. A few discrepancies do exist between the various methods, a potential result of the sustained changepoint definition implemented and number of states assumed in our HMM. However, the settings used to achieve the results seem valid given the oceanographic application and are more intuitive in controlling changepoint results compared to abstract tuning parameters and penalisation terms in other changepoint methods.

Results on a variety of simulated data also indicate that the methodology works well in quantifying the uncertainty of changepoint characteristics. Comparisons with a time-domain approach demonstrate the real advantage of our proposed methodology lies in considering piecewise MA processes which are not readily analysed using the HMM framework in the time-domain without some approximation taking place.

The current LSW framework assumes that the observed time series is mean zero and constant with prior detrending occurring before analysis is performed. However, as non-stationarity can also arise from changes in mean, future work would consider a modified version of the LSW framework such that changes in mean are also accounted for. This would thus provide a potentially powerful unified framework in which changes in mean and second-order structure

are analysed simultaneously.

Acknowledgements

The authors would like to thank Shell International Exploration Production for supplying the oceanographic data, and Richard Davis and Haeran Cho for providing respective codes of AutoPARM and CF algorithms.

References

Aston, J. A. D., J. Y. Peng, and D. E. K. Martin (2011). Implied distributions in multiple change point problems. *Statistics and Computing 22*, 981–993.

Baum, L. E., T. Petrie, G. Soules, and N. Weiss (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics 41*(1), pp. 164–171.

Cappé, O., E. Moulines, and T. Rydén (2005). *Inference in Hidden Markov Models*. Springer Series in Statistics.

Chen, J. and A. K. Gupta (2000). *Parametric Statistical Change Point Analysis*. Birkhauser.

Chib, S. (1998). Estimation and comparison of multiple change-point models. *Journal of Econometrics 86*, 221–241.

Cho, H. and P. Fryzlewicz (2012). Multiscale and multilevel technique for consistent segmentation of nonstationary time series. *Statistica Sinica 21*, 671–681.

Choi, H., H. Ombao, and B. Ray (2008). Sequential change-point detection methods for nonstationary time series. *Technometrics 50*(1), 40–52.

Davis, R. A., T. C. M. Lee, and G. A. Rodriguez-Yam (2006). Structural break estimation for nonstationary time series models. *Journal of the American Statistical Association 101*, 223–239.

Del Moral, P., A. Doucet, and A. Jasra (2006). Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society Series B 68*(3), 411–436.

Eckley, I., P. Fearnhead, and R. Killick (2011). Analysis of changepoint models. In D. Barber, A. Cemgil, and S. Chiappa (Eds.), *Bayesian Time Series Models*, pp. 215–238. Cambridge University Press.

Frühwirth-Schantter, S. (2005). *Finite Mixture and Markov Switching Models*. Spring Series in Statistics.

Fryzlewicz, P. and G. Nason (2006). Haar–fisz estimation of evolutionary wavelet spectra. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68*(4), 611–634.

Gott, A. and I. Eckley (2013). A note on the effect of wavelet choice on the estimation of the evolutionary wavelet spectrum. *Communications in Statistics–Simulation and Computation 42*, 393–406.

Grimmett, G. and D. Stirzaker (2001). *Probability and Random Processes* (3rd ed.). Oxford university press.

MacDonald, I. L. and W. Zucchini (1997). *Monographs on Statistics and Applied Probability 70: Hidden Markov and Other Models for Discrete-valued Time Series*. Chapman & Hall/CRC.

Nam, C. F. H., J. A. D. Aston, and A. M. Johansen (2012a). Parallel Sequential Monte Carlo samplers and estimation of the number of states in a Hidden Markov model. *CRiSM Research Report 12-23*.

Nam, C. F. H., J. A. D. Aston, and A. M. Johansen (2012b). Quantifying the uncertainty in change points. *Journal of Time Series Analysis 33*(5), 807–823.

Nason, G., R. Von Sachs, and G. Kroisandt (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *Journal of the Royal Statistical Society: Series B 62*(2), 271–292.

Nason, G. P. (2008). *Wavelet Methods in Statistics with R*. Springer Verlang.

Percival, D. B. and A. T. Walden (2007). *Wavelet Methods for Time Series Analysis*. Cambridge Series in Statistical and Probabilistic Mathematics.

Scott, S. (2002). Bayesian methods for hidden Markov models: Recursive computing in the 21st century. *Journal of the American Statistical Association 97*(457), 337–351.

Vidakovic, B. (1999). *Statistical Modeling by Wavelets*. Wiley Series in Probability and Statistics.

Viterbi, A. (1967, April). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *Information Theory, IEEE Transactions on 13*(2), 260 – 269.

Wang, Y. (1995). Jump and sharp cusp detection by wavelets. *Biometrika 82*(2), 385–397.

Whitcher, B., P. Guttorp, and D. Percival (2000). Multiscale detection and location of multiple variance changes in the presence of long memory. *Journal of Statistical Computation and Simulation 68*(1), 65–87.

Zhou, Y., A. Johansen, and J. Aston (2012). Bayesian model comparison via path-sampling sequential Monte Carlo. In *Statistical Signal Processing Workshop (SSP), 2012 IEEE*, pp. 245–248. IEEE.

A Appendix

A.1 Joint density of $\mathbf{I}_k = (I_{1,k}, I_{2,k}, \ldots, I_{J^*,k})$

The following section considers the generalised version of computing the density of a transformed random vector. X and Y denote standard random vectors here with no connections to the HMM or wavelet setup. This material is from Grimmett and Stirzaker (2001). As $(Y_1 = X_1^2, Y_2 = X_2^2) = T(X_1, X_2)$ is a many-to-one mapping, direct application of a standard change of variable via the Jacobian argument (Grimmett and Stirzaker, 2001, p. 109) is not permissible. In the one-dimensional case, the following proposition is proposed.

Proposition A-1. *Let $I_1, I_2, \ldots, I_n$ be intervals which partition $\mathbb{R}^2$, and suppose that $Y = g(x)$ where g is strictly monotone and continuously differentiable on every I_i. For each i, the function $g : I_i \to \mathbb{R}$ is invertible on $g(I_i)$ with the inverse function h_i. Then*

$$f_Y(y) = \sum_{i=1}^{n} f_X(h_i(y))|h_i'(y)|,$$

with the convention that the ith summand is 0 if h_i is not defined at y, and $h_i'(\cdot)$ is the first derivative of $h_i(\cdot)$.

Proof. See page 112 of Grimmett and Stirzaker (2001). □

Therefore,

Proposition A-2. *For $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n) = (X_1^2, X_2^2, \ldots, X_n^2)$*

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{Y}}(y_1, \ldots, y_n) = \frac{1}{2^n \prod_{i=1}^{n} |x_i|} \sum_{a_1,\ldots,a_n \in \{+,-\}} f_{\mathbf{X}}\left(a_1|x_1|, \ldots, a_n|x_n|\right).$$

Proof. Applications of Propositions A-1 and Corollary 4 (Grimmett and Stirzaker, 2001, p. 109). □

A.2 Computing Σ_k^D, the covariance structure of $\mathbf{D}_k$

This section outlines how the covariance structure of $\mathbf{D}_k = (D_{1,j} \ldots, D_{J^*,k})$, can be computed from the Evolutionary Wavelet Spectrum $W_j^2(\frac{k}{n})$.

Proposition A-3. *The autocovariance structure for the observation process, Y_t, can be characterised by the Evolutionary Wavelet Spectrum as follows:*

$$\mathrm{Cov}(Y_t, Y_{t-v}) = \sum_{l} \sum_{m} W_l^2 \left(\frac{m}{n}\right) \psi_{l,m-t} \psi_{l,m-t+v}.$$

Proof. See proof of Proposition 1 in Nason et al. (2000). □

Proof of Proposition 2. As LSW processes are assumed to have mean zero, $\mathbb{E}[Y_t] = 0$, then it follows that the wavelet coefficients are mean zero themselves since they can be seen as a linear combination of Gaussian observations. Thus $\mathbb{E}[D_{j,k}] = \mathbb{E}[D_{j',k'}] = 0$. Then

$$\begin{aligned}
\mathrm{Cov}(D_{j,k}, D_{j',k'}) &= \mathbb{E}[D_{j,k}D_{j',k'}] - \mathbb{E}[D_{j,k}]\mathbb{E}[D_{j',k'}] = \mathbb{E}[D_{j,k}D_{j',k'}] \\
&= \mathbb{E}\left[\left(\sum_t Y_t \psi_{j,k-t}\right)\left(\sum_s Y_s \psi_{j',k'-s}\right)\right] \\
&= \mathbb{E}\left[\sum_t \left(\sum_l \sum_m W_l\left(\frac{m}{n}\right)\psi_{l,m-t}\xi_{l,m}\right)\psi_{j,k-t} \sum_s \left(\sum_p \sum_q W_p\left(\frac{q}{n}\right)\psi_{p,q-s}\xi_{p,q}\right)\psi_{j',k'-s}\right] \\
&= \sum_{t,l,m,s,p,q} W_l\left(\frac{m}{n}\right)\psi_{l,m-t}\psi_{j,k-t}W_p\left(\frac{q}{n}\right)\psi_{p,q-s}\psi_{j',k'-s}\mathbb{E}[\xi_{l,m}\xi_{p,q}]/
\end{aligned}$$

$$\text{By definition,} \qquad \mathbb{E}[\xi_{l,m}\xi_{p,q}] = \begin{cases} 1, & \text{iff } l = p,\ m = q; \\ 0, & \text{otherwise.} \end{cases}$$

Thus,

$$\begin{aligned}
\mathrm{Cov}(D_{j,k}, D_{j',k'}) &= \sum_{t,l,s,m} W_l^2\left(\frac{m}{n}\right)\psi_{l,m-t}\psi_{l,m-s}\psi_{j,k-t}\psi_{j',k'-s} \\
&= \sum_t \psi_{j,k-t} \sum_s \psi_{j',k'-s} \sum_l \sum_m W_l^2\left(\frac{m}{n}\right)\psi_{l,m-t}\psi_{l,m-s}.
\end{aligned}$$

Let $s = t - v$, then

$$\begin{aligned}
\mathrm{Cov}(D_{j,k}, D_{j',k'}) &= \sum_t \psi_{j,k-t} \sum_{t+v} \psi_{j',k'-t+v} \sum_l \sum_m W_l^2\left(\frac{m}{n}\right)\psi_{l,m-t}\psi_{l,m-t+v} \\
&= \sum_t \psi_{j,k-t} \sum_v \psi_{j',k'-t+v} \sum_l \sum_m W_l^2\left(\frac{m}{n}\right)\psi_{l,m-t}\psi_{l,m-t+v} \\
&= \sum_t \sum_v \psi_{j,k-t}\psi_{j',k'-t+v}\mathrm{Cov}(Y_t, Y_{t-v}).
\end{aligned}$$

Thus,

$$\mathrm{Cov}(D_{j,k}, D_{j',k'}) = \sum_t \sum_v \psi_{j,k-t}\psi_{j',k-t+v}\mathrm{Cov}(Y_t, Y_{t-v}). \tag{10}$$

□

A.3 Determining how much of the EWS one needs to know to compute Σ_k^D

In determining how much of the EWS needs to be known when computing the covariance structure at location k, we consider the following lines of logic. Let $\mathcal{L}_j$ denote the support for the wavelet at scale j (number of non-zero filter coefficients in ψ_j). The number of non-zero

product filtering coefficients, $\psi_{l,m-t}\psi_{l,m-t+v}$, is greatest when we consider the variance of the wavelet coefficients or observations process and no lag is present ($v = 0$). We thus consider $\text{Var}(D_{j,k})$ and $\text{Var}(Y_t)$. In addition, the number of non-zero product terms will be greatest for the coarsest scale considered, J^*, with corresponding support $\mathcal{L}_{J^*}$

$\text{Var}(D_{j,k})$ will be dependent on observations $Y_k, \ldots, Y_{k-(\mathcal{L}_j-1)}$ for any scale $j = 1, \ldots, J^*$. Thus for the coarsest scale $\text{Var}(D_{J^*,k})$ will be dependent on observations $Y_k, \ldots, Y_{k-(\mathcal{L}_{J^*}-1)}$. The variance for the most distant observation $Y_{k-(\mathcal{L}_{J^*}-1)}$ is dependent on the power from the following locations: $k-(\mathcal{L}_j-1)-(\mathcal{L}_j-1), \ldots, k-(\mathcal{L}_j-1)$, for scale j. The coarsest scale requires the most power feeding into it: $W^2_{J^*}\left(\frac{k-2(\mathcal{L}_{J^*}-1)}{n}\right), \ldots, W^2_{J^*}\left(\frac{k-(\mathcal{L}_{J^*}-1)}{n}\right)$. For the most recent observation Y_k at the coarsest scale, the following power needs to be known $W^2_{J^*}\left(\frac{k-(\mathcal{L}_{J^*}-1)}{n}\right), \ldots, W^2_{J^*}\left(\frac{k}{n}\right)$.

Thus to compute Σ^D_k, the covariance structure of the wavelet coefficients at location k, we must record the power from the locations $k-2(\mathcal{L}_j-1), \ldots, k$ for scale $j = 1, \ldots, J^*$.

A.4 Order of HMM with respect to analysing wavelet and J^*

We briefly comment on the behaviour of the order of the HMM as we consider more scales and different choices in analysing wavelet. Recall that the order of the HMM is associated with the analysing wavelet considered and J^*, the number of scales considered. More specifically, the HMM order is $2\mathcal{L}_{J^*} - 1$.

For the case of the Haar wavelet, where $\mathcal{L}_j = 2, 4, 8, 16$ for $j = 1, 2, 3, 4$, the corresponding order of the induced HMM is $3, 7, 15, 31$ for $J^* = 1, 2, 3, 4$. Similarly, Daubechies Extrexmal Phase wavelets with two vanishing moment has the following supports $\mathcal{L}_j = 4, 10, 22, 46$ for $j = 1, 2, 3, 4$. The induced order of HMM is thus $7, 19, 43, 91$ for $J^* = 1, 2, 3, 4$ scale processes respectively. Thus by considering coarser scales and smoother analysing wavelets, the order of the induced HMM grows exponentially which causes computational problems eventually. The use of a Haar wavelet and only considering a few finer scale processes is thus advocated.

A.5 SMC samplers example implementation

This section describes more explicitly the SMC samplers implementation described in Section 3.2. Defining $l(\mathbf{d}^2_{1:n}|\theta, H)$ as the likelihood, and $p(\theta|H)$ as the prior of the model parameters, we can define the following sequence of distributions,

$$\pi_b(\theta) \propto l(\mathbf{d}^2_{1:n}|\theta, H)^{\gamma_b} p(\theta|H) \qquad b = 1, \ldots, B, \tag{11}$$

where $\{\gamma_b\}_{b=1}^B$ is a non-decreasing tempering schedule such that $\gamma_1 = 0$ and $\gamma_B = 1$. We could therefore sample from the sequence of distribution $\{\pi_b\}_{b=1}^B$ as follows:

Initialisation, Sampling from $\pi_1 = p(\theta|H)$: Assume independence between the transition probability matrix, $\mathbf{P}$ and the state dependent power, W^2.

$$p(\theta|H) = p(\mathbf{P}|H)p(W^2|H). \tag{12}$$

Transition Probability matrix, P: Sample each of the H transition probability rows $p_r = (p_{r1}, \ldots, p_{rH}), r = 1, \ldots, H$ independently from a Dirichlet prior distribution. As HMMs are typically associated with persistent behaviour in the same underlying state, asymmetric priors encouraging persistent behaviour are generally implemented. That is,

$$p_r \overset{\text{iid}}{\sim} \text{Dir}(\alpha_r) \qquad r = 1, \ldots, H$$
$$p(\mathbf{P}|H) = \prod_{r=1}^{H} p(p_r|H),$$

where α_r is the associated hyperparameter encouraging persistency.

State Dependent Power, W^2: Sample each of the state dependent inverse power for each scale independently from a Gamma distribution. That is,

$$\lambda_{j,r} = \frac{1}{w_{j,r}^2} \overset{\text{iid}}{\sim} \text{Gamma}(\alpha_\lambda, \beta_\lambda) \qquad j = 1, \ldots, J^*, r = 1, \ldots, H$$
$$p(\Lambda = \frac{1}{W^2}|H) = \prod_{j=1}^{J^*} \prod_{r=1}^{H} p(\frac{1}{w_{j,r}^2}|H),$$

where α_λ and β_λ are associated shape and scale hyperparameters.

Mutation and Reweighting, approximating π_b from π_{b-1}: We consider Random Walk Metropolis Hastings proposal kernels on different domains given the constraints of the parameters; $\mathbf{P}$ is a stochastic matrix, $w_{j,r}^2$ are non-negative. We consider mutating and updating components of θ separately, using the most recent value of the components (akin to Gibbs sampling). In particular, we consider the following mutation strategies to move from θ_{b-1}^i to θ_b^i, for particle i at iteration b.

Transition Probability matrix, P: Consider each of the H transition probability rows p_r separately, and mutate on the logit scale. That is, we propose moving from p_r to

p_r^P via:

$$\text{Define the current logits:} \qquad l_r = \left(l_{r1} = \log \frac{p_{r1}}{p_{rH}}, \ldots, l_{rH} = \log \frac{p_{rH}}{p_{rH}} = 0 \right), \quad (13)$$

$$\text{Proposal logits:} \qquad l_r^P = l_r + \epsilon_l \qquad \epsilon_l \sim \text{MVN}(0, \Sigma_l), \quad \text{with } l_{rH}^P = 0, \quad (14)$$

$$\text{Proposal probability vectors:} \quad p_r^P = \left(\frac{\exp l_{r1}^P}{\sum_{n=1}^H \exp l_{rn}^P}, \ldots, \frac{\exp l_{rH}^P}{\sum_{n=1}^H \exp l_{rn}^P} \right), \quad (15)$$

where Σ_l is a suitable $H \times H$ proposal covariance matrix.

State Dependent Power, W^2: Consider each of the state dependent inverse powers for each scale independently, and mutate on the log scale. That is we propose moving from $\lambda_{j,r}$ to $\lambda_{j,r}^P$ via:

$$\lambda_{j,r}^P = \exp(\log \lambda_{j,r} + \epsilon_\lambda) \qquad \epsilon_\lambda \sim \text{N}(0, \sigma_\lambda^2), j = 1, \ldots, J^*, r = 1, \ldots, H, \quad (16)$$

where σ_λ^2 is a suitable proposal variance.

Reweighting: One can show that under general conditions of SMC samplers, the re-weighting formula for particle i to approximate π_b is:

$$U_b^i = \frac{U_{b-1}^i \tilde{u}_b(\theta_{b-1}^i, \theta_b^i)}{\sum_{i=1}^N U_{b-1}^i \tilde{u}_b(\theta_{b-1}^i, \theta_b^i)} \quad (17)$$

$$\text{with } \tilde{u}_b(\theta_{b-1}^i, \theta_b^i) = \frac{\pi_b(\theta_{b-1}^i)}{\pi_{b-1}(\theta_{b-1}^i)} = \frac{l(\mathbf{d}_{1:n}^2 | \theta_{b-1}^i, H)^{\gamma_b}}{l(\mathbf{d}_{1:n}^2 | \theta_{b-1}^i, H)^{\gamma_{b-1}}}. \quad (18)$$

Final Output: We have a weighted cloud of N particles approximating the parameter posterior:

$$p(\theta | \mathbf{d}_{1:n}^2, H) \approx \{\theta_B^i, U_B^i | H\}_{i=1}^N \equiv \{\theta^i, U^i | H\}_{i=1}^N. \quad (19)$$

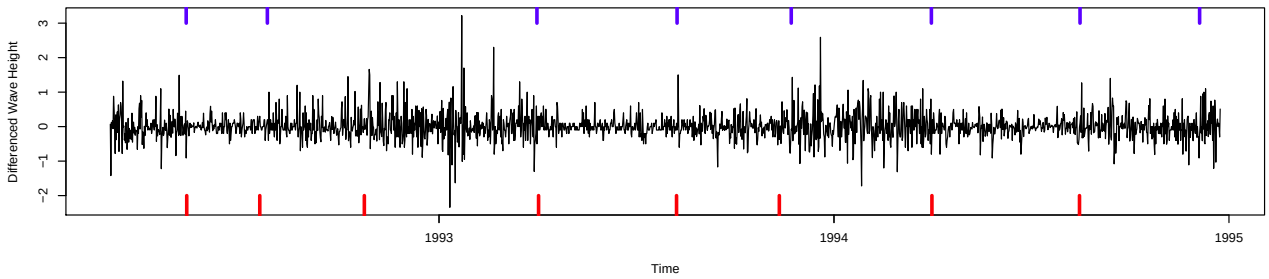

Figure 1: Changepoint estimates on wave height data from existing autocovariance changepoint methods (CF= blue top ticks, AutoPARM= red bottom ticks). The plot highlights that different changepoint methods will often provide different results, and thus there is a need to account for the discrepancies in methods by quantifying the uncertainty.

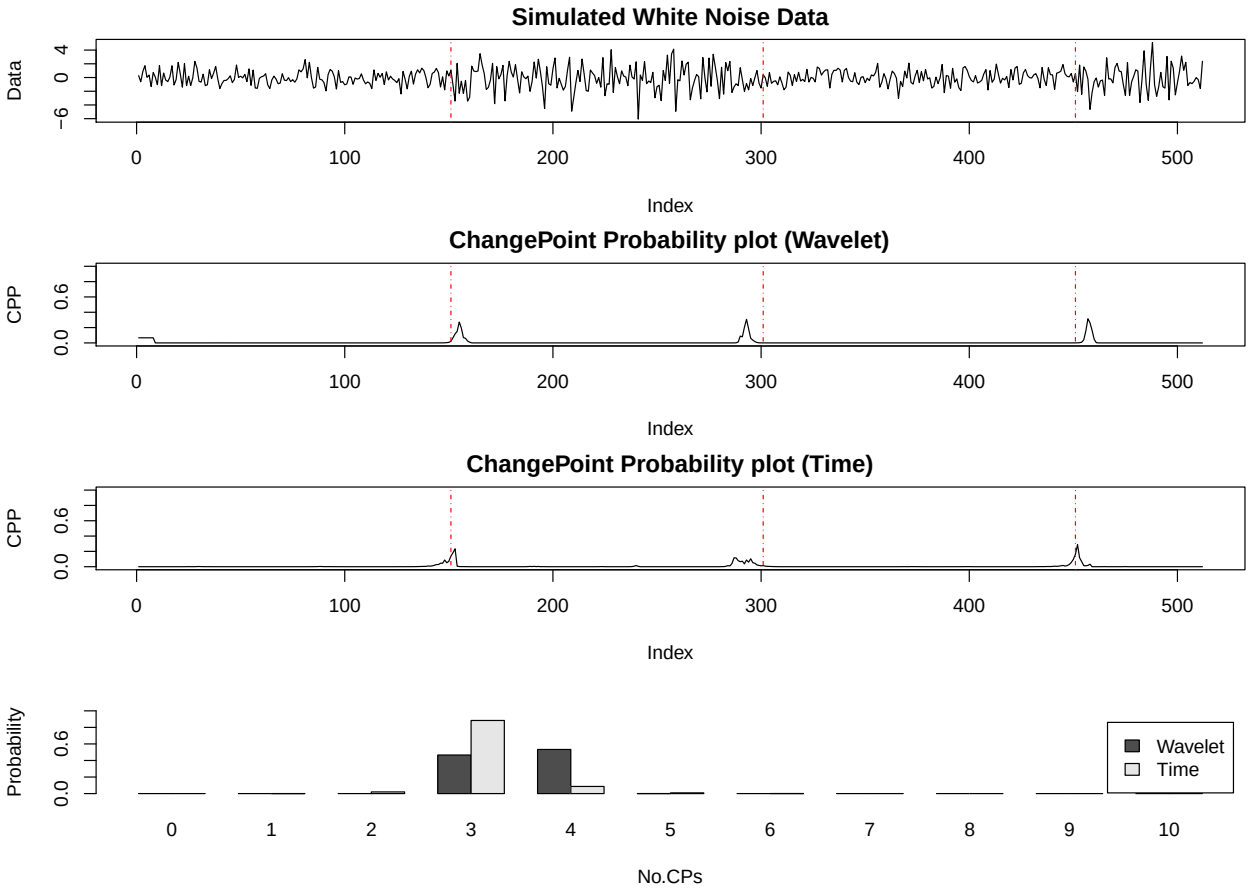

Figure 2: Changepoint results (CPP plot and distribution of number of changepoints) for simulated Gaussian white noise data with a change in variance (1 and 4). 1st panel presents the simulated data analysed. 2nd and 3rd panel displays the CPP plots under the wavelet and time-domain approaches respectively. 4th panel presents the distribution of number of changepoints. The proposed methodology compliments the time-domain approach and concurs with the truth.

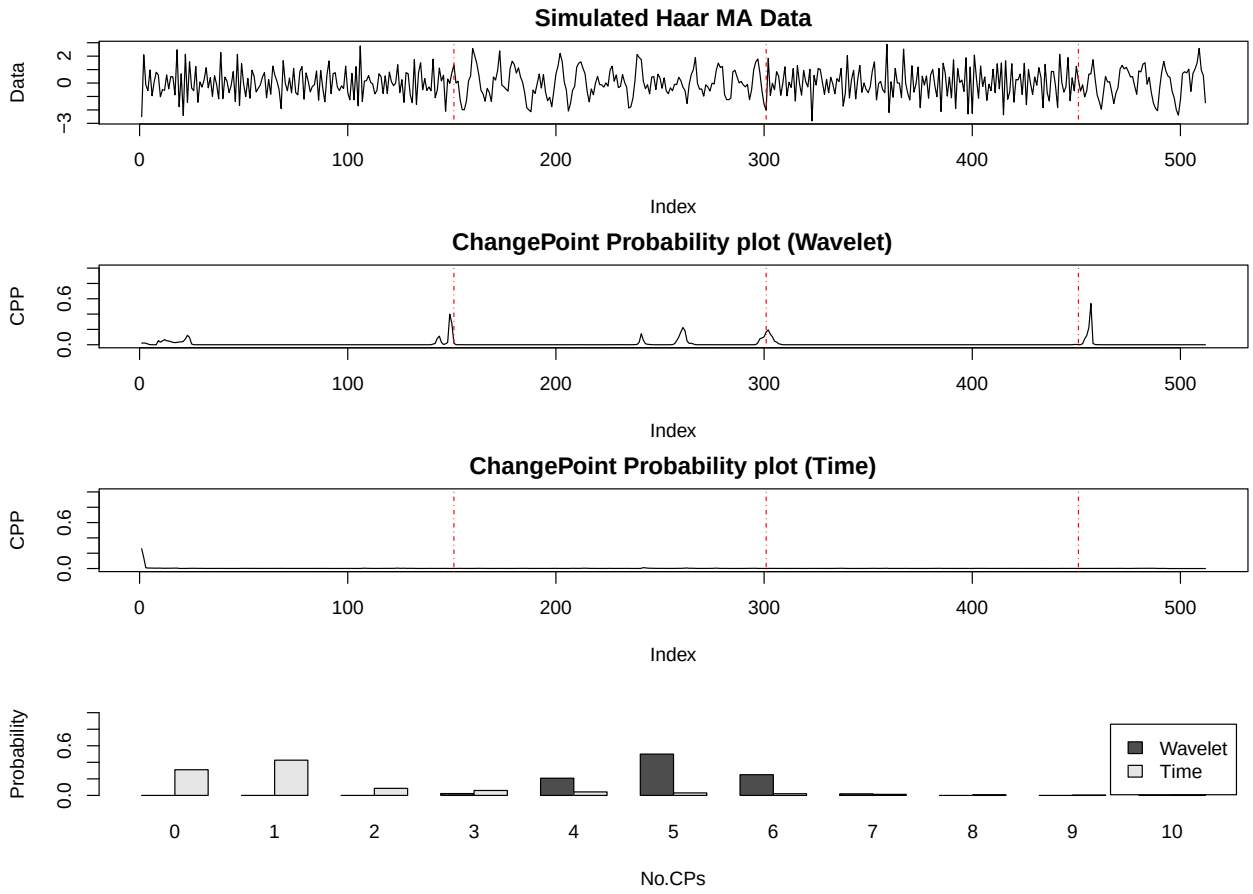

Figure 3: Changepoint results for piecewise Haar MA data with a change in order, constant variance (change in power across scales $\rightarrow$ MA(1) $\leftrightarrow$ MA(7), $\sigma^2 = 1$). 1st panel presents the simulated data analysed. 2nd and 3rd panel displays the CPP plots under the wavelet and time-domain approaches respectively. 4th panel presents the distribution of number of changepoints. The wavelet-domain approach is successfully able to identify the defined changepoint locations, in addition to other changepoints. This is reflected in the distribution of the number of changepoints. The time-domain fails to identify the changepoint characteristics however due to high autocorrelation present in the data and the change within it.

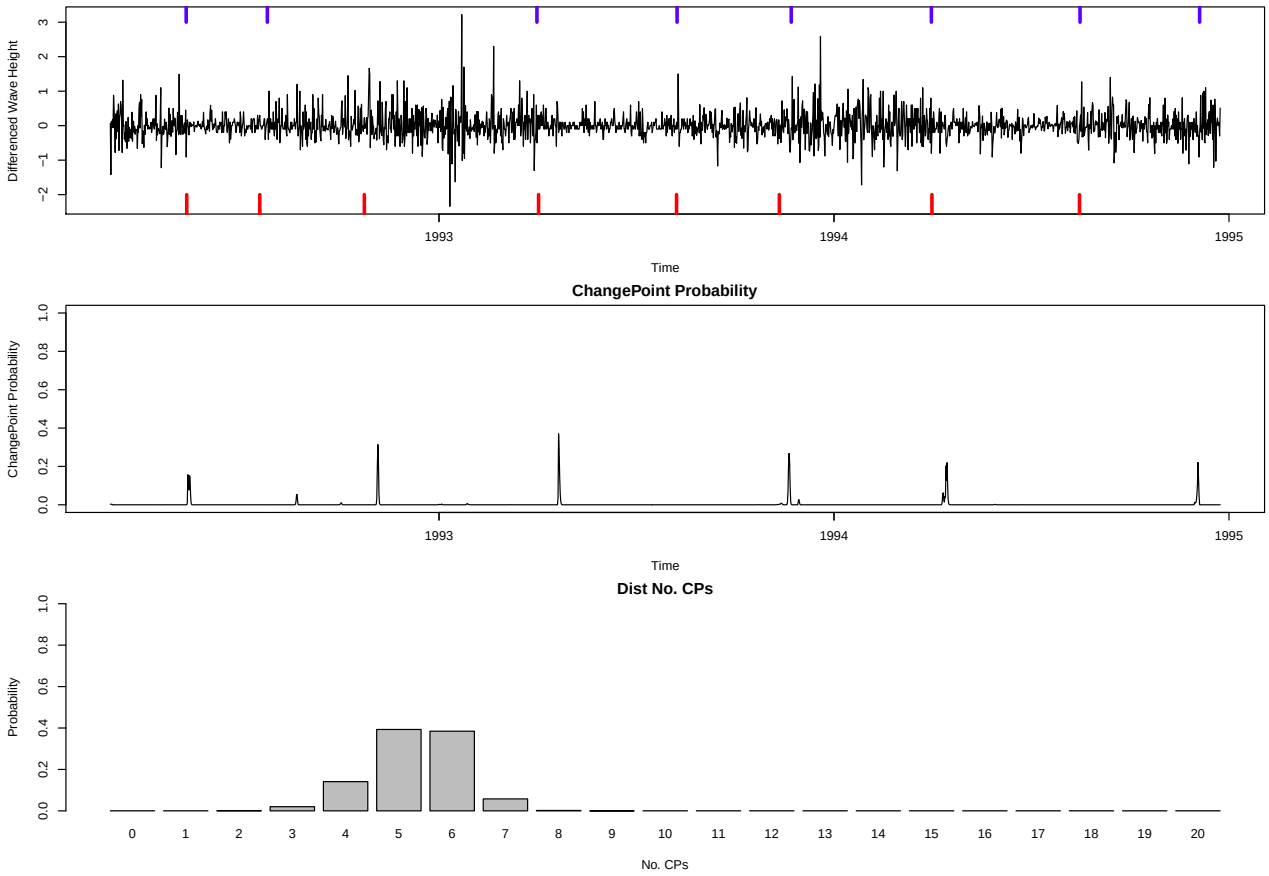

Figure 4: Changepoint results for North Sea data. Top panel displays the analysed data and the changepoint estimates from existing approaches (CF= blue top ticks, AutoPARM= red bottom ticks). Middle and bottom panel display the CPP plot and distribution of number of changepoints respectively under the proposed methodology. This corresponds to the start of storm seasons and the number of them. Analysis considers the two finest scales of the wavelet periodogram ($J^* = 2$), and assumes two underlying states ($H = 2$) reflecting "stormy" and "non-stormy" seasons.