A nonparametric approach to detecting changes in variance in locally stationary time series

J.-L. Chapman[1], I. A. Eckley[2], R. Killick[2*]

April 7, 2019

Abstract

This article proposes a nonparametric approach to detecting changes in variance within a time series which we demonstrate is resilient to departures from the assumption of Normality or presence of outliers. Our method is founded on a local estimate of the variance provided by the Locally Stationary Wavelet (LSW) framework. Within this setting, the structure of this local estimate of the variance will be piecewise constant if a time series has piecewise constant variance. Consequently, changes in the variance of a time series can be detected in a non-parametric setting.

*Corresponding author: r.killick@lancaster.ac.uk.

[1]STOR-i Centre for Doctoral Training, Lancaster University, Lancaster, U.K.;

[2]Department of Mathematics and Statistics, Lancaster University, Lancaster, U.K.

In addition, using a simulation study, we explore the robustness of our approach against the typical assumption of Normality and to the presence of outliers. We illustrate the application of the approach to changes in variability of wind speeds at a location in the UK.

Keywords: changepoints, local stationarity, wavelets, wind speed.

1 Introduction

Since its introduction in the context of quality control by Page (1954), changepoint analysis has become a very active area of research. Many important contributions have been made in recent years, especially in the independent and identically distributed (i.i.d.) setting, see Eckley et al. (2011) for examples.

Changepoint methods have been used extensively to derive insight for a number of important environmental and ecological applications. See, for example, the important work of Andersen et al. (2009); Evans et al. (2016); Hilborn et al. (2017); Richardson et al. (2018). In this article, we will consider the problem of detecting changes in variance within wind speed data related to challenges arising within the renewable energy sector. Specifically, in recent years there has been an increasing focus on detecting damage in wind turbine blades. As Chou et al. (2013) report, damage to these blades can cause up to 19.4% of wind turbine damage. Such damage can be caused by various factors including severe environmental conditions such as gusty winds, lightening strikes and storms (Herr and Heidenreich, 2015; Hoell and Omenzetter, 2015). Amongst a variety of different analyses one might undertake, it may be of interest to segment the wind speed observerd at a given location into regions of differing variability to allow better understanding of the wind gusts experienced by the turbine. Data of this form may be heavy tailed, and subject to outliers. This is the approach we adopt, proposing a novel method for the detection of change in variance that is resilient to these challenges.

The challenge of detecting changes in a sequence $\{x_i\}_{i=1,\ldots,n}$ of observations, reduces to the problem of finding the values of m and $\{\tau_i\}_i$ which minimise the following expression

$$\sum_{i=1}^{m+1} \left[\mathcal{C}(x_{(\tau_{i-1}+1):\tau_i})\right] + \beta f(m). \tag{1}$$

Here we have m changepoints with associated ordered positions $0 = \tau_0, \tau_1, \ldots,$ $\tau_m, \tau_{m+1} = n$. As a result, having m changepoints causes the data to be split into $m+1$ segments such that segment i contains the observations $x_{(\tau_{i-1}+1):\tau_i}$ and has variance $\sigma^2_{(i)} := \sigma^2_{\tau_{i-1}+1} = \ldots = \sigma^2_{\tau_i}$. The first term in equation (1) is a cost function for the segment $x_{(\tau_{i-1}+1):\tau_i}$. The second term in equation (1) is a penalty which guards against over fitting. The function $f(m)$ is often taken to be simply the number of changepoints m, resulting in a penalty that is linear in the number of changepoints. β is a tuning parameter where small values result in more changes than larger values. Different methods can be adopted in order to minimise equation (1). Examples of exact methods include PELT (Killick et al., 2012) and Segment Neighbourhood (Auger and Lawrence, 1989); whilst Binary Segmentation is a commonly used approximate method (Scott and Knott, 1974).

Within the literature a number of different variance cost functions have been proposed, one of the best known is the penalized likelihood approach, see (Chen and Gupta, 2012) for a description. Within this setting the log likelihood is calcuated for each segment and summed over segments to give the log likelihood for the data sequence. To avoid over fitting, the penalty is subtracted from the log likelihood as in (1). Alternatives to the likelihood based approach to changepoint detection include the important work of Inclan and Tiao (1994), who use a nonparametric Cumulative Sums of Squares (CSS) approach, and Fearnhead (2006) who develops a Bayesian posterior odds approach.

In practice data sequences are often prone to outliers and/or heavy tail structures which the majority of approaches are intolerant to. Typically some pre-processing of the data is often performed in an attempt to mitigate these effects (Candemir and Oğuz, 2017). In some cases this is a straightforward adaptation, however given the unprecedented volume of data now being generated, pre-processing is becoming increasingly impractical and often subjective (Taleb et al., 2015). This motivates the need for new methods that are inherently resilient to such features.

The novel contribution of this article is a non-parametric method for detecting changes in variance in the presence of outliers and heavy tails. We develop this non-parametric method using the Locally Stationary Wavelet (LSW) model, due to Nason et al. (2000), to provide a local estimate of the variance of a time series. The article is organised as follows: in Section 2 we describe the Locally Stationary Wavelet (LSW) model and our method for detecting changes in variance. The method is then assessed under various simulation scenarios (Section

3). Lastly, Section 4 applies our method to wind speed data collected from a site in the UK.

2 A nonparametric approach to detecting changes in variance

In this section we describe our non-parametric method for detecting changes in variance. Our approach is based on the key insight that detecting a change in variance in the time domain can be transformed into detecting a change in mean in a transformed domain, given a suitable transformation. We are by no means the first to consider this, see for example, Darkhovski (1994); Inclan and Tiao (1994). In contrast to this earlier work we adopt a wavelet based approach which we described below. Prior to describing our approach we provide a brief introduction to the locally stationary wavelet modelling framework that we use.

2.1 Locally Stationary Wavelet Framework

Our method for detecting changes in variance relies upon capturing the local behaviour of a time series' variance. This could be achieved using a rolling window estimate of the variance, but would require choice of a window size. Instead we choose to adopt the locally stationary wavelet (LSW) framework which is built upon non-decimated wavelets.

The advantage of the LSW framework is that it encompasses many common time series processes, such as moving average and autoregressive processes. Of particular interest for this work, we can use the LSW framework to attain a local time-varying measure of the variance. Below, we provide a brief introduction to the LSW approach to time series which have piecewise stationary second order structure and also introduce a time-varying measure of the variance.

Following Fryzlewicz and Nason (2006), a triangular stochastic array $\{X_{t,N}\}_{t=0,\ldots,N-1}$, for a dyadic length of time $N = 2^J \geq 1$, is a locally stationary wavelet (LSW) process if there is a mean-square representation

$$X_{t,N} = \sum_{j=1}^{\infty} \sum_{k} W_j(k/N) \psi_{j,k-t} \xi_{j,k}, \qquad j \in \{1, 2, \ldots\}, k \in \mathbb{Z}, \tag{2}$$

where $\{\psi_{j,k-t}\}_{j,k}$ is a set of discrete compactly supported non-decimated wavelets and the $\xi_{j,k}$ are zero-mean, orthonormal, identically distributed random variables. The function $W_j(z) : [0, 1] \to \mathbb{R}$ is real-valued and piecewise constant with some finite unknown number of jumps. Let $\mathcal{J}_j$ denote the total magnitude of jumps in $W_j^2(z)$. Then the functions, $W_j(z)$, satisfy

1. $\sum_{j=1}^{\infty} W_j^2(z) < \infty$ uniformly in z and
2. $\sum_{j=1}^{\infty} 2^j \mathcal{J}_j < \infty$.

This definition of an LSW process (2) is a modification of that by Nason et al. (2000) in which the Lipschitz continuity constraint is replaced by that of total variation. This allows a process with piecewise constant second order structure to be modelled.

The use of the notation $X_{t,N}$ rather than the traditional X_t is to emphasize the triangular stochastic array across different N, although in practice dependence on N is often suppressed within notation.

To enable various theoretical model properties to be estalished, Nason et al. (2000) adopted the rescaled-time approach proposed by Dahlhaus (1997). For a given LSW series $\{X_{t,N}\}_{t=1,\ldots,N}$, its time-varying spectrum is defined as $S_j(k/N) = |W_j(k/N)|^2$. To estimate the spectrum, define $d_{j,k} = \sum_{t=1}^{N} X_t \psi_{j,k-t}$, to be the empirical wavelet coefficients of an LSW process $X_{t,N}$. They then demonstrate that the corrected wavelet periodogram $\mathbf{L}_j(z) = \sum_{l=1}^{J} A_{j,l}^{-1} |d_{l,z}|^2$ is an asymptotically unbiased estimator of the time-varying spectrum, $S_j(k/N)$. Here, $A_{j,l}$ is the discrete autocorrelation wavelet inner product

$$A_{j,l} = \langle \Psi_j, \Psi_l \rangle = \sum_{\tau} \sum_{k} \psi_{j,k} \psi_{j,k+\tau} \sum_{m} \psi_{l,m} \psi_{l,m+\tau}. \tag{3}$$

We note that the wavelet periodogram is an inconsistent estimator of the evolutionary wavelet spectrum (Proposition 4 Nason et al. (2000)). Therefore in order to obtain a consistent estimator we would traditionally smooth the periodogram.

Using representation (2), the variance of X_t is given by

$$\mathrm{var}(X_t) = \sum_{j,k} W_j^2(k/N) \psi_{j,k-t}^2. \tag{4}$$

The dependence on time in equation (4) is introduced indirectly via the compact support of the wavelet. Using the wavelet spectrum, Nason et al. (2000) introduce the **localized variance function** for a LSW process of length $N = 2^J$. This is defined to be

$$\sigma^2(z) = \sum_{j=1}^{J} S_j(z), \tag{5}$$

where $z = k/N \in (0,1)$ is rescaled time. If a time series has a constant variance, then the dependence on z in equation (5) is lost and the localised measure becomes a global one.

The time-varying estimate of the variance (5) can be interpreted as a windowed rolling estimate of the variance of the time series. However, unlike a usual rolling estimate, no consideration of the window length is required. The benefit of a wavelet approach is that a variety of window sizes are used in the wavelet transform. Through the compact support of the wavelets the representation in (4) is unique given the wavelet (Nason et al., 2000).

Figure 1 shows an example of a process with (a) constant variance and (b) piecewise variance and their associated smoothed and unsmoothed local variance functions in (c), (d) and (e), (f) respectively. Figure 1(c) demonstrates that smoothing the spectral estimate masks the abrupt change that is clearly visible in (b) and (f). For this reason, the following section presents a method based on the unsmoothed localised variance.

[Figure 1 about here.]

2.2 The NPLE Method

As previously described, if a time series is second order stationary then its evolutionary wavelet spectrum will be constant across each scale. Similarly, if a time series is piecewise second order stationary, then the spectrum will be piecewise constant (Fryzlewicz and Nason, 2006). Consequently, as the localised variance function in equation (5) is the sum of the spectrum over scales, this means that the localised variance function will also be piecewise constant. In order to exploit this property for changepoint detection, we need to translate it into a practical setting. Thus our estimate of the unsmoothed local variance function is defined as

$$\hat{\sigma}^2(z) = \sum_{j=1}^{J} \sum_{l=1}^{J} A_{j,l}^{-1} d_{l,z}^2. \tag{6}$$

Due to the compact support of the wavelets it is clear that, for a signal with piecewise constant variance, this estimate is also piecewise constant. The following section outlines the method for detecting these changes in the localised variance.

2.2.1 The nonparametric model

The localised variance function (6) is a sum of correlated χ^2 random variables. In practice it is difficult to obtain the distribution for this (Gordon and Ramig, 1983). We choose to adopt a non parametric approach to this changepoint detection problem.

We consider the localized variance, σ^2, and model its cumulative distribution function, $G(u) = \mathbb{P}(\sigma^2 \leq u)$, for quantile u using the empirical CDF

$$\hat{G}(u) = \frac{1}{n} \left(\sum_{t=1}^{n} \mathbb{I}_{\{\hat{\sigma}_t^2 < u\}} + \frac{1}{2} \mathbb{I}_{\{\hat{\sigma}_t^2 = u\}} \right), \tag{7}$$

where the $\hat{\sigma}_t^2$ are assumed to be independent. With this distribution function in mind, following Zou et al. (2014) the maxiumum log likelihood of $G(u)$ is given by

$$n\{\hat{G}(u) \log \hat{G}(u) + (1 - \hat{G}(u)) \log (1 - \hat{G}(u))\}, \tag{8}$$

because for a fixed value of u, we have $n\hat{G}(u) \sim \mathrm{Bin}(n, G(u))$.

Recall that in order to identify changepoints, we aim to minimise the following

$$\sum_{i=1}^{m+1} \left[\mathcal{C}(\hat{\sigma}^2_{\{\tau_{i-1}+1\}:\tau_i}) \right] + \beta f(m). \tag{9}$$

where the cost function for segment i is given by the negative of the empirical log likelihood of the CDF of the localised variance estimate:

$$-\mathcal{L}(\hat{\sigma}^2_{\{\tau_{i-1}+1\}:\tau_i}; u) = (\tau_i - \tau_{i-1}) \times \left[\hat{G}_i(u) \log \hat{G}_i(u) + (1 - \hat{G}_i(u)) \log \left(1 - \hat{G}_i(u)\right) \right]. \tag{10}$$

Zou et al. (2014) recommend an integrated form of the cost function (10)

$$\int_{-\infty}^{\infty} -\mathcal{L}(\hat{\sigma}^2_{\{\tau_{i-1}+1\}:\tau_i}; u) \; dw(u), \tag{11}$$

where $w(\cdot)$ is a weight function, dependent upon the CDF of the data set, such that the integral is finite. The consistency of this approach is detailed in Zou et al. (2014). This allows information across all time points to be incorporated into the cost function.

The computational cost of the cost function suggested by Zou et al. (2014) is of order $\mathcal{O}(Mn^2 + n^3)$, where M is a specified maximum number of changepoints (Haynes et al., 2017b). Zou et al. (2014) suggest a screening step to help reduce this computational time; however this jeopardizes the accuracy of the locations of the changepoints. Haynes et al. (2017b) suggest an improved segment cost that involves approximating the integral in (11) by a sum with some fixed number of terms K. This improves the computational time taken to calculate the cost for a given segment to $\mathcal{O}(\log n)$. The suggested approximation is as follows.

Following Haynes et al. (2017b), we fix a K and define $\gamma = \frac{-\log(2n-1)}{K}$. Time is then rescaled according to quantiles dependent upon the choice of K. Let $\{t_k\}_{k=1,\ldots,K.}$ be equal to the $(1 + (2n-1)\exp\{\gamma(2k-1)\})^{-1}$ empirical quantile of the data. The approximation to the integral in equation (11) is then given by:

$$\mathcal{C}_K(\hat{\sigma}^2_{\{\tau_{i-1}+1\}:\tau_i}) = \frac{2\log(2n-1)}{K} \sum_{k=1}^{K} \mathcal{L}(\hat{\sigma}^2_{\{\tau_{i-1}+1\}:\tau_i}; t_k). \tag{12}$$

We could use any search function in order to identify the changepoints using (12). However, Haynes et al. (2017b) show that this cost function is compatible with PELT (Killick et al., 2012), a computationally efficient search for changepoints. We therefore use this search method in our simulation study in Section 3.

Based upon the above description, we choose to call the method outlined here Non-Parametric change in variance detection using Localised Estimates, abbreviated to NPLE.

2.3 Penalty choice

Penalty choice is a practical challenge in many changepoint settings. We choose to take an adaptive approach to penalty selection following that of Lavielle (2005). Intuitively, this approach involves selecting the segmentation which causes the most significant decrease in the cost function. This can be presented graphically in an analogous way to a scree plot used in Principal Components analysis (Jolliffe, 2002). Figure 2 shows an example plot of a cost function against the number of changepoints identified for a model with 2 true changepoints. It is visible that the true segmentation occurs at the point of maximum curvature, or ‘elbow’, of the plot; where the largest relative decrease in the cost function occurs. The procedure of identifying this ‘elbow’ can be formalized, and automatized, as follows.

[Figure 2 about here.]

In line with Lavielle (2005), let m_{MAX} be an upper bound on the number of changepoints in the model. The PELT search algorithm results in a single optimal segmentation for a given penalty value. In order to obtain segmentations for a range of penalty values efficiently we utilize the CROPS method (Haynes et al., 2017a). From this range of segmentations we then wish to determine $\hat{m}$; the estimated number of changepoints in the model. Following Lavielle (2005), we obtain $\hat{m}$ using the following procedure:

1. For $0 \leq m \leq m_{\mathrm{MAX}}$ let

$$\tilde{\mathcal{J}}_m = \frac{\mathcal{J}_{m_{\mathrm{MAX}}} - \mathcal{J}_m}{\mathcal{J}_{m_{\mathrm{MAX}}} - \mathcal{J}_0} m_{\mathrm{MAX}} + 1, \tag{13}$$

   where $\mathcal{J}_m$ is the cost for the segmentation corresponding to m changepoints at locations $\tau_{1:m}$. The associated costs have now been normalised between 1 and $m_{\mathrm{MAX}} + 1$.

2. Then, for $1 \leq m \leq m_{\mathrm{MAX}} - 1$, let

$$D_m = \tilde{\mathcal{J}}_{m-1} - 2\tilde{\mathcal{J}}_m + \tilde{\mathcal{J}}_{m+1}, \tag{14}$$

   and $D_1 = \infty$.

3. The estimate for the true location of the changepoint is given by the largest value of m such that the second derivative of $\mathcal{J}_m$, D_m, is greater than some threshold S,

$$\hat{m} = \max\{0 \leq m \leq m_{\mathrm{MAX}} - 1 | D_m > S\}. \tag{15}$$

The above procedure has also been implemented for penalty choice in a wavelet context by Killick et al. (2013). The intuition behind this approach is that true changes will be added to the segmentation first as they will result in the largest improvement to the cost function. Following this we will start to add spurious changes to the data, which are just due to noise, and so the improvement in fit will be small. The aim of the choice of S is to put a threshold on the rate of change in the scaled test statistic as the number of changes increases. For an individual dataset we would do this using the changepoint equivalent of a scree plot.

3 Simulation study

In the following simulation study we test the robustness of NPLE against the log likelihood of a Normal distribution with changing variance (MLvar) (Chen and Gupta, 2012) and the non parametric Cumulative Sums of Squares (CSS) (Inclan and Tiao, 1994). This allows for a comparison between both a parametric and non-parametric method. Each of these are implemented using the **changepoint** package (Killick et al., 2016; Killick and Eckley, 2014) in R (R Core Team, 2018). For the calculation of the localised variance estimate we utilize the **wavethresh**

(Nason, 2012) and `changepoint.np` (Haynes, 2016) packages. The study also considers departures from the idealised Normal distribution change in variance setting. Specifically, the simulations study provides a pratical assessment of the proposed approach's resilience to departures from Normality, including outliers and heavy tailed dependence structure.

3.1 Random Outliers

In this first study we seek to test how each of the methods performs with varying degrees of outliers. To this end, we simulate time series with different proportions of outliers. Specifically, we simulate epidemic changes in variance, $\sigma = (1, 3, 1, 3, 1, 3)$ from a Normal distribution of length 2048 with changes at $365i$ for $i = 1, \ldots, 5$. The timing of the outliers are simulated from a $\text{Unif}(1, 2048)$ distribution. To create outliers at these time points, we add a fixed constant, 15, to the existing observations. We repeat this for $P = 0.01\%, 1\%$ and 5% density of outlying observations within each data set as well as the no outlier case for comparison. The choice to use additive outliers instead of multiplicative outliers means that the size of the outliers will vary less across segments with differing variances.

[Table 1 about here.]

Table 1 shows the number of changepoints detected by each of the methods for the four values of P over 500 repetitions. As expected, the performance of each method degrades as the percentage of outlying values increases. However, this degredation is not uniform across the methods. NPLE detects the correct number of changepoints 63% of the time when 5% of observations are outliers, in comparison, CSS achieves a similar rate when only 0.01% of observations are outliers.

[Figure 3 about here.]

Figure 3 shows the density of detected changepoint locations for each of the methods for P equal to 0.01%, 1% and 5%. NPLE maintains accurate changepoint locations as P increases, whereas the other methods are drawn to the outliers.

The results of these simulations demonstrate that NPLE is less sensitive to outliers than the other methods. When using MLvar, and similarly CSS, the outliers contribute to both the likelihood and the sum of squares directly and distort the estimates.

In the next simulation study, we consider another model with outliers, however they are located at fixed points in time.

3.2 Fixed Outliers

In this section we test the robustness of the model for increasingly sized changes in variance, using variance changes that are more difficult to detect than those in Section 3.1.We simulated 500 repetitions of a Normal distribution of length 2048 with changepoints at $365i$ for $i = 1, \ldots, 5$ and $\sigma = (1, 1.6, 1, 1.8, 1, 2)$. We also consider the effect of proximity of outliers to changepoint locations. Hence, we introduce multiplicative outliers located at times (361, 462, 723, 924, 1244, 1630, 1881) with inflation factors (20, -20, 16, 18, 20, 10, 7).

Figure 4a shows a realisation of this model where it is important to note the location of the outlier in relation to the location of the changepoint. The first and third outliers occur close to changepoint locations; whereas the remaining outliers are firmly within segments. Despite the locations of the outliers being fixed, in comparison to the uncertain locations in Section 3.1, the size of the outliers are more variable as a consequence of their multiplicative nature.

[Figure 4 about here.]

Figure 4b shows the density of detected changepoints and Table 2 gives the corresponding numbers of changepoints detected. NPLE detects the true number of changes 87% of the time, whereas MLvar and CSS achieve only 13% and 14% respectively.

Turning consideration to the locations of the changes. For the first change, for MLvar and CSS, the presence of the outliers near the changepoint means that there are two distinct peaks corresponding the location of the change. This is not the case for NPLE, but the outlier appears to result in the change is detected slightly early. At the third change, MLvar and CSS often detect a change either side of the true changepoint location.

All three methods perform similarly when detecting the second change.

Despite being the largest changes, the last three are detected correctly the least by MLvar and CSS, this is probably a consequence of the methods detecting a larger number of changes elsewhere, induced by the outliers. The large outliers at 462, 924, and 1244 have clearly resulted in spurious changes for both MLvar and CSS.

[Table 2 about here.]

Our final simulation study considered data which instead of having outliers, exhibits heavy tail behaviour.

3.3 Heavy Tail Structure

In this section we consider data which is generated from a Generalised Extreme Value (GEV) distribution, with zero mean ($\mathbb{E}(X_t) = 0$), that exhibits varying changes in variance. The changes are located at times $256i$, $i = 1 \ldots 7$ and the sequence of standard deviations is given by $\sigma = (1, 1.6, 1, 1.8, 1, 2, 1, 2.5)$. We consider three values of the shape parameter for the GEV distribution: $0, 0.25$ and 0.45. Note that as σ is a function of the shape and scale parameters, we keep the shape constant and only modify the scale across the segments to obtain the required σ. The tails become heavier as the shape parameter, ξ, increases across simulations.

[Table 3 about here.]

Table 3 shows the number of changepoints detected and Figure 5 show the corresponding densities for the locations. As expected, as the tails become heavier the detection rate decreases for all methods. Whilst they all perform similarly for $\xi = 0$ as the shape parameter increases, NPLE is most resilient to the heavy tails providing around double the detection rate as $\xi = 0.25$ and 0.45.

[Figure 5 about here.]

This illustrates NPLE's reduced sentitivity to heavy tailed distributions. For MLvar we are using a Gaussian assumption and so we expect this method to perform poorly but CSS does not have any tail assumptions.

4 Application to Wind Speed Characterization

We now turn to consider the detection of variance changepoints within a time series of wind speeds. The data we analyse were obtained at a UK wind farm location during November 2005. Each measurement represents the average wind speed obtained from an anemometer at the farm. The series contains 4261 observations, as depicted in Figure 6a. The data that support the findings of this study are available from the corresponding author upon reasonable request.

To explore whether any changes in variance exist within this wind speed data, we begin by taking first differences to remove the mean behaviour. The resulting series has a very clear, non-constant variance structure (Figure 6b). There also appear to be some anomalous observations that could potentially affect changepoint estimation. Note that whilst first differences were sufficient for this data sequence, wind speeds are periodic and taking a longer stretch of data would require a different approach to remove this. Next, we apply both the NPLE and MLvar methods to the differenced wind speeds. To provide a fair comparison between the methods we use the Lavielle (2005) method for penalty choice for both methods. The diagnostic plots are give in Figure 7 where it is clear that the elbow in the curve for NPLE is at 9 changes and for MLvar is at

8 changes.

The resulting changepoint plots for NPLE and MLvar are given in Figure 8. Note, in particular, how MLvar appears to be inflating the variance estimate for the first segment of data in response to the anomalous points. This results in a later changepoint than the NPLE method which chooses to use two changepoints to capture the period of smaller variability. For operational decisions the segmentation provided by NPLE is preferred.

[Figure 6 about here.]

[Figure 7 about here.]

[Figure 8 about here.]

5 Conclusion

In this article we have introduced a novel changepoint detection procedure to detect changes in variance (NPLE). The key benefits of our nonparametric approach are its capacity to provide changepoint estimates that are resilient to outliers and departures from normality. This method is shown to perform well against an established nonparametric method (CSS) and penalised likelihood approaches (MLvar) in all simulated settings. We also considered the utility of NPLE on data obtained from a UK wind farm. In future work we plan to extend our approach to detect changes in local autocovariance. Such a development could prove useful to a wide variety of environmental applications.

References

Andersen, T., Carstensen, J., Hernndez-Garca, E., and Duarte, C. M. (2009). Ecological thresholds and regime shifts: approaches to identification. *Trends in Ecology & Evolution*, 24(1):49 – 57.

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

Candemir, C. and Oğuz, K. (2017). A comparative study on parameter selection and outlier removal for change point detection in time series. In *2017 European Conference on Electrical Engineering and Computer Science (EECS)*, pages 218–224. IEEE.

Chen, J. and Gupta, A. (2012). *Parametric Statistical Change Point Analysis*. Birkhäuser Boston.

Chou, J.-S., Chiu, C.-K., Huang, I.-K., and Chi, K.-N. (2013). Failure analysis of wind turbine blade under critical wind loads. *Engineering Failure Analysis*, 27:99–118.

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. 25:1–37.

Darkhovski, B. S. (1994). Nonparametric methods in change-point problems: A general approach and some concrete algorithms. *Lecture Notes-Monograph Series*, pages 99–107.

Eckley, I. A., Fearnhead, P., and Killick, R. (2011). Analysis of changepoint models. *Bayesian Time Series Models*, pages 205–224.

Evans, A. L., Singh, N. J., Friebe, A., Arnemo, J. M., Laske, T. G., Fröbert, O., Swenson, J. E., and Blanc, S. (2016). Drivers of hibernation in the brown bear. *Frontiers in Zoology*, 13(1):7.

Fearnhead, P. (2006). Exact and efficient bayesian inference for multiple changepoint problems. *Statistics and computing*, 16(2):203–213.

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

Gordon, N. and Ramig, P. (1983). Cumulative distribution function of the sum of correlated chisquared random variables: The sum of correlated chisquared random variables. *Journal of Statistical Computation and Simulation*, 17(1):1–9.

Haynes, K. (2016). *changepoint.np: Methods for Nonparametric Changepoint Detection*. R package version 0.0.2.

Haynes, K., Eckley, I. A., and Fearnhead, P. (2017a). Computationally efficient changepoint detection for a range of penalties. *Journal of Computational and Graphical Statistics*, 26(1):134–143.

Haynes, K., Fearnhead, P., and Eckley, I. A. (2017b). A computationally efficient nonparametric approach for changepoint detection. *Statistics and Computing*, 27(5):1293–1305.

Herr, D. and Heidenreich, D. (2015). How turbulent winds abuse wind turbine drivetrains. *Wind Power Engineering*.

Hilborn, R., Amoroso, R. O., Bogazzi, E., Jensen, O. P., Parma, A. M., Szuwalski, C., and Walters, C. J. (2017). When does fishing forage species affect their predators? *Fisheries Research*, 191:211 – 221.

Hoell, S. and Omenzetter, P. (2015). Damage detection in wind turbine blades based on time series correlations. In *7th International Conference on Structural Health Monitoring of Intelligent Infrastructure, SHMII 2015*.

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.

Jolliffe, I. (2002). *Principal component analysis*. Wiley Online Library.

Killick, R. and Eckley, I. (2014). changepoint: An r package for changepoint analysis. *Journal of Statistical Software*, 58(3):1–19.

Killick, R., Eckley, I., Jonathan, P., et al. (2013). A wavelet-based approach for detecting changes in second order structure within nonstationary time series. *Electronic Journal of Statistics*, 7:1167–1183.

Killick, R., Fearnhead, P., and Eckley, I. (2012). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, 107(500):1590–1598.

Killick, R., Haynes, K., and Eckley, I. A. (2016). *changepoint: An R package for changepoint analysis.* R package version 2.2.2.

Lavielle, M. (2005). Using penalized contrasts for the change-point problem. *Signal processing*, 85(8):1501–1510.

Nason, G. (2012). wavethresh: Wavelets statistics and transforms. r package version 4.5.

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 (Statistical Methodology)*, 62(2):271–292.

Page, E. (1954). Continuous inspection schemes. *Biometrika*, pages 100–115.

R Core Team (2018). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

Richardson, A. D., Hufkens, K., Milliman, T., Aubrecht, D. M., Chen, M., Gray, J. M., Johnston, M. R., Keenan, T. F., Klosterman, S. T., Kosmala, M., Melaas, E. K., Friedl, M. A., and Frolking, S. (2018). Tracking vegetation phenology across diverse north american biomes using phenocam imagery. *Scientific Data*, 5(180028).

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

Taleb, I., Dssouli, R., and Serhani, M. A. (2015). Big data pre-processing: a quality framework. In *Big Data (BigData Congress), 2015 IEEE International Congress on*, pages 191–198. IEEE.

Zou, C., Yin, G., Feng, L., Wang, Z., et al. (2014). Nonparametric maximum likelihood approach to multiple change-point problems. *The Annals of Statistics*, 42(3):970–1002.

List of Figures

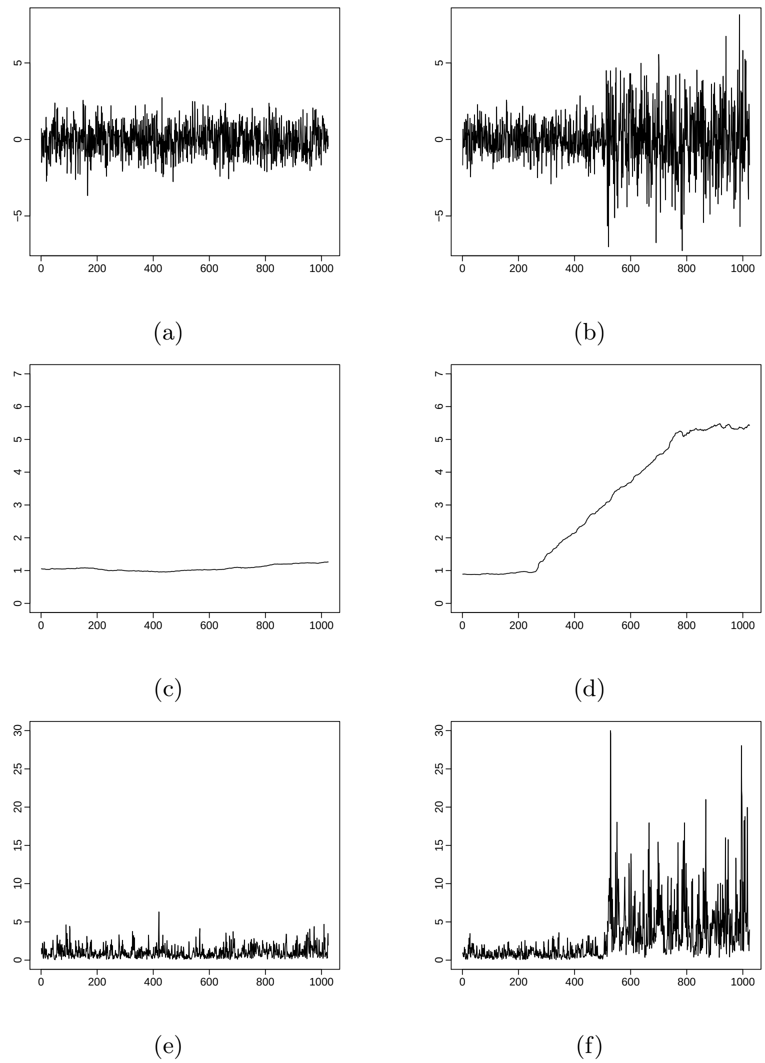


Figure 1: A time series with (a) constant variance (b) piecewise variance with their associated smoothed local variance function in (c) and (d) respectively, and their unsmoothed local variance function in (e) and (f) respectively.

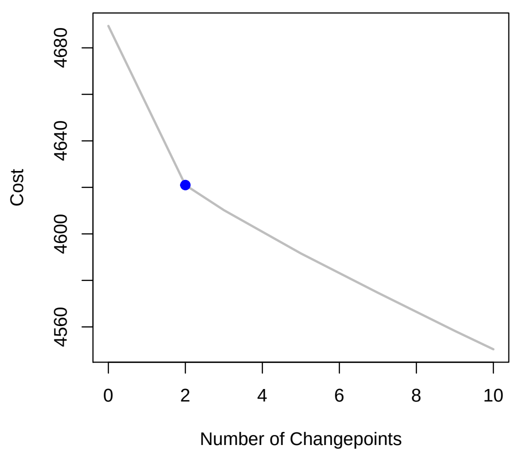
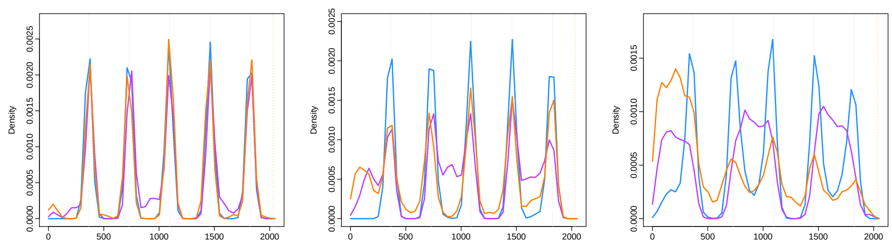
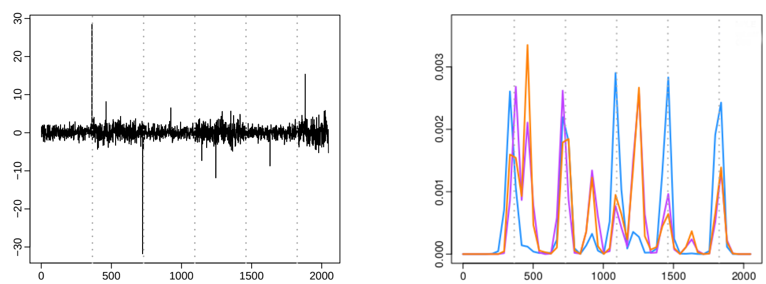


Figure 2: Example plot of the number of changepoints against the cost function for a model with two changes in variance. From the plot we can correctly identify the true number of changes to be two.



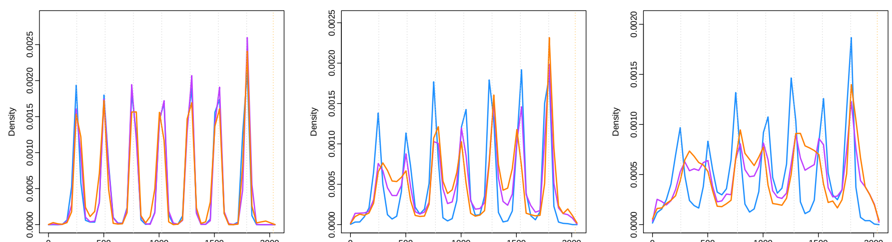
(a) $P = 0.01\%$ (b) $P = 1\%$ (c) $P = 5\%$

Figure 3: Density of detected changepoint locations using (blue) NPLE, (purple) MLvar and (orange) CSS, when the percentage of outliers is equal to (a) 0.01% (b) 1% and (c) 5%.



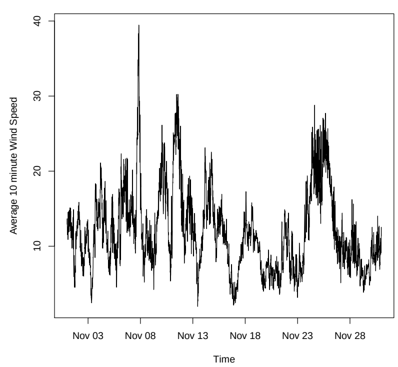
(a) Realisation (b) Changepoint Locations

Figure 4: For the outliers model, (a) a realisation of the data and (b) density plots for detected changes in variance using (blue) NPLE (purple) MLvar and (orange) CSS for the outliers model.

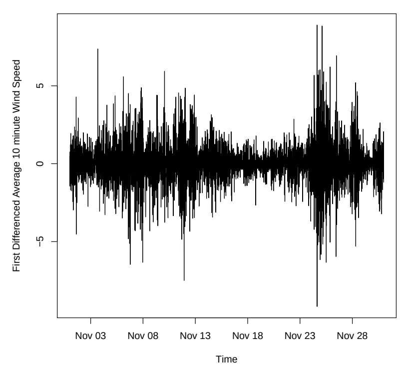


(a) $\xi = 0$ (b) $\xi = 0.25$ (c) $\xi = 0.45$

Figure 5: Density plots for detected changes in variance using (blue) NPLE (purple) MLvar and (orange) CSS for 500 realisations of simulated Generalised Extreme Value data.

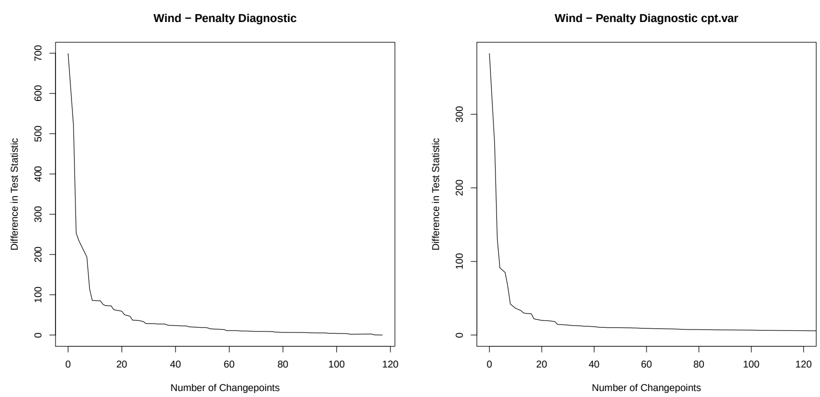


(a)



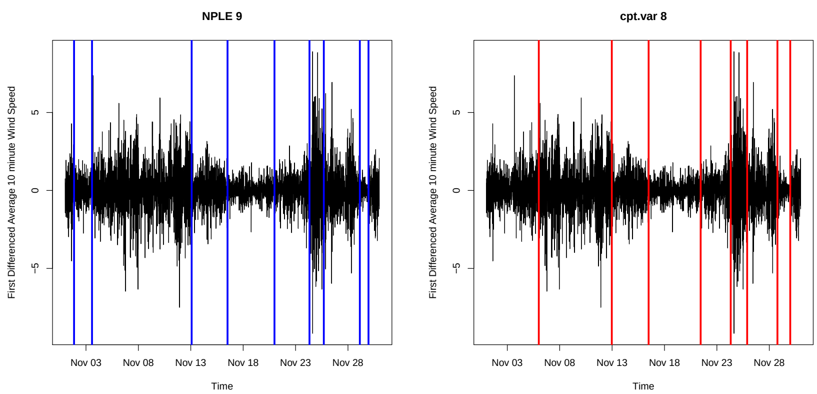
(b)

Figure 6: (a) Plot of the original Wind Speed data, (b) Difference of the data from (a).



(a) (b)

Figure 7: Diagnostic plots for (a) NPLE and (b) MLvar following the method in Lavielle (2005).



(a) (b)

Figure 8: Changepoint plots for (a) NPLE with 9 changes and (b) MLvar with 8 changes following the method in Lavielle (2005).

List of Tables

	P	0	1	2	3	4	5	6	7	≥ 8
	0.00%	0.00	0.00	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00
	0.01%	0.00	0.00	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00
NPLE	1.00%	0.00	0.00	0.00	0.00	0.02	**0.98**	0.00	0.00	0.00
	5.00%	0.01	0.00	0.01	0.08	0.07	**0.63**	0.08	0.12	0.00
	0.00%	0.00	0.00	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00
	0.01%	0.00	0.00	0.00	0.01	0.01	**0.74**	0.04	0.20	0.00
MLvar	1.00%	0.07	0.03	0.06	0.14	0.13	**0.32**	0.13	0.12	0.00
	5.00%	0.28	0.00	0.21	0.06	0.17	**0.06**	0.11	0.11	0.00
	0.00%	0.00	0.00	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00
	0.01%	0.00	0.00	0.01	0.28	0.00	**0.66**	0.00	0.01	0.04
CSS	1.00%	0.00	0.00	0.12	0.20	0.02	**0.28**	0.05	0.10	0.23
	5.00%	0.00	0.01	0.14	0.11	0.04	**0.10**	0.00	0.07	0.53

Table 1: Proportion of changepoints detected for different percentages of outliers across 500 repetitions.

	0	1	2	3	4	5	6	7	≥ 8
NPLE	0.00	0.00	0.00	0.02	0.00	**0.82**	0.01	0.15	0.00
MLvar	0.02	0.01	0.19	0.04	0.30	**0.13**	0.15	0.15	0.00
CSS	0.09	0.00	0.22	0.02	0.21	**0.14**	0.11	0.21	0.00

Table 2: Proportion of changepoints detected for the outliers model across 500 repetitions.

	ξ	0	1	2	3	4	5	6	7	≥ 8
	0.00	0.00	0.00	0.00	0.01	0.01	0.11	0.02	**0.85**	0.00
NPLE	0.25	0.00	0.00	0.00	0.03	0.04	0.19	0.08	**0.57**	0.08
	0.45	0.00	0.02	0.00	0.06	0.08	0.16	0.15	**0.31**	0.22
	0.00	0.00	0.01	0.00	0.02	0.01	0.13	0.02	**0.80**	0.01
MLvar	0.25	0.00	0.03	0.02	0.11	0.04	0.20	0.10	**0.30**	0.20
	0.45	0.00	0.06	0.08	0.12	0.08	0.20	0.11	**0.15**	0.2
	0.00	0.00	0.00	0.00	0.01	0.00	0.06	0.00	**0.87**	0.05
CSS	0.25	0.02	0.08	0.01	0.12	0.02	0.17	0.02	**0.33**	0.22
	0.45	0.06	0.07	0.09	0.15	0.05	0.16	0.07	**0.15**	0.19

Table 3: Proportion of changepoints detected for the simulated Generalised Extreme Value data.