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

J.-L. Chapman¹ | I. A. Eckley² | R. Killick²

¹STOR-i Centre for Doctoral Training, Lancaster University, Lancaster, UK
²Department of Mathematics and Statistics, Lancaster University, Lancaster, UK

Correspondence
R. Killick, Department of Mathematics and Statistics, Lancaster University, Lancaster, LA1 4YF, UK.
Email: r.killick@lancaster.ac.uk

Funding information
STOR-i Centre for Doctoral Training,
Grant/Award Number: EP/H023151/1;
The Office for National Statistics Data Science Campus

Abstract
This paper proposes a nonparametric approach to detecting changes in variance within a time series that 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 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 nonparametric setting. In addition, using a simulation study, we explore the robustness of our approach against the typical assumption of normality and presence of outliers. We illustrate the application of the approach to changes in variability of wind speeds at a location in the United Kingdom.

KEYWORDS
change points, 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 setting (see Eckley, Fearnhead, & Killick, 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 works of Andersen, Carstensen, Hernández-García, & Duarte, 2009; Evans et al., 2016; Hilborn et al., 2017; Richardson et al., 2018). In this paper, 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 increasing focus on detecting damage in wind turbine blades. As Chou, Chiu, Huang, and Chi (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, lightning strikes, and storms (Herr & Heidenreich, 2015; Hoell & Omenzetter, 2015). Among a variety of different analyses one might undertake, it may be of interest to segment the wind speed observed 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 \( \{x_i\}_{i=1}^n \) of observations reduces to the problem of finding the values of \( m \) and \( \{\tau_i\}_i \), which minimize the following expression:

\[
\sum_{i=1}^{m+1} \left[ C \left( x_{(\tau_i-1)+1}; \tau_i \right) \right] + \beta f(m).
\] (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 \( \sigma^2_{(i)} := \sigma^2_{\tau_i, \tau_i+1} = \cdots = \sigma^2_{\tau_m, \tau_{m+1}} \). 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 that guards against overfitting. 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. \( \beta \) is a tuning parameter where small values result in more changes than larger values. Different methods can be adopted in order to minimize Equation (1). Examples of exact methods include PELT (Killick, Fearnhead, & Eckley, 2012) and segment neighborhood (Auger & Lawrence, 1989), whereas binary segmentation is a commonly used approximate method (Scott & Knott, 1974).

Within the literature, a number of different variance cost functions have been proposed, one of the best known being the penalized likelihood approach (see Chen & Gupta, 2012 for a description). Within this setting, the log likelihood is calculated for each segment and summed up over segments to give the log likelihood for the data sequence. To avoid overfitting, the penalty is subtracted from the log likelihood as in (1). Alternatives to the likelihood-based approach to changepoint detection include the important works of Inclán 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-tailed structures to which a majority of approaches are intolerant. Typically, some preprocessing of the data is often performed in an attempt to mitigate these effects (Candemir & Oğuz, 2017). In some cases, this is a straightforward adaptation; however, given the unprecedented volume of data now being generated, preprocessing is becoming increasingly impractical and often subjective (Taleb, Dssouli, & Serhani, 2015). This motivates the need for new methods that are inherently resilient to such features.

The novel contribution of this paper is a nonparametric method for detecting changes in variance in the presence of outliers and heavy tails. We develop this nonparametric method using the locally stationary wavelet (LSW) model, due to Nason, Von Sachs, and Kroisandt (2000), to provide a local estimate of the variance of a time series. This paper is organized as follows: In Section 2, we describe the 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 United Kingdom.

## 2 | A NONPARAMETRIC APPROACH TO DETECTING CHANGES IN VARIANCE

In this section, we describe our nonparametric 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; Inclán & 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 LSW modeling framework that we use.

### 2.1 | LSW framework

Our method for detecting changes in variance relies upon capturing the local behavior 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 LSW framework, which is built upon nondecimated 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 has a piecewise stationary second-order structure, and 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 \( N = 2^I \geq 1 \), is an LSW process if there is a mean-square representation

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

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

1. \( \sum_{j=1}^{\infty} W_j^2(z) < \infty \) uniformly in \( z \) and
2. \( \sum_{j=1}^{\infty} J_f < \infty \).

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 a piecewise constant second-order structure to be modeled.

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

To enable various theoretical model properties to be established, 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 \( d_{j,k} = \sum_{t=1}^{N_j} X_t\psi_{j,k-t} \) to be the empirical wavelet coefficients of an LSW process \( X_{t,N} \). They then demonstrate that the corrected wavelet periodogram \( L_j(z) = \sum_{j=1}^{\infty} A_{j,l}^2 |d_{j,z}|^2 \) 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

\[
A_{j,l} = \langle \psi_j, \psi_l \rangle = \sum_{t} \sum_{k} \psi_{j,k} \psi_{j,k+t} \sum_{m} \psi_{l,m} \psi_{l,m+t}.
\]

We note that the wavelet periodogram is an inconsistent estimator of the evolutionary wavelet spectrum (see proposition 4 in 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

\[
\text{var}(X_t) = \sum_{j,k} W_j^2(k/N)\psi_{j,k-t}^2.
\]

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 an LSW process of length \( N = 2^I \). This is defined to be

\[
\sigma^2(z) = \sum_{j=1}^{J} S_j(z),
\]

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 localized 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 localized variance.

### 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 & Nason, 2006). Consequently, as the localized variance function in Equation (5) is the sum of the spectrum over scales, this means that the localized variance function will also be piecewise constant. In order to
FIGURE 1 A time series with (a) constant variance and (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.

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

$$\hat{\sigma}^2(z) = \sum_{j=1}^{J} \sum_{l=1}^{L} A_{j,l} d_{j,l}^2.$$  \hspace{1cm} (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 localized variance.
2.2.1 The nonparametric model

The localized variance function (6) is a sum of correlated \( \chi^2 \) random variables. In practice, it is difficult to obtain the distribution for this (Gordon & Ramig, 1983). We choose to adopt a nonparametric approach to this changepoint detection problem.

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

\[
\hat{G}(u) = \frac{1}{n} \left( \sum_{i=1}^{n} \mathbb{I}(\hat{\sigma}_i^2 < u) + \frac{1}{2} \mathbb{I}(\hat{\sigma}_i^2 = u) \right).
\]

(7)

where \( \hat{\sigma}_i^2 \) are assumed to be independent. With this distribution function in mind, following Zou, Yin, Feng, and Wang (2014), the maximum log likelihood of \( G(u) \) is given by

\[
n\{ \hat{G}(u) \log \hat{G}(u) + (1 - \hat{G}(u)) \log(1 - \hat{G}(u)) \},
\]

(8)

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

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

\[
\sum_{i=1}^{m+1} \left[ C \left( \hat{\sigma}^2_{(\tau_{i-1}+1):\tau_i}; u \right) \right] + \beta f(m),
\]

(9)

where the cost function for segment \( i \) is given by the negative of the empirical log likelihood of the CDF of the localized variance estimate, that is,

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

(10)

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

\[
\int_{-\infty}^{\infty} -\mathcal{L} \left( \hat{\sigma}^2_{(\tau_{i-1}+1):\tau_i}; u \right) \, dw(u),
\]

(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 \( \mathcal{O}(Mn^2 + n^3) \), where \( M \) is a specified maximum number of changepoints (Haynes, Fearnhead, & Eckley, 2017). 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. (2017) 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. (2017), we fix \( K \) and define \( \gamma = \frac{-\log(2n - 1)}{K} \). Time is then rescaled according to quantiles dependent upon the choice of \( K \). Let \( \{ t_k \}_{k=1,...,K} \) 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

\[
C_K \left( \hat{\sigma}^2_{(\tau_{i-1}+1):\tau_i}; \gamma \right) = \frac{2 \log(2n - 1)}{K} \sum_{k=1}^{K} \mathcal{L} \left( \hat{\sigma}^2_{(\tau_{i-1}+1):\tau_i}; t_k \right).
\]

(12)

We could use any search function in order to identify the changepoints using (12). However, Haynes et al. (2017) 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 Nonparametric change in variance detection using Localized Estimates, abbreviated to NPLE.
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 that 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 component analysis (Jolliffe, 2002). Figure 2 shows an example plot of a cost function against the number of changepoints identified for a model with two 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 automated, as follows.

In line with Lavielle (2005), let \( m_{\text{MAX}} \) 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, Eckley, & Fearnhead, 2017). 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 \( 0 \leq m \leq m_{\text{MAX}} \), let
   \[
   \tilde{J}_m = \frac{J_{m_{\text{MAX}}} - J_m}{J_{m_{\text{MAX}}} - J_0} m_{\text{MAX}} + 1,
   \]
   where \( J_m \) is the cost for the segmentation corresponding to \( m \) changepoints at locations \( r_{1:m} \). The associated costs have now been normalized between 1 and \( m_{\text{MAX}} + 1 \).

2. Then, for \( 1 \leq m \leq m_{\text{MAX}} - 1 \), let
   \[
   D_m = \tilde{J}_{m-1} - 2\tilde{J}_m + \tilde{J}_{m+1},
   \]
   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 \( J_m \), \( D_m \), is greater than some threshold \( S \), that is,
   \[
   \hat{m} = \max \{ 0 \leq m \leq m_{\text{MAX}} - 1 | D_m > S \}. \]

The above procedure has also been implemented for penalty choice in a wavelet context by Killick, Eckley, and Jonathan (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 data set, 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 & Gupta, 2012) and the nonparametric CSS (Inclán & Tiao, 1994). This allows for a comparison between both a parametric and a nonparametric method. Each of these is implemented using the changepoint package (Killick, Haynes, & Eckley, 2016; Killick & Eckley, 2014) in R (R Core Team, 2018). For the calculation of the localized variance estimate, we utilize the wavethresh (Nason, 2012) and changepoint.np (Haynes, 2016) packages. The study also considers departures from the idealized normal distribution change in variance setting. Specifically, the simulation study provides a practical assessment of the proposed approach’s resilience to departures from normality, including outliers and a 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 a 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 365\( i \) for \( i = 1, \ldots, 5 \). The timing of the outliers is simulated from a 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 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 degradation 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 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

| \( P \) | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | \( \geq 8 \) |
|-------|---|---|---|---|---|---|---|---|-------|
| 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 |
| MLvar | 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 | 0.01 | 0.01 | 0.74 | 0.04 | 0.20 | 0.00 |
| 1.00% | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.74 | 0.04 | 0.20 | 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 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 |
| 0.01% | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.28 | 0.00 | 0.66 | 0.00 | 0.04 |
| CSS | 0.00% | 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 | 0.00 |

Note: Detection of a variance change using, NPLE: Nonparametric Localized Estimates, MLvar = likelihood assuming Normality, CSS = cumulative sums of squares.
Density of detected changepoint locations using NPLE (Nonparametric Localized Estimates; blue), MLvar (likelihood assuming Normality; purple), and CSS (cumulative sums of squares; orange), when the percentage of outliers is equal to (a) 0.01%, (b) 1%, and (d) 5%.

A realization of the data and density plots for detected changes in variance using NPLE (Nonparametric Localized Estimates; blue), MLvar (likelihood assuming Normality; purple), and CSS (cumulative sums of squares; orange) for the outliers model 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 realization 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 is more variable as a consequence of their multiplicative nature.

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 to the location of the change. This is not the case for NPLE; however, the outlier appears to result in the change being detected slightly early. At the third change, MLvar and CSS often detect a change in 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.

Our final simulation study considered data that, instead of having outliers, exhibits heavy-tail behavior.
### Table 2
Proportion of changepoints detected for the outliers model 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 |

*Note. NPLE = Nonparametric Localized Estimates; MLvar = likelihood assuming Normality; CSS = cumulative sums of squares.*

### 3.3 Heavy-tailed structure

In this section, we consider data generated from a generalized extreme value distribution, with zero mean ($E(X_t) = 0$), 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 $\sigma = (1, 1.6, 1, 1.8, 1, 2, 1.25)$. We consider three values of the shape parameter for the generalized extreme value distribution: 0, 0.25, and 0.45. Note that as $\sigma$ 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 $\sigma$. The tails become heavier as the shape parameter, $\xi$, increases across simulations.

Table 3 shows the number of changepoints detected, and Figure 5 shows the corresponding densities for the locations. As expected, as the tails become heavier, the detection rate decreases for all methods. While 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.

### Table 3
Proportion of changepoints detected for the simulated generalized extreme value data

| $\xi$   | 0   | 1   | 2   | 3   | 4   | 5   | 6   | 7   | ≥8 |
|--------|-----|-----|-----|-----|-----|-----|-----|-----|----|
| NPLE   | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.11 | 0.02 | **0.85** | 0.00 |
|        | 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 |
| MLvar  | 0.00 | 0.00 | 0.01 | 0.00 | 0.02 | 0.01 | 0.13 | 0.02 | **0.80** | 0.01 |
|        | 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 |
| CSS    | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.06 | 0.00 | 0.00 | **0.87** | 0.05 |
|        | 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 |

*Note. NPLE = Nonparametric Localized Estimates; MLvar = likelihood assuming Normality; CSS = cumulative sums of squares.*

### Figure 5
Density plots for detected changes in variance using NPLE (Nonparametric Localized Estimates; blue), MLvar (likelihood assuming Normality; purple), and CSS (cumulative sums of squares; orange) for 500 realizations of simulated generalized extreme value data. (a) $\xi = 0$, (b) $\xi = 0.25$, (c) $\xi = 0.45$
This illustrates NPLE’s reduced sensitivity 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 analyze 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 4,261 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 behavior. The resulting series has a very clear, nonconstant variance structure (Figure 6b). There also appear to be some anomalous observations that could potentially affect changepoint estimation. Note that while 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

![Figure 6](image_url)  
(a) Plot of the original wind speed data. (b) First difference of the data from (a)

![Figure 7](image_url)  
(a) Diagnostic plots for (a) NPLE (Nonparametric Localized Estimates) and (b) MLvar (likelihood assuming Normality) following the method in Lavielle (2005)
methods. The diagnostic plots are given in Figure 7, where it is clear that the elbow in the curve for NPLE is at nine changes and that for MLvar is at eight 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.

5 | CONCLUSION

In this paper, 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 penalized 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.

ACKNOWLEDGEMENTS

Chapman gratefully acknowledges the financial support of EPSRC, through the STOR-i Centre for Doctoral Training (EP/H023151/1), and The Office for National Statistics Data Science Campus.

ORCID

I. A. Eckley
https://orcid.org/0000-0001-9957-2460
R. Killick
https://orcid.org/0000-0003-0583-3960

REFERENCES

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

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

Candemir, C., & 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)* (pp. 218–224). Bern, Switzerland: IEEE.

Chen, J., & Gupta, A. (2012). *Parametric statistical change point analysis*. Boston, MA: Birkhäuser Boston.
