Confidence curves for change points in hydrometeorological time series

Zhou, Changrang; van Nooijen, Ronald; Kolechkina, Alla; van de Giesen, Nick

DOI
10.1016/j.jhydrol.2020.125503

Publication date
2020

Document Version
Final published version

Published in
Journal of Hydrology

Citation (APA)
Zhou, C., van Nooijen, R., Kolechkina, A., & van de Giesen, N. (2020). Confidence curves for change points in hydrometeorological time series. *Journal of Hydrology*, *590*, 1-19. Article 125503. https://doi.org/10.1016/j.jhydrol.2020.125503

Journal of Hydrology 590 (2020) 125503

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

Confidence curves for change points in hydrometeorological time series

Changrang Zhou[a,*], Ronald van Nooijen[a], Alla Kolechkina[b], Nick van de Giesen[a]

[a] Water Management Department, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, Netherlands
[b] Delft Center for Systems and Control, Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, Delft, Netherlands

ARTICLE INFO

This manuscript was handled by A. Bardossy, Editor-in-Chief, with the assistance of Anne-Catherine Favre, Associate Editor

Keywords:
Approximate empirical likelihood ratio
Parametric likelihood ratio
Confidence curves
Confidence sets
Similarity index
Change point detection

ABSTRACT

In this paper, a method based on Approximate Empirical likelihood ratio and a Deviance function combined with bootstrapping (AED-BP) is proposed to construct a confidence curve for the location of a change point. The method is compared with a method based on parametric Profile Likelihood and a Deviance function combined with Monte Carlo simulation (PLD-MC). A confidence curve provides a representation of the uncertainty in the outcome of the change point analysis. To evaluate the practical usability of confidence curves constructed by AED-BP, its properties were examined and its performance was compared to that of PLD-MC. The methods were applied to both synthetic and real data. Synthetic data were generated from three parametric distributions: Fréchet with a constant shape parameter, log-normal, and gamma distributions. The real data are the hydrometeorological data analysed in other studies. The change points found in the original publications are used as a reference in this present paper. The results show that AED-BP has a performance that is similar to PLD-MC, but has an advantage in that it is not necessary to select a distribution family for the data. The AED-BP results on the Annual Maximum Runoff series for the stations Yichang and Hankou along the Yangtze river are among the first that show a possible effect of the presence of the Three Gorges dam.

1. Introduction

While it is clear that climate change affects the hydrological cycle (Donat et al., 2017) and that there is an increased risk of extremes in precipitation (Lehmann et al., 2015), discharge (Hirabayashi et al., 2013) and temperature, the effects on a regional scale may vary considerably (Tamaddun et al., 2016). The analysis of time series of precipitation, temperature, discharge and other variables is an important tool in the search for and examination of such changes. However, in order to be effective, the analysis must allow for non-stationarity. Roughly speaking, there are two types of non-stationarity in hydrological processes to be considered: gradual change and abrupt change. The main sources of these changes are human interventions and climate variability (Haddeland et al., 2013). Detecting change points contributes to detecting changes in the water cycle due to human and natural causes during the Anthropocene, a component of some important open questions in hydrology (Blöschl et al., 2019). The need for more hydrological data that is mentioned in McMillan et al. (2016) makes it more important than ever to determine whether or not known changes have impacted system response. With a good understanding of the size of the impact, better use can be made of long hydrological time series that otherwise would need to be treated as two shorter series. To do so, it is necessary to establish whether or not a known change has caused detectable impact, for instance, in the form of a change point.

The examination of changes in catchment behaviour is not a purely academic exercise: future catchment behaviour is a major factor in all decisions on future water management. If one wishes to analyse non-stationarity, then a first essential step is finding the abrupt changes, because any abrupt change will interfere with the search for gradual changes and other statistical properties of the series. Therefore, this paper focusses on the detection of abrupt changes in the hydrometeorological processes through the analysis of time series.

The concept of an abrupt change is formalized as follows: a time series is said to have a change point when the statistical characteristics of the series before and after the change point show a significant difference. Finding change points has attracted attention from many fields, for instance, in oceanography (Killick et al., 2010), economics, finance (Chen and Gupta, 2012), biology (Brodsky and Darkhovsky, 1993) and meteorology (Beaulieu et al., 2012; Gocic and Trajkovic, 2013). In hydrology, change point detection plays an indispensable role in homogeneity tests for hydrological observations (Kundzewicz and Robson, 2004).

A number of methods have been developed to find change points; some require a parametric description of the probability distribution of

* Corresponding author.
E-mail address: C.Zhou-1@tudelft.nl (C. Zhou).

https://doi.org/10.1016/j.jhydrol.2020.125503
Received 2 March 2020; Received in revised form 30 July 2020; Accepted 31 August 2020
Available online 11 September 2020

the data points in the time series (Chen and Gupta, 2012), others do not (Pettitt, 1979; Lee et al., 2003; Hawkins and Zamba, 2005; Gurevich and Vexler, 2010). Traditional change point detection methods are often designed to accept or reject the null hypothesis at a given significance level. If it is rejected, then a point estimate of the location is obtained more or less as a by-product. This approach does not offer much room for the communication of degrees of uncertainty. Moreover, its use of the traditional *p*-value based approach is a potential weakness (Wasserstein et al., 2019).

In some cases, change point detection may deliver unexpected results, for instance, when events have taken place that lead hydrologists to expect a change, but for the given *p*-value the null hypothesis is not rejected. To be more specific, one might find that the known information is that a dam or reservoir was constructed upstream of a gauging station at a given year, but no change point is detected in the time series beyond that point. With just a hypothesis test no further insight is available. This is particularly problematical, because, for some of the tests used in hydrology, results may change when different combinations of starting and ending year are used (Zhou et al., 2019). It is therefore important to examine new methods for change point detection that provide more information on the uncertainty of the results.

In the current study, a new method is developed that represents the uncertainty about the location of a change point by providing confidence sets at all confidence levels. A confidence set is a generalization of a confidence interval. Our method was inspired by the work on confidence curves in connection with change point detection in Cunen et al. (2018). Their 'method B', which constructs a confidence curve for the location of a change point by using a parametric profile likelihood function to construct a deviance function, shows considerable promise. However, it presupposes that it is known to which family of distributions the data points belong; this knowledge is used both in the formulation of the deviance function and in a Monte Carlo (MC) procedure that draws from that family to approximate the distribution of the deviance function. In hydrology, it is not always clear which family should be chosen. In addition, the method also involves optimizing a fairly large number of profile likelihoods. For some distribution families, this may be costly. The method presented in this paper avoids these potential drawbacks by using an empirical likelihood ratio instead of a parametric profile likelihood function and bootstrapping (BP) samples from the original sample.

There are alternative approaches that can be used to represent the uncertainty in the change point location. One is the use of confidence intervals instead of point estimates for a given level of significance, but this still limits the available information to that for one level of significance. Another approach would be to use Bayesian techniques. Bayesian techniques are particularly attractive in hydrology (Coles and Tawn, 1996; Renard et al., 2010) because of the non-repeatability of hydrological observations. An example of a Bayesian change point analysis method can be found in Perreault et al. (2000). However, in addition to the need to find a proper distribution family for hydrological records, Bayesian approaches also need to find a suitable prior.

The remainder of this paper is organized as follows: first two different methodologies for confidence curve construction are presented, the parametric 'method B' from Cunen et al. (2018) and the non-parametric method proposed in this study, and indicators are defined that can be used to evaluate and compare the performance of the curves. Then, the results of the application of the methods to synthetic data are analysed. Next, the non-parametric method is applied to several hydrometeorological time series and the outcomes are compared to results found in the literature. Finally, we discuss the results and present our conclusions. Mathematical details can be found in the Appendices (A–D).

2. Methodology

This study introduces a new method to represent and analyse uncertainties in change point detection. It should therefore present the basic principle behind the method, test its performance, and examine its application to real data. To ensure a clear exposition, a simple formulation of the change point detection problem will be used. For details of the notation and definitions see Appendix A.

The formulation of the method and the numerical experiments will be limited to the case where there is at most one change (AMOC). All time series will be modelled as a vector Y of n independent continuous random variables $Y_1, Y_2, \ldots, Y_n$. The *null hypothesis* H_0 will be that the Y_i are independent identically distributed (i.i.d.) random variables. The alternative hypothesis H_1 will be that there is an index τ that splits the series into two sub-series: a left sub-series where $Y_1, Y_2, \ldots, Y_\tau$ are i.i.d. random variables and a right sub-series where $Y_{\tau+1}, Y_{\tau+2}, \ldots, Y_n$ are i.i.d. random variables, but the distributions of the left sub-series and the right sub-series of the series are different.

Furthermore, it will be assumed that the corresponding distributions are from the same family of distributions, and that the individual members can be fully specified by a vector θ of parameters that may change and a vector ζ of parameters that stay the same. The *probability density function* (pdf) of a member of the distribution family will be referred to as $f(\cdot;\theta, \zeta)$, and the *cumulative distribution function* (cdf) will be denoted by $F(\cdot;\theta, \zeta)$. We will use θ_L for the parameter vector corresponding to the left sub-series and θ_R for the parameter vector corresponding to the right sub-series. In the remainder of the paper, Y will represent the random vector defined here; y will represent a realization of Y, and y_{obs} will represent the actual observed time series.

Both the parametric and the non-parametric change point methods need to sample from the sub-series to the left and to the right of the change point; for short sub-series this is likely to cause problems. Moreover, our non-parametric method which uses an approximate empirical likelihood is related to the empirical likelihood method discussed in Zou et al. (2007) where it is stated that the empirical likelihood may not exist for short sub-series, and it is recommended to consider only a subset of the possible change points. Finally, the approximation we use for the empirical log-likelihood does not hold everywhere, but it does hold for change points at $n_{tr}, n_{tr}+1, \ldots, n-n_{tr}$, when

$$n_{tr} = \lfloor 2\log n \rfloor \tag{1}$$

where $\lfloor\cdot\rfloor$ denotes rounding down towards the nearest integer.

2.1. Confidence curves based on the profile log-likelihood

Cunen et al. (2018) presented a method to construct confidence curves (Definition 3 of Appendix A) based on the log-likelihood function ℓ. In the case of change point detection ℓ is given by

$$\ell(\tau, \theta_L, \theta_R, \zeta; y) = \sum_{i=1}^{\tau} \log f(y_i;\theta_L, \zeta) + \sum_{i=\tau+1}^{n} \log f(y_i;\theta_R, \zeta) \tag{2}$$

As a first step in the derivation of the method, they introduce a profile log-likelihood. In general, a profile log-likelihood is used when only part of the parameter vector is of interest. For instance, a vector $\nu = (\lambda, \eta)$ where only λ is of interest, η is a *nuisance parameter*. In such a case, one can take the supremum of the log-likelihood over all η, which is then called the *profile log-likelihood* for λ. According to Murphy and van der Vaart (2000), the 'profile likelihood may be used to a considerable extent as a full likelihood' for the parameter of interest. If an estimate for τ in (2) is needed, then τ is the parameter of interest and $\theta_L, \theta_R, \zeta$ are nuisance parameters. Therefore, they define the profile log-likelihood by

$$\ell_{prof}(\tau;y) = \sup_{\theta_L,\theta_R,\zeta} \ell(\tau, \theta_L, \theta_R, \zeta;y) \tag{3}$$

The notation $\hat{\theta}_L(\tau;y)$, $\hat{\theta}_R(\tau;y)$ and $\hat{\zeta}(\tau;y)$ is used to denote a combination of values of θ_L, θ_R and ζ where ℓ_{prof} attains its global maximum, so

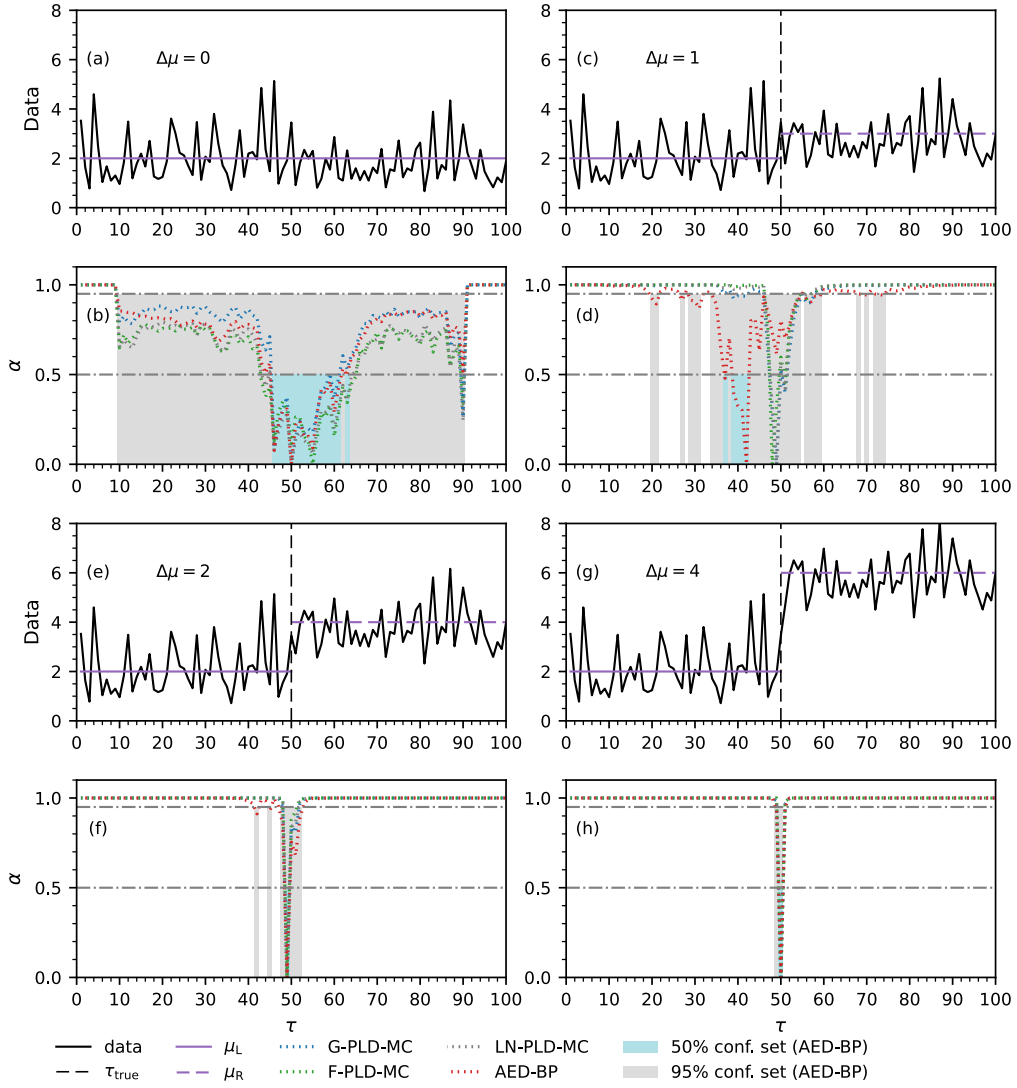

Fig. 1. Confidence curves for change point location in synthetic data from a log-normal distribution.

$$\ell_{\text{prof}}(\tau;y) = \ell(\tau, \hat{\theta}_{\text{L}}(\tau;y), \hat{\theta}_{\text{R}}(\tau;y), \hat{\zeta}(\tau;y);y) \tag{4}$$

The value of τ for which ℓ_{prof} is maximal will be denoted by $\hat{\tau}(y)$. The values $\hat{\theta}_{\text{L}}(\hat{\tau}(y);y)$, $\hat{\theta}_{\text{R}}(\hat{\tau}(y);y)$, and $\hat{\zeta}(\hat{\tau}(y);y)$ will be used as estimators for the parameters θ_{L}, θ_{R}, and ζ respectively.

Next, the *deviance function* D is introduced

$$D(\tau, y) = 2(\ell_{\text{prof}}(\hat{\tau}(y);y) - \ell_{\text{prof}}(\tau;y)) \tag{5}$$

This is then used to define random variables $D(\tau, Y)$ with $\tau = 1, 2, \ldots, n-1$. For a given τ its distribution is estimated by

$$\forall\, r \in \mathbb{R}\colon K_\tau(r) = \Pr(D(\tau, Y) < r | \tau, \theta_{\text{L}} = \hat{\theta}_{\text{L}}(\hat{\tau}(y);y), \theta_{\text{R}} = \hat{\theta}_{\text{R}}(\hat{\tau}(y);y), \zeta = \hat{\zeta}(\hat{\tau}(y);y)) \tag{6}$$

In the case of a discrete parameter τ, no exact or approximate expression is known for the distribution K_τ, so it needs to be approximated by simulation. Note that by definition $D(\hat{\tau}(y), y) = 0$. Now for each sample there will be at least one k, namely $k = \hat{\tau}(y)$ for which $D(\tau, y) = 0$. As there are only a finite number of values that τ can take, this implies that $\Pr(D(\tau, Y) = 0) = 0$ cannot hold for all τ. Therefore, there is at least one τ' with $\Pr(D(\tau', Y) = 0) > 0$, and so the distribution of $D(\tau', Y)$ has a positive point probability at 0. This implies that $D(\tau', Y)$ is never uniformly distributed, but it is part of the definition of a confidence curve that $\text{cc}(\tau_{\text{true}};Y)$ is uniformly distributed on $[0, 1]$ when τ_{true} is the true value of τ (Appendix A). Nevertheless, as in Cunen et al. (2018), D will be used to define the function that will be referred to as a confidence curve

$$\text{cc}(\tau;y_{\text{obs}}) = K_\tau(D(\tau, y_{\text{obs}}))$$

where y_{obs} is the observed sample.

The simulations needed to obtain K_τ are performed as follows:

1. Obtain estimates of the distribution parameters τ, θ_{L}, θ_{R}, and ζ by determining $\hat{\tau}(y_{\text{obs}})$, $\hat{\theta}_{\text{L}}(\hat{\tau}(y_{\text{obs}});y_{\text{obs}})$, $\hat{\theta}_{\text{R}}(\hat{\tau}(y_{\text{obs}});y_{\text{obs}})$, and $\hat{\zeta}(\hat{\tau}(y_{\text{obs}});y_{\text{obs}})$ respectively.
2. For each $k \in \{n_{\text{tr}}, n_{\text{tr}}+1, \ldots, n-n_{\text{tr}}\}$ and $j = 1, 2, \ldots, N$, generate a sample $y^{(j,k)}$ where the $y_i^{(j,k)}$ with $i = 1, 2, \ldots, k$ are distributed with $\theta = \hat{\theta}_{\text{L}}(\hat{\tau}(y_{\text{obs}});y_{\text{obs}})$, $\zeta = \hat{\zeta}(\hat{\tau}(y_{\text{obs}});y_{\text{obs}})$, and the $y_i^{(j,k)}$ with $i = k+1, k+2, \ldots, n$ are distributed with $\theta = \hat{\theta}_{\text{R}}(\hat{\tau}(y_{\text{obs}});y_{\text{obs}})$, $\zeta = \hat{\zeta}(\hat{\tau}(y_{\text{obs}});y_{\text{obs}})$; n_{tr} is used to avoid calculation of profile log-likelihoods based on a handful of points.
3. Approximate the curve $\text{cc}(\tau;y_{\text{obs}}) = K_\tau(D(\tau, y_{\text{obs}}))$ by

$$K_{\tau,N}(D(\tau, y_{\text{obs}})) = \frac{1}{N}\sum_{j=1}^{N} \left[\!\left[D(\tau, y^{(j,\tau)}) < D(\tau, y_{\text{obs}}) \right]\!\right] \tag{7}$$

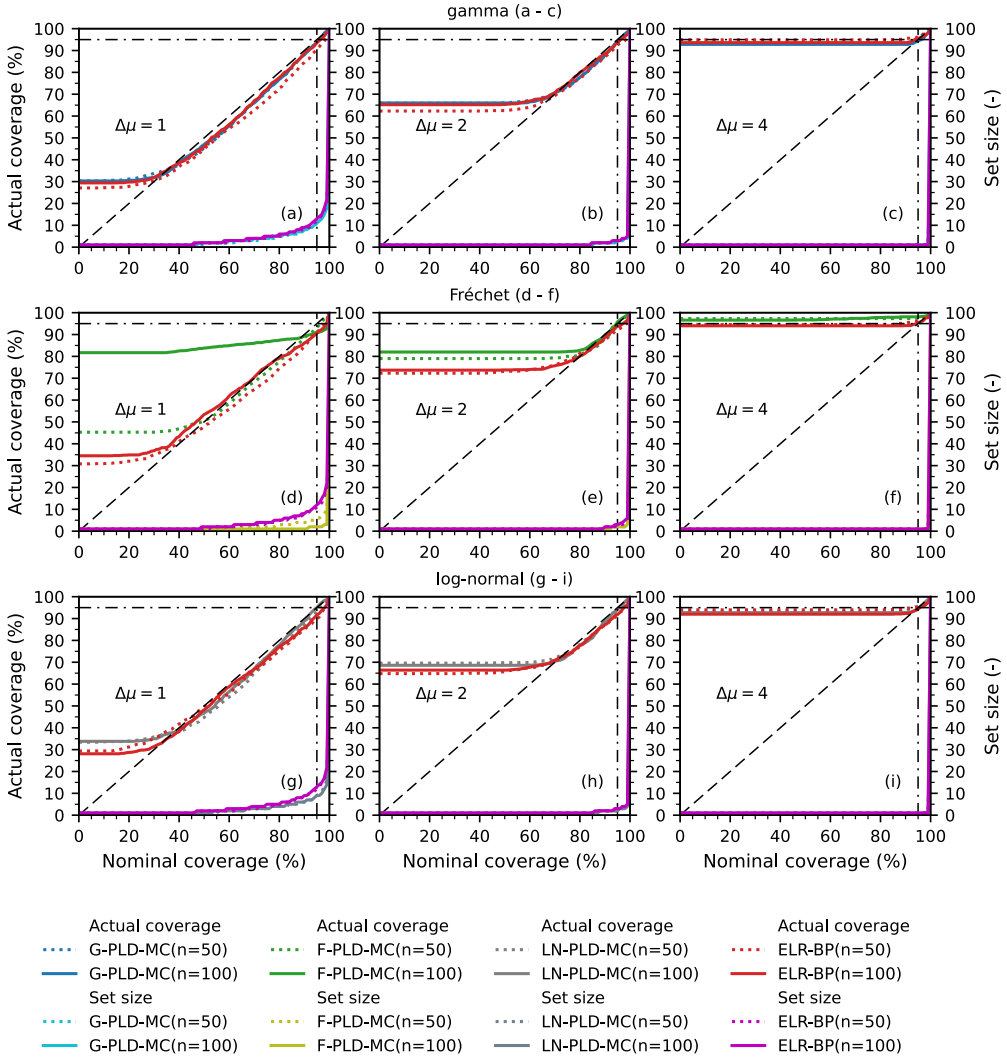

Fig. 2. Actual coverage probabilities and confidence set size as a function of nominal coverage probabilities for synthetic data.

This method of confidence curve construction uses a parametric Profile Likelihood based Deviance function with Monte Carlo simulation (PLD-MC).

2.2. Confidence curves based on the approximate empirical likelihood ratio

To show the relation between PLD-MC and the method proposed in this study, it is necessary to take several intermediate steps. These steps are presented in detail in Appendix B. The end result is that the role of the profile log-likelihood in the deviance function used in the construction of the confidence curve is taken over by an approximation ℓ_{apn} of the empirical log-likelihood given by

$$\ell_{\text{apn}}(\tau;y) = \frac{\tau(n-\tau)}{n} \frac{\left(\frac{1}{\tau}\sum_{i=1}^{\tau} y_i - \frac{1}{n-\tau}\sum_{i=\tau+1}^{n} y_i\right)^2}{\frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \frac{1}{n}\sum_{j=1}^{n} y_j\right)^2} \tag{8}$$

To define the corresponding deviation function D_{apn}, we need to introduce $\hat{\tau}_{\text{apn}}(y)$, the value of τ for which $\ell_{\text{apn}}(\tau, y)$ attains its maximum. Now D_{apn} is

$$D_{\text{apn}}(\tau;y) = 2(\ell_{\text{apn}}(\hat{\tau}_{\text{apn}}(y);y) - \ell_{\text{apn}}(\tau;y)) \tag{9}$$

To determine the distribution $K_{\text{apn},\tau}(r)$ of $D_{\text{apn}}(\tau;Y)$, formally given by

$$K_{\text{apn},\tau}(r) = \Pr(D_{\text{apn}}(\tau;Y) < r) \tag{10}$$

we use the following procedure:

1. Determine $\tau_0 = \hat{\tau}_{\text{apn}}(y_{\text{obs}})$, and split y_{obs} into a left part and a right part at τ_0.
2. For each candidate position $\tau \in \{n_{\text{tr}}, n_{\text{tr}}+1, \ldots, n-n_{\text{tr}}\}$, use bootstrapping to resample y_{obs} and get N new samples $y_{\text{res}}^{(j)}$ $(j = 1, 2, \ldots, N)$. For each j, $y_{\text{res}}^{(j)}$ is composed of a sequence of τ values drawn from the left part of y_{obs} followed by a sequence of $(n-\tau)$ values drawn from the right part of y_{obs} (Hall and Martin, 1988).
3. Approximate the curve $\text{cc}(\tau;y_{\text{obs}}) = K_{\text{apn},\tau}(D_{\text{apn}}(\tau, y_{\text{obs}}))$ by

$$\frac{1}{N}\sum_{j=1}^{N}\left[\!\left[D_{\text{apn}}(\tau, y_{\text{res}}^{(j)}) < D_{\text{apn}}(\tau, y_{\text{obs}}) \right]\!\right] \tag{11}$$

Here n_{tr} is used to avoid calculation of approximate empirical likelihoods based on a handful of points.

Thus the newly developed method is based on the Approximate Empirical likelihood ratio that is used in a Deviance function combined with Bootstrapping to calculate the confidence curves (AED-BP).

Table 1
Actual coverage probability for given confidence levels (conservative coverage is marked by a grey background).

α			0.90		0.95		0.99	
Distribution	$\Delta\mu$	n	PLD-MC	AED-BP	PLD-MC	AED-BP	PLD-MC	AED-BP
gamma	1	50	0.880	0.845	0.935	0.907	0.992	0.966
		100	0.880	0.887	0.935	0.937	0.991	0.982
	2	50	0.877	0.871	0.933	0.925	0.991	0.958
		100	0.877	0.886	0.937	0.937	0.985	0.975
	4	50	0.937	0.956	0.953	0.964	0.993	0.976
		100	0.928	0.936	0.943	0.955	0.982	0.980
Fréchet	1	50	0.884	0.849	0.931	0.912	0.980	0.970
		100	0.893	0.861	0.911	0.906	0.927	0.954
	2	50	0.882	0.876	0.942	0.923	0.991	0.951
		100	0.898	0.888	0.958	0.937	0.990	0.965
	4	50	0.973	0.954	0.975	0.962	0.990	0.973
		100	0.981	0.941	0.981	0.955	0.990	0.977
log-normal	1	50	0.873	0.850	0.933	0.898	0.983	0.959
		100	0.889	0.858	0.951	0.916	0.987	0.963
	2	50	0.873	0.871	0.928	0.919	0.990	0.955
		100	0.886	0.872	0.942	0.918	0.989	0.962
	4	50	0.933	0.946	0.944	0.956	0.981	0.970
		100	0.928	0.923	0.949	0.943	0.988	0.972

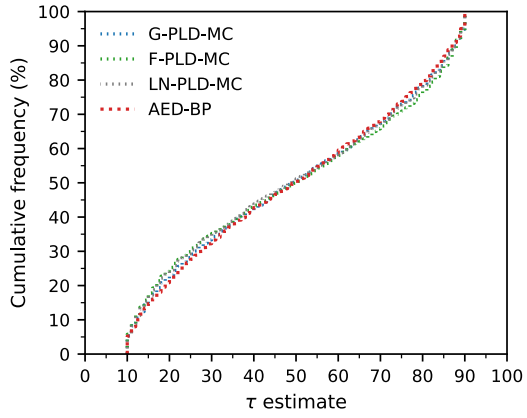

Fig. 3. Frequency distribution of the change point estimate for the different methods applied to log-normal samples when H_0 holds.

2.3. Properties of confidence curves

To evaluate the performance of AED-BP and PLD-MC and to compare confidence curves, metrics are needed. For synthetic data series objective criteria can be formulated because the true change point is known. The results of AED-BP when applied to real data series can only be judged by the amount of uncertainty in the result and by comparison with other published results.

The following properties will be examined:

- Actual versus nominal coverage probability for the confidence sets produced by the curves at several confidence levels for synthetic data. For the definition of actual and nominal coverage see Appendix A.
- The cumulative distribution of $\hat{\tau}(y)$ for PLD-MC and $\hat{\tau}_{\text{apn}}(y)$ for AED-BP for the synthetic data experiments when the null hypothesis H_0 holds. This provides a measure of possible bias for reporting certain

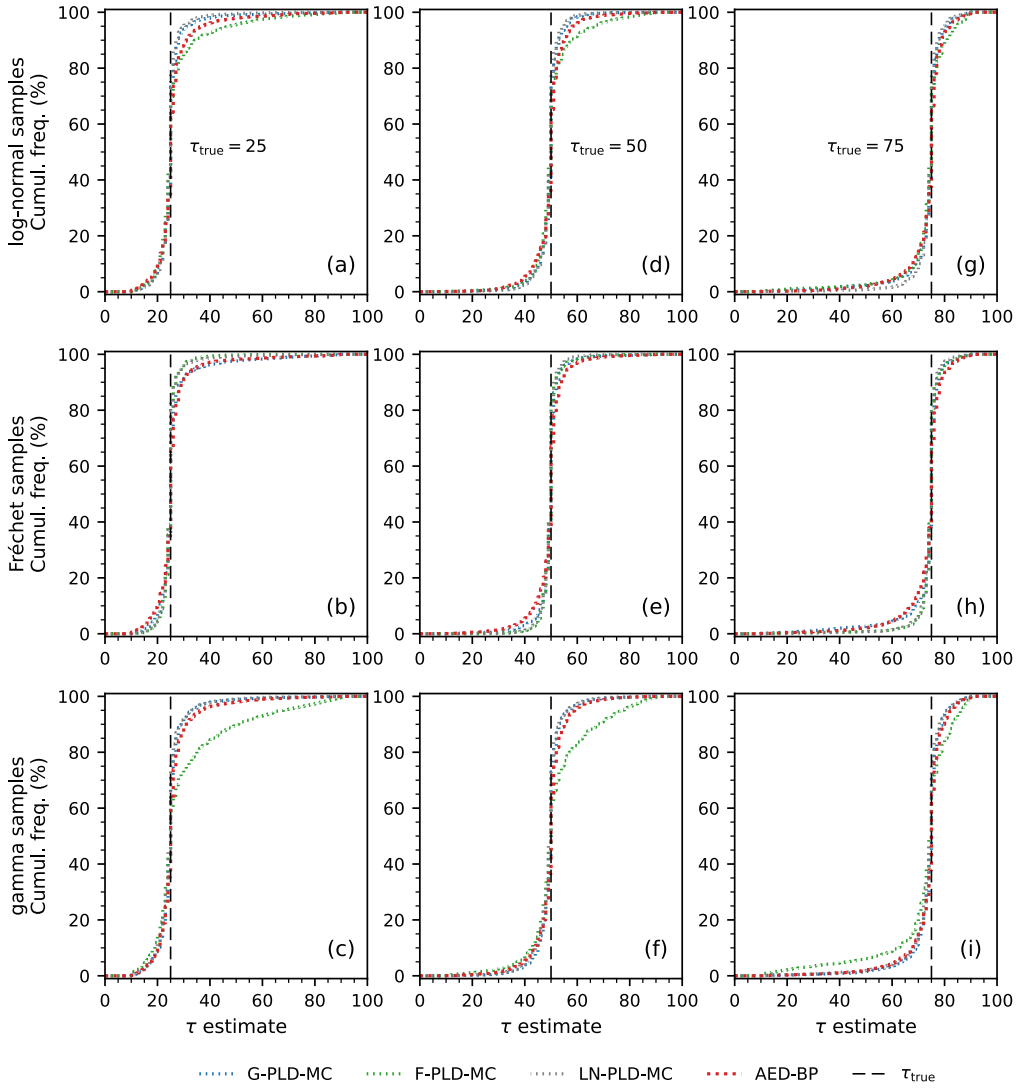

Fig. 4. Frequency distribution of the change point estimate for the different methods under the alternative hypothesis H_1 with $\tau = 25, 50, 75$ and with magnitudes of change $\Delta\mu = 1$.

locations when a type I error is made. Ideally, the distribution should be uniform.

- The cumulative distribution of $\hat{\tau}(y)$ and $\hat{\tau}_{\text{apn}}(y)$ for the synthetic data experiments when the alternative hypothesis H_1 holds. This provides a measure of how well the point estimators work. Ideally, the distribution should be a step function with the step at the true change point location.
- The *similarity index* for two confidence curves. In Appendix C a derivation is given for

$$\widetilde{J} = \frac{\sum_{\tau=1}^{n} \min(1 - \text{cc}(\tau, y_{\text{obs}}), 1 - \text{cc}'(\tau, y_{\text{obs}}))}{\sum_{\tau=1}^{n} \max(1 - \text{cc}(\tau, y_{\text{obs}}), 1 - \text{cc}'(\tau, y_{\text{obs}}))} \tag{12}$$

which will be used to compare the similarity of pairs of confidence curves.

- The *slimness of the confidence sets* associated with a confidence curve as the ratio between the actual number of points contained in a confidence set at confidence level α and the number αn

$$\text{slimness} = \frac{\#R_\alpha}{\alpha n} \tag{13}$$

where $\#R_\alpha$ is the number of points in the set R_α defined in (A.12). For methods that exclude the first and last $n_{\text{tr}} - 1$ points of the series a modified form should be used

$$\text{slimness} = \frac{\#R_\alpha}{\alpha(n - 2(n_{\text{tr}} - 1))} \tag{14}$$

3. Analysis results for synthetic data

In order to evaluate the performance of PLD-MC and AED-BP, synthetic time series from three different distributions are generated: the log-normal distribution, the gamma distribution, and the Fréchet distribution with a constant shape parameter. Moreover, three variants of PLD-MC will be considered. One using the log-normal pdf (LN-PLD-MC), one using the gamma pdf (G-PLD-MC), and one using the Fréchet pdf (F-PLD-MC). The parametric distribution functions of the three distributions can be found in Appendix D. Notice that in this study, the shape parameter of the Fréchet distribution is fixed to avoid over-fitting to hydrometeorological data.

3.1. Data generation

For each combination consisting of a distribution, a change point at $\tau = 25, 50, 75$, and a change in the mean $\Delta\mu = 1, 2, 4$, a set of 1000

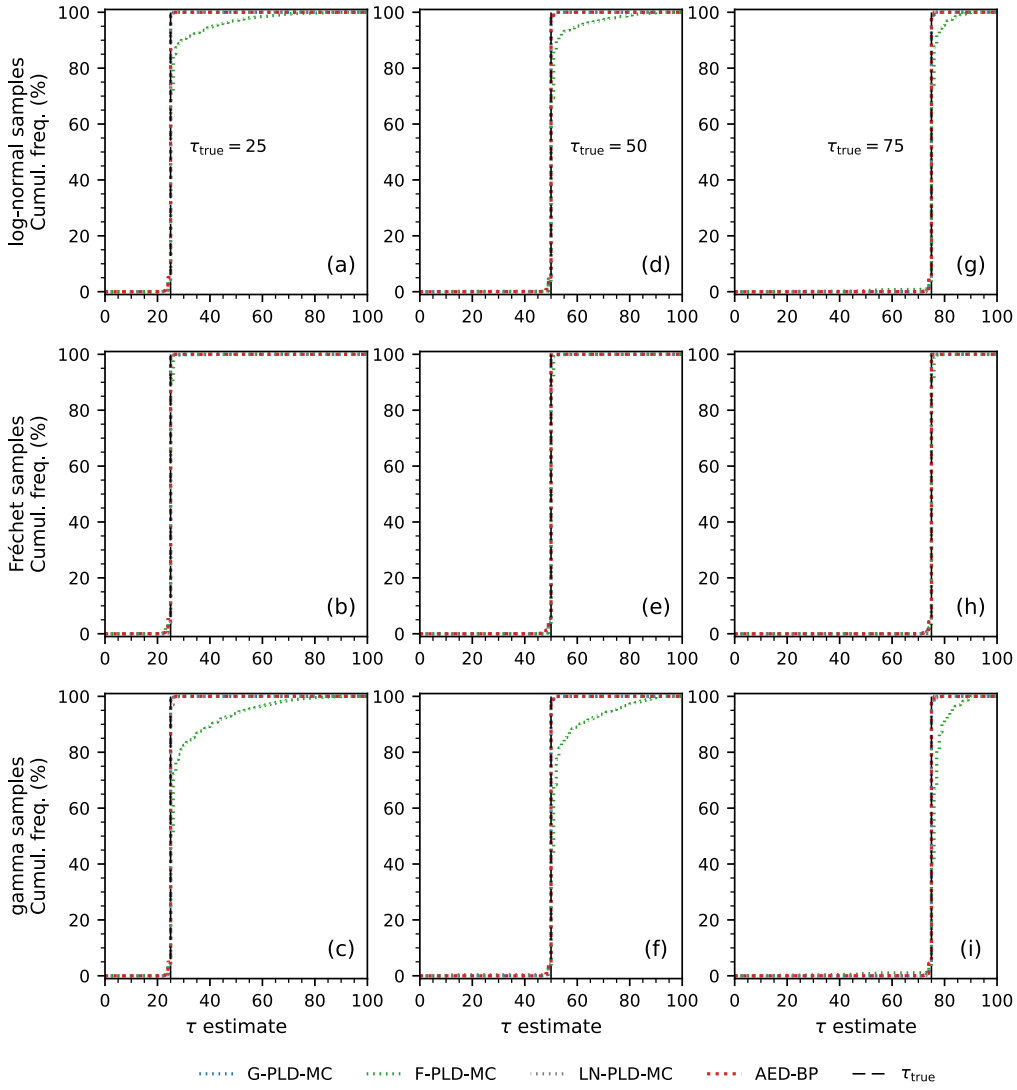

Fig. 5. Frequency distribution of the change point estimate for the different methods under the alternative hypothesis H_1 with $\tau = 25, 50, 75$ and with magnitudes of change $\Delta\mu = 4$.

synthetic time series of length $n = 100$ were generated. For the coverage analysis, an additional 1000 synthetic time series of length $n = 50$ with a change at were generated for changes in the mean of $\Delta\mu = 1, 2, 4$. The τ used in the generation of the time series will be referred to as τ_{true}.

The mean of the distribution for the sub-series up to τ will be denoted by μ_{L}, and the mean of the distribution for the sub-series beyond τ will be denoted by μ_{R}. Similarly, σ_{L} and σ_{R} will refer to the standard deviation of these distributions. For all series we have $\sigma_{\text{L}} = \sigma_{\text{R}} = 1$, $\mu_{\text{L}} = 2$, and $\mu_{\text{R}} = \mu_{\text{L}} + \Delta\mu$. The scale of change is measured by $\Delta\mu/\sigma$, and in our setup $\sigma_L = \sigma_R = 1$, therefore, for the synthetic data the relative size of the change in the sample mean at a change

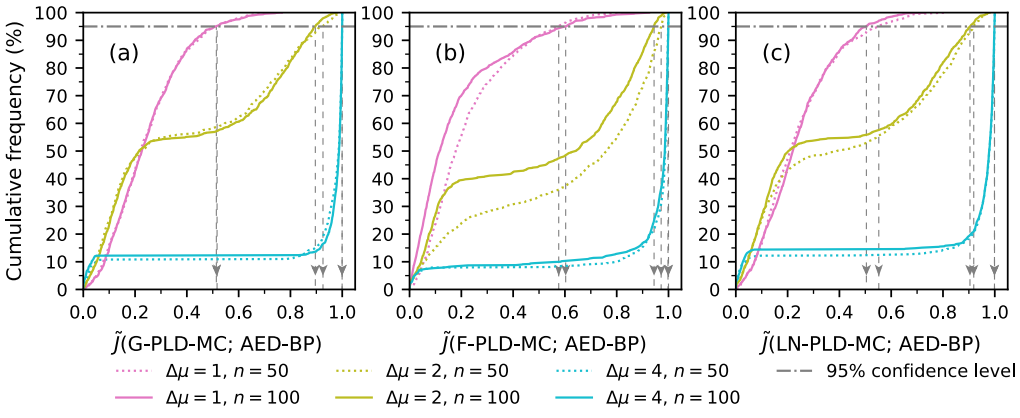

Fig. 6. Similarity index $\tilde{J}$ between confidence curves constructed by PLD-MC and AED-BP.

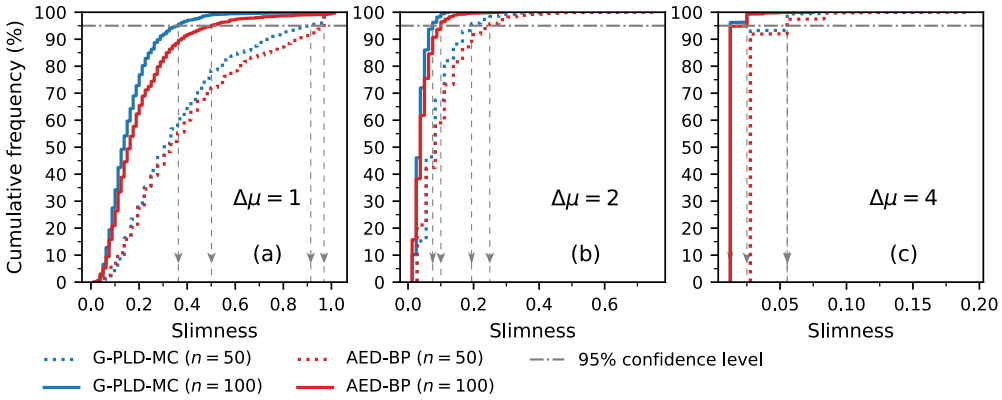

Fig. 7. An example to show the cumulative frequency distribution of the slimness of 95% confidence sets for synthetic time series generated from the gamma distribution.

point can be represented by $\Delta\mu$.

3.2. Some examples of confidence curves for synthetic data

Fig. 1 shows four synthetic data sets (a, c, e, g) of length $n = 100$ drawn from the log-normal distribution, and the corresponding confidence curves (b, d, f, h) for the different methods. To illustrate how uncertainty is represented by confidence curves, the 95% confidence sets for AED-BP are shown in Fig. 1 (b, d, f, h). Series (a) was generated with $\Delta\mu = 0$; series (c, e, g) were generated with a change in the mean of $\Delta\mu = 1, 2, 4$ respectively at $\tau = 50$. Note that Fig. 1 shows information for just four data sets, so it cannot be used to draw conclusions about the relative performances of the methods. Fig. 1 (b) suggests that when the null hypothesis H_0 holds, the confidence sets are much larger than when the null hypothesis does not hold, see Fig. 1 (d, f, h). When $\Delta\mu$ is small, in general when $\Delta\mu/\sigma$ is small, both methods find larger confidence sets at the higher confidence levels. It is important to keep in mind that confidence intervals at levels below 0.5 are of limited usefulness as they need only contain the true change point in less than half of all experiments.

3.3. Performance of confidence curve methods for synthetic data

For synthetic data, the actual coverage probability will be examined as well as the distribution of the estimate of the change point both when the null hypothesis H_0 holds and when it does not hold. The slimness of the confidence sets for $\alpha = 0.95$ will be examined as well. These properties will be used to determine the relative merits of the methods. Finally, the similarity of the curves generated by PLD-MC and AED-BP will be examined.

3.3.1. Actual versus nominal coverage probability of confidence curves

Fig. 2 presents plots of both confidence set size and actual coverage as a function of confidence level. Plots (c, f, i) show clearly that for certain coverage levels there is no corresponding set with an actual coverage close to the nominal coverage. This can be explained as follows. The change point location is an integer, therefore the smallest non-empty confidence set is a set that contains just one point. For a one point set, the confidence level of the set can never be lower than the probability that this point is the change point. Plots (c, f, i) show that for $\Delta\mu = 4$ this probability is often above 90%. For $\Delta\mu = 1, 2$ and confidence levels of 80% or higher, both PLD-MC and AED-BP deliver reasonable actual coverage probabilities.

For the nominal coverage probabilities $\alpha = 0.90, 0.95, 0.99$, estimates of the actual coverage probability of the confidence curves constructed by the PLD-MC variants and AED-BP are listed in Table 1. For $\Delta\mu = 1$, the actual coverage probabilities are somewhat lower than the nominal coverage probabilities, so the sets are permissive. For

Table 2
The mean value of slimness of confidence sets for PLD-MC and AED-BP at three different confidence levels.

α			0.90		0.95		0.99	
Distribution	$\Delta\mu$	n	PLD-MC	AED-BP	PLD-MC	AED-BP	PLD-MC	AED-BP
gamma	1	50	0.29	0.32	0.37	0.40	0.55	0.57
		100	0.12	0.16	0.16	0.20	0.26	0.32
	2	50	0.07	0.08	0.09	0.10	0.14	0.16
		100	0.03	0.03	0.04	0.04	0.06	0.07
	4	50	0.03	0.03	0.03	0.03	0.04	0.04
		100	0.01	0.01	0.01	0.01	0.02	0.02
Fréchet	1	50	0.15	0.31	0.19	0.39	0.31	0.55
		100	0.02	0.16	0.03	0.21	0.04	0.33
	2	50	0.04	0.08	0.06	0.12	0.09	0.19
		100	0.02	0.03	0.02	0.05	0.03	0.09
	4	50	0.03	0.03	0.03	0.03	0.04	0.05
		100	0.01	0.01	0.01	0.01	0.02	0.02
log-normal	1	50	0.23	0.33	0.30	0.41	0.47	0.56
		100	0.10	0.16	0.13	0.21	0.20	0.33
	2	50	0.06	0.08	0.08	0.11	0.12	0.17
		100	0.03	0.03	0.03	0.05	0.05	0.08
	4	50	0.03	0.03	0.03	0.03	0.04	0.04
		100	0.01	0.01	0.01	0.01	0.02	0.02

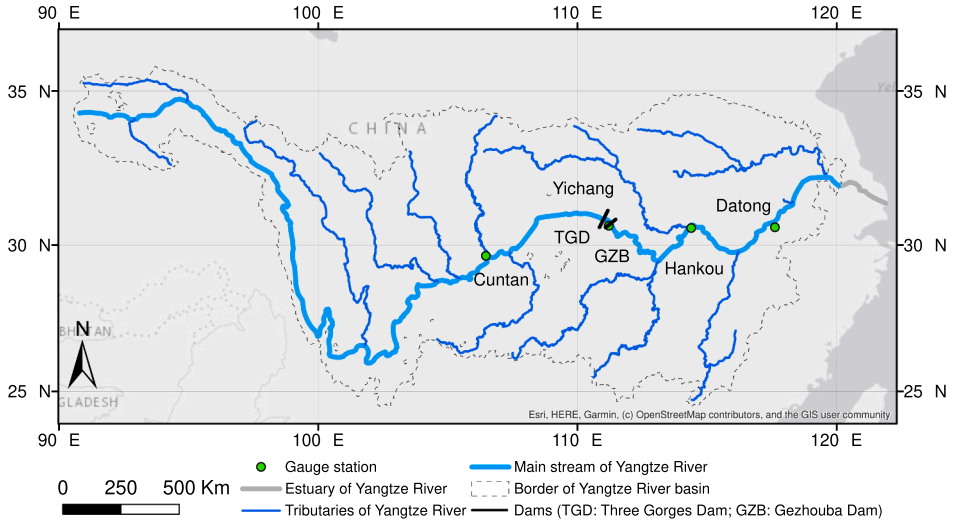

Fig. 8. Location of four gauge stations and two dams on the Yangtze River (China).

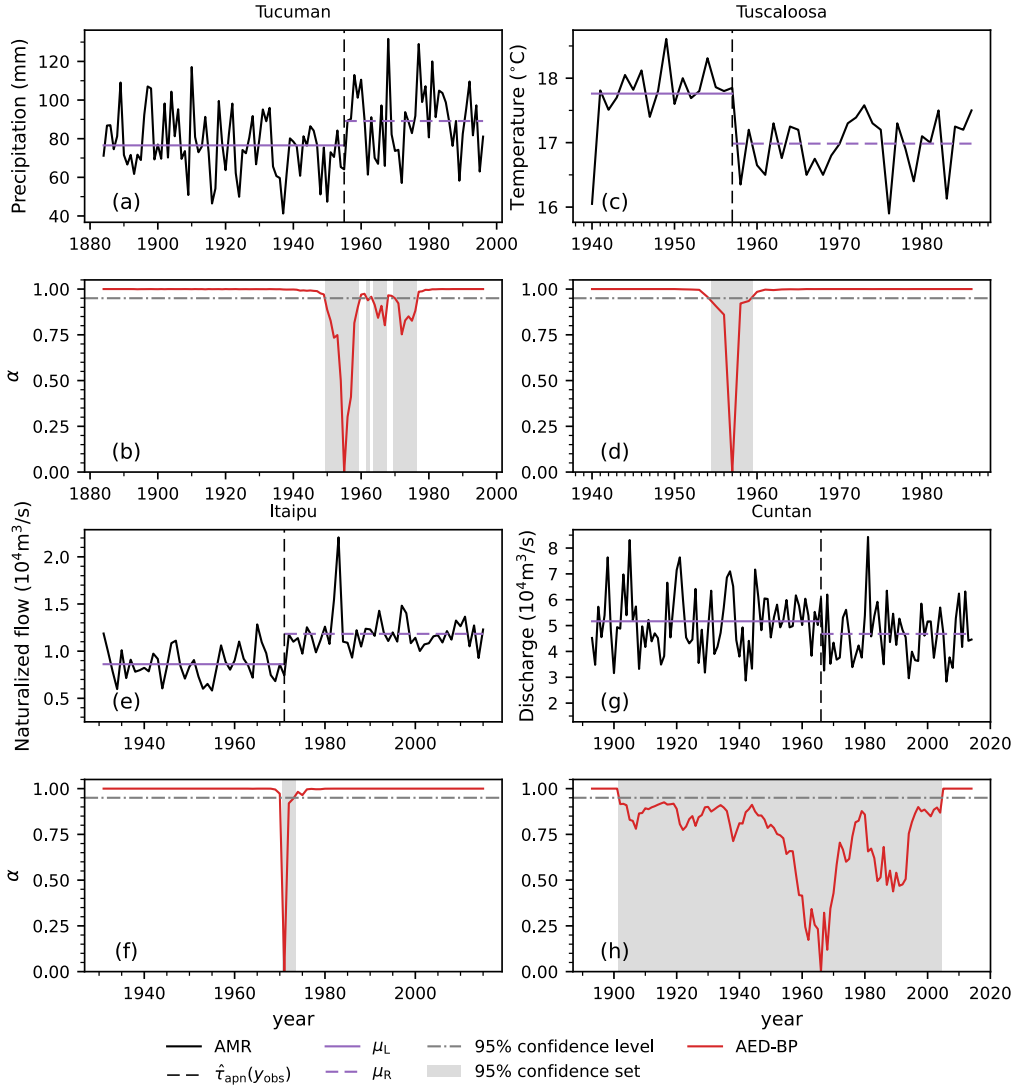

Fig. 9. Change point analysis of hydrometeorological time series from published papers.

Table 3
Change points found in the hydrometeorological series and statistical properties of the series.

Time series	from	to	τ_{ref}	$\hat{\tau}_{apn}(y_{obs})$	σ	$\frac{\Delta\mu}{\sigma}$	slimness
Tucumán	1884	1996	1956	1955	18 mm	0.76	0.24
Tuscaloosa	1940	1986	1957	1957	0.61 °C	− 1.3	0.16
Itaipu	1931	2015	1971	1971	2.5×10^3 m^3/s	1.3	0.046
Cuntan	1893	2014	not found	1966	12×10^3 m^3/s	− 0.45	1.0
Yichang	1946	2014	1962, 1966	2005	8.6×10^3 m^3/s	− 1.3	0.10
Yichang 'short'	1946	2010	n/a	1962	8.3×10^3 m^3/s	− 0.87	0.83
Hankou	1952	2014	not found	2005	8.9×10^3 m^3/s	− 0.89	0.49
Datong	1950	2014	not found	2003	11×10^3 m^3/s	− 0.76	1.0

$\Delta\mu = 4$, the actual coverage for $\alpha = 0.90$ and $\alpha = 0.95$ is often higher than the nominal coverage, the corresponding sets are conservative (see Appendix A). Actual coverage tends to be closer to the nominal value for longer time series. The actual coverage was estimated as follows: for $m = 1, 2, \ldots, M$ synthetic time series were generated, and for each time series m the confidence curve and the confidence set $R_{\alpha,m}$ at confidence level α were determined. The actual coverage was estimated by dividing the number k of sets for which $\tau_{true} \in R_{\alpha,m}$ by M. If the actual coverage were equal to the nominal coverage, then the number k of sets $R_{\alpha,m}(\alpha)$ that contained the true change point would be distributed according to a binomial distribution

$$\Pr(k) = \binom{M}{k} \alpha^k (1-\alpha)^{M-k} \tag{15}$$

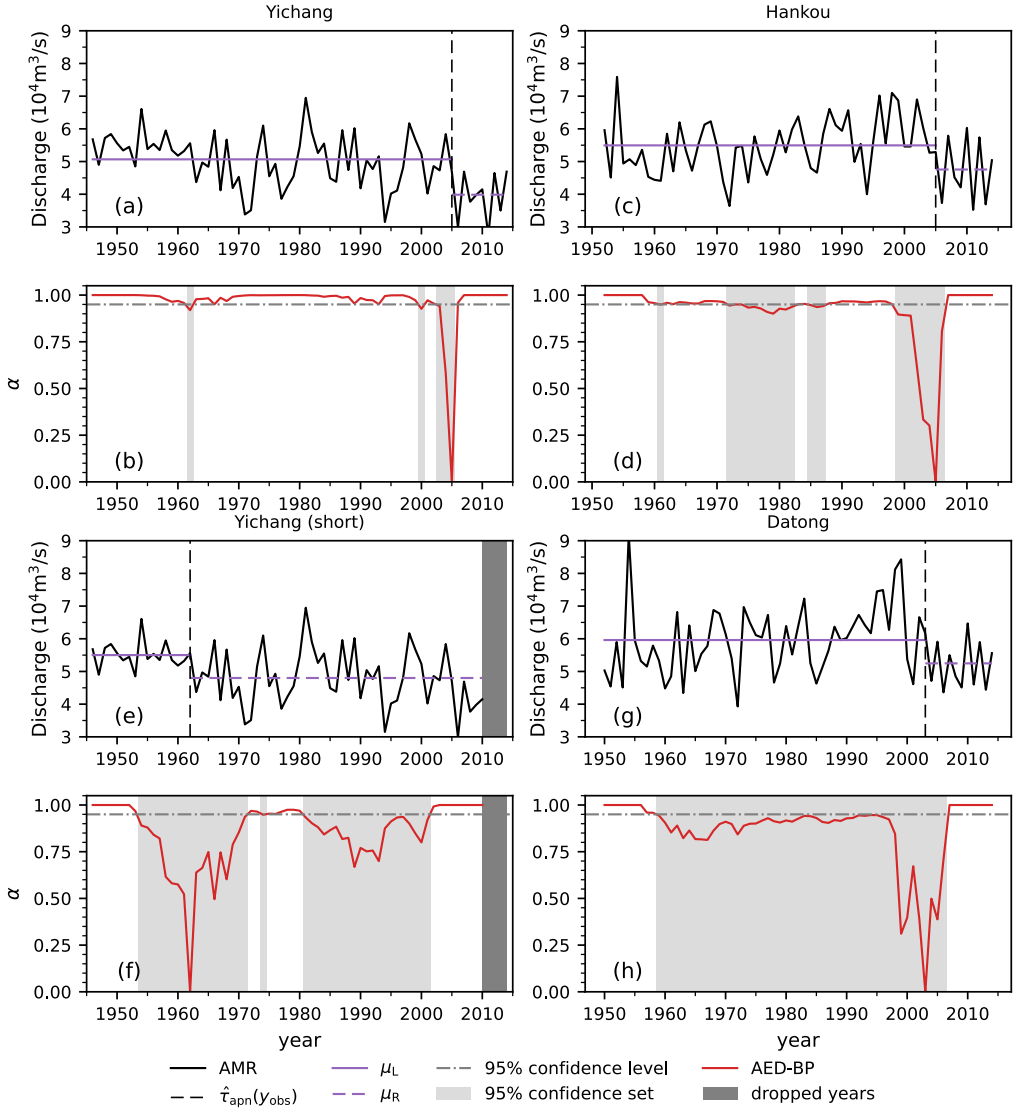

Fig. 10. Change point analysis of annual maximum discharge time series in Yangtze River.

so the distribution of the coverages k/M found by an experiment with M realizations is known. For the binomial distribution the variance is $M\alpha(1-\alpha)$, so the standard deviation of k/M is $\sqrt{\alpha(1-\alpha)/M}$. For $M = 1000$, the standard deviation of the distribution of k for $\alpha = 0.9$ is 0.009; for $\alpha = 0.95$ it is 0.007, and for $\alpha = 0.99$ it is 0.003. When combining this information with Table 1, please keep in mind that the location of the change point is a discrete random variable, so for some confidence levels it might not be possible to define a confidence set with that exact coverage.

3.3.2. The frequency distribution of the estimated change points when the null hypothesis holds

Fig. 3 shows the frequency distribution of the change point estimates when the null hypothesis H_0 holds ($\Delta\mu = 0$) for all methods. Results are shown for log-normal samples; the results for other sample types are very similar. For both PLD-MC and AED-BP, the frequency distribution is close to uniform, except near the endpoints of the series. Moreover, under the null hypothesis, the type of parametric distribution used in the PLD-MC method has little or no effect on the outcome for the distributions considered here.

3.3.3. The frequency distribution of the estimated change points when the alternative hypothesis holds

Figs. 4 and 5 show the frequency distribution of the change point estimates when the alternative hypothesis holds for the different methods for $\Delta\mu = 1$ and $\Delta\mu = 4$ respectively. The plots for $\Delta\mu = 2$ were omitted as the curves lie between those for $\Delta\mu = 1$ and $\Delta\mu = 4$. The results are very similar for all three change point locations ($\tau = 25, 50, 75$). Moreover, the frequency distributions of $\hat{\tau}_{\text{apn}}(y)$ for the AED-BP are very close to $\hat{\tau}(y)$ for the PLD-MC variant based on the distribution that matches the data source.

When we compare Figs. 4 and 5, it is clear that LN-PLD-MC, G-PLD-MC, and AED-BP perform very well for $\Delta\mu = 1$ and $\Delta\mu = 4$ on all data, while F-PLD-MC struggles with data from the log-normal and the gamma distribution even for $\Delta\mu = 4$. Because we expected this effect, samples were taken from two distributions with a shared fixed support $[0, \infty)$, namely the log-normal and the gamma distribution, and one distribution with a parameter dependent support, the Fréchet distribution.

3.3.4. Similarity of confidence curves

The similarity of PLD-MC and AED-BP confidence curves was measured by the similarity index $\widetilde{J}$ (12). Fig. 6 shows examples of the resulting frequency distribution. The sample length has very limited influence on the similarity index, but the size of the change strongly influences the result. When $\Delta\mu = 1$, for half of the time series the similarity index is below 0.25 and according to Fig. 6, this is very low. When $\Delta\mu = 4$, for most of the time series, the similarity index is above 0.8, which indicates that both methods obtain very similar results. For $\Delta\mu = 2$ results vary, for 30% of the time series the similarity index exceeds 0.7.

3.3.5. The slimness of confidence sets

The slimness of the confidence sets R_α provides an indication of the reduction of uncertainty that the methods are capable of. By construction a set R_α at a higher confidence level will include the sets at lower confidence levels, so from $0 < \alpha < \alpha' < 1$ it follows that $R_\alpha \subseteq R_{\alpha'}$. As mentioned in Appendix A, a set with αn points randomly chosen from $\{1, 2, \ldots, n\}$ will have coverage probability α. Therefore, a slimness value well below one indicates that the curve is more informative than a random guess, while a slimness value of one or more indicates it does not do better than a random guess at that confidence level. Slimness of confidence sets at confidence levels of 90%, 95% and 99% are considered for synthetic data, but only 95% confidence sets are taken into account for the real data. We calculated the slimness of confidence sets (13) at different confidence levels for 1000 synthetic time series. Fig. 7 gives the frequency distribution of the slimness of the confidence sets with $\alpha = 0.95$ constructed for gamma distributed samples with $\Delta\mu = 1, 2, 4$ and sample length $n = 50, 100$. Fig. 7 shows that the confidence sets for $n = 100$ are slimmer than those for $n = 50$, so the sets tend to be slimmer for longer time series. If we compare (a), (b) and (c) then we see that the slimness decreases rapidly with increasing $\Delta\mu$ and the remaining uncertainty for $\Delta\mu = 4$ is very low. In fact, if we convert the slimness for $\Delta\mu = 4$ back to the set size, then we see that for $n = 50$ and $n = 100$, most 95% confidence sets contain 3 or fewer points. The differences in slimness between G-PLD-MC and AED-BP are minor. In Table 2 mean values for slimness are presented. For each distribution, the corresponding PLD-MC method was applied. For $\Delta\mu = 4$ all methods deliver similar mean slimness. For $\Delta\mu = 1$ results were similar for G-PLD-MC and AED-BP, but on samples from the Fréchet and log-normal distribution the matching parametric method did better than AED-BP.

4. Analysis results for hydrometeorological data

The results of AED-BP for synthetic data series are promising. The next step is the application of the method to hydrometeorological time series that have been examined in previous studies and the comparison of our results with those of previous studies.

4.1. Data source

The time series used for analysis are the following:

- The annual average rainfall data from Tucumán in Argentina for the years 1884 to 1996. The time series is well documented, and in Jandhyala et al. (2010), a change point in the time series was found near 1956 by a Bayesian method. Wu et al. (2001) also studied this series, and they state: '[C.] Lamelas [a meteorologist from the Agricultural Experimental Station Obispo Colombres, Tucumán] believes that there was a change in the mean, caused by the construction of a dam in Tucumán from 1952 to 1962'.
- The annual average temperature time series from a station in Tuscaloosa, Alabama (USA). The time series in Tuscaloosa from 1940 to 1986 was selected because during this period, there was only one documented reason for a change point, namely in November 1957 (Reeves et al., 2007). All eight methods used in that study found a change point in the year of 1957.
- Conte et al. (2019) used the bootstrap Pettitt test to detect change points in the annual average naturalized flow of the Itaipu Hydroelectric Plant in Brazil from 1931 to 2015. They found a significant change point for the naturalized flow in 1971.
- Time series of annual maximum run-off (AMR) for four stations on the Yangtze River in China were analysed in Zhou et al. (2019). The four stations are of interest because they are located on Yangtze River, a river that has gone through many alterations over the past 100 years, notably the construction of the Three Gorges project. The stations are: Cuntan (1893–2014) upstream of the Three Gorges dam, and Yichang (1946–2014), Hankou (1952–2014), and Datong (1950–2014) downstream of the Gezhouba dam, see Fig. 8. In Zhou et al. (2019) of the four stations only Yichang station yielded change points. The paper applied three methods to this series: the Pettitt method, a method based on the Cramér von Mises test, and a variant on the CUSUM method. CUSUM found a change point in 1962 with $\Delta\mu/\sigma = -0.91$ and the other two methods found a change point in 1966 with $\Delta\mu/\sigma = -0.84$.

4.2. Analysis results

For the Tucumán, Tuscaloosa, and Itaipu, time series shown in Fig. 9 (a, c, e), the confidence curves generated by AED-BP are shown in Fig. 9 (b, d, f). These curves show that, for each of the change points, the uncertainty is low. Table 3 lists the change points τ_{ref} found in the references, $\hat{\tau}_{\text{apn}}(y_{\text{obs}})$ for AED-BP, and the slimness of the 95% confidence sets found by AED-BP. For Tuscaloosa and Itaipu, the estimate $\hat{\tau}_{\text{apn}}(y_{\text{obs}})$ coincides with the point found in the references. For Tucumán $\hat{\tau}_{\text{apn}}(y_{\text{obs}})$ is off by one year, but well within the 95% confidence set.

For Cuntan, Zhou et al. (2019) did not find a significant change point. This agrees with the results of AED-BP: the shape of the confidence curve in Fig. 9 (h) suggests there is little or no reliable information on the change point location. The difficulty in finding a change point might also be due to the relative smallness of the putative change, see Table 3.

For Yichang station, AED-BP strongly suggests that there is a change point near 2005, see Fig. 10 (a). For this station, the slimness is acceptable, according to Fig. 10 (b) or Table 3, but the discrepancy between the current and the earlier study is intriguing. To further examine it, a sub-series of the time series was analysed. Fig. 10 (f) shows that for the time series from 1946 to 2010 (Yichang 'short') AED-BP found a change point in 1962, but the size of the 95% confidence set is now much larger. The size of the change is also smaller. To see whether there might be a second change point, years were successively dropped from the series. It turned out that 2005 was selected until it was masked by n_{tr} (which was 8 in this case). Table 3 shows that when 2005 was masked, 1962 was found, but with a much wider 95% confidence set and therefore uncertainty.

For Hankou, downstream of Yichang, Fig. 10 (d) suggests there may be a change point in 2005. The uncertainty is bigger than for Yichang. The putative change is a bit smaller, which may explain part of the additional uncertainty (Table 3). The methods used in the reference did not find a significant change point at this station.

Further downstream lies Datong station, but, while there is a drop in the confidence curve around 2003, the 95% confidence interval with slimness 1.0 (Table 3) is completely uninformative, see also Fig. 10 (h). The methods used in the reference did not find a significant change point at this station.

From Table 3 and Fig. 10 (d, f, h) it would seem that, if the construction of the dam did indeed cause changes in extreme discharges, then these are less visible further downstream.

5. Conclusion

This study provides a distribution-free way to construct confidence curves for change points. The method is based on an Approximation of the Empirical likelihood function, which is used to construct a Deviance function. The bootstrap method is used to construct an approximation of the probability distribution of the deviance (AED-BP). The method introduced by Cunen et al. (2018) is used as an alternate source of confidence curves. It combines a Parametric Likelihood function with a Deviance function and Monte Carlo simulation (PLD-MC). Both methods intrinsically provide confidence sets at all confidence levels that quantify the uncertainty in the results of change point detection. This is an advantage over classical change point detection methods that do not have this feature. Bayesian methods do provide a representation of uncertainty, but they need a prior distribution. The advantage of AED-BP over PLD-MC is that it is non-parametric. This frees the user from the need to select of a distribution family for the time series.

Simulations with synthetic data show that the confidence curves can correctly represent the uncertainty in results of change point detection. Moreover, the performance of the AED-BP is similar to that of PLD-MC. The similarity between confidence curves constructed by AED-BP and PLD-MC is very high when the jump in the mean is large. For the experiments done in this paper, the sample length does not have much influence on the similarity between two confidence curves. For both parametric and non-parametric methods, uncertainty of the change point results decreases with increasing series length. In the experiments with synthetic data, the uncertainty also decreases as the ratio of the change in the mean to the standard deviation increases.

Experiments with real data show that the AED-BP is applicable for hydrometeorological data, but as most non-parametric methods, it may be somewhat less effective than a parametric method with the correct underlying distribution. This needs further investigation. The AED-BP results on the AMR series for the stations Yichang and Hankou along the Yangtze river are among the first that show a possible change point due to the Three Gorges dam on the AMR after the first generator became operational in 2003. From the results of the real data, it seems that there might be multiple change points in a time series. Therefore, we plan to extend the distribution-free method to a multiple change point problem in a future study.

6. Funding

This work was supported by the China Scholarship Council under Grant No. 201706710004.

CRediT authorship contribution statement

Changrang Zhou: Writing - original draft, Writing - review & editing, Investigation, Conceptualization, Formal analysis, Data curation, Methodology, Software. **Ronald van Nooijen:** Supervision, Formal analysis, Software, Writing - review & editing, Validation, Conceptualization. **Alla Kolechkina:** Formal analysis, Validation, Visualization, Writing - review & editing. **Nick van de Giesen:** Supervision, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was partially developed within the framework of the Panta Rhei research initiative of the International Association of Hydrological Sciences (IAHS) by the members of the working group on 'Natural and man-made control systems in water resources' (Montanari et al., 2013). The authors would like to thank J. Reeves (University of Georgia), K. Jandhyala (Washington State University), and D. M. Bayer (Federal University of Rio Grande do Sul) for sharing their data, which allowed us to test our method. Moreover, they are especially grateful to A. Bardossy (editor) and two anonymous reviewers for the detailed comments and suggestions that helped to improve the paper.

Appendix A. Notation and definitions

A.1. Basic notation

There is a wide range of notations in use in statistics. Here, the notation and terminology used in this paper are specified. Random variables are denoted by capital letters and realizations of random variables by the corresponding lower case letters. Parameters of distributions are denoted by

lower case Greek letters. If E is an event then

$$\Pr(E) \tag{A.1}$$

denotes the probability of that event. A sequence of n independent identically distributed random variables $X_1, X_2, \ldots, X_n$ is a *random sample* of size n. The sample as a whole may be referred to as X.

Traditionally, probability theory and statistics make use of the indicator function of a set, which is a function that takes the value one on points in the set and zero elsewhere. There is a simpler and more general approach proposed by Knuth (1992), who in turn cites Iverson (1962) as the original source of the idea. This approach uses special brackets to translate an expression that is false or true into 0 or 1 respectively. Here $[\![]\!]$are used. Examples are:

$$[\![1 \leq 4 \leq 3]\!] = 0 \tag{A.2}$$

$$[\![1 \leq 2 \leq 3]\!] = 1 \tag{A.3}$$

$$\left[\!\!\left[1 \leqslant x \leq 3 \right]\!\!\right] = \begin{cases} 0 & x < 1 \\ 1 & 1 \leqslant x \leq 3 \\ 0 & x > 3 \end{cases} \tag{A.4}$$

The indicator function of a set A applied to a variable x can now be written as $[\![x \in A]\!]$. For the empirical cumulative distribution function (ecdf) F_n of a random sample of size n, one can write

$$F_n(t) = \frac{1}{n} \sum_{i=1}^{n} \left[\!\!\left[X_i \leqslant t \right]\!\!\right] \tag{A.5}$$

Please note that for each fixed value of t the expression $F_n(t)$ is itself a random variable.

A.2. Definitions of confidence curves

In the literature confidence curves have been defined in several different ways. The following definition provides a starting point. As in Cunen et al. (2018), α is used to denote a confidence level.

Definition 1. A *confidence interval* with *confidence level* (also known as *confidence coefficient*) α for a statistic λ of a random sample X is an interval with random endpoints $u(X)$ and $v(X)$ such that for the true value λ_0

$$\Pr(u(X) \leqslant \lambda_0 \leqslant v(X)) = \alpha \tag{A.6}$$

For a confidence interval, the *nominal coverage probability* equals the confidence level. If one of the assumptions used in the derivation of the endpoints does not hold, then the *actual coverage probability* may well be different.

While very useful, traditional confidence intervals are somewhat restrictive. For instance, if we have a bimodal distribution, then a combination of two intervals, each centred on a mode, may contain fewer values and therefore be more informative than any single interval at the same confidence level. Therefore a more general concept was introduced: the confidence set.

Definition 2. A *confidence set* with confidence level α for a parameter λ is a random set $R(X)$ such that

$$\Pr(\lambda \in R(X)) = \alpha \tag{A.7}$$

Here α is the *nominal coverage probability of the set*. In the calculation of α assumptions are made on the distribution of X that may or may not hold for a specific application. If they do not hold, then it becomes necessary to distinguish between the nominal and the actual coverage probability. The *actual coverage probability of the set* is the probability that the parameter is in the set for a given application. It can be approximated by a Monte Carlo experiment. If the actual coverage probability exceeds α, then the true value lies in the set with probability greater than α. Usually, this means we will err on the side of caution. In this case, the set is called *conservative*. If the actual coverage probability is lower than α, then the set is called *anti-conservative* or *permissive*.

Definition 2 contains an undefined term, namely 'random set'. A general definition can be found in, for instance, Molchanov (2017). For the purposes of this study a definition by analogy is perhaps more helpful. Just like a random variable represents an aspect of an event as a real number, a random set represents an aspect of an event as a set, for instance, a set of real numbers. Note, that a confidence interval is a special case of a random set.

The confidence curve concept has evolved over time. An early definition was given by Birnbaum (1961) who defined a confidence curve as 'a set of upper and lower confidence limits, at each confidence coefficient from 0.5 to 1, inclusive'. As stated earlier, in some cases it might be advantageous to use confidence sets instead of confidence intervals. To that end, Schweder and Hjort (2016, Definition 4.3) gave a more general abstract definition of a confidence curve. Here we give a variation on that definition.

Definition 3. Suppose X is a random sample of size n, and λ is a property of the underlying distribution with values in a value set V. A function $g(\lambda, x)$ with range $[0, 1]$ that is continuous in x for fixed λ is a *confidence curve* when:

1. There is a point estimator $\hat{\lambda}$ for λ such that

$$\min_{\lambda \in V} g(\lambda, x) = g(\hat{\lambda}(x), x) = 0 \tag{A.8}$$

for all realizations x of X.

2. For the true value λ_{true} of the property λ, the random variable $g(\lambda_{\text{true}}, X)$ has the uniform distribution on the unit interval.

It is important to note that point 2 in Definition 3 means that the value of $g(\lambda_{\text{true}}, X)$ need not be the minimum of $g(\lambda, x)$. It is not the minimum of the curve, but the curve as a whole that is meaningful. The estimate $\hat{\lambda}(x)$ is merely a reference point that, in the case of a confidence curve with only one minimum, has a role similar to that of the median in the case of a probability distribution for λ.

Example 1. If X is a sample of size n from the normal distribution then λ could, for instance, be the mean or the variance, and the values of x, realizations of X, would lie in $\mathbb{R}^n$. If we take λ to be the mean, and the underlying distribution is a normal distribution with unknown mean μ_{true} and known variance σ, then g could, for instance, be

$$g(\lambda;x) = \begin{cases} 1 - 2\Phi\left(\frac{\lambda - \frac{1}{n}\sum_{i=1}^{n} x_i}{\sigma/\sqrt{n}}\right) & \lambda < \frac{1}{n}\sum_{i=1}^{n} x_i \\ 2\Phi\left(\frac{\lambda - \frac{1}{n}\sum_{i=1}^{n} x_i}{\sigma/\sqrt{n}}\right) - 1 & \lambda > \frac{1}{n}\sum_{i=1}^{n} x_i \end{cases} \tag{A.9}$$

where Φ is the cumulative distribution function of the standard normal distribution and

$$\hat{\lambda}(x) = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{A.10}$$

For the case with unknown σ, see, for example, Schweder and Hjort (2016, page 73).

If $\text{cc}(\cdot,\cdot)$ is a confidence curve according to Definition 3, then for fixed λ_0, the function $\text{cc}(\lambda_0, X)$ is measurable and a random variable. Hence, we can speak of the distribution of $\text{cc}(\lambda_0, X)$. For a given confidence level $\alpha \in [0, 1]$ and a given property value λ, it is possible to determine

$$\Pr(\text{cc}(\lambda, X) \leqslant \alpha) \tag{A.11}$$

Next, define the sets

$$R_\alpha(x) = \{\lambda\colon \text{cc}(\lambda, x) \leqslant \alpha\} \tag{A.12}$$

If λ_{true} is the true value of the parameter, then according to Definition 3, the random variable $\text{cc}(\lambda_{\text{true}}, X)$ is uniformly distributed on $[0, 1]$, and therefore

$$\Pr(\text{cc}(\lambda_{\text{true}}, X) \leqslant \alpha) = \alpha \tag{A.13}$$

Next, consider the probability $\Pr(\lambda_{\text{true}} \in R_\alpha(X))$. By definition, $\lambda_{\text{true}} \in R_\alpha(X)$, if and only if $\text{cc}(\lambda_{\text{true}}, x) \leqslant \alpha$. It follows that

$$\Pr(\lambda_{\text{true}} \in R_\alpha(X)) = \Pr(\text{cc}(\lambda_{\text{true}}, X) \leqslant \alpha) \tag{A.14}$$

Combined with the fact that $\text{cc}(\lambda_{\text{true}}, X)$ is uniformly distributed on $[0, 1]$, it now follows that the $R_\alpha(X)$ is a confidence set with confidence level α. This suggests that in practice one way to test the validity of a confidence curve is to obtain a large number m of independent realizations of the sample X, say $x^{(1)}, x^{(2)}, \ldots, x^{(m)}$ from a distribution with known $\lambda = \lambda_0$, and check that

$$\frac{1}{m}\sum_{j=1}^{m} \left[\!\left[\text{cc}(\lambda_0, x^{(j)}) \leqslant \alpha\right]\!\right] - \alpha \tag{A.15}$$

goes to zero as m increases.

In the case of a change point in a time series of length n, the property of interest is the location τ of the change which is an element of the set $\{1, 2, \ldots, n-1\}$. This τ takes the role of λ. There is only a finite number of subsets of $V = \{1, 2, \ldots, n-1\}$, so only a finite number of possible choices for $R_\alpha(x)$. Moreover, the sets $R_\alpha(x)$ derived from a confidence curve are nested, which further limits the number of available sets. As each subset will correspond to one confidence level α, and there are infinitely many confidence levels, the best we can hope to achieve is $\Pr\{\text{cc}(\lambda_{\text{true}}, X) \leqslant \alpha\} \approx \alpha$, so

$$\Pr\{\text{cc}(\lambda_{\text{true}}, X) \leqslant \alpha\} - \alpha \tag{A.16}$$

cannot be zero for all α, and therefore (A.15) cannot go to zero for all α, but should be small.

Please keep in mind, that confidence curves represent confidence in an outcome, and this is not the same as probability. That being said, if we have a small set with high confidence, then the particular sample strongly suggests that we look for the change point in that set.

One way to illustrate this relation is the following. If the series contains a change point, and a set S of approximately $\alpha(n-1)$ points is selected at random from $\{1, 2, \ldots, n-1\}$, then the probability that the actual change point τ_{true} lies in that set is approximately α. This suggests that a set $R_\alpha(x)$ that contains more points than $\alpha(n-1)$ indicates large uncertainty at that confidence level, while sets $R_\alpha(x)$ that are much smaller correspond to low uncertainty at that confidence level. In such a way the size of the sets $R_\alpha(x_{\text{obs}})$ for the observed sample x_{obs} can be linked to the uncertainty in the location of the change.

Appendix B. From confidence curves based on parametric likelihood to confidence curves based on approximate empirical likelihood

To show the relations between the method proposed by Cunen et al. (2018) and the method proposed in this study, it is necessary to make a few intermediate steps. The first step is to relate the deviance function to the log-likelihood ratio.

B.1. The log-likelihood ratio

It is useful to start with the log-likelihood ratio for the parametric case, which is also used in change point detection (Csörgö and Horváth, 1997). The likelihood ratio for the AMOC problem is given by

$$\Lambda_\tau(y) = \frac{\sup_{\theta,\zeta} \prod_{i=1}^{n} f(y_i;\theta,\zeta)}{\sup_{\theta_L,\theta_R,\zeta} \prod_{i=1}^{\tau} f(y_i;\theta_L,\zeta) \prod_{i=\tau+1}^{n} f(y_i;\theta_R,\zeta)} \tag{B.1}$$

Note, that the numerator represents the null hypothesis of no change, and the denominator represents one of $n-1$ alternative hypotheses, namely the one where the change occurs at τ.

Csörgö and Horváth (1997) state that it is now natural to consider

$$Z_n(y) = \max_{\tau=1,2,\ldots,n-1} -2\log\Lambda_\tau(y) \tag{B.2}$$

and reject the null hypothesis of no change when this is large. From (2) and (B.1) it follows that

$$-2\log\Lambda_\tau(y) = 2\left(\sup_{\theta_L,\theta_R,\zeta} \ell(\tau,\theta_L,\theta_R,\zeta;y) - \sup_{\theta,\zeta} \ell(\tau,\theta,\theta,\zeta;y)\right) \tag{B.3}$$

or, using (3),

$$-2\log\Lambda_\tau(y) = 2\left(\ell_{\text{prof}}(\tau;y) - \sup_{\theta,\zeta} \ell(\tau,\theta,\theta,\zeta;y)\right) \tag{B.4}$$

The value of

$$\sup_{\theta,\zeta} \ell(\tau,\theta,\theta,\zeta;y) \tag{B.5}$$

is independent of τ, so

$$-2\log\Lambda_{\hat\tau(y)}(y) + 2\log\Lambda_\tau(y) = 2\left(\ell_{\text{prof}}(\hat\tau(y);y) - \sup_{\theta,\zeta} \ell(\hat\tau(y),\theta,\theta,\zeta;y)\right) - 2\left(\ell_{\text{prof}}(\tau;y) - \sup_{\theta,\zeta} \ell(\tau,\theta,\theta,\zeta;y)\right) = 2\left(\ell_{\text{prof}}(\hat\tau(y);y) - \ell_{\text{prof}}(\tau;y)\right) = D(\tau,y) \tag{B.6}$$

One problem that needs to be addressed is that for τ close to the start or end of the series, the optimization problem may not have a solution. It is therefore necessary to avoid calculations near the start or end of the series.

B.2. Confidence curves based on the empirical likelihood ratio

The parametric form of the distribution underlying an environmental time series is not known, therefore the approach based on the profile likelihood always involves a choice of distribution family. There is an alternative: an approach based on the empirical likelihood (Owen, 1988; Owen, 1990). For a change point in the mean, such an approach is presented, for instance, in Zou et al. (2007) and Shen (2013). In Hall and La Scala (1990) the empirical likelihood for a distribution property λ is defined as follows. Suppose $X_1, X_2, \ldots, X_n$ form a random sample of size n. To define the empirical likelihood we need the set

$$\mathscr{S} = \left\{ p \in [0,1]^n : \sum_{i=1}^{n} p_i = 1 \right\} \tag{B.7}$$

of all probability mass functions on the set $\{1, 2, \ldots, n\}$. Now suppose $\hat\lambda(p, x)$ is an estimator for λ when $x_1, x_2, \ldots, x_n$ is a sample from a *discrete* distribution, where x_i has probability of occurrence p_i. The *empirical likelihood* L for a given value λ_0 of λ is defined as

$$L(\lambda_0, x) = \max_{p\in\mathscr{S}} \left\{ \prod_{i=1}^{n} p_i : \hat\lambda(p,x) = \lambda_0 \right\} \tag{B.8}$$

The empirical likelihood ratio is derived by dividing L by

$$\max_{p\in\mathscr{S}} \prod_{i=1}^{n} p_i \tag{B.9}$$

which is achieved at $p_1 = p_2 = \ldots = p_n = 1/n$, this follows from the arithmetic geometric mean inequality. Therefore the *empirical likelihood ratio* is

$$\Lambda_{\text{emp}}(\lambda_0, x) = \max_{p\in\mathscr{S}} \left\{ \prod_{i=1}^{n} np_i : \hat\lambda(p,x) = \lambda_0 \right\} \tag{B.10}$$

Example 2. Suppose the distribution property of interest is the mean. In that case

$$\hat\lambda(p,x) = \sum_{i=1}^{n} p_i x_i \tag{B.11}$$

and

$$L(\lambda_0, x) = \max_{p\in\mathscr{S}} \left\{ \prod_{i=1}^{n} p_i : \sum_{i=1}^{n} p_i x_i = \lambda_0 \right\} \tag{B.12}$$

with likelihood ratio

$$\max_{p\in\mathscr{S}} \left\{ \prod_{i=1}^{n} np_i : \sum_{i=1}^{n} p_i x_i = \lambda_0 \right\} \tag{B.13}$$

For the change point problem with a change in the mean, Zou et al. (2007) proposed the empirical likelihood ratio

$$\Lambda_{\text{emp}}(\tau;y) = \frac{\sup_{p\in\mathscr{S}_\tau} \{\prod_{i=1}^{n} p_i : \sum_{i=1}^{\tau} p_i y_i = \sum_{i=\tau+1}^{n} p_i y_i\}}{\sup_{p\in\mathscr{S}_\tau} \prod_{i=1}^{n} p_i} \tag{B.14}$$

where

$$\mathscr{S}_\tau = \left\{ p \in [0,1]^n : \sum_{i=1}^{\tau} p_i = 1 \,;\, \sum_{i=\tau+1}^{n} p_i = 1 \right\} \tag{B.15}$$

As in the parametric case, the numerator represents the null hypothesis of no change, and the denominator represents one of $n-1$ alternative hypotheses, namely, the one where a change occurs at τ.

The optimization problem in the denominator of (B.14) has as its solution $p_1 = p_2 = \ldots, p_\tau = 1/\tau$ and $p_{\tau+1} = p_{\tau+2} = \ldots, p_n = 1/(n-\tau)$. Note, that the optimization problem in the numerator is solvable only if the convex hull of $\{y_1, y_2, \ldots, y_k\}$ and $\{y_{k+1}, y_{k+2}, \ldots, y_n\}$ overlap.

They define the empirical log-likelihood ratio as

$$\ell_{\text{emp}}(\tau;y) = -2\log\Lambda_{\text{emp}}(\tau;y) \tag{B.16}$$

and their statistic is

$$Z_{\text{emp}} = \max_{1\leqslant\tau<n} \ell_{\text{emp}}(\tau;y) \tag{B.17}$$

The link between $2\log\Lambda_\tau(y)$ and $D(\tau, y)$ in the parametric likelihood case now suggests that it might be possible to build a confidence curve by taking $\hat{\tau}_{\text{emp}}(y)$ to be the value of τ for which $\ell_{\text{emp}}(\tau;y)$ attains the maximum value, and then defining an *empirical deviation function*

$$D_{\text{emp}}(\tau, y) = 2(\ell_{\text{emp}}(\hat{\tau}_{\text{emp}}(y);y) - \ell_{\text{emp}}(\tau;y)) \tag{B.18}$$

But this leaves a problem: determining the distribution function $K_{\text{emp},\tau}$ of $D_{\text{emp}}(\tau, Y)$ that is the values of

$$K_{\text{emp},\tau}(r) = \Pr(D_{\text{emp}}(\tau, Y) < r) \tag{B.19}$$

If we approximate $K_{\text{emp},\tau}$ by repeated sampling, then this involves solving many optimization problems that may or may not have a solution. This makes it attractive to search for an alternative to ℓ_{emp}. Shen (2013) derived the following approximation formula for the logarithm of the empirical likelihood ratio for scalar y_i

$$-2\log\Lambda_{\text{emp}}(\tau;y) = \frac{\tau(n-\tau)}{n} \frac{\left(\frac{1}{\tau}\sum_{i=1}^{\tau} y_i - \frac{1}{n-\tau}\sum_{i=\tau+1}^{n} y_i\right)^2}{\frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \frac{1}{n}\sum_{j=1}^{n} y_j\right)^2} + O_p(n_{\text{tr}}^{-1/2}) \tag{B.20}$$

for $n_{\text{tr}} < \tau < n - n_{\text{tr}}$ where n_{tr} tends to infinity as n tends to infinity. The approximate formula holds under the assumption that the higher-order moments of Y exists: $E\|Y\|^3 < \infty$ ($\|\cdot\|$ is the Euclidean norm). The term $O_p(n_{\text{tr}}^{-1/2})$ is present because the approximation does not hold for τ near the start or the end of the series. We will use n_{tr} as given by (1). For the distributions used in the tests in this paper the condition on the third moment is always satisfied for the log-normal and the gamma distribution; for Fréchet as parametrized in (D.5) it holds because $\xi = 0.139 < 1/3$.

We felt it would be interesting to see what would happen if we introduced the approximation ℓ_{apn} of ℓ_{emp} given by

$$\ell_{\text{apn}}(\tau;y) = \frac{\tau(n-\tau)}{n} \frac{\left(\frac{1}{\tau}\sum_{i=1}^{\tau} y_i - \frac{1}{n-\tau}\sum_{i=\tau+1}^{n} y_i\right)^2}{\frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \frac{1}{n}\sum_{j=1}^{n} y_j\right)^2}$$

One reason to assume that this might work is that a similar formula is given as the basis for a test statistic for change point detection in Csörgö and Horváth (1997, page 85).

Appendix C. Similarity index between randomly generated confidence curves

If different methods are applied to the same data, it can be of interest to compare the resulting confidence curves. This is of special interest for the case of real data. For the real time series, we wish to know whether the methods agree or not: that is how similar the confidence curves are. As a starting point, we take the Jaccard index, see Schubert and Telcs (2014) who in turn refer to Jaccard (1901). For two sets, $A = \{a_1, a_2, \ldots, a_{n_A}\}$ and $B = \{b_1, b_2, \ldots, b_{n_B}\}$, the *Jaccard index* is given by

$$J\left(A, B\right) = \frac{\#(A \cap B)}{\#(A \cup B)} \tag{C.1}$$

where $\#S$ denotes the number of elements in a finite set S.

For two confidence curves cc(.,.) and cc′(.,.) and a fixed α this index can serve to compare the sets $R_\alpha = \{\tau\text{: } \mathrm{cc}(\tau, y_{\mathrm{obs}}) \leqslant \alpha\}$ and $R'_\alpha = \{\tau\text{: } \mathrm{cc}'(\tau, y_{\mathrm{obs}}) \leqslant \alpha\}$ as follows

$$R_\alpha \cap R'_\alpha = \{\tau\text{: } \max(\mathrm{cc}(\tau, y_{\mathrm{obs}}), \mathrm{cc}'(\tau, y_{\mathrm{obs}})) \leqslant \alpha\} \tag{C.2}$$

and

$$R_\alpha \cup R'_\alpha = \{\tau\text{: } \min(\mathrm{cc}(\tau, y_{\mathrm{obs}}), \mathrm{cc}'(\tau, y_{\mathrm{obs}})) \leqslant \alpha\} \tag{C.3}$$

One way to extend this to the entire curve is to integrate over α. For $R_\alpha \cap R'_\alpha$ this results in

$$\begin{aligned}
&\textstyle\sum_{\alpha=0}^{1} \#(R_\alpha \cap R'_\alpha) = \int_0^1 \#(R_\alpha \cap R'_\alpha)d\alpha = \\
&\int_0^1 \#\{\tau\text{: } \max(\mathrm{cc}(\tau, y_{\mathrm{obs}}), \mathrm{cc}'(\tau, y_{\mathrm{obs}})) \leqslant \alpha\}d\alpha \\
&= \int_0^1 \textstyle\sum_{\tau=1}^{n} [\![\max(\mathrm{cc}(\tau, y_{\mathrm{obs}}), \mathrm{cc}'(\tau, y_{\mathrm{obs}})) \leqslant \alpha]\!] d\alpha \\
&= \textstyle\sum_{\tau=1}^{n} \int_0^1 [\![\max(\mathrm{cc}(\tau, y_{\mathrm{obs}}), \mathrm{cc}'(\tau, y_{\mathrm{obs}})) \leqslant \alpha]\!] d\alpha
\end{aligned}$$

where the summation could be moved through the integral because the individual terms in the sum under the integral are integrable, so linearity of integration could be used. A single integral in this expression can be rewritten as follows

$$\begin{aligned}
&\int_0^1 [\![\max(\mathrm{cc}(\tau, y_{\mathrm{obs}}), \mathrm{cc}'(\tau, y_{\mathrm{obs}})) \leqslant \alpha]\!] d\alpha \\
&\quad = \int_{\max(\mathrm{cc}(\tau, y_{\mathrm{obs}}), \mathrm{cc}'(\tau, y_{\mathrm{obs}}))}^{1} d\alpha \\
&= 1 - \max(\mathrm{cc}(\tau, y_{\mathrm{obs}}), \mathrm{cc}'(\tau, y_{\mathrm{obs}}))
\end{aligned}$$

and

$$1 - \max(a, b) = \min(1 - a, 1 - b)$$

so

$$\int_0^1 \#\left(R_\alpha \cap R'_\alpha\right) d\alpha = \sum_{\tau=1}^{n} \min(1 - \mathrm{cc}(\tau, y_{\mathrm{obs}}), 1 - \mathrm{cc}'(\tau, y_{\mathrm{obs}}))$$

A similar approach can be applied to the denominator, and we get the following *similarity index*

$$\widetilde{J} = \frac{\sum_{\tau=1}^{n} \min(1 - \mathrm{cc}(\tau, y_{\mathrm{obs}}), 1 - \mathrm{cc}'(\tau, y_{\mathrm{obs}}))}{\sum_{\tau=1}^{n} \max(1 - \mathrm{cc}(\tau, y_{\mathrm{obs}}), 1 - \mathrm{cc}'(\tau, y_{\mathrm{obs}}))} \tag{C.4}$$

which will be used to compare the similarity of pairs of confidence curves. It is similar to the Ružička index (Schubert and Telcs, 2014). This index is one for identical curves and smaller than one for curves that differ.

To get an impression of how the value of similarity index (C.4) relates to similarity, the following experiment was performed. For n = 10, 20, …,100 we generated 5000 pairs of i.i.d. samples of size n drawn from a uniform distribution on [0, 1]. While such a sample may not bear much resemblance to a confidence curve, they share domain and range. The distribution of $\widetilde{J}$ for these pairs provides some indication of the range of values of $\widetilde{J}$ that may occur for curves that were constructed to be unrelated to each other.

Fig. C.11 shows the cumulative frequency distribution of $\widetilde{J}$ for different sample lengths n. For a larger n, the distribution of $\widetilde{J}$ approaches a step function. Fig. C.12 shows the 90%, 95%, and 99% quantiles for $\widetilde{J}$ as a function of sample size. To aid in the interpretation of Fig. C.12, Table C.4 is provided. For instance, if we have two random data sets with sample length n = 100, then the 95% quantile of the similarity index $\widetilde{J}$ is 0.55. The actual similarity index $\widetilde{J}_{\mathrm{actual}}$ between two confidence curves with a sample length n = 100 follows from (12). If $\widetilde{J}_{\mathrm{actual}}$ is higher than 0.55, then we can be reasonably confident that the curves are similar.

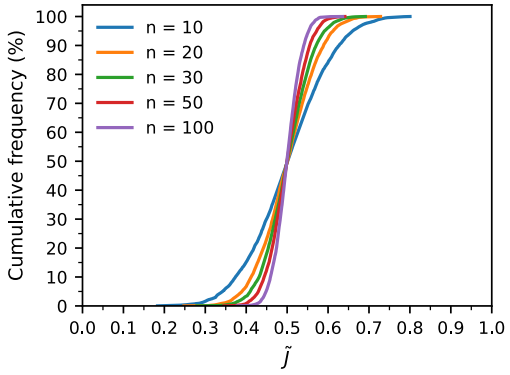

Fig. C.11. Cumulative frequency for values of similarity index $\widetilde{J}$ for pairs of randomly generated curves for different sample lengths.

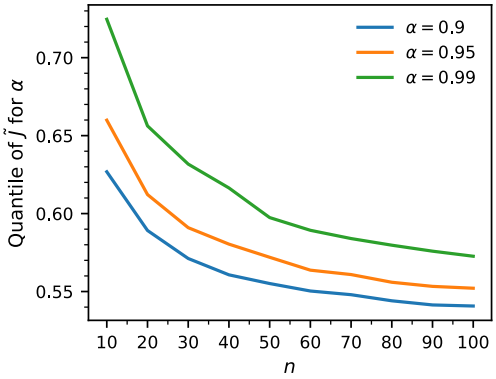

Fig. C.12. Quantiles for values of similarity index $\tilde{J}$ for pairs of randomly generated curves for different sample lengths.

Table C.4
Quantiles of similarity index $\tilde{J}$ for random pairs of curves for different sample lengths.

$\alpha \backslash n$	10	20	30	40	50	60	70	80	90	100
0.9	0.62	0.59	0.57	0.56	0.55	0.55	0.55	0.55	0.54	0.54
0.95	0.66	0.61	0.59	0.58	0.57	0.57	0.56	0.56	0.55	0.55
0.99	0.72	0.66	0.63	0.61	0.6	0.59	0.59	0.58	0.57	0.57

Appendix D. Parametric distribution functions for log-normal, gamma and Fréchet distributions

The probability density functions of the three distributions: log-normal, gamma and Fréchet are given together with the relation between the distribution parameters and the mean and the standard deviation of the distributions.

- For the log-normal distribution (LN), the pdf is a

$$f(x, \varphi, \psi) = \begin{cases} 0 & x \le 0 \\ \frac{1}{x\psi\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \varphi)^2}{2\psi^2}\right) & x > 0 \end{cases} \tag{D.1}$$

The parameters are linked to the mean μ and standard deviation σ by

$$\varphi = \ln\mu - \frac{1}{2}\ln\left(\frac{\sigma^2}{\mu^2} + 1\right); \psi = \sqrt{\ln\left(\frac{\sigma^2}{\mu^2} + 1\right)} \tag{D.2}$$

- For the gamma distribution (G), the pdf is

$$f(x, k, \theta) = \begin{cases} 0 & x \le 0 \\ \frac{1}{\Gamma(k)\theta^k} x^{k-1} \exp\left(-\frac{x}{\theta}\right) & x > 0 \end{cases} \tag{D.3}$$

The parameters are linked to the mean and standard deviation by

$$k = \frac{\mu^2}{\sigma^2}; \theta = \frac{\sigma^2}{\mu} \tag{D.4}$$

- For the Fréchet distribution (F) in the Generalized Extreme Value parametrization, the pdf is

$$f\left(x, m, s, \xi\right) = \begin{cases} 0 & x \leqslant m \\ \frac{\left(1 + \xi\frac{x-m}{s}\right)^{-1-\frac{1}{\xi}}}{s} \exp\left(-\left(1 + \xi\frac{x-m}{s}\right)^{-\frac{1}{\xi}}\right) & x > m \end{cases} \tag{D.5}$$

where we fix the shape parameter to $\xi = 0.139$, a value recommended by Ragulina and Reitan (2017) for a global data set of extreme precipitation data. As $0.139 < 1/3$ the third moment exists. The parameters are linked to the mean and standard deviation by

$$m = \mu - \sigma\frac{\Gamma(1-\xi) - 1}{\sqrt{\Gamma(1-2\xi) - (\Gamma(1-\xi))^2}} \tag{D.6}$$

$$s = \sigma \frac{\xi}{\sqrt{\Gamma(1-2\xi) - (\Gamma(1-\xi))^2}} \tag{D.7}$$

References

Beaulieu, C., Chen, J., Sarmiento, J.L., Mar 2012. Change-point analysis as a tool to detect abrupt climate variations. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 370 (1962), 1228–1249.

Birnbaum, A., 1961. Confidence curves: an omnibus technique for estimation and testing statistical hypotheses. Journal of the American Statistical Association 56, 246–249.

Blöschl, G., Bierkens, M.F., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., Kirchner, J. W., McDonnell, J.J., Savenije, H.H.G., Sivapalan, M., Stumpp, C., Toth, E., Volpi, E., Carr, G., Lupton, C., Salinas, J., Széles, B., et al., A.V., 2019. Twenty-three unsolved problems in hydrology (UPH) – a community perspective. Hydrological Sciences Journal 64 (10), 1141–1158.

Brodsky, E., Darkhovsky, B.S., 1993. Nonparametric Methods in Change Point Problems. Springer Science & Business Media.

Chen, J., Gupta, A.K., 2012. Parametric Statistical Change Point Analysis: With Applications to Genetics, Medicine, and Finance, 2nd Edition. Birkhäuser/Springer, New York.

Coles, S.G., Tawn, J.A., 1996. A Bayesian analysis of extreme rainfall data. Journal of the Royal Statistical Society. Series C (Applied Statistics) 45 (4), 463–478.

Conte, L.C., Bayer, D.M., Bayer, F.M., Jul 2019. Bootstrap Pettitt test for detecting change points in hydroclimatological data: case study of Itaipu hydroelectric plant. Brazil. Hydrological Sciences Journal 64 (11), 1312–1326.

Csörgö, M., Horváth, L., 1997. Limit Theorems in Change-Point Analysis, vol. 18 John Wiley Sons Inc.

Cunen, C., Hermansen, G., Hjort, N.L., 2018. Confidence distributions for change-points and regime shifts. Journal of Statistical Planning and Inference 195, 14–34.

Donat, M.G., Lowry, A.L., Alexander, L.V., O'Gorman, P.A., Maher, N., 2017. Addendum: More extreme precipitation in the world's dry and wet regions. Nature Climate Change 7 (2), 154–158.

Gocic, M., Trajkovic, S., 2013. Analysis of changes in meteorological variables using Mann-Kendall and Sen's slope estimator statistical tests in Serbia. Global and Planetary Change 100, 172–182.

Gurevich, G., Vexler, A., 2010. Retrospective change point detection: From parametric to distribution free policies. Communications in Statistics – Simulation and Computation 39 (5), 899–920.

Haddeland, I., Heinke, J., Biemans, H., Eisner, S., Flörke, M., Hanasaki, N., Konzmann, M., Ludwig, F., Masaki, Y., Schewe, J., Stacke, T., Tessler, Z.D., Wada, Y., Wisser, D., Dec 2013. Global water resources affected by human interventions and climate change. Proceedings of the National Academy of Sciences 111 (9), 3251–3256.

Hall, P., La Scala, B., 1990. Methodology and algorithms of empirical likelihood. International Statistical Review/ Revue Internationale de Statistique 58 (2), 109–127.

Hall, P., Martin, M., 1988. On the bootstrap and two-sample problems. Australian Journal of Statistics 30A (1), 179–192.

Hawkins, D.M., Zamba, K.D., 2005. A change-point model for a shift in variance. Journal of Quality Technology 37 (1), 21–31.

Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H., Kanae, S., 2013. Global flood risk under climate change. Nature Climate Change 3 (9), 816–821.

Iverson, K.E., 1962. A Programming Language. John Wiley & Sons Inc, New York, NY, USA.

Jaccard, P., 1901. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. Bulletin del la Société Vaudoise des Sciences Naturelles 37, 547–579.

Jandhyala, V.K., Fotopoulos, S.B., You, J., Feb 2010. Change-point analysis of mean annual rainfall data from Tucumán, Argentina. Environmetrics 21 (7–8), 687–697.

Killick, R., Eckley, I.A., Ewans, K., Jonathan, P., 2010. Detection of changes in variance of oceanographic time-series using changepoint analysis. Ocean Engineering 37 (13), 1120–1126.

Knuth, D.E., May 1992. Two notes on notation. American Mathematical Monthly 99 (5), 403–422.

Kundzewicz, Z.W., Robson, A.J., 2004. Change detection in hydrological records – a review of the methodology/ Revue méthodologique de la détection de changements dans les chroniques hydrologiques. Hydrological Sciences Journal 49 (1), 7–19.

Lee, S., Ha, J., Na, O., Na, S., 2003. The cusum test for parameter change in time series models. Scandinavian Journal of Statistics 30 (4), 781–796.

Lehmann, J., Coumou, D., Frieler, K., Oct 2015. Increased record-breaking precipitation events under global warming. Climatic Change 132 (4), 501–515.

McMillan, H., Montanari, A., Cudennec, C., Savenije, H., Kreibich, H., Krueger, T., Liu, J., Mejia, A., Loon, A.V., Aksoy, H., Baldassarre, G.D., Huang, Y., Mazvimavi, D., Rogger, M., Sivakumar, B., Bibikova, T., Castellarin, A., Chen, Y., Finger, D., Gelfan, A., Hannah, D.M., Hoekstra, A.Y., Li, H., Maskey, S., Mathevet, T., Mijic, A., Acuña, A.P., Polo, M.J., Rosales, V., Smith, P., Viglione, A., Srinivasan, V., Toth, E., van Nooyen, R., Xia, J., Mar 2016. Panta Rhei 2013–2015: Global perspectives on hydrology, society and change. Hydrological Sciences Journal 61 (7), 1174–1191.

Molchanov, I., 2017. Theory of Random Sets, second ed. Springer-Verlag, London.

Montanari, A., Young, G., Savenije, H.H.G., Hughes, D., Wagener, T., Ren, L.L., Koutsoyiannis, D., Cudennec, C., Toth, E., Grimaldi, S., Blöschl, G., Sivapalan, M., Beven, K., Gupta, H., Hipsey, M., Schaefli, B., Arheimer, B., Boegh, E., Schymanski, S.J., Baldassarre, G.D., Yu, B., Hubert, P., Huang, Y., Schumann, A., Post, D.A., Srinivasan, V., Harman, C., Thompson, S., Rogger, M., Viglione, A., McMillan, H., Characklis, G., Pang, Z., Belyaev, V., Jul 2013. Panta Rhei—Everything Flows: Change in hydrology and society—the IAHS scientific decade 2013–2022. Hydrological Sciences Journal 58 (6), 1256–1275.

Murphy, S.A., van der Vaart, A.W., 2000. On profile likelihood. Journal of the American Statistical Association 95 (450), 449–485.

Owen, A.B., 1988. Empirical likelihood ratio confidence intervals for a single functional. Biometrika 75 (2), 237–249.

Owen, A., 1990. Empirical likelihood ratio confidence regions. Annals of Statistics 18 (1), 90–120.

Perreault, L., Bernier, J., Bobée, B., Parent, E., 2000. Bayesian change-point analysis in hydrometeorological time series. Part 1. The normal model revisited. Journal of Hydrology 235 (3), 221–241.

Pettitt, A.N., 1979. A non-parametric approach to the change-point problem. Journal of the Royal Statistical Society. Series C (Applied Statistics), 126–135.

Ragulina, G., Reitan, T., 2017. Generalized extreme value shape parameter and its nature for extreme precipitation using long time series and the Bayesian approach. Hydrological Sciences Journal 62 (6), 863–879.

Reeves, J., Chen, J., Wang, X.L., Lund, R., Lu, Q.Q., 2007. A review and comparison of changepoint detection techniques for climate data. Journal of Applied Meteorology and Climatology 46 (6), 900–915.

Renard, B., Kavetski, D., Kuczera, G., Thyer, M., Franks, S.W., 2010. Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. Water Resources Research 46 (W05521), 1–22.

Schubert, A., Telcs, A., 2014. A note on the Jaccardized Czekanowski similarity index. Scientometrics 98 (2), 1397–1399.

Schweder, T., Hjort, N.L., 2016. Confidence, likelihood, probability. In: Statistical Inference with Confidence Distributions. Cambridge University Press, Cambridge.

Shen, G., 2013. On empirical likelihood inference of a change-point. Statistics & Probability Letters 83 (7), 1662–1668.

Tamaddun, K., Kalra, A., Ahmad, S., 2016. Identification of streamflow changes across the continental united states using variable record lengths. Hydrology 3 (2).

Wasserstein, R.L., Schirm, A.L., Lazar, N.A., 2019. Moving to a world beyond "$p < 0.05$". The American Statistician 73 (sup1), 1–19.

Wu, W.B., Woodroofe, M., Mentz, G., 2001. Isotonic regression: Another look at the changepoint problem. Biometrika 88 (3), 793–804.

Zhou, C., van Nooijen, R., Kolechkina, A., Hrachowitz, M., 2019. Comparative analysis of nonparametric change-point detectors commonly used in hydrology. Hydrological Sciences Journal 64 (14), 1690–1710.

Zou, C., Liu, Y., Qin, P., Wang, Z., 2007. Empirical likelihood ratio test for the change-point problem. Statistics & Probability Letters 77 (4), 374–382.