Environmental and Ecological Statistics
https://doi.org/10.1007/s10651-021-00485-5

Modified information criterion for regular change point models based on confidence distribution

Suthakaran Ratnasingam[1] · Wei Ning[2,3]

Received: 13 March 2020 / Revised: 23 December 2020 / Accepted: 4 January 2021

Abstract
In this article, we propose procedures based on the modified information criterion and the confidence distribution for detecting and estimating changes in a three-parameter Weibull distribution. Corresponding asymptotic results of the test statistic associated the detection procedure are established. Moreover, instead of only providing point estimates of change locations, the proposed estimation procedure provides the confidence sets for change locations at a given significance level through the confidence distribution. In general, the proposed procedures are valid for a large class of parametric distributions under Wald conditions and the certain regularity conditions being satisfied. Simulations are conducted to investigate the performance of the proposed method in terms of powers, coverage probabilities and average lengths of confidence sets with respect to a three-parameter Weibull distribution. Corresponding comparisons are also made with other existing methods to indicate the advantages of the proposed method. Rainfall data is used to illustrate the application of the proposed method.

Keywords Change point · Confidence distribution · Information criterion · Weibull distribution

1 Introduction

Change point analysis plays an important role in identifying points in time when the probability distribution of stochastic processes or time series changes. When a change

Handling Editor: Luiz Duczmal

✉ Wei Ning
wning@bgsu.edu

1 Department of Mathematics, California State University, San Bernardino, San Bernardino, CA 92407, USA

2 Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH 43403, USA

3 School of Mathematics and Statistics, Beijing Institute of Technology, Beijing 100081, China

point exists, it is not advisable to perform a statistical analysis without taking into account of the existence of that change point because it could lead to misleading results. The change point analysis attempts to identify the number of change point(s) and the corresponding location(s). Change-point analysis has been extensively explored since Page (1954, 1955). Sen and Srivastava (1975a, b) derived the exact and asymptotic distribution of their test statistic for testing a single change in the mean of a sequence of normal random variables. Worsley (1979) studied the power of likelihood ratio and cumulative sum tests for a change in a binomial probability model. Vostrikova (1981) proposed the binary segmentation procedure to detect multiple changes in the data set. This approach has the advantage of detecting multiple structural changes recursively and saving a great deal of computing time. Srivastava and Worsley (1986) studied the multiple changes in the multivariate normal mean and approximated the null distribution of the likelihood ratio test statistic based on an improved Bonferroni inequality. Chen and Gupta (1995) studied the likelihood procedure for testing the change point hypothesis under the multivariate Gaussian model. Asymptotic results of the test based on the likelihood ratio test can be found in Csörgő and Horváth (1997). Gurevich and Vexler (2005) investigated the change-point problem in logistic regression. Wu (2008) provided a simultaneous change-point analysis and variable selection in a regression problem. Ning and Gupta (2009) studied the change-point problem for the generalized lambda distribution. Ramanayake and Gupta (2010) considered the problem of detecting a change-point in an exponential distribution with repeated values. Chen and Gupta (2012) discussed change-point problems for various parametric models with different approaches. Arellano-Valle and C. L. and Loschi, R. H. (2013) presented a Bayesian approach to study the change-point problem of the skew normal distribution. Ngunkeng and Ning (2014) studied the different change-point scenarios for the skew normal distribution from the viewpoint of the model selection. Alghamdi et al. (2018) proposed the change-point detection procedure for the Rayleigh Lomax distribution using Schwartz and modified information criterion. Recently, Ratnasingam and Ning (2020) studied Confidence Distribution based approach for a skew normal change-point model incorporating modified information criterion.

Among all the distributions, the Weibull distribution is one of the most widely used lifetime distributions in reliability engineering. Due to its versatility, it can take on the characteristics of other types of distributions based on the value of the shape parameter. The three parameters Weibull distribution is defined as follows.

$$f_X(x) = \begin{cases} \dfrac{\alpha}{\beta}\left(\dfrac{x-\theta}{\beta}\right)^{\alpha-1} \exp\left(-\left(\dfrac{x-\theta}{\beta}\right)^{\alpha}\right); & x \geq \theta \\ 0\,; & x < \theta \end{cases} \tag{1}$$

and the cumulative distribution function is,

$$F_X(x) = 1 - \exp\left(-\left(\frac{x-\theta}{\beta}\right)^{\alpha}\right), \tag{2}$$

where θ is the location parameter, $\beta > 0$ is the scale parameter, and $\alpha > 0$ is the shape parameter. On short hand, we denote $W(\theta, \alpha, \beta)$. When $\theta = 0$, it reduces to two-parameter Weibull distribution.

Jandhyala et al. (1999) proposed a change-point methodology for identifying changes in the scale and shape parameters of a two-parameter Weibull distribution. The asymptotic results of the likelihood ratio test (LRT) statistic for detecting unknown changes in the parameters were derived as well as maximum likelihood estimate (MLE) of the unknown change point were obtained. They applied such a Weibull change-point model to model minimal daily temperatures measured in Uppsala. Juruskova (2007) investigated the asymptotic behavior of a log-likelihood ratio statistic for testing a change in a three-parameter Weibull distribution. In this paper, we propose a detection procedure based on the modified information criterion (MIC) to detect simultaneous changes in parameters of a three-parameter Weibull distribution. Moreover, instead of only providing point estimates of change locations, we propose a method based on the confidence distribution (CD) to construct the confidence sets for change locations at a given significance level. The proposed procedures can be applicable for a large class of parametric distributions under Wald conditions and the certain regularity conditions being satisfied. The interested readers may refer to Chen et al. (2006) for further information.

This paper is organized as follows. In Sect. 2, the detecting procedure based on the modified information criterion (MIC) is proposed. The asymptotic results of the test statistic associated with the detecting procedure is shown to be a χ^2 distribution. In Sect. 3, we develop a change point detection procedure for the three-parameter Weibull distribution. The procedure of constructing the confidence set of a change through the confidence distribution is provided in Section 4. Simulations to investigate the performance of the proposed procedures and compare them with other existing methods in terms of powers, coverage probabilities, and lengths of confidence sets are conducted in Sect. 5. A real data application is given in Sect. 6 to illustrate the proposed procedure. Some discussion is provided in Sect. 7.

2 Methodology

2.1 Modified information criterion

In general, the change point problem involves hypothesis testing and parameter estimation. More specifically, we need to test the null hypothesis of no change point versus the alternative hypothesis of having at least one change. Further, we need to estimate the corresponding location of the change point, if there is any. One of the most popular methods of detecting change points is the use of model selection criteria. The Schwarz information criterion (SIC) Schwarz (1978) is one of the popular criteria for model selection. Zhang and Siegmund (2007) noted that the conventional SIC could detect change points more effectively when changes take place in the middle of the data. However, as Chen et al. (2006) pointed out, the conventional SIC method did not consider the complexity of the model which may cause the redundancy of the parameter space, especially a change occurring near the beginning or the end of data. To tackle this issue, Chen et al. (2006) proposed the modified information criterion (MIC) by adjusting the penalty term in SIC so that it reflects the contributions of change-point locations to model complexity. This approach assigns a larger penalty when the change point location is close to the first or the last observation in the data set.

Let $x_1, \ldots, x_n$ be a random sample drawn from the density function $f(x; \Theta)$. The Schwarz information criterion (SIC) proposed by Schwarz (1978) is given as follows.

$$\text{SIC} = -2\ell_n(\hat{\Theta}) + \dim(\hat{\Theta}) \log(n), \tag{3}$$

where $\ell_n(\cdot)$ is the log-likelihood function of the random sample, $\hat{\Theta}$ is the maximum likelihood estimate (MLE) of the parameter Θ and $\dim(\hat{\Theta})$ is the dimension of the parameter space. We denote Θ_L, Θ_R to be the pre-change and post-change parameters respectively and $\hat{\Theta}_L$, $\hat{\Theta}_R$ to be the MLEs of the pre-change and post-change parameters. Let k be the unknown change point location. The SIC in the context of having at least one change point can be written as

$$\text{SIC}(k) = -2\ell_n(\hat{\Theta}_L(k), \hat{\Theta}_R(k), k) + \left\{2\dim(\hat{\Theta}_L(k)) + 1\right\} \log(n), \tag{4}$$

where $1 \leq k < n$. Further, Eq. (3) defines the SIC under the null hypothesis of no change, which we denote as SIC(n). However, as Chen et al. (2006) pointed out that the SIC(k) defined in (4) does not consider the change location to be a parameter, which may cause the redundancy of the parameter space when the change occurs near the beginning, or the end of data. Therefore, the modified information criterion (MIC) proposed by Chen et al. (2006) is given as follows. Under the null hypothesis of no change, the MIC is defined as,

$$\text{MIC}(n) = -2\ell_n(\hat{\Theta}) + \dim(\hat{\Theta}) \log(n), \tag{5}$$

where $\hat{\Theta}$ maximizes $\ell_n(\Theta)$. Therefore, under H_0, both SIC(n) and MIC(n) are same. Under the alternative hypothesis, the MIC is defined as,

$$\text{MIC}(k) = -2\ell_n(\hat{\Theta}_L(k), \hat{\Theta}_R(k), k) + \left\{2\dim(\hat{\Theta}_L(k)) + \left(\frac{2k}{n} - 1\right)^2\right\} \log(n), \tag{6}$$

where $1 \leq k < n$. The difference between (4) and (6) is that (6) considers the contribution of the change location k to the model as a parameter. If $\text{MIC}(n) > \min_{1 \leq k < n} \text{MIC}(k)$, then we select the model with a change point and the estimate of the change point is given by

$$\hat{k} = \underset{1 \leq k < n}{\arg\min}\{\text{MIC}(k)\}. \tag{7}$$

Moreover, for the purpose of verifying the statistical significance of the detected change point, the associated MIC-based test statistic is defined as,

$$S_n = \text{MIC}(n) - \min_{1 \leq k < n} \text{MIC}(k) + \dim(\Theta) \log(n), \tag{8}$$

where MIC(n) and MIC(k) are defined in (5) and (6) respectively. In particular, the above standardization eliminates the constant term $\dim(\Theta)\log n$ in the difference of MIC(n) and MIC(k). Chen et al. (2006) showed, under Wald conditions and the regularity conditions, as $n \longrightarrow \infty$,

$$S_n \longrightarrow \chi_d^2, \tag{9}$$

in distribution under H_0, where d is the dimension of Θ. For the purpose of comparison later, we also provide the test statistic associated with the conventional SIC procedure as follows.

$$T_n = \text{SIC}(n) - \min_{1 \le k < n} \text{SIC}(k) + \dim(\Theta)\log n. \tag{10}$$

Chen and Gupta (1997) and Csörgő and Horváth (1997) pointed out the asymptotic distribution of the related statistic for the SIC is found to have type I extreme value distribution. We point out here that the test statistic S_n is constructed based on the modified information criteria (MIC) of the null and the alternative hypotheses. When we use the information criterion such as SIC and MIC, to detect changes by comparing the values under the null hypothesis and the values under the alternative hypothesis directly, one may raise the question that the small differences among SICs and MICs might be caused by the fluctuation of the data, therefore, there may be no change at all, especially when the SICs and MICs are very close. Therefore, to make the conclusion about the change point statistically convincing, the constructed S_n is proposed with the given significance level α and the associated critical value c_α. Moreover, the test statistic T_n is constructed based on the conventional information criterion SIC under the null and the alternative hypotheses. $SIC(n)$ is defined same as the Eq. (3), and $SIC(k)$ is given in the Eq. (4). The last constant term $\dim(\Theta)\log n$ in the Eq. (10) is to remove the constant term $\dim(\Theta)\log n$ in the difference between $SIC(n)$ and $SIC(k)$. The structure of S_n is similar. The main difference between S_n and T_n is that S_n incorporates the contribution of the change location k associated with the complexity of the model, while T_n does not. As a result, S_n is desirable when the change point location k is close to the first or the last observation in the data set.

3 MIC-based detection procedure for three-parameter weibull distribution

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent random variables belong to a three-parameter Weibull distribution. The change point problem for a three-parameter Weibull distribution is defined as follows.

$$X_i \sim \begin{cases} W(\theta_L, \alpha_L, \beta_L) & i = 1, \ldots, k \\ W(\theta_R, \alpha_R, \beta_R) & i = (k+1), \ldots, n. \end{cases} \tag{11}$$

where the pdf and cdf of three-parameter Weibull distribution are given in (1) and (2). We are testing the following hypotheses.

$$\begin{aligned} H_0 : \theta_1 &= \theta_2 = \cdots = \theta_n = \theta \\ \alpha_1 &= \alpha_2 = \cdots = \alpha_n = \alpha \\ \beta_1 &= \beta_2 = \cdots = \beta_n = \beta, \end{aligned}$$

versus

$$\begin{aligned} H_1 :\ & \underbrace{\theta_1 = \cdots = \theta_k}_{\theta_L} \neq \underbrace{\theta_{k+1} = \cdots = \theta_n}_{\theta_R} \\ & \underbrace{\alpha_1 = \cdots = \alpha_k}_{\alpha_L} \neq \underbrace{\alpha_{k+1} = \cdots = \alpha_n}_{\alpha_R} \\ & \underbrace{\beta_1 = \cdots = \beta_k}_{\beta_L} \neq \underbrace{\beta_{k+1} = \cdots = \beta_n}_{\beta_R}, \end{aligned}$$

where (α, θ, β), $(\alpha_L, \theta_L, \beta_L)$ and $(\alpha_R, \theta_R, \beta_R)$ are unknown parameters and need to be estimated. Parameter k is the unknown change location and needs to be estimated as well. Under the null hypothesis, the log-likelihood function is given as,

$$\ell_n(\theta, \alpha, \beta) = n\log(\alpha) - n\log(\beta) + (\alpha - 1)\sum_{i=1}^{n}\log\left(\frac{x_i - \theta}{\beta}\right) - \sum_{i=1}^{n}\left(\frac{x_i - \theta}{\beta}\right)^{\alpha} \tag{12}$$

The maximum likelihood estimators (MLEs) of θ, α and β can be obtained by setting these partial derivatives equal to zero.

$$\begin{aligned} \frac{\partial \ell_n(\theta, \alpha, \beta)}{\partial \alpha} &= \frac{n}{\alpha} + \sum_{i=1}^{n}\log\left(\frac{x_i - \theta}{\beta}\right) - \sum_{i=1}^{n}\left(\frac{x_i - \theta}{\beta}\right)^{\alpha}\log\left(\frac{x_i - \theta}{\beta}\right) \\ \frac{\partial \ell_n(\theta, \alpha, \beta)}{\partial \beta} &= \frac{-n\alpha}{\beta} + \frac{\alpha}{\beta}\sum_{i=1}^{n}\left(\frac{x_i - \theta}{\beta}\right)^{\alpha+1} \\ \frac{\partial \ell_n(\theta, \alpha, \beta)}{\partial \theta} &= (\theta - 1)\sum_{i=1}^{n}\left(\frac{1}{x_i - \theta}\right) + \frac{\alpha}{\beta}\sum_{i=1}^{n}\left(\frac{x_i - \theta}{\beta}\right)^{\alpha-1} \end{aligned}$$

The $\mathrm{MIC}(n)$ is defined as,

$$\mathrm{MIC}(n) = -2\ell_n(\hat{\theta}, \hat{\alpha}, \hat{\beta}) + 3\log(n), \tag{13}$$

where $\hat{\theta}$, $\hat{\alpha}$ and $\hat{\beta}$ are the MLEs of θ, α and β respectively. Similarly, under the alternative hypothesis, the log-likelihood function is,

$$\begin{aligned}
\ell_{H_1} = l(k, \theta_L, \alpha_L, \beta_L, \theta_R, \alpha_R, \beta_R) &= \sum_{i=1}^{k} \log(f(x_i, \theta_L, \alpha_L, \beta_L) \\
&+ \sum_{i=k+1}^{n} \log(f(x_i, \theta_R, \alpha_R, \beta_R) \\
&= \Bigg\{k\log(\alpha_L) - k\log(\beta_L) + (\alpha_L - 1)\sum_{i=1}^{k}\log\left(\frac{x_i - \theta_L}{\beta_L}\right) \\
&- \sum_{i=1}^{k}\left(\frac{x_i - \theta_L}{\beta_L}\right)^{\alpha_L}\Bigg\} + \Bigg\{(n-k)\log(\alpha_R) - (n-k)\log(\beta_R) \\
&+ (\alpha_R - 1)\sum_{i=k+1}^{n}\log\left(\frac{x_i - \theta_R}{\beta_R}\right) - \sum_{i=k+1}^{n}\left(\frac{x_i - \theta_R}{\beta_R}\right)^{\alpha_R}\Bigg\}
\end{aligned} \tag{14}$$

The MLEs of the pre-change parameters θ_L, α_L and β_L can be obtained by solving the following equations.

$$\begin{aligned}
\frac{\partial \ell_{H_1}}{\partial \alpha_L} &= \frac{k}{\alpha_L} + \sum_{i=1}^{k}\log\left(\frac{x_i - \theta_L}{\beta_L}\right) - \sum_{i=1}^{k}\left(\frac{x_i - \theta_L}{\beta_L}\right)^{\alpha_L}\log\left(\frac{x_i - \theta_L}{\beta_L}\right) \\
\frac{\partial \ell_{H_1}}{\partial \beta_L} &= \frac{-k\alpha_L}{\beta_L} + \frac{\alpha_L}{\beta_L}\sum_{i=1}^{k}\left(\frac{x_i - \theta_L}{\beta_L}\right)^{\alpha_L + 1} \\
\frac{\partial \ell_{H_1}}{\partial \theta_L} &= (\theta_L - 1)\sum_{i=1}^{k}\left(\frac{1}{x_i - \theta_L}\right) + \frac{\alpha_L}{\beta_L}\sum_{i=1}^{k}\left(\frac{x_i - \theta_L}{\beta_L}\right)^{\alpha_L - 1}
\end{aligned}$$

and the MLEs of the post-change parameters θ_R, α_R and β_R are the solutions of the following equations.

$$\begin{aligned}
\frac{\partial \ell_{H_1}}{\partial \alpha_R} &= \frac{(n-k)}{\alpha_R} + \sum_{i=k+1}^{n}\log\left(\frac{x_i - \theta_R}{\beta_R}\right) - \sum_{i=k+1}^{n}\left(\frac{x_i - \theta_R}{\beta_R}\right)^{\alpha_R}\log\left(\frac{x_i - \theta_R}{\beta_R}\right) \\
\frac{\partial \ell_{H_1}}{\partial \beta_R} &= \frac{-(n-k)\alpha_R}{\beta_R} + \frac{\alpha_R}{\beta_R}\sum_{i=k+1}^{n}\left(\frac{x_i - \theta_R}{\beta_R}\right)^{\alpha_R + 1} \\
\frac{\partial \ell_{H_1}}{\partial \theta_R} &= (\theta_R - 1)\sum_{i=k+1}^{n}\left(\frac{1}{x_i - \theta_R}\right) + \frac{\alpha_R}{\beta_R}\sum_{i=k+1}^{n}\left(\frac{x_i - \theta_R}{\beta_R}\right)^{\alpha_R - 1}
\end{aligned}$$

Now MIC(k) is given by,

$$\mathrm{MIC}(k) = -2l(k, \hat{\theta}_L, \hat{\alpha}_L, \hat{\beta}_L, \hat{\theta}_R, \hat{\alpha}_R, \hat{\beta}_R) + \left\{6 + \left(\frac{2k}{n} - 1\right)^2\right\} \log n. \tag{15}$$

where $(\hat{\theta}_L, \hat{\alpha}_L, \hat{\beta}_L)$ and $(\hat{\theta}_R, \hat{\alpha}_R, \hat{\beta}_R)$ are MLEs of the parameters before and after the change respectively. If there is a change, the change point k is estimated by the equation (7). Furthermore, the MIC-based test statistics S_n can be obtained from (8). Under the Wald conditions and the regularity conditions provided by Chen et al. (2006), we have the following theorems.

Theorem 1 *As* $n \to \infty$,

$$S_n \longrightarrow \chi^2_3 \,,$$

in distribution under the null hypothesis.

Theorem 1 implies that S_n defined in (8) is consistent when there is a fixed amount of change in the Weibull parameters at k, so that, $k = n$ has a limit in (0,1), the model with change point is chosen with the probability approaching to one.

Theorem 2 *As* $n \to \infty$, *the change point satisfies* $0 < k/n < 1$. *Then the change point estimator* $\hat{k}$ *satisfies*

$$\hat{k} - k = O_p(1) \,.$$

The proofs of both Theorems are similar to ones in Chen et al. (2006). Theorem 2 implies that the change point $\hat{k}$ achieves the best convergence rate.

4 Confidence curve for three parameter weibull distribution

In this section, we provide steps to construct a confidence curve for the change point in a three-parameter Weibull distribution based on MIC. Most existing literature on change-point problem focused on providing the point estimate of the change location. Recently, Cunen et al. (2018) proposed the confidence curve along with the confidence sets for the change point estimate through the confidence distribution (CD). The concept of a CD has its roots in Fisher's fiducial distribution. A CD is similar to a point estimator or an interval estimator but it uses a sample-dependent distribution function on the parameter space to estimate the parameter of interest. It also can provide confidence intervals of all nominal levels for a parameter of interest through confidence curves. More details and recent developments are referred to Xie and Singh (2013).

Cunen et al. (2018) used the traditional log-likelihood function to obtain the point estimate of the change location, then constructed the confidence curves and confidence sets at given nominal levels through CD. However, as we mentioned in Sect. 2, this method does not consider the complexity of the model. Therefore, it is not effective due to the possible redundancy of the parameter space, especially when the change point is near the beginning or the end of data set. Thus, we modify their approach by estimating the change location $\hat{k}$ using (7). The confidence curve for a three-parameter Weibull distribution can be obtained as follows.

Step 1: The profile log-likelihood function can be obtained by maximizing the log-likelihood function (14) over the parameters for each candidate value of k where $1 \leq k < n-1$.

$$\begin{aligned}\ell_{prof}(k) &= \max\Big(l(k, \theta_L, \alpha_L, \beta_L, \theta_R, \alpha_R, \beta_R)\Big) \\ &= l(k, \hat{\theta}_L, \hat{\alpha}_L, \hat{\beta}_L, \hat{\theta}_R, \hat{\alpha}_R, \hat{\beta}_R)\,.\end{aligned}$$

Step 2: The deviance is given by,

$$D(k, \boldsymbol{x}) = 2\{\ell_{prof}(\hat{k}) - \ell_{prof}(k)\}\,, \tag{16}$$

where $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$ and $x_1, x_2, \cdots, x_k$ is a sample coming from the distribution $W(\theta_L, \alpha_L, \beta_L)$ and $x_{k+1}, \cdots, x_n$ coming from $W(\theta_R, \alpha_R, \beta_R)$. In particular, $\ell_{prof}(\hat{k}) = \max_{1 \leq k < n}(\ell_{prof}(k))$ and $\hat{k}$ is obtained by using (7) which is different from Cunen et al. (2018).

Step 3: To construct a confidence curve for k based on the deviance function, we consider the estimated distribution of $D(k, \boldsymbol{x})$ at position k as follows.

$$\Psi_k(x) = P_{k, \hat{\Theta}_L, \hat{\Theta}_R}\{D(k, \boldsymbol{x}) < x\}\,, \tag{17}$$

where $x \in \mathbb{R}$. In the case of continuous parameters, Wilks theorem states that $\Psi_k(x)$ is approximately the distribution function of a χ^2_1. However, Wilks theorem does not hold for a discrete parameter k. Therefore, we compute Ψ_k through the simulations. The confidence curve can be constructed as,

$$cc(k, \boldsymbol{x}_{obs}) = \Psi_k(D(k, \boldsymbol{x}_{obs})) = P_{k, \hat{\Theta}_L, \hat{\Theta}_R}\{D(k, \boldsymbol{x}) < D(k, \boldsymbol{x}_{obs})\}\,. \tag{18}$$

The probability that $cc(k, \boldsymbol{x}_{obs}) < \alpha$, under the true value of k, is often approximated well with α. Then, the confidence sets for k can be visualized using the plot $cc(k, \boldsymbol{x}_{obs})$. The $cc(k, \boldsymbol{x}_{obs})$ is the acceptance probability for k, or one minus the p-value for testing that value of k by using the deviance-based test which rejects the null hypothesis for high values of $D(k, \boldsymbol{x})$. We compute Ψ_k and hence $cc(k, \boldsymbol{x}_{obs})$ by simulations as follows.

$$cc(k, \boldsymbol{x}_{obs}) = \frac{1}{B}\sum_{j=1}^{B} I\big\{D\big(k, \boldsymbol{x}^*_j\big) < D(k, \boldsymbol{x}_{obs})\big\}, \tag{19}$$

for large number of B of simulated copies of data set $\boldsymbol{x}^*$. For each possible value of k, we simulate data $\boldsymbol{x}^*_j$, $j = 1, \cdots, B$ from $f(x, \Theta_L)$ and $f(x, \Theta_R)$ to the left and right side of k respectively. See Cunen et al. (2018) for more details. Our proposed procedure uses MIC defined in (6) to construct the confidence curves. This is different from Cunen et al. (2018) approach where they estimate the change point location k by maximizing the profile-likelihood function over all possible values of k, however, we estimate k using (7) by considering the impact of the locations of changes. The MIC-based statistics S_n in (8) can be used to confirm a significant change statistically to avoid the fluctuations caused by noise.

5 Simulation study

In this section, we conduct simulations at various values of the change point location k with different sample sizes $n = \{50, 100, 150\}$. The pre-change distribution is set to be $W(1, 1, 2)$ and the post-change distribution after the change are set to be $W(1.25, 1.25, 2.25)$, $W(1.5, 1.5, 2.5)$, and $W(1.75, 1.75, 2.75)$ respectively. Since S_n defined in (8) and T_n defined in (10) have different probabilities of Type I error, therefore, we cannot simply compare the power of S_n and T_n directly. Thus, to make fair comparisons, we consider the powers of T_n^* which is the statistic T_n after the probability of Type I error of being adjusted to be equal to the corresponding S_n by increasing (decreasing) its critical values.

First, we verify the null asymptotic distribution of S_n stated in Theorem 1 numerically. For different sample sizes $n = \{100, 200, 400\}$ we sketch the χ_3^2 quantile-quantile(Q-Q) plot for S_n values in Figs. 1, 2 and 3. From the plots, we observe that the null asymptotic distribution of S_n can be approximated to χ_3^2 reasonably when the sample size increases. This confirms the result given in Theorem 1.

Second, we investigate the convergence of the change point estimator $\hat{k}$ in terms of the empirical probability distribution of $|\hat{k} - k| \leq \delta$ at various sample sizes and different values of the parameters, where k is the true value of change location and δ is set to be the difference between the estimated and the true value of k. The pre-change distribution is set to be $W(1, 1, 2)$ and the post-change distribution after the change are set to be $W(1.25, 1.25, 2.25)$, $W(1.5, 1.5, 2.5)$, $W(1.75, 1.75, 2.75)$, and $W(2.75, 2.75, 3.75)$ respectively. We consider various values for δ, including $\delta = \{1, 2, 3, 4, 5\}$. The results are summarized in Tables 1, 2, and 3. The results indicate the convergence rate closer to 1 with the large values of δ and the differences in parameters. We also observe that the proposed method can provide better convergence rates with relatively large sample sizes such as $n = 100$ and $n = 150$ comparing to relatively small sample size such as $n = 50$.

The power comparisons between S_n and T_n^* with different values of parameters, samples sizes and change locations are reported in Tables 4, 5 and 6. The true change point k considering $\{10, 15, 25\}$, $\{20, 40, 50\}$, $\{25, 50, 75\}$ for sample size 50, 100, 150 respectively. Because of the symmetric property of the performance, in our simulation, we only consider the true change-point locations below or equal to the midpoint of the data set. From the tables, it can be clearly seen, in both procedures, the power increases as the difference between the parameters increases. Further, the power tends to increase when sample size increases. Unlike T_n^*, the S_n method considers the impact of the change point location. Consequently, the power based on the S_n method is higher than the T_n^* method when the change occurs near the beginning or the end of the data. For example, with the sample size 50 and the post-change distribution $W(1.25, 1.25, 2.25)$, at the change locations $k = 10$ and $k = 15$, the powers of T_n^* are 0.452 and 0.499 respectively, which are lower than 0.490 and 0.542 of S_n. The results of these two methods are graphed in Fig. 4.

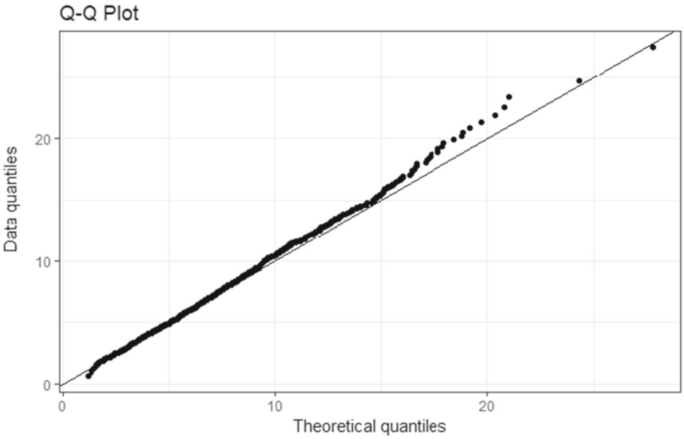


Fig. 1 The Chi-square Q-Q plot of S_n for $n = 100$

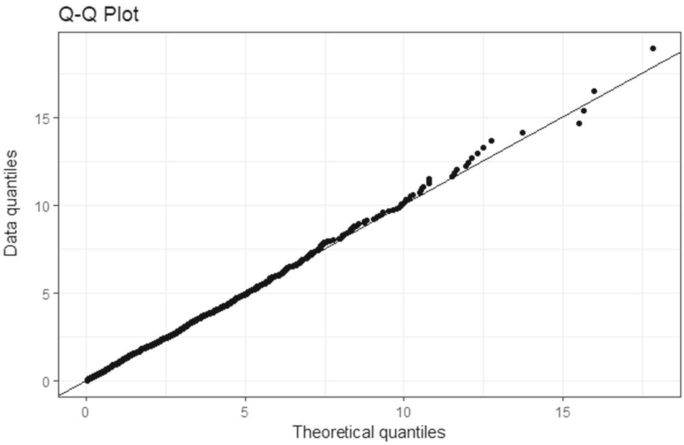


Fig. 2 The Chi-square Q-Q plot of S_n for $n = 200$

We also conduct simulations to compare the performance of the proposed method based on MIC in Sect. 3 and the one proposed by Cunen et al. (2018) in terms of the coverage probabilities and average sizes of confidence sets of change points at various scenarios. A confidence set of a change point can be defined as $\{k : cc(k, x) \leq \alpha\}$. The size of a confidence set is determined by the number of k belonging to the confidence set with a given nominal level. For sample size $n = 50$ and 100, we generate the data set with the true change point locations $k = \{10, 15, 25\}$ and $k = \{20, 40, 50\}$ respectively. The results are summarized in Tables 7, 8, 9 and 10. The MIC-based method provides better coverage probabilities when the change point is located near the beginning or the end of the data set. Not surprisingly, both methods provide approximately same coverage probabilities when the change occurs in the middle. Further, when the change point location k is near the beginning or the end of the data, MIC-based approach provides thinner confidence sets compare to the

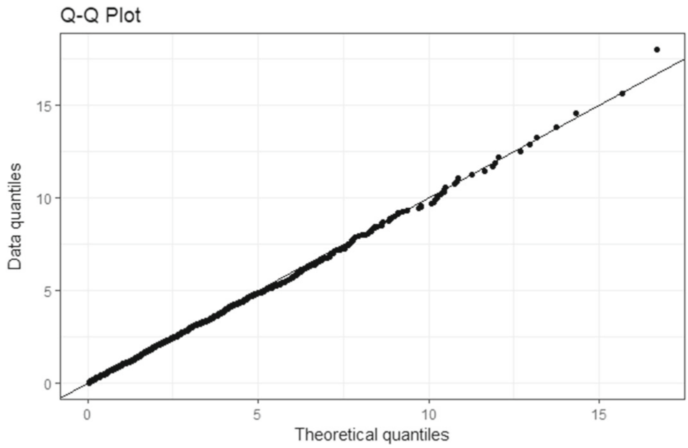


Fig. 3 The Chi-square Q-Q plot of S_n for $n = 400$

Table 1 Probability distribution of $|\hat{k} - k| \leq \delta$ and the sample size is $n = 50$ at various change point locations k

	$k \setminus \delta \rightarrow$	$P(\lvert\hat{k} - k\rvert \leq \delta)$				
		1	2	3	4	5
W(1.25, 1.25, 2.25)	10	0.156	0.246	0.308	0.573	0.634
	15	0.150	0.233	0.302	0.568	0.668
	25	0.167	0.242	0.307	0.558	0.610
W(1.5, 1.5, 2.5)	10	0.365	0.508	0.594	0.766	0.810
	15	0.341	0.488	0.586	0.749	0.896
	25	0.350	0.500	0.578	0.737	0.885
W(1.75, 1.75, 2.75)	10	0.485	0.642	0.731	0.883	0.929
	15	0.552	0.708	0.788	0.837	0.967
	25	0.532	0.692	0.759	0.805	0.936
W(2.75, 2.75, 3.75)	10	0.905	0.931	0.939	0.945	0.977
	15	0.904	0.930	0.939	0.941	0.968
	25	0.861	0.881	0.885	0.896	0.951

traditional log-likelihood based approach used by Cunen et al. (2018). Similarly, if the change occurs in the center, the average sizes of the confidence sets are roughly equal for both methods.

6 Application

In this section, we use the annual maximum rainfall data at one rain gauge in Fort Collins, Colorado from 1900 through 1999. The data is available under extRemes package in R software, see Gilleland and Katz (2016). The data consist of 100 obser-

Table 2 Probability distribution of $|\hat{k} - k| \leq \delta$ and the sample size is $n = 100$ at various change point locations k

	$k \setminus \delta \rightarrow$	$P(\vert\hat{k} - k\vert \leq \delta)$				
		1	2	3	4	5
W(1.25, 1.25, 2.25)	20	0.222	0.342	0.417	0.771	0.828
	40	0.268	0.385	0.464	0.720	0.870
	50	0.269	0.352	0.436	0.787	0.830
W(1.5, 1.5, 2.5)	20	0.508	0.620	0.702	0.863	0.994
	40	0.487	0.625	0.712	0.863	0.992
	50	0.537	0.649	0.709	0.855	0.982
W(1.75, 1.75, 2.75)	20	0.673	0.786	0.845	0.975	0.995
	40	0.673	0.786	0.828	0.948	0.964
	50	0.659	0.765	0.799	0.932	0.941
W(2.75, 2.75, 3.75)	20	0.868	0.939	0.954	0.984	0.999
	40	0.857	0.924	0.947	0.961	0.972
	50	0.842	0.904	0.918	0.949	0.956

Table 3 Probability distribution of $|\hat{k} - k| \leq \delta$ and the sample size is $n = 150$ at various change point locations k

	$k \setminus \delta \rightarrow$	$P(\vert\hat{k} - k\vert \leq \delta)$				
		1	2	3	4	5
W(1.25, 1.25, 2.25)	25	0.315	0.431	0.512	0.767	0.813
	50	0.273	0.380	0.467	0.738	0.887
	75	0.276	0.388	0.470	0.728	0.861
W(1.5, 1.5, 2.5)	25	0.529	0.691	0.768	0.815	0.958
	50	0.547	0.659	0.735	0.881	0.912
	75	0.503	0.629	0.694	0.834	0.957
W(1.75, 1.75, 2.75)	25	0.680	0.802	0.849	0.874	0.987
	50	0.665	0.748	0.792	0.815	0.921
	75	0.603	0.697	0.734	0.852	0.965
W(2.75, 2.75, 3.75)	25	0.927	0.958	0.975	0.987	0.998
	50	0.913	0.945	0.972	0.988	0.994
	75	0.905	0.922	0.957	0.962	0.978

vations. The graph of the data is sketched in Fig. 5. Figure 6 shows the autocorrelation function (ACF) of the data.

To ensure the independence of the data, we use the Portmanteau test statistic to check the independence and normality of the dataset which is given as below.

$$Q_k = n \sum_{i=1}^{k} r_i^2,$$

Table 4 Power comparison of T_n^* and S_n, $n = 50$ and $k = 10, 15, 25$

	k	W(1.25, 1.25, 2.25)	W(1.5, 1.5, 2.5)	W(1.75, 1.75, 2.75)
T_n^*	10	0.452	0.531	0.691
	15	0.499	0.686	0.801
	25	0.462	0.748	0.907
S_n	10	0.490	0.568	0.716
	15	0.542	0.724	0.821
	25	0.523	0.786	0.927

Table 5 Power comparison of T_n^* and S_n, $n = 100$ and $k = 20, 40, 50$

	k	W(1.25, 1.25, 2.25)	W(1.5, 1.5, 2.5)	W(1.75, 1.75, 2.75)
T_n^*	20	0.657	0.736	0.964
	40	0.752	0.927	0.998
	50	0.714	0.932	0.999
S_n	20	0.710	0.768	0.970
	40	0.806	0.939	0.998
	50	0.767	0.953	0.999

Table 6 Power comparison of T_n^* and S_n, $n = 150$ and $k = 25, 50, 75$

	k	W(1.25, 1.25, 2.25)	W(1.5, 1.5, 2.5)	W(1.75, 1.75, 2.75)
T_n^*	25	0.765	0.796	0.992
	50	0.895	0.970	1.000
	75	0.878	0.990	1.000
S_n	25	0.811	0.829	0.994
	50	0.916	0.979	1.000
	75	0.902	0.991	1.000

where r_i are the autocorrelation coefficient at lag i, and k is the lag up to which the autocorrlation coefficient function of the data. Under the null hypothesis $Q_k \sim \chi_k^2$. Using the Portmanteau test, we get $Q_{20} = 100\sum_{i=1}^{20} r_i^2 = 16.025735 < \chi_{0.95}^2(20) = 31.41043$. Thus, we fail to reject the null hypothesis which leads to the independence of the data. Since $MIC(n) = MIC(100) = 1156.3874 > \min_{1\le k<n} MIC(k) = MIC(2) = 1143.4111$, the estimated change location is $\hat{k} = 2$. The corresponding test statistic $S_n = 27.6097$ with the critical value $\chi_{0.05}^2(3) = 7.815$ and the p-value 4.3858×10^{-6}. It confirms the change in the data. With the conventional SIC, we have $\min_{1\le k<n} SIC(k) = SIC(2)$. Therefore, two methods provide the same conclusion. For potential multiple changes in the data, the binary segmentation method by Vostrikova (1981) is applied. Such a method decomposes the detecting procedure into several steps by assuming at most one change at each step. This process is repeated

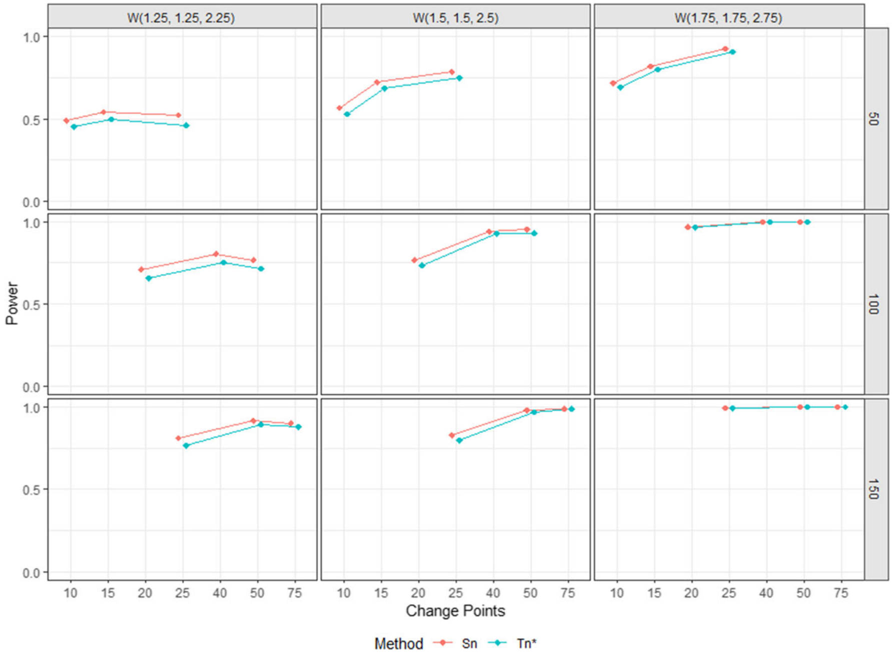


Fig. 4 Power comparison of T_n^* and S_n with various change point locations and different sample sizes $n = 50, 100, 150$

Table 7 Comparison of the coverage probabilities, $k = 10, 15, 25$ and $n = 50$

	α	$k = 10$		$k = 15$		$k = 25$	
		MIC	loglik	MIC	loglik	MIC	loglik
W(1.25, 1.25, 2.25)	0.50	0.61	0.61	0.68	0.67	0.66	0.67
	0.90	0.80	0.78	0.82	0.80	0.81	0.80
	0.95	0.83	0.81	0.86	0.84	0.85	0.84
	0.99	0.89	0.87	0.90	0.88	0.91	0.90
W(1.5, 1.5, 2.5)	0.50	0.69	0.68	0.73	0.72	0.72	0.72
	0.90	0.83	0.82	0.85	0.84	0.88	0.88
	0.95	0.88	0.87	0.89	0.88	0.91	0.91
	0.99	0.92	0.91	0.94	0.93	0.96	0.95
W(1.75, 1.75, 2.75)	0.50	0.70	0.70	0.72	0.72	0.75	0.76
	0.90	0.82	0.82	0.86	0.86	0.88	0.89
	0.95	0.87	0.87	0.90	0.90	0.91	0.92
	0.99	0.92	0.92	0.96	0.95	0.96	0.96

Table 8 The comparisons of the average sizes of the confidence sets, $k = 10, 15, 25$ and $n = 50$

	α	$k = 10$		$k = 15$		$k = 25$	
		MIC	loglik	MIC	loglik	MIC	loglik
W(1.25, 1.25, 2.25)	0.50	17.39	17.66	18.13	18.48	18.41	18.96
	0.90	25.26	25.28	25.57	25.75	25.86	26.22
	0.95	28.25	28.18	28.36	28.40	28.67	28.93
	0.99	33.51	33.26	33.84	33.64	34.22	34.11
W(1.5, 1.5, 2.5)	0.50	13.37	13.96	13.35	13.94	13.69	14.02
	0.90	18.51	19.03	18.36	19.01	18.72	18.99
	0.95	20.82	21.33	20.57	21.22	21.03	21.25
	0.99	26.12	26.48	25.79	26.36	26.21	26.30
W(1.75, 1.75, 2.75)	0.50	10.53	10.75	10.26	10.47	10.48	10.66
	0.90	14.18	14.42	13.41	13.63	13.83	14.00
	0.95	15.92	16.19	14.92	15.15	15.38	15.54
	0.99	20.35	20.61	18.87	19.08	19.21	19.30

Table 9 Comparison of the coverage probabilities, $k = 20, 40, 50$ and $n = 100$

	α	$k = 20$		$k = 40$		$k = 50$	
		MIC	loglik	MIC	loglik	MIC	loglik
W(1.25, 1.25, 2.25)	0.50	0.73	0.72	0.76	0.75	0.75	0.76
	0.90	0.88	0.87	0.89	0.88	0.90	0.90
	0.95	0.91	0.90	0.93	0.92	0.93	0.90
	0.99	0.96	0.94	0.97	0.96	0.96	0.96
W(1.5, 1.5, 2.5)	0.50	0.75	0.73	0.76	0.74	0.78	0.78
	0.90	0.90	0.89	0.91	0.90	0.91	0.91
	0.95	0.94	0.93	0.95	0.95	0.95	0.94
	0.99	0.97	0.96	0.98	0.98	0.98	0.98
W(1.75, 1.75, 2.75)	0.50	0.79	0.78	0.80	0.80	0.78	0.78
	0.90	0.91	0.91	0.90	0.90	0.90	0.90
	0.95	0.94	0.94	0.93	0.93	0.95	0.95
	0.99	0.97	0.97	0.97	0.97	0.98	0.98

until no more change point is detected. With the binary segmentation method, the multiple change locations in the data are {2, 6, 17, 39, 94}. The confidence sets for the change point estimates $\hat{k} = 2, 6, 17, 39, 94$ are {2}, {5, 6, 7, 8}, {17, 18, 39, 40}, {34, 35, 37, 38, 39, 47, 50}, {3, 4, 44, 47, 71, 72, 73, 74, 75, 94, 95} respectively. Figures 7, 8 and 9 show the confidence curves for estimated change locations. The 95% confidence set is marked by the horizontal red dashed line. Figure 10 shows all the estimated changes in the data set.

Table 10 Comparison of the mean sizes of the confidence sets, $k = 20, 40, 50$ and $n = 100$

	α	$k = 20$		$k = 40$		$k = 50$	
		MIC	loglik	MIC	loglik	MIC	loglik
W(1.25, 1.25, 2.25)	0.50	27.64	30.09	26.84	29.05	27.08	28.61
	0.90	38.45	40.78	36.57	38.85	37.34	38.92
	0.95	43.09	45.24	41.01	43.22	41.71	43.27
	0.99	54.08	55.74	50.91	52.79	51.89	53.26
W(1.5, 1.5, 2.5)	0.50	18.66	19.33	17.61	18.03	18.08	18.19
	0.90	24.34	25.10	22.24	22.68	23.03	23.10
	0.95	26.95	27.73	24.42	24.85	25.20	25.25
	0.99	33.27	34.06	29.71	30.15	30.71	30.69
W(1.75, 1.75, 2.75)	0.50	12.56	12.66	13.38	13.47	14.27	14.26
	0.90	15.84	15.97	16.03	16.15	16.98	16.98
	0.95	17.43	17.57	17.28	17.40	18.26	18.26
	0.99	21.32	21.48	20.50	20.63	21.34	21.33

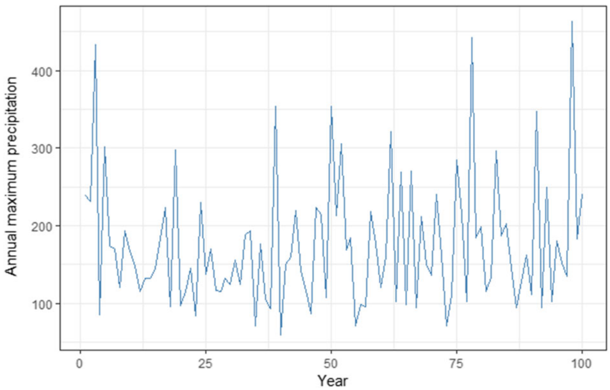
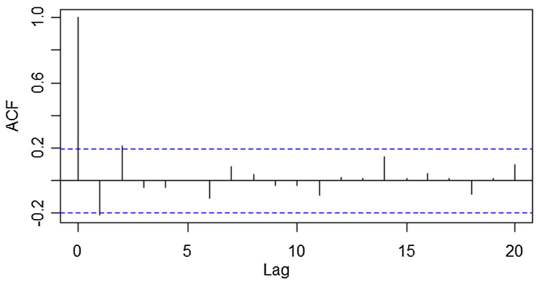


Fig. 5 The annual maximum rainfall data at one rain gauge in Fort Collins, Colorado

Fig. 6 The auto-correlation plot for annual maximum rainfall data at one rain gauge in Fort Collins, Colorado



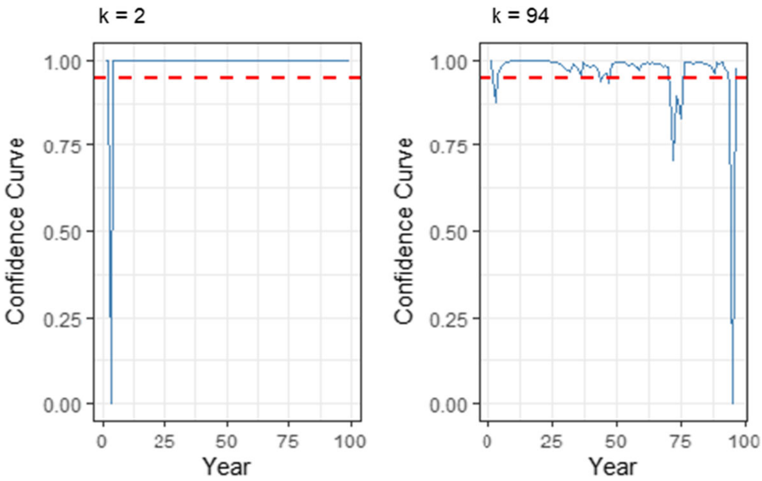


Fig. 7 Left: Confidence curve for change point at $\hat{k} = 2$, Right: Confidence curve for the second subset above ($2 < k \leq 100$), $\hat{k} = 94$

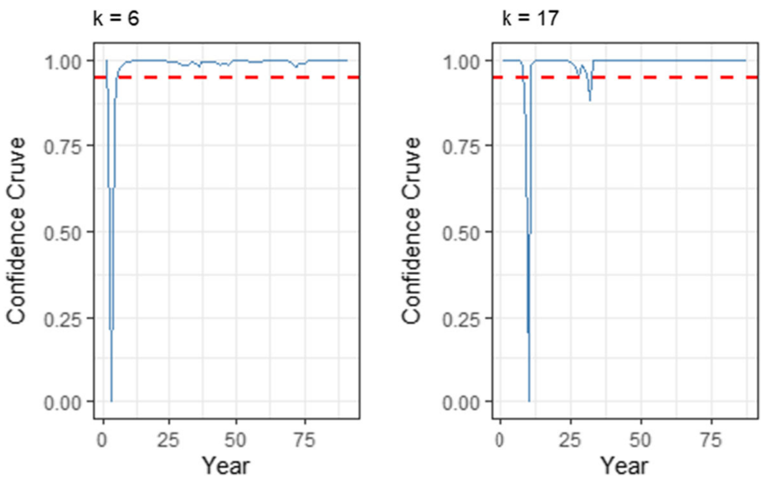


Fig. 8 Left: Confidence curve for the third subset below ($2 < k \leq 94$), the $\hat{k} = 6$, Right: Confidence curve for the fourth subset ($7 \leq k \leq 94$), $\hat{k} = 17$

7 Conclusion

In this paper, we propose a change point detection method for a three-parameter Weibull distribution simultaneous based on the modified information criterion (MIC). The simultaneous changes in the parameters are considered. The asymptotic properties for the associated test statistic have been established. Moreover, we propose a modified approach to construct the confidence sets for change locations at a given significance level through confidence distributions. In terms of powers, coverage probabilities and average sizes of confidence sets at various scenarios, simulations are conducted to

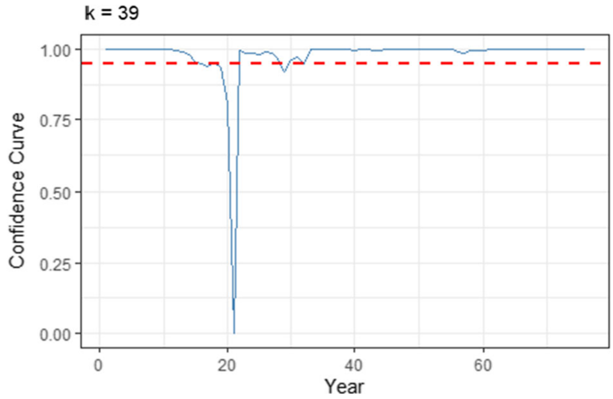


Fig. 9 Confidence curve for the fifth subset ($17 < k \leq 94$), $\hat{k} = 39$

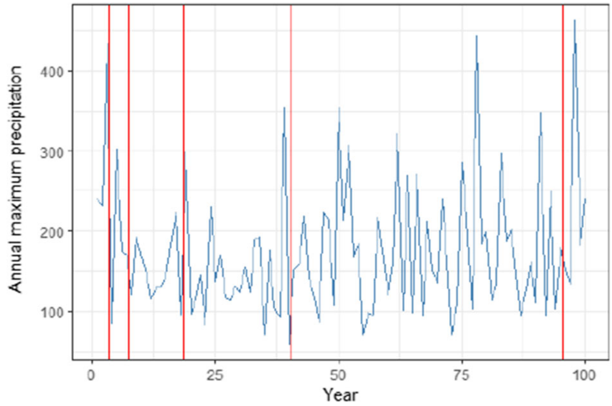


Fig. 10 The annual maximum rainfall data at one rain gauge in Fort Collins, Colorado with change point locations

compare the proposed method with the one based on the conventional Schwarz information criterion (SIC) and the method provided in Cunen et al. (2018). Simulations indicate that our method is competitive to other existing methods and even better when the change happens near the beginning or the end of the data. Along with the binary segmentation method, the proposed method is applied to detect multiple change points and construct corresponding confidence sets for the annual maximum rainfall data.

Acknowledgements The authors would like to thank three anonymous reviewers and the Associate Editor for their constructive comments and suggestions, which helped to improve this manuscript significantly.

References

Alghamdi A, Ning W, Gupta AK (2018) An information approach for the change point problem of the rayleigh lomax distribution. Int J Intell Technol Appl Stat 11(4):233–254

Arellano-Valle RB, Castro L, Loschi RH (2013) Change point detection in the skew-normal model parameters. Commun Stat Theory Methods 42:603–618

Chen J, Gupta AK (1995) Likelihood procedure for testing change points hypothesis for multivariate gaussian model. Random Oper Stoch Equ 3:235–244

Chen J, Gupta AK (1997) Testing and locating variance change points with application to stock prices. J Am Stat Assoc 92:739–747

Chen J, Gupta AK (2012) Parametric statistical change point analysis with applications to genetics, medicine, and finance, 2nd edn. Birkhauser, Boston

Chen J, Gupta AK, Pan J (2006) Information criterion and change point problem for regular models. Indian J Stat 68(2):252–282

Csörgő M, Horváth L (1997) Limit theorems in change-point analysis. Wiley, New York

Cunen C, Hermansen G, Hjort NL (2018) Confidence distributions for change-points and regime shifts. J Stat Plann Inference 195:14–34

Gilleland E, Katz RW (2016) extremes 2.0: an extreme value analysis package in r. J Stat Softw 72(8):1–39

Gurevich G, Vexler A (2005) Change point problems in the model of logistic regression. J Stat Plann Inference 131:313–331

Jandhyala VK, Fotopoulos SB, Evaggelopoulos N (1999) Change point methods for weibull models with applications to detection of trends in extreme temperatures. Environmetrics 10:547–564

Juruskova D (2007) Maximum log-likelihood ratio test for a change in three parameter weibull distribution. J Stat Plann Inference 137:1805–1815

Ngunkeng G, Ning W (2014) Information approach for the change-point detection in the skew normal distribution and its applications. J Seq Anal 33(4):475–490

Ning W, Gupta AK (2009) Change point analysis for generalized lambda distributions. Commun Stat Simul Comput 38:1789–1802

Page ES (1954) Continue inspection schemes. Biometrika 41:100–135

Page ES (1955) A test for a chance in a parameter occurring at an unknown point. Biometrika 42:523–527

Ramanayake A, Gupta AK (2010) Testing for a change point in a sequence of exponential random variables with repeated values. J Stat Comput Simul 80:191–199

Ratnasingam S, Ning W (2020) Confidence distributions for skew normal change-point model based on modified information criterion. J Stat Theory Pract 14(3):1–21

Schwarz G (1978) Estimating the dimension of a model. Ann Stat 6:461–464

Sen AK, Srivastava MS (1975a) On tests for detecting change in mean. Ann Stat 3:98–108

Sen AK, Srivastava MS (1975b) Some one-sided tests on change in level. Technometrics 17:61–64

Srivastava MS, Worsley KJ (1986) Likelihood ratio tests for a change in the multivariate mean. J Am Stat Assoc 81:199–204

Vostrikova LJ (1981) Detecting "disorder" in multidimensional random processes. Soviet Math Dokl 24:55–59

Worsley K (1979) On the likelihood ratio test for a shift in location of normal populations. J Am Stat Assoc 74:365–367

Wu Y (2008) Simultaneous change point analysis and variable selection in a regression problem. J Multiv Anal 99:2154–2171

Xie M, Singh K (2013) Condence distribution, the frequentist distribution estimator of a parameter: a review. Int Stat Rev 81(1):2–39

Zhang N, Siegmund D (2007) A modified bayes information criterion with applications to the analysis of comparative genomic hybridization data. Biometrics 63:22–32

Dr. Suthakaran Ratnasingam currently is Assistant Professor in the Department of Mathematics at California State University, San Bernardino, California, USA.

Dr. Wei Ning currently is Professor of Statistics in the Department of Mathematics and Statistics at Bowling Green State University, Ohio, USA.