A wavelet-based approach for changepoint detection in nonstationary time series

Killick, R. & Eckley, I. A.
Department of Mathematics & Statistics, Lancaster University, Lancaster, LA1 4YF, UK
Jonathan, P.
Shell Technology Centre Thornton, P.O. Box 1, Chester, UK

June 9, 2012

Abstract

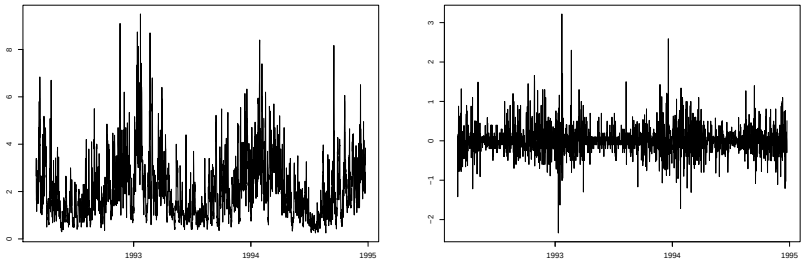
This article proposes a test to detect changes in general autocovariance structure in nonstationary time series motivated by a problem arising from ocean engineering. Our approach is founded on the locally stationary wavelet (LSW) process model for time series which has previously been used for classification and segmentation of time series. The proposed test can be performed when there is little information of the form of the autocovariance structure. We illustrate an application of our test with oceanographic data from the North Sea where we seek to identify seasons when storms are prevalent.

Keywords: changepoint, segmentation, autocovariance, local stationarity, significant wave height

1 Introduction

The work presented in this paper is motivated by a problem encountered in ocean engineering. Short-term offshore operations, such as inspection and maintenance of marine structures, are typically performed outside seasons when storms are prevalent to minimise risk. As such they require knowledge of the date of onset of the storm season, particularly if this is changing predictably in time. There is evidence (e.g. from phenology: Molau et al. (2005); Linderholm (2006); from ice cover: Hidgkins et al. (2002)) that the onset of seasons is shifting ("season creep") as a result of climate change; this is often assessed using wave height data. Often the measurements of wave heights are assumed to be independent (e.g. Ferreira and Guedes-Soares (2002)) which in many energy time series is unrealistic. As such, methods for segmenting wave height time series which may be correlated in time are required. This is the problem which we consider.

Consider, by way of example, significant wave heights for a location in the central North Sea. Figure 1(a) depicts the measured wave height at 12 hour intervals for the period March 1992 - December 1994. Ocean engineers note that whilst certain features such as the cyclic nature of the series or the fact that wave heights tend to be lower during the summer than the winter are clear, other features are harder to discern. Specifically the points at which the storm season



(a) (b)

Figure 1: (a) Significant wave heights (b) Differenced significant wave heights, for a location in the central north sea.

starts and ends are much more challenging to identify. For example methods of modelling the trend alone do not produce sufficiently accurate estimates of the storm season start and end. Use of established change in mean or change in regression methods such as those implemented in Killick and Eckley (2010) confirm this. However ocean engineers have suggested that looking at the second order structure could be more helpful since the variability of wave heights is expected to be small during calm and larger during storm seasons.

To identify whether a change in second order structure has occurred we must first remove the trend from the data. We choose to remove the trend by differencing but other methods of trend removal could be used. The resulting series is depicted in Figure 1(b). The change in second order structure is now more evident and the series appears to have a locally stationary structure. The question we therefore consider is how to detect the changes in autocovariance structure in this series and whether these changes correspond to the start and end of the storm season.

Previous work. The statistical and econometric literature contains various parametric approaches for identifying changes in second order structure. For example, Davis et al. (2006) and Gombay (2008) use the likelihood to detect changes in autoregressive (AR) models. More specifically, Davis et al. (2006) introduce the Auto-PARM algorithm for segmenting time series with a piecewise AR structure. The test statistic used in Auto-PARM is a penalised likelihood ratio, the penalty being the Minimum Description Length (MDL, Rissanen (1978)). The multiple changepoint solution space is then traversed using a genetic algorithm (see Holland (1992)) to identify both the number and location of changepoints. In practice Auto-PARM is powerful in detecting changepoints due to the likelihood-based test statistic but has varied results when the assumptions of piecewise AR structure or independence between segments are violated. Similarly, the genetic algorithm, as it is based on random mutation, can produce different results each time it is applied to a dataset.

Other, non-parametric methods, for detecting changes in second order structure have also been proposed. See for example Ombao et al. (2001) and Ahamada et al. (2004). These non-parametric approaches make less restrictive assumptions but typically have less power compared to their parametric counterparts.

Our approach. In recent years several researchers have sought to develop locally stationary time series models, see Dahlhaus (2012) for a comprehensive review of this area. The changepoint method which we propose is founded on one such modelling framework – the Locally Stationary Wavelet time series model of Nason et al. (2000). The foundational work of Nason et al. (2000), and further refined by Fryzlewicz and Nason (2006), provides a rigorous framework for modelling the type of time series one might encounter in changepoint problems of the type which we consider. Specifically we propose a likelihood-based method to identify changes in the autocovariance structure in piecewise second-order stationary time series which can be represented within the LSW framework. Such an approach is, we believe, beneficial as it enables a test to be implemented without being forced to assume that the same dependence structure, e.g. AR model, is appropriate for all locations.

Our approach is reminiscent of recent work by Cho and Fryzlewicz (2011) who propose a nonparametric test for autocovariance changepoint detection. However, in contrast to the approach which we propose, Cho and Fryzlewicz (2011) construct a test statistic assuming that the variance of the summary statistics is constant across proposed changepoint locations. Given the test we propose is parametric in nature, we would anticipate it has greater power when compared to Cho and Fryzlewicz (2011).

The article is structured as follows: Section 2 briefly reviews LSW processes before we proceed to describe our method for detecting changes in autocovariance structure in Section 3. The performance of the method is compared against Auto-PARM and the method from Cho and Fryzlewicz (2011) using simulated data in Section 4, and for detection of storm seasons from oceanographic time series in Section 5.

2 Background to locally stationary wavelet processes

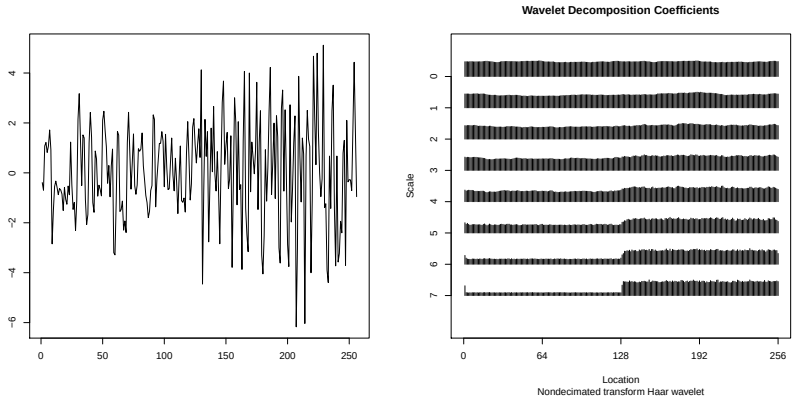
We begin by providing a brief review of the LSW modelling paradigm. Recall that a wavelet is a (compactly supported) oscillating function with several properties that allow efficient location-scale decompositions; see Vidakovic (1999) and Nason (2008) for accessible introductions.

Since its inception many authors have used the LSW framework as a building block for a variety of methodological advances including forecasting (Fryzlewicz et al., 2003) and classification (Fryzlewicz and Ombao, 2009). Below we give a brief review of the LSW approach and the properties which we will utilise in Section 3.

Following Cho and Fryzlewicz (2011), a triangular stochastic array $\{X_{t,n}\}_{t=0}^{n-1}$ for $n = 1, 2, \ldots$, is in a class of Locally Stationary Wavelet (LSW) processes if there exists a mean-square representation

$$X_{t,n} = \sum_{j=1}^{\infty} \sum_{k} W_j\left(\tfrac{k}{n}\right) \psi_{j,k-t} \xi_{jk}. \tag{1}$$

with j in $\{1, 2, \ldots\}$ and $k \in \mathbb{Z}$ as scale and location parameters, respectively. The $\psi_j = (\psi_{j,0}, \ldots, \psi_{i,\mathcal{L}_j-1})$ are discrete, real-valued, compactly supported, non-decimated wavelet vectors with support lengths $\mathcal{L}_j = \mathcal{O}(2^j)$, and the $\xi_{j,k}$



(a) (b)

Figure 2: Example of a piecewise stationary MA process (model H) with changepoint at 128 (a) Original Time Series, (b) Average Evolutionary Wavelet Spectrum estimate over 100 realisations.

are zero-mean, orthonormal, identically distributed random variables. For each $j \geq 1$, the time-varying amplitudes, $W_j(z) : [0,1] \to \mathbb{R}$ are real valued, piecewise constant functions with a finite (but unknown) number of jumps. Let $\mathcal{M}_j$ denote the total magnitude of jumps in $W_j^2(z)$, then the variability of functions $W_j(z)$ is controlled so that

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

Recall from classical time series theory that the second order structure of a time series is characterised by the spectrum. The spectrum for a LSW process is defined as $S_j(k/n) = |W_j(k/n)|^2$ and changes over time as well as scale. Nason et al. (2000) show that if X_t is a stationary process, i.e. does not change over time, then $W_j^2(k/n)$ is constant across each scale j. Thus if the second order structure of a time series is piecewise stationary, this manifests itself as piecewise constant $W_j^2(k/n)$, see Figure 2 for an example.

The specific relationship between the autocovariance structure of X_t and $W_j^2(k/n)$ is such that it is easily shown that

$$\operatorname{cov}(X_t, X_{t+v}) = \sum_{j,k} W_j^2\left(\tfrac{k}{n}\right) \psi_{j,k-t} \psi_{j,k-t-v}.$$

In the next section we utilise this decomposition to construct a test for a change in second order structure.

3 Detecting changes in second order structure

The Locally Stationary Wavelet (LSW) process can capture many dependence structures. Of particular interest in changepoint applications is the fact that

a piecewise second-order time series will have its structure encoded as piecewise constant sequences in the local wavelet periodogram - a feature noted by Cho and Fryzlewicz (2011). Consequently when using the LSW framework for changepoint methods we do not need to be prescriptive about the structure of the dependence beyond the requirements of the LSW definition.

In this article we restrict our consideration to LSW processes where the innovations $\{\xi_{j,k}\}$ in (1) are Gaussian although extensions to other distributions are possible.

Below we seek to identify a finite number of changes in autocovariance structure. This is achieved by first expressing our hypothesis for a single autocovariance change in the traditional setting. We then re-formulate the hypothesis in terms of the $W_j^2(k/n)$ parameters from the LSW model. Following this, we derive the form of the likelihood of a Gaussian LSW process under the null and alternative hypotheses which allows us to construct a test statistic in the usual way. We then conclude this section by extending the single changepoint model to multiple changes.

3.1 Detecting a single change in autocovariance structure

The traditional hypothesis formulation for testing a single autocovariance change is founded on a time $\tau \in \{1, \ldots, n\}$ where $\text{cov}\,(X_\tau, X_{\tau-v}) \neq \text{cov}\,(X_{\tau+1}, X_{\tau-v+1})$. Following Berkes et al. (2009) we formally define a change at a fixed lag v as follows.

Definition 3.1 *The hypothesis to test for a single change in second order structure at any lag $v \geq 0$ is*

$$\begin{aligned}
\boldsymbol{H_0} &: \text{cov}\,(X_0, X_{0-v}) = \text{cov}\,(X_1, X_{1-v}) = \ldots = \text{cov}\,(X_{n-1}, X_{n-1-v}) = \rho_{0,v} \\
\boldsymbol{H_1} &: \rho_{1,v} = \text{cov}\,(X_0, X_{0-v}) = \ldots = \text{cov}\,(X_\tau, X_{\tau-v}) \\
&\qquad \neq \text{cov}\,(X_{\tau+1}, X_{\tau+1-v}) = \ldots = \text{cov}\,(X_{n-1}, X_{n-1-v}) = \rho_{n,v}.
\end{aligned}$$

For notational simplicity we assume that $X_{-v}, \ldots, X_{-1}$ are observed.

Typically an upper bound is placed on v to ensure that identifiability problems do not occur but this should not be so small that the model fails to capture the second order structure adequately (e.g. using $v = 0$ when there is dependence). Auto-PARM and similar time domain methods require the practitioner to decide how many lags, v should be considered. However, if we assume that the time series X_t can be well modelled by a LSW process then the autocovariance structure across all lags is defined in terms of the $W_j^2(k/n)$. In addition, if the time series X_t is stationary (H_0) then $W_j^2(k/n) = \gamma_j$ at each scale j. In other words, under the null hypothesis the spectrum is constant over time and hence only one parameter per scale need be estimated. We therefore propose an equivalent form of the hypothesis provided in Definition 3.1, given here:

Definition 3.2 *The hypothesis test for a single change in autocovariance structure for a LSW time series can be stated as follows:*

$$\begin{aligned}
\boldsymbol{H_0} &: W_j^2\left(\tfrac{0}{n}\right) = W_j^2\left(\tfrac{1}{n}\right) = \ldots = W_j^2\left(\tfrac{n-1}{n}\right) = \gamma_{0,j} \qquad \forall j \\
\boldsymbol{H_1} &: \gamma_{1,j} = W_j^2\left(\tfrac{0}{n}\right) = \ldots = W_j^2\left(\tfrac{\tau}{n}\right) \neq W_j^2\left(\tfrac{\tau+1}{n}\right) = \ldots = W_j^2\left(\tfrac{n-1}{n}\right) = \gamma_{n,j},
\end{aligned}$$

for some $j \in \{1, 2, \ldots, J = \lfloor \log_2 n \rfloor\}$.

Given this alternative definition we must now derive the form of the likelihood under the null and alternative hypotheses (Section 3.1.2). Before we consider this, let us first describe the likelihood of a general LSW process with Gaussian innovations.

3.1.1 Likelihood of a LSW process with Gaussian innovations

Let $\boldsymbol{x} = \{x_1, \ldots, x_n\}$ be an observed time series which is a LSW process with Gaussian innovations $\{\xi_{j,k}\}$. Then the log-likelihood of the time series can be expressed as follows:

$$\ell(W|\boldsymbol{x}) = \tfrac{n}{2}\log 2\pi - \tfrac{1}{2}\log|\Sigma_W| - \tfrac{1}{2}\boldsymbol{x}'\Sigma_W^{-1}\boldsymbol{x}, \tag{2}$$

where the variance-covariance matrix, Σ_W, has elements;

$$\Sigma_W(k,k') = \operatorname{cov}(X_k, X_{k'}) = \sum_{l,m} W_l^2\left(\tfrac{m}{n}\right)\psi_{l,m-k}\psi_{l,m-k'}.$$

Here $\Sigma_W(k,k')$ denotes the element in the k-th row and k'-th column of the Σ_W matrix. Henceforth, for notational simplicity, we shall omit the indexing of Σ by W although its dependence on W will still be implicit.

We now turn to consider the formulation of the likelihood under the null and alternative hypotheses.

3.1.2 Likelihood-based test statistic

Recall our equivalent form of the single changepoint hypothesis (Definition 3.2). Under the null hypothesis the spectrum of a stationary process is constant, i.e. $W_j^2(k/n) = \gamma_j$ for all locations k and scales j. Similarly under the alternative hypothesis the spectrum is piecewise constant. Hence, given a time series $\boldsymbol{x} = \{x_t\}_{t=1}^n$ which can be modelled as a LSW process with Gaussian innovations the likelihood ratio test statistic for a change in autocovariance structure is given by

$$\lambda_\tau = \max_{J<\tau<n-J}\left\{\log\left|\hat{\Sigma}_0\right| + \boldsymbol{x}'\hat{\Sigma}_0^{-1}\boldsymbol{x} - \log\left|\hat{\Sigma}_1\right| - \boldsymbol{x}'\hat{\Sigma}_1^{-1}\boldsymbol{x}\right\}$$

where $J = \lfloor \log_2 n \rfloor$. Here $\hat{\Sigma}_0$ and $\hat{\Sigma}_1$ are the maximum likelihood estimates of the null and alternative variance-covariance matrices respectively. The entries of these matrices are given by,

$$\hat{\Sigma}_0(k,k') = \sum_l \sum_m \hat{\gamma}_{0,l}\psi_{l,m-k}\psi_{l,m-k'}, \tag{3}$$

and

$$\hat{\Sigma}_1(k,k') = \sum_l \left[\sum_{m\le\tau} \hat{\gamma}_{1,l}\psi_{l,m-k}\psi_{l,m-k'} + \sum_{m>\tau} \hat{\gamma}_{n,l}\psi_{l,m-k}\psi_{l,m-k'}\right], \tag{4}$$

where $\hat{\gamma}_{0,l}$, $\hat{\gamma}_{1,l}$ and $\hat{\gamma}_{n,l}$ are the maximum likelihood estimates of the null, pre-change and post-change transfer function, $W_l^2(m/n)$ respectively.

Unfortunately, due to the structure of Σ, the maximum likelihood estimates of $W_j^2(k/n)$ have no closed form. However several authors (Nason et al., 2000;

Fryzlewicz and Nason, 2006; Van Bellegem and von Sachs, 2008) have derived alternative means of estimating $\{W_j^2(k/n)\}$. We turn to these to provide plug-in estimates as an alternative for $\hat{\gamma}$. In what follows we estimate $W_j^2(k/n)$ by averaging the empirical estimates from Nason et al. (2000) within each segment and using these as plug-in estimates in 3 and 4. Due to the use of plug-in estimates the test statistic is thus a likelihood-based test rather than a formal likelihood ratio test.

3.2 Multiple changes in second order structure

Thus far we have considered the problem of detecting a single changepoint. More generally we are interested in detecting (potentiall) multiple changes. Several algorithms have been proposed in the changepoint literature which extend single changepoint tests to multiple changes (e.g. Scott and Knott (1974); Auger and Lawrence (1989); Inclan and Tiao (1994); Bai and Perron (2003); Killick et al. (2011)). Typically the consistency of these search methods are proved for additive error whereas the LSW model has a multiplicative error structure. However, Cho and Fryzlewicz (2011) extend the consistency of the Binary Segmentation algorithm (Scott and Knott, 1974) to multiplicative errors and thus we implement this algorithm here.

In essence, the Binary Segmentation algorithm (Algorithm 1) applies the single changepoint test and upon identifying a change, iteratively implements the test statistic on sub-segments of the time series.

In practice, we require an upper bound on the number of changepoints, m in order that the model remains identifiable. Thus we assume that there exists an upper bound, Q, on the number of changepoints in the series such that $J(Q+1) < n = 2^J$, which when rearranged gives $Q < (2^J - J)/J$.

Input: A set of data of the form, $(x_1, x_2, \ldots, x_n)$.
A test statistic $\lambda(\cdot)$ dependent on the data.
An estimator of changepoint position $\hat{\tau}(\cdot)$.
A rejection threshold (penalty) c.

Initialise: Let $\mathcal{C} = \emptyset$, and $\mathcal{S} = \{[1, n]\}$

Iterate while $\mathcal{S} \neq \emptyset$

1. Choose an element of $\mathcal{S}$; denote this element as $[s, t]$.
2. If $\lambda(x_{s:t}) < c$, remove $[s, t]$ from $\mathcal{S}$.
3. If $\lambda(x_{s:t}) \geq c$ then:
   (a) remove $[s, t]$ from $\mathcal{S}$;
   (b) calculate $r = \hat{\tau}(x_{s:t}) + s - 1$, and add r to $\mathcal{C}$;
   (c) if $r \neq s$ add $[s, r]$ to $\mathcal{S}$;
   (d) if $r \neq t - 1$ add $[r + 1, t]$ to $\mathcal{S}$.

Output the set of changepoints recorded $\mathcal{C}$.

Algorithm 1: The Binary Segmentation Algorithm.

3.3 Penalty Choice

It is common in the changepoint literature to introduce a penalty to guard against overfitting (too many changepoints). When consideration extends to (potentially) multiple changepoints the penalty typically takes the form of a function dependent on the number of changepoints, e.g. $\beta f(m)$ where β is a constant and $f(\cdot)$ is a function of the number of changepoints m. See Lavielle (2005); Chen et al. (2006); Harchaoui and Levy-Leduc (2010) for examples.

The approach we adopt follows the work of Lavielle (2005); a data driven method which is intuitive whilst also conveying the complexity of the decision through a graphical manner. The intuition behind the approach is the same as that of parameter selection in linear modelling, i.e. when a true changepoint (informative parameter) is included the negative log-likelihood will decrease by a larger amount than when a false changepoint (uninformative parameter) is included. Hence, for a given sequence of data, we seek to identify the point at which the negative log-likelihood ceases to reduce significantly. More formally, let K_{max} be an upper bound on the number of segments; $1 \leq K \leq K_{max}$; $\boldsymbol{\tau}$ be the ordered locations of changepoints; $J(\boldsymbol{\tau}, \boldsymbol{x})$ be the negative log-likelihood for time series $\boldsymbol{x}$ with changepoints $\boldsymbol{\tau}$; and $\hat{\boldsymbol{\tau}}_K$ the locations of K changepoints that minimise $J(\boldsymbol{\tau}_K, \boldsymbol{x})$. Then, the approach can be summarised as follows:

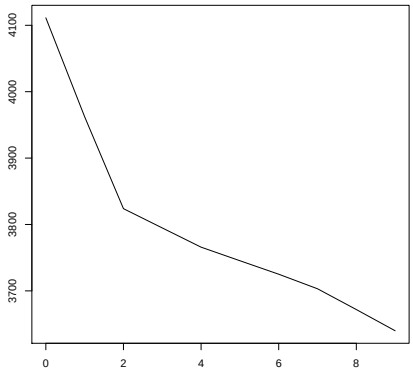
1. For each $K \in \{0, \ldots, K_{max}\}$ calculate $\hat{\boldsymbol{\tau}}_K$ and $J(\hat{\boldsymbol{\tau}}_K, \boldsymbol{x})$.
2. Determine, $l_K = J(\hat{\boldsymbol{\tau}}_K, \boldsymbol{x}) - J(\hat{\boldsymbol{\tau}}_{K+1}, \boldsymbol{x})$ for $K \in \{0, \ldots, K_{max} - 1\}$.
3. The number of changepoints is the largest K such that $l_i \gg l_j$ for $j > i$.

A simple graphical method can also be implemented for the above which may prove more intuitive in practice. This involves producing a plot of $J(\boldsymbol{\tau}_K)$ against K and visually identifying the point of maximum curvature. Figure 3(a) shows an example, from model B in Section 4, of the type of graphic produced by this method where the true number of changes is 2. For this example the lengths of the stable intervals are 211, 186, 24, 24, ... from which it is clear that $l_i \gg l_j$ for $i = 2$. This is the procedure which we implement in Section 4. As such this is reminiscent of using a scree plot in Principal Components Analysis to identify the number of principal components.

4 Simulation study

In this section we compare the performance of the method presented in Section 3 against the approaches proposed by Cho and Fryzlewicz (2011) and Davis et al. (2006). Our study follows that of Cho and Fryzlewicz (2011) (CF) who compare their non-parametric approach with Auto-PARM (AP) proposed by Davis et al. (2006). We replicate a selection of these simulation examples to demonstrate how the proposed method, henceforth referred to as WL (wavelet-based likelihood), compares with CF and AP. In addition, new simulation examples are introduced to highlight the flexibility of the WL method.

Following Cho and Fryzlewicz (2011), in all simulations we report results obtained for Haar wavelets. For comparison we used the default values specified in Cho and Fryzlewicz (2011) and Davis et al. (2006) for implementation of the CF and AP methods respectively. For the WL method, the number of



(a)

Figure 3: An example plot, from model B, of the negative log-likelihood against number of changepoints for the purpose of identifying the number of changepoints.

changepoints were identified using the method described in Section 3.3 and the practitioner implementing the approach had no knowledge of the true number of changepoints.

For each simulation case we consider (i) the number and (ii) the location of the identified changepoints. In particular, as Eckley et al. (2011) observe, consideration of both the number *and* distribution of changepoint estimates is helpful in assessing the benefits of the different methods. Tables 1, 2 and 3 report the number of changepoints detected by simulation scenario whilst Figure 4 displays the density of identified changepoints for scenarios where the methods are materially different. In each of the scenarios detailed below the innovations, ϵ_t have standard Normal distribution unless specified otherwise.

(A) Stationary AR(1) process with no changepoints This scenario is designed to assess the performance of the test when there are no changepoints. We fit an AR(1) model for a variety of parameter values, a (as used by Cho and Fryzlewicz (2011)). More specifically we simulate from,

$$X_t = aX_{t-1} + \epsilon_t \qquad \text{for } 1 \leq t \leq 1024.$$

Results from Table 1 show that as expected, Auto-PARM outperforms both CF and WL. However, the WL method has similar performance to Auto-PARM for all but the highest positive auto-covariances.

(B) Piecewise stationary AR process with clearly observable changes The 100 datasets are simulated from,

$$X_t = \begin{cases} 0.9X_{t-1} + \epsilon_t & \text{if } 1 \leq t \leq 512, \\ 1.68X_{t-1} - 0.81X_{t-2} + \epsilon_t & \text{if } 513 \leq t \leq 768, \\ 1.32X_{t-1} - 0.81X_{t-2} + \epsilon_t & \text{if } 769 \leq t \leq 1024, \end{cases}$$

Table 1: Results for scenario A from our method (WL), the method from Cho and Fryzlewicz (2011) (CF) and Auto-PARM (AP). We report the percentage of repetitions that identified that number of changepoints. True number of changepoints is in bold.

	Model A											
a	-0.7			-0.1			0.4			0.7		
no. cpts	WL	CF	AP	WL	CF	AP	WL	CF	AP	WL	CF	AP
0	**100**	**71**	**100**	**100**	**89**	**100**	**100**	**94**	**100**	**91**	**92**	**100**
1	0	24	0	0	11	0	0	5	0	9	7	0
≥ 2	0	5	0	0	0	0	0	1	0	0	1	0

i.e. the changes occur at dyadic points in the time series. This example highlights the increased power the WL method has over the traditional likelihood when the changepoints are at dyadic points in the time series. This is because there are more scales where the changepoint can be identified and thus more evidence for a change. Table 2 shows that we have comparable performance to Auto-PARM and outperform CF by 28% in this type of scenario.

(C) Piecewise stationary AR process with less clearly observable changes The 100 datasets are simulated from,

$$X_t = \begin{cases} 0.4X_{t-1} + \epsilon_t & \text{if } 1 \leq t \leq 400, \\ -0.6X_{t-1} + \epsilon_t & \text{if } 401 \leq t \leq 612, \\ 0.5X_{t-1} + \epsilon_t & \text{if } 613 \leq t \leq 1024, \end{cases}$$

i.e. the changes no longer occur at dyadic points in the time series. As with scenario B, our method has a similar performance to that of Auto-PARM whilst CF appears to over estimate the number of changepoints 24% of the time.

(D) Piecewise stationary AR process with a short segment The 100 datasets are simulated from,

$$X_t = \begin{cases} 0.75X_{t-1} + \epsilon_t & \text{if } 1 \leq t \leq 50, \\ -0.5X_{t-1} + \epsilon_t & \text{if } 51 \leq t \leq 1024. \end{cases}$$

The results for this example demonstrate that with relatively short segments the three methods perform well.

(E) Piecewise stationary AR process with high autocorrelation The 100 datasets are simulated from,

$$X_t = \begin{cases} 1.399X_{t-1} - 0.4X_{t-2} + \epsilon_t, & \epsilon_t \sim \mathcal{N}(0, 0.8^2) & \text{if } 1 \leq t \leq 400, \\ 0.999X_{t-1} + \epsilon_t, & \epsilon_t \sim \mathcal{N}(0, 1.2^2) & \text{if } 401 \leq t \leq 750, \\ 0.699X_{t-1} + 0.3X_{t-2} + \epsilon_t & \epsilon_t \sim \mathcal{N}(0, 1) & \text{if } 751 \leq t \leq 1024. \end{cases}$$

The changepoints in this model are hard to identify because the autocorrelation function does not change very much between segments. For number of

Table 2: Results for scenarios B-D from our method (WL), the method from Cho and Fryzlewicz (2011) (CF) and Auto-PARM (AP). We report the percentage of repetitions that identified that number of changepoints. True number of changepoints is in bold.

no. of	Model B			Model C			Model D		
cpts	WL	CF	AP	WL	CF	AP	WL	CF	AP
0	0	0	0	0	0	0	4	2	0
1	0	0	0	0	0	0	**94**	**83**	**100**
2	**98**	**70**	**94**	**94**	**76**	**100**	2	15	0
3	2	27	6	6	22	0	0	0	0
4	0	3	0	1	1	0	0	0	0
≥ 5	0	0	0	0	1	0	0	0	0

changepoints, the CF method outperforms both Auto-PARM and WL due to its variance stabilizing techniques. However, when we consider location of changepoints Figure 4(a) shows that the performance of our method is perhaps closer to CF than first appears. The distribution of changepoints for CF and our method are similar in shape and noticably different from that of Auto-PARM which detects spurious changepoints towards the start of the data.

(F) Piecewise stationary ARMA(1,1) process The 100 datasets are simulated from,

$$X_t = \begin{cases} 0.7X_{t-1} + \epsilon_t + 0.6\epsilon_{t-1} & \text{if } 1 \leq t \leq 125, \\ 0.3X_{t-1} + \epsilon_t + 0.3\epsilon_{t-1} & \text{if } 126 \leq t \leq 352, \\ 0.9X_{t-1} + \epsilon_t & \text{if } 353 \leq t \leq 704, \\ 0.1X_{t-1} + \epsilon_t - 0.5\epsilon_{t-1} & \text{if } 705 \leq t \leq 1024. \end{cases}$$

Figure 4(b) shows that all the methods identify the final changepoint more often than the first two. This is because the autocovariance function has greater difference between the second and third segments than between the first and second. For the remaining two changepoints, the WL method identifies both with similar probability whereas Auto-PARM prefers the second and CF the first changepoint. This again shows that the WL method is, in some sense, a compromise between the general structure of CF and the likelihood power of Auto-PARM.

(G) Piecewise stationary MA process The datasets are simulated from,

$$X_t = \begin{cases} \epsilon_t + 0.8\epsilon_{t-1} & \text{if } 1 \leq t \leq 128, \\ \epsilon_t + 1.68\epsilon_{t-1} - 0.81\epsilon_{t-2} & \text{if } 129 \leq t \leq 256, \end{cases}$$

In this scenario the WL method outperforms Auto-PARM and CF methods. Although Auto-PARM detects the largest number of correct changepoints, Figure 4(c) shows that the location of these changepoints is split across two modes; around 125 and 135. The CF method detects the least correct changes and the accuracy of the identified changes is poor. Figure 4(c) shows that the CF method has its primary mode just prior to location 150 when the true changepoint is 128. There are other modes with similar weightings around 125, 170,

Table 3: Results for scenarios E-G from our method (WL), the method from Cho and Fryzlewicz (2011) (CF) and Auto-PARM (AP). We report the percentage of repetitions that identified that number of changepoints. True number of changepoints is in bold.

no. of	Model E			Model F			Model G ($N = 1,000$)		
cpts	WL	CF	AP	WL	CF	AP	WL	CF	AP
0	0	0	0	0	0	0	0	0	0
1	26	9	9	20	12	51	**99**	**85**	**100**
2	**45**	**75**	**33**	22	36	33	1	15	0
3	26	15	31	**35**	**45**	**16**	0	0	0
4	3	1	15	22	6	0	0	0	0
≥ 5	0	0	12	1	1	0	0	0	0

200, 225 and 250. The use of additional simulations is to highlight the extra modes which would otherwise appear negligible. In contrast to both these methods, the WL method has its mode around 127 and retains good detection of the correct number of changepoints.

5 Storm season identification

We now return to considering storm season identification using significant wave heights. Recall that there is evidence (see for example Hidgkins et al. (2002)) that the onset of storm seasons is shifting. This is of interest as short-term operations, such as inspection and maintenance of marine structures, are typically performed outside seasons when storms are prevalent to minimise risk. As such knowledge of the date of onset of the storm season, e.g. of winter in the North Sea, or of hurricane season in the Gulf of Mexico, is particularly important if this is changing. For example, within the oil and gas industry, the current popularity of liquefying natural gas on an offshore production platform and offloading it to a shuttle gas carrier, requires knowledge of environmental winds, waves and currents that excite vessel motions, limiting floating liquified natural gas (FLNG) operations. If the time periods that allow production and offloading are sufficient, the system has the potential to work safely and efficiently, and the economic benefit of FLNG can be realised, maximising the time periods during which FLNG platforms can be offloaded. Similarly, for offshore wind turbines, environmental conditions frequently cause costly delays in the installation and maintenance process. A systematic approach to identify approximately homogeneous periods in the ocean environment, and transitions between calm and stormy seasons in particular, is therefore highly desirable.

Recall that first differences were applied to the original data in Figure 1(a) to remove the trend. We thus apply the WL, AP and CF methods to detect changes in auto-covariance of the differenced data shown in Figure 1(b). Applying the WL method from Section 3 we identify changepoints at 12th May 1992, 5th October 1992, 1st April 1993, 19th November 1993, 10th April 1994 and 10th August 1994. Each of these dates is indicated by a dashed vertical line in Figure 5. The changepoints identified by the CF and AP methods are indicated by vertical lines at the top and bottom of Figure 5 respectively. It is clear that the automatic penalty selection contained within the CF and AP methods results in

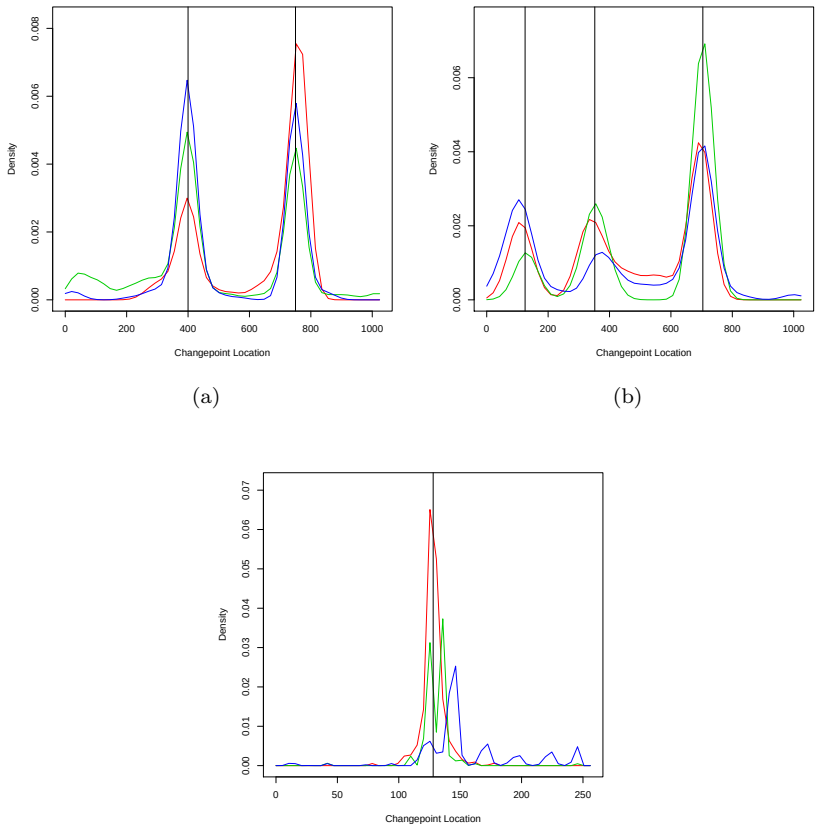


Figure 4: Density plots of the changepoints identified for (a) model E, (b) model F and (c) model G. In each plot the red line is the WL method; the blue line is the CF method and the green line is the AP method.

more changepoints being identified than storm seasons. Both methods identify changes at approximately the same times as the WL method with additions between seasons. The exception to this is the CF method which does not identify a change around October 1992 where the WL and AP methods do. The results for the three methods were (blindly) presented to ocean engineers who selected the WL method as the set of changepoints which most closely matched their own assessment of the start and end of the storm season.

6 Concluding remarks

As discussed in Section 1, oceanographic applications require methods for segmenting time series where the observations cannot be assumed to be independent. The work presented in this paper has addressed this issue by developing a test for a change in autocovariance structure using the LSW framework and

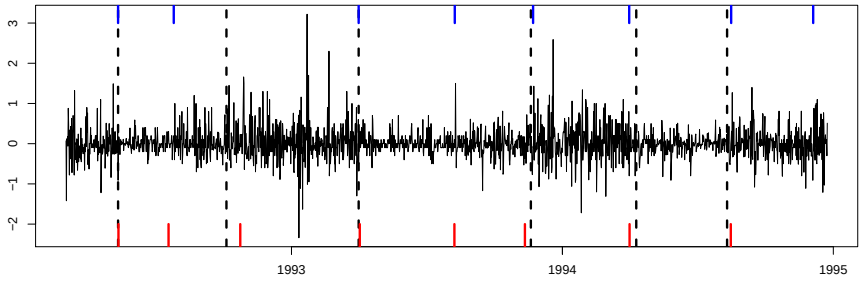


Figure 5: Central North sea application: First differenced significant wave heights with identified changepoints. Vertical dotted lines: WL method, short lines at the top of the plot: CF method, short lines at the bottom of the plot: AP method.

assuming no prior knowledge of the structure of the autocovariance. A benefit of this approach is that it does not assume a specific structure for the autocovariance.

The proposed approach was compared with both the Auto-PARM method of Davis et al. (2006) and a wavelet-based method recently proposed by Cho and Fryzlewicz (2011). The results showed that, as expected, when the Auto-PARM assumptions of piecewise AR model and independent segments are valid, it tends to outperform both CF and the proposed method. However, when the underlying assumption of a piecewise AR model is not valid, there are cases where the proposed method outperforms both Auto-PARM and CF. Results also highlighted that when considering performance of a changepoint method, the location of the changes should be taken into account in addition to the number of changepoints detected. The simulation study in section 4 used the Haar wavelet. The same tests have also been conducted using other wavelet families and similar results were obtained.

The proposed method was also applied to oceanographic data encountered by an industrial collaborator. The results demonstrate that the proposed method gives an automatic approach for segmenting series into seasons when storms are prevalent for further analysis or planning of maintenance. An added advantage of the WL method is that a specific number of changepoints can be identified whereas the CF and AP methods would require postprocessing procedures for selecting the changes that correspond to storm seasons.

Acknowledgements

R. Killick gratefully acknowledges financial support from Shell Research Limited and EPSRC. P. Jonathan was financially supported by Shell International Exploration Production. The authors are grateful to Richard Davis and Haeran Cho for providing the code for Auto-PARM and CF algorithms respectively.

References

Ahamada, I., Jouini, J., and Boutahar, M. (2004). Detecting multiple breaks in time series covariance structure: a non-parametric approach based on the evolutionary spectral density. *Applied Economics*, 36(10):1095–1101.

Auger, I. E. and Lawrence, C. E. (1989). Algorithms for the optimal identification of segment neighborhoods. *Bulletin of Mathematical Biology*, 51(1):39–54.

Bai, J. and Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, 18(1):1–22.

Berkes, I., Gombay, E., and Horvath, L. (2009). Testing for changes in the covariance structure of linear processes. *Journal of Statistical Planning and Inference*, 139(6):2044 – 2063.

Chen, J., Gupta, A. K., and Pan, J. (2006). Information criterion and change point problem for regular models. *Sankhya*, 68:252–282.

Cho, H. and Fryzlewicz, P. (2011). Multiscale and multilevel technique for consistent segmentation of nonstationary time series. *Statistica Sinica*, 22:207–229.

Dahlhaus, R. (2012). Locally stationary processes. In *Handbook of Statistics*, volume 30, pages 351–413.

Davis, R. A., Lee, T. C. M., and Rodriguez-Yam, G. A. (2006). Structural break estimation for nonstationary time series models. *Journal of the American Statistical Association*, 101(473):223–239.

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

Ferreira, J. A. and Guedes-Soares, C. (2002). Modelling bivariate distributions of significant wave height and mean wave period. *Applied Ocean Research*, 21:31–45.

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

Fryzlewicz, P. and Ombao, H. (2009). Consistent classification of nonstationary time series using stochastic wavelet representation. *Journal of the American Statistical Association*, 104(485):299–312.

Fryzlewicz, P., Van Bellegem, S., and von Sachs, R. (2003). Forecasting nonstationary time series by wavelet process modelling. *Annals of the Institute of Statistical Mathematics*, 55:737–764.

Gombay, E. (2008). Change detection in autoregressive time series. *Journal of Multivariate Analysis*, 99(3):451–464.

Harchaoui, Z. and Levy-Leduc, C. (2010). Multiple change-point estimation with a total variation penalty. *Journal of the American Statistical Association*, 105(492):1480–1493.

Hidgkins, G. A., James, I. C., and Huntington, T. G. (2002). Historical changes in lake ice-out dates as indicators of climate change in new england, 1850–2000. *International Journal Climatology*, 22:1819–1827.

Holland, J. H. (1992). *Adaptation in natural and artificial systems*. MIT Press, Cambridge, MA, USA.

Inclan, C. and Tiao, G. C. (1994). Use of cumulative sums of squares for retrospective detection of changes of variance. *Journal of the American Statistical Association*, 89(427):913–923.

Killick, R. and Eckley, I. A. (2010). *R package- changepoint: Analysis of Changepoint Models*. Lancaster University, Lancaster, UK.

Killick, R., Fearnhead, P., and Eckley, I. A. (2011). Optimal detection of changepoints with a linear computational cost. *In Submission*.

Lavielle, M. (2005). Using penalized contrasts for the change-point problem. *Signal Processing*, 85:1501–1510.

Linderholm, H. W. (2006). Growing season changes in the last century. *Agricultural and Forest Meteorology*, 137:1–14.

Molau, U., Nordenhaell, U., and Eriksen, B. (2005). Onset of flowering and climate variability in an alpine landscape: A 10-year study from swedish lapland. *American Journal of Botany*, 92:422–431.

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

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

Ombao, H. C., Raz, J. A., von Sachs, R., and Malow, B. A. (2001). Automatic Statistical Analysis of Bivariate Nonstationary Time Series. In Memory of Jonathan A. Raz. *Journal of the American Statistical Association*, 96(454):543–560.

Rissanen, J. (1978). Modeling by shortest data description. *Automatica*, 14(5):465 – 471.

Scott, A. J. and Knott, M. (1974). A cluster analysis method for grouping means in the analysis of variance. *Biometrics*, 30(3):507–512.

Van Bellegem, S. and von Sachs, R. (2008). Locally adaptive estimation of evolutionary wavelet spectra. *Annals of Statistics*, 36(4):1879–1924.

Vidakovic, B. (1999). *Statistical Modeling by Wavelets*. Wiley, New York.