Capturing the uncertainty about a sudden change in the properties of time series with confidence curves

Zhou, Changrang; van Nooijen, Ronald; Kolechkina, Alla

DOI
10.1016/j.jhydrol.2023.129092

Publication date
2023

Document Version
Final published version

Published in
Journal of Hydrology

Citation (APA)
Zhou, C., van Nooijen, R., & Kolechkina, A. (2023). Capturing the uncertainty about a sudden change in the properties of time series with confidence curves. *Journal of Hydrology*, *617*, Article 129092. https://doi.org/10.1016/j.jhydrol.2023.129092

Journal of Hydrology 617 (2023) 129092

Contents lists available at ScienceDirect

Journal of Hydrology

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

Research papers

Capturing the uncertainty about a sudden change in the properties of time series with confidence curves

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

[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 Fateh Chebana, Associate Editor.

Keywords:
Confidence curves
Pseudolikelihood
Likelihood
Change point detection
Structural break
L-moments
Method of moments
Uncertainty

ABSTRACT

The representation of uncertainty in results is an important aspect of statistical techniques in hydrology and climatology. Hypothesis tests and point estimates are not well suited for this purpose. Other statistical tools, such as confidence curves, are better suited to represent uncertainty. Therefore three parametric methods to construct confidence curves for the location of a sudden change in the properties of a time series, a change point (CP), are analyzed for three distributions: log-normal, gamma, and Gumbel. Two types of change are considered: a change in the mean and a change in the standard deviation. A question that confidence curves do not answer is how likely the null hypothesis of 'no change' is. A possible statistic to help answer this question, denoted by Un, is introduced and analyzed. It is compared to the statistic that underlies the Pettitt test. All methods perform well in terms of coverage and confidence set size. One method is based on the profile likelihood for a CP, the other two, first defined in this article, on the pseudolikelihood for a CP. The main advantage of the pseudolikelihood over the profile likelihood lies in the much lower computational cost. The confidence curves generated by the three methods are very similar. In a limited test on time series of measurements found in the literature, the methods gave results that largely matched those reported elsewhere. Some results are also given for an order one autoregressive series with a lognormal marginal distribution.

1. Introduction

Today, the need to take into account climate variability and the results of human interventions in water management and hydrology seems clear (Kolokytha et al., 2017; Teegavarapu, 2018). To do so, it is necessary to combine statistical information, obtained from hydrological and climatological time series, with investigations of how the natural variations in the behavior of the physical system and human alterations of that system could result in changes in those time series, and link the changes suggested by statistics to physical causes (Tao et al., 2011). While this will often be a search for trends or periodic changes, the time series in question must also be tested for abrupt changes, either to find real changes (Tao et al., 2011; Harrigan et al., 2014), or to see whether it is necessary to split a series into two parts for further analysis (Cong et al., 2017). Beaulieu et al. (2012) mention that an abrupt change in the statistical properties of a time series could signal an undocumented change in the measurement procedure. A general overview of change detection is given in Kundzewicz and Robson (2004).

In this article the emphasis is on abrupt changes. But please keep in mind that, for instance, the initial filling of a reservoir may take several years, so it may look as a trend on a daily scale and as a jump in the time series of annual maximum flows. There have been many publications on the detection of an abrupt change, or change point (CP), in hydrology and climatology (Beaulieu et al., 2012; Conte et al., 2019; Xie et al., 2014; Xiong and Guo, 2004). Theoretical work on CP detection in general was done, for example, by Pettitt (1979), Chen and Gupta (2001, 2011), or Brodsky and Darkhovsky (2013).

There are many statistical tools that can be used to detect the presence of CPs. Ideally, such a tool should provide information on the uncertainty in the location of the CP. The traditional tests, such as the one presented in Pettitt (1979), focus on the acceptance or rejection of the null hypothesis that there is no CP at a given significance level, a form of *Null-Hypothesis Significance Testing* (NHST). If the null hypothesis is rejected, then the CP is assumed to be at the location that results in the largest value for the test statistic. Such a point estimate gives no indication of the probability that this is the true CP location. In the literature the emphasis tends to be on determining whether or not there is a CP and delivering as point estimate of the CP location. Attention for the uncertainty in its location is limited. Based on Web of Science search it seems that less than 10% of the papers dealing

* Corresponding author.
E-mail address: r.r.p.vannooyen@tudelft.nl (R. van Nooijen).

https://doi.org/10.1016/j.jhydrol.2023.129092
Received 5 August 2021; Received in revised form 24 October 2022; Accepted 5 January 2023
Available online 13 January 2023

with CPs uses a method that does more than deliver a point estimate. This also holds when only papers dealing with environmental times series are considered. Of those relatively rare papers that deal with the uncertainty in the CP location most use Bayesian methods. Examples of Bayesian methods are easier to find, see, for instance, Perreault et al. (1999, 2000) or, for a hierarchical Bayesian version, Belisle et al. (1998), Chu and Zhao (2004). Bayesian methods require prior distributions for all parameters, and they can be computationally intensive. A rare example of a frequentist paper that looks into uncertainty is Hušková and Kirch (2008) where a bootstrap based method is used to construct confidence intervals for CPs.

Strictly speaking, the methods discussed in this article serve a different purpose than NHST, and they are not designed to reject or not reject the null hypothesis. However, experiments showed that from the confidence curves a number may be calculated that may serve the same purpose as the original p-value, namely to indicate data 'worthy of a second look' (Nuzzo, 2014). This is of interest in situations where a large set of time series is studied and it would not be feasible to analyze all confidence curves by eye. Different thresholds for that value could then be used to separate the set into three groups: curves that provide clear information on the location of a CP, curves that provide no information on the location of a CP, and curves that need to be investigated further. A possible candidate for such a number is the function Un introduced in this paper.

As in all of statistics, there are parametric and nonparametric methods. The nonparametric methods avoid the choice of a distribution for the time series, but they tend to specialize in detecting changes in either the mean or the standard deviation, not both at the same time (Eastwood, 1993). The parametric tests can look for changes in all parameters of the underlying distribution. For hydrological, meteorological, or climatological time series the distribution may or may not be known. Sheskin (2003, pp. 97–98) states that while parametric tests generally provide a more powerful test of the alternative hypothesis, they may lose that advantage if the assumptions underlying the test are violated. An example of a nonparametric detection method using confidence curves can be found in Zhou et al. (2020). It therefore makes sense to examine both types of CP tests.

The current article examines two parametric approaches. While the emphasis is on detection of changes in the mean, additional experiments showed that the same algorithm is equally sensitive to changes in the standard deviation. Both parametric approaches belong to the domain of parametric statistics and represent uncertainty by a confidence curve. Both use a likelihood where the location of the CP, the distribution parameters to the left of the CP, and the distribution parameters to the right of the CP are free variables. One then introduces a profile likelihood, the other introduces a pseudolikelihood.

A method based on the first approach can be found in Cunen et al. (2018) where it is called 'method B'. Method B is based on a calculation of the log-likelihood of the time series for all possible CP locations. In this calculation, the parameters of the distribution to the left and to the right of the CP are so-called 'nuisance parameters'; their values are needed to calculate the log-likelihood, but are not of intrinsic interest. A profile log-likelihood approach is used to calculate the log-likelihoods. The resulting log-likelihood values for the potential CPs are used to construct a deviance function. Next, *Monte Carlo* (MC) simulation is used to approximate the distribution of the values of the deviance function for each potential CP. This approximate distribution is then used to build a confidence curve. In the remainder of this study, this method will be referred to as *Confidence curve based on Maximum Likelihood parameter estimation* (CML). A potential problem with this method is that it is very computationally intensive. Even in the case of just one CP, two optimizations of a log-likelihood need to be done for each possible CP location to determine the profile log-likelihood. Moreover, a MC simulation is needed to determine an approximate distribution for each possible CP location. This MC simulation needs to repeat the profile log-likelihood calculation as often as is needed to obtain an approximate distribution. As shown in Appendix E, this leads to a computational complexity linearly proportional to the number of samples in the MC simulation and proportional to the cube of the time series length. Run-times on a desktop workstation may take hundreds of seconds for a single sample. Removal of the maximum likelihood optimization can reduce the computational cost considerably; this is the chief reason to examine the second approach.

The current study presents two methods based on the second approach which uses pseudolikelihood. More information on pseudolikelihood can be found in, for example, Gong and Samaniego (1981). Distribution parameters are estimated by the *Method of Moments* (MoM) or *L-Moments* (LMo); this reduces the computational cost of the likelihood calculations. Moreover, fast code for these methods is often easier to obtain than for log-likelihood optimization. These methods will be referred to as *Confidence curve based on Method of Moments parameter estimation* (CMoM) and *Confidence curve based on L-Moments parameter estimation* (CLMo), respectively. As the experimental results for CMoM and CLMo were very similar, only CML and CLMo results are reported in this study.

To verify that CLMo (CMoM) works, it should be demonstrated that the results obtained are similar to those of CML. As hydrological time series are relatively short, asymptotic results on the performance of the methods may not be valid. Therefore, it is necessary to generate statistics on the performance of all methods through computer experiments. In this study, this has been done for several two-parameter distributions where the cost of the maximum likelihood calculations is still manageable: log-normal (LN), gamma (GA), and Gumbel (GU). For ease of interpretation of the results, clarity of method representation, and to keep the computing time needed down to a manageable level, only the case of *At Most One Change* (AMOC) is considered.

The equations used in the methods are derived for independent identically distributed (i.i.d.) random variables (RVs). The methods are based on pseudolikelihood, and in principle they can be extended to time series with short range correlation, see also Aue and Horváth (2012). For other deviations from the i.i.d. assumption such as periodic components or trends, methods that have been applied in combination with other CP detection schemes should work here as well. Preliminary results on series with short range dependence show that the methods with formulas based on i.i.d. still work, but information about uncertainty decreases in quality, see Appendix G. Finally, with regard to long range dependence, Aue and Horváth (2012) state that tests for structural stability that are designed for the i.i.d. or short range correlation case are not robust against long memory. Moreover, time series with short range correlation and change points may be difficult to distinguish from time series with long memory (Berkes et al., 2006).

The remainder of this paper is organized as follows. First, the two approaches to confidence curve construction are presented. Next, indicators are defined that can be used to evaluate the method performance and compare the confidence curves. Then the results of the application of the methods to synthetic data are analyzed. After that, the methods are applied to several hydrological and climatological time series for which CP detection results are available in the literature. Finally, we discuss the results and present our conclusions. Mathematical details can be found in Appendices A–E. In Appendix F several time series of annual maximum flows for different stations downstream of the site of the Three Gorges dam are checked for CPs.

2. Methodology

All time series will be modeled as a vector Y of n independent RVs $Y_1, Y_2, \ldots, Y_n$. In the remainder of the paper, y will represent a realization of Y, and y_{obs} will represent the actual observed time series.

The *null hypothesis* H_0 will be that the Y_i are i.i.d. RVs. The *alternative hypothesis* H_1 will be that there is an index $\tau \in \{1, 2, \ldots, n-1\}$ such that the original random vector is split into two subseries: a subseries with i.i.d. RVs $\{Y_1, Y_2, \ldots, Y_\tau\}$ and a subseries with i.i.d. RVs

$\{Y_{\tau+1}, Y_{\tau+2}, \ldots, Y_n\}$, but the distributions of the variables in the two subseries are different. Furthermore, it is assumed that all distributions are from the same family, so they differ only in the parameters used in the shared *probability density function* (pdf) and *cumulative distribution function* (cdf). The cdf of Y_i will be referred to as $F(\cdot;\theta)$ and the pdf as $f(\cdot;\theta)$ where θ is a vector. The parameter vectors for the left and the right subseries will be θ_L and θ_R respectively. The null hypothesis can now be expressed as $\theta_L = \theta_R$.

Both approaches need to (approximately) solve maximum likelihood problems for the subseries to the left and to the right of the change point τ. Intuitively it is clear that parameter estimation for very small samples will be difficult. Some studies considering this are Lettenmaier and Burges (1982), Delicado and Goria (2008), Landwehr et al. (1979). These suggest that for short subseries, the results may vary considerably from sample to sample. A minimum subseries length $n_{\min}$ will therefore be used. As a result, only a subset of CP locations, given by

$$L_{CP} = \{n_{\min}, n_{\min}+1, \ldots, n-n_{\min}\} \tag{1}$$

was considered, and no parameter estimates for subseries shorter than $n_{\min}$ were carried out. The choice of $n_{\min}$ is to a certain extent arbitrary. Here we take a minimum subseries length

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

where the notation $\lfloor . \rfloor$ denotes rounding down towards the nearest integer; $n_{\min} = 1$ corresponds to considering all possible CPs. One reason to consider trimming is that the variance in parameter estimates tends to decrease with increasing sample size. As a result, a short sequence may lead too much 'wilder' parameter estimates than a long sequence (Landwehr et al., 1979). This would seem undesirable when looking for parameter changes.

2.1. A description of the two approaches

The starting point for both approaches is the log-likelihood function ℓ for a CP problem. The value of ℓ for a CP τ, distribution parameter vectors θ_L and θ_R, and a realization y of the time series is

$$\ell(\tau, \theta_L, \theta_R; y) = \sum_{i=1}^{\tau} \log f(y_i; \theta_L) + \sum_{j=\tau+1}^{n} \log f(y_j; \theta_R); \quad \tau \in \{1, 2, \ldots, n-1\} \tag{3}$$

Here, θ_L and θ_R are *vectors of nuisance parameters* and τ is the parameter of interest. A common way of dealing with nuisance parameters is the following. For each τ take the supremum (least upper bound) of (3) over all θ_L, θ_R; the resulting function is called the *profile log-likelihood*

$$\ell_{\text{prof}}(\tau; y) = \sup_{\theta_L, \theta_R} \ell(\tau, \theta_L, \theta_R; y) \tag{4}$$

For a closed bounded parameter set, the supremum coincides with the maximum. For a given τ, let $\hat{\theta}_L(\tau, y)$ and $\hat{\theta}_R(\tau, y)$ stand for the values of θ_L and θ_R, respectively, for which $\ell(\tau, \theta_L, \theta_R; y)$ attains the maximum value. With this notation (4) is equivalent to

$$\ell_{\text{prof}}(\tau; y) = \ell(\tau, \hat{\theta}_L(\tau, y), \hat{\theta}_R(\tau, y); y) \tag{5}$$

In CML, $\hat{\theta}_L(\tau, y)$ and $\hat{\theta}_R(\tau, y)$ are calculated whenever needed. The smallest value of $\tau \in L_{CP}$ for which ℓ_{prof} is maximal will be denoted by $\hat{\tau}(y)$,

$$\hat{\tau}(y) = \min\left(\arg\max_{\tau \in L_{CP}} \ell(\tau, \hat{\theta}_L(\tau, y), \hat{\theta}_R(\tau, y); y) \right) \tag{6}$$

The minimum is taken to allow for the, highly unusual, case where there are multiple maxima. In CLMo (CMoM), instead of a profile log-likelihood ℓ_{prof}, a pseudo log-likelihood ℓ_{pseu} is used. To obtain ℓ_{pseu}, the LMo (MoM) estimates $\tilde{\theta}_L(\tau, y)$ and $\tilde{\theta}_R(\tau, y)$ of the nuisance parameters are inserted in (3)

$$\ell_{\text{pseu}}(\tau; y) = \ell(\tau, \tilde{\theta}_L(\tau, y), \tilde{\theta}_R(\tau, y); y) \tag{7}$$

These estimates are assumed to be acceptable approximations of the maximum likelihood estimation results. The smallest value of $\tau \in L_{CP}$ for which ℓ_{pseu} is maximal will be denoted by $\tilde{\tau}(y)$,

$$\tilde{\tau}(y) = \min\left(\arg\max_{\tau \in L_{CP}} \ell(\tau, \tilde{\theta}_L(\tau, y), \tilde{\theta}_R(\tau, y); y) \right) \tag{8}$$

From this point onwards, all methods follow the same path towards a confidence curve. A *deviance function* for CML is defined as the deviance of ℓ_{prof} from the maximum value it attains at $\hat{\tau}(y)$

$$D_{\text{prof}}(\tau, y) = 2\left\{ \ell_{\text{prof}}(\hat{\tau}(y), y) - \ell_{\text{prof}}(\tau, y) \right\} \tag{9}$$

and a deviance function for CLMo (CMoM) is defined as the deviance of ℓ_{pseu} from the maximum value it attains at $\tilde{\tau}(y)$

$$D_{\text{pseu}}(\tau, y) = 2\left\{ \ell_{\text{pseu}}(\tilde{\tau}(y), y) - \ell_{\text{pseu}}(\tau, y) \right\} \tag{10}$$

For all $\tau \in L_{CP}$, the distribution of the deviance function for CML follows from

$$\forall r \in \mathbb{R} : K_{\tau,\text{prof}}(r) = \Pr\left(D_{\text{prof}}(\tau, Y) < r \mid \tau, \theta_L = \hat{\theta}_L(\hat{\tau}(y), y), \theta_R = \hat{\theta}_R(\hat{\tau}(y), y) \right) \tag{11}$$

and for CLMo (CMoM) from

$$\forall r \in \mathbb{R} : K_{\tau,\text{pseu}}(r) = \Pr\left(D_{\text{pseu}}(\tau, Y) < r \mid \tau, \theta_L = \tilde{\theta}_L(\tilde{\tau}(y), y), \theta_R = \tilde{\theta}_R(\tilde{\tau}(y), y) \right) \tag{12}$$

No exact or approximate expression for K_τ is available. Therefore, an MC simulation will be used to approximate K_τ.

In Cunen et al. (2018) and in this paper, a *confidence curve* (see also Appendix A) is defined using the distribution of the deviance function

$$\text{cc}(\tau, y_{\text{obs}}) = K_\tau\left(D(\tau, y_{\text{obs}}) \right) \tag{13}$$

where y_{obs} is an observation of the random sample Y. The MC approximation of K_τ is obtained as follows:

1. Estimate parameters τ, θ_L, θ_R by first solving for $\tau^*(y_{\text{obs}})$, and then calculating $\theta_L^*(\tau^*(y_{\text{obs}}), y_{\text{obs}})$ and $\theta_R^*(\tau^*(y_{\text{obs}}), y_{\text{obs}})$.
2. For each possible location $\tau \in L_{CP}$, and $j = 1, 2, \ldots, N$, draw a new sample $y^{(j,k)}$ where the components $y_i^{(j,k)}$ $(k = 1, 2, \ldots, \tau)$ are distributed according to the distribution $F(.,\theta)$ with $\theta = \theta_L^*(\tau^*(y_{\text{obs}}), y_{\text{obs}})$ and the components $y_i^{(j,k)}$ $(k = \tau+1, \tau+2, \ldots, n)$ are distributed according to the distribution $F(.,\theta)$ with $\theta = \theta_R^*(\tau^*(y_{\text{obs}}), y_{\text{obs}})$.
3. Approximate the confidence curve $\text{cc}(\tau, y_{\text{obs}}) = K_\tau(D(\tau, y_{\text{obs}}))$ by

$$K_{\tau,N}(D(\tau, y_{\text{obs}})) = \frac{1}{N} \sum_{j=1}^{N} \mathbf{1}_{D(\tau,y) < D(\tau, y_{\text{obs}})} \tag{14}$$

Where $\mathbf{1}$ is the indicator function

$$\mathbf{1}_{a<b} = \begin{cases} 0 & a \geq b \\ 1 & a < b \end{cases} \tag{15}$$

and τ^* is $\hat{\tau}$ for CML and $\tilde{\tau}$ for CLMo (CMoM), and θ^* is $\hat{\theta}$ for CML and $\tilde{\theta}$ for CLMo (CMoM). Standard software is not yet available for the methods, a custom implementation in Matlab® was written. Where necessary, random numbers were obtained from the 'threefry' type random number generator in Matlab.

2.2. Properties of confidence curves

The performance of CML and CLMo (CMoM) methods will be examined and compared by exploring some properties (Zhou et al., 2020) of confidence curves constructed by the two methods. They are:

- The cumulative frequency distribution of the $\hat{\tau}(y)$ for CML and $\tilde{\tau}(y)$ for CLMo (CMoM) based on synthetic data when the null hypothesis H_0 (there is no CP) holds. In this case, the distribution should be close to uniform. If it is not uniform, then it indicates that there is a bias for certain locations when a type I error (incorrect rejection of the null hypothesis) occurs.

- The cumulative frequency distribution of the $\hat{\tau}(y)$ and $\tilde{\tau}(y)$ for synthetic data when the alternative hypothesis H_1 (there is a CP) holds. While the point where the deviance function is zero is not necessarily the true CP, it is contained in all confidence sets that follow from the confidence curve. If these sets are narrow, then this point should be near the true CP. For a definition of a confidence set, see Definition 1 in Appendix A.
- The *actual* versus *nominal* coverage probability for the confidence sets produced by the curves at all confidence levels for synthetic data. The actual coverage probability at a given confidence level (nominal coverage probability) indicates the probability of a confidence set containing the true value of the parameter of interest. For detailed definitions of actual and nominal coverage probability, see Definition 1 in Appendix A.
- A summary of the *uncertainty* about the CP associated with a confidence curve cc is defined in Appendix B by (B.7)

$$\mathrm{Un}\,(\mathrm{cc}) = \frac{\left(\sum_{k=n_{\min}}^{n-n_{\min}} \mathbf{1}_{\mathrm{cc}(k)\le\gamma_{\max}}\right) - 1}{n - 2n_{\min}}$$

where

$$\gamma_{\max} = \frac{n - 2n_{\min}}{n - 2n_{\min} + 1}$$

- The *similarity index* is used to measure the similarity of two confidence curves

$$\tilde{J}\left(\mathrm{cc}, \mathrm{cc}'\right) = \frac{\sum_{k=n_{\min}}^{n-n_{\min}} \min\left(1 - \mathrm{cc}\,(k), 1 - \mathrm{cc}'\,(k)\right)}{\sum_{k=n_{\min}}^{n-n_{\min}} \max\left(1 - \mathrm{cc}\,(k), 1 - \mathrm{cc}'\,(k)\right)} \tag{16}$$

where cc and cc′ are a pair of confidence curves. This index was proposed in Zhou et al. (2020) and resembles the Ružička index (Schubert and Telcs, 2014). It is one for identical curves and smaller than one for curves that differ.

2.3. Synthetic time series generation and examples of confidence curves

The CML and CLMo (CMoM) methods were implemented for three distributions (LN, GA, GU). To evaluate the performance of CML and CLMo (CMoM), synthetic data were generated from the underlying distributions. The distributions were selected because they are commonly used in hydrology (Hamed and Rao, 2019; Thompson, 2017; Haktanir, 1991; Karim and Chowdhury, 1995). The pdfs and the relations between the parameters and moments and L-moments are given in Appendix D. The change in statistical properties of the synthetic data was a change in the mean μ for CML (CMoM) and a change in the mean or in the standard deviation σ for CLMo.

For each distribution and each combination of a change in the mean $\Delta\mu = 1, 2, 4$ and a series length $n = 40, 100$, a set of $M = 1000$ artificial time series of length n with standard deviation $\sigma = 1$, $\tau = n/4, n/2, 3n/4$, and a jump $\Delta\mu$ in the mean between τ and $\tau + 1$ was generated. The location of the CP during sample generation will be referred to as τ_{true} in this study. The mean of a specific distribution for the subseries up to τ was $\mu_{\mathrm{L}} = 2$, and the mean for the subseries after τ was $\mu_{\mathrm{R}} = \mu_{\mathrm{L}} + \Delta\mu$, where $\Delta\mu = 1, 2, 4$. Examples of synthetic data sets with $\Delta\mu = 0, 1, 2$ and the corresponding confidence curves for CML and CLMo are given in Fig. 1(a–f).

The standard deviation for the left and right subseries will be referred to as σ_{L} and σ_{R} respectively. Additional experiments were done for CLMo with $\mu_{\mathrm{R}} = \mu_{\mathrm{L}}$, $\sigma_{\mathrm{R}} = \sigma_{\mathrm{L}} + \Delta\sigma$, where $\Delta\sigma = 1, 2, 3$. Examples of data series synthetic data sets with $\Delta\sigma = 0, 1, 2$ are given in Fig. 1(g–l). The effect of shifting (different μ_{L}) or scaling (different σ_{L}) a time series is discussed in Appendix C. Where necessary, random numbers to generate samples were obtained from the 'twister' type random number generator in Matlab.

3. Evaluation of the methods for synthetic data

The performance of the confidence curves produced by CML and CLMo (CMoM), as represented by the properties listed in Section 2.2, are examined. For all methods, $N = 1000$ MC simulations were used to generate the approximate confidence curve. The majority of the experiments involved a change in the mean. However, one advantage of a parametric method is that it looks for changes in all parameters at the same time, so some experiments were performed for synthetic series with a change in the standard deviation as well.

3.1. The cumulative frequency distribution of the CP estimate

The methods produce confidence curves instead of point estimates. These confidence curves are characterized by their shape and the location of their minimum. In the remainder of the paper, the minimum of the confidence curve will be referred to as the (point) estimate of the CP.

3.1.1. The cumulative frequency distribution of the CP estimate when the null hypothesis holds

Fig. 2 shows the cumulative frequency distribution of the CP estimates found by CML and CLMo when the null hypothesis holds (no CP). In this case, if the method is forced to select a CP, then it should not display a preference for any particular CP. The possible candidates are the elements of the set L_{CP} defined in (1). The black line in Fig. 2(a–c) shows the corresponding uniform frequency distribution on that set. The experimental results do not match this exactly, but do approximate it. The methods have a slight preference for points near the middle of the time series. For LN, GA, and GU the methods CML and CLMo give similar results.

3.1.2. The cumulative frequency distribution of the CP estimate when the alternative hypothesis holds

Figs. 3(a–f) show the frequency distribution of detected CPs by CMLo when the alternative hypothesis holds for $\tau_{\mathrm{true}} = n/4, n/2, 3n/4$ and $n = 40, 100$. Figs. 3(g–l) show the equivalent results for $\Delta\sigma = 1, 2$.

For CLMo, the minima of the confidence curves for the different samples are spread around the true CP. Results for CML with a change in the mean are similar. The spread decreases with increasing $\Delta\mu$ ($\Delta\sigma$) and n. For example, for $\Delta\mu = 1$ and $n = 100$, about 90% of the estimates lie within ± 10 points of the actual CP. For $\Delta\mu = 2$, the spread reduces to ± 5 points. On average the CP estimates are closer to the true value for changes in the mean than for changes in the standard deviation. The frequency distributions found by all methods for synthetic time series drawn from the three distributions are very similar.

3.2. Actual versus nominal coverage probability

The difference between the actual and the nominal coverage of the confidence sets defined by the confidence curve is quite important for their practical use. If the actual coverage of a confidence set is lower than the nominal one, then it is *permissive* Appendix A. This may cause problems, because it suggests too much certainty about the CP location; if the set were a person, then that person would be overconfident. If the actual coverage probability exceeds the nominal coverage, then the set is *conservative*; while this is less problematical than permissiveness, it suggests too much uncertainty; the set would please an overcareful person. The actual coverage was estimated as follows: synthetic time series with indices $m = 1, 2, \ldots, M$ were generated, and for each time series m, the confidence curve and the confidence set $R_{\gamma,\mathrm{m}}$ at confidence level γ were determined. Finally, the number k of sets for which $\tau_{\mathrm{true}} \in R_{\gamma,\mathrm{m}}$ was divided by M. In Fig. 4, plots of the actual versus nominal coverage are shown.

When interpreting Fig. 4, it is important to recall that if a CP is present, then there is only a finite number of possible locations for

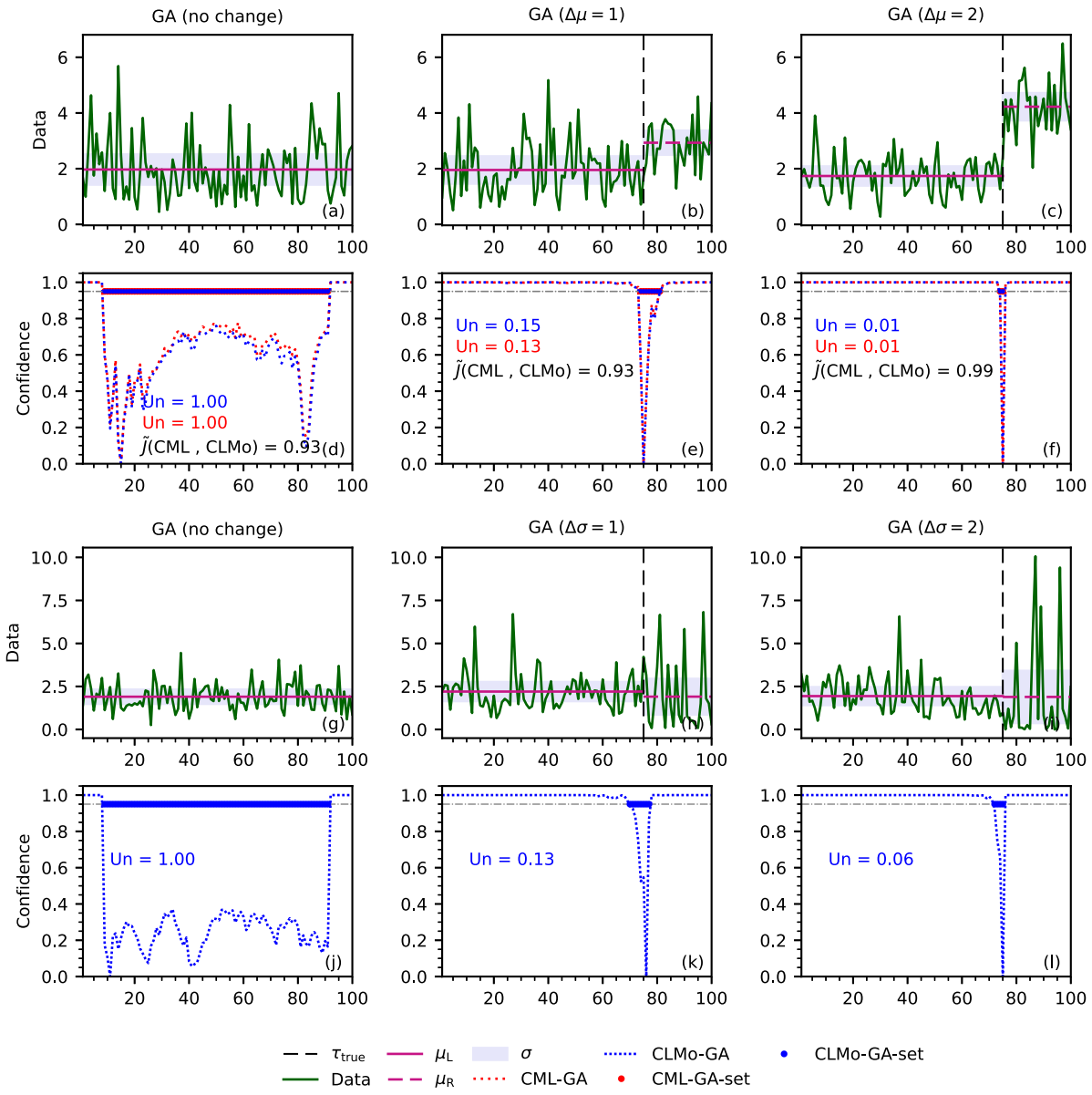

Fig. 1. Synthetic GA distributed data sets with a change $\Delta\mu$ in the mean or $\Delta\sigma$ in the standard deviation, and the corresponding confidence curves.

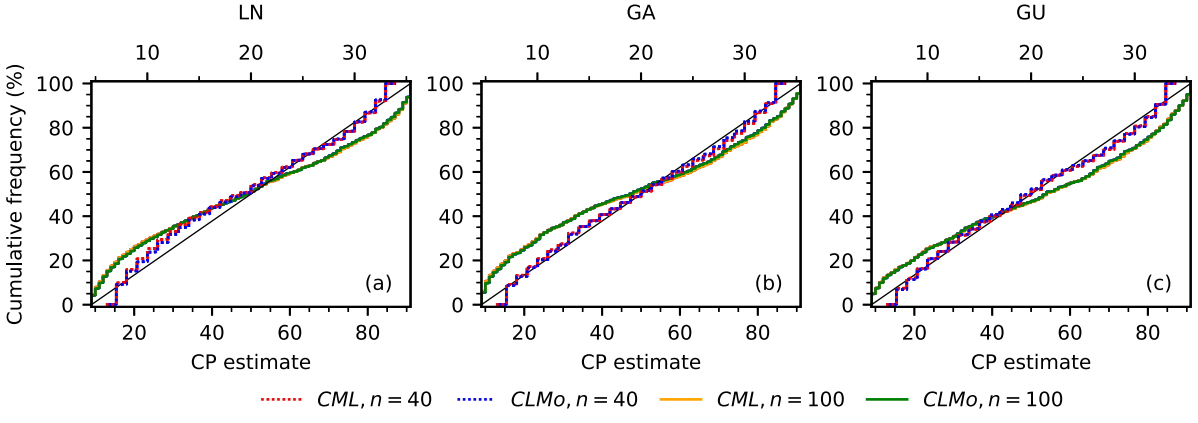

Fig. 2. The cumulative frequency distribution of CPs for $n = 40, 100$ when H_0 holds.

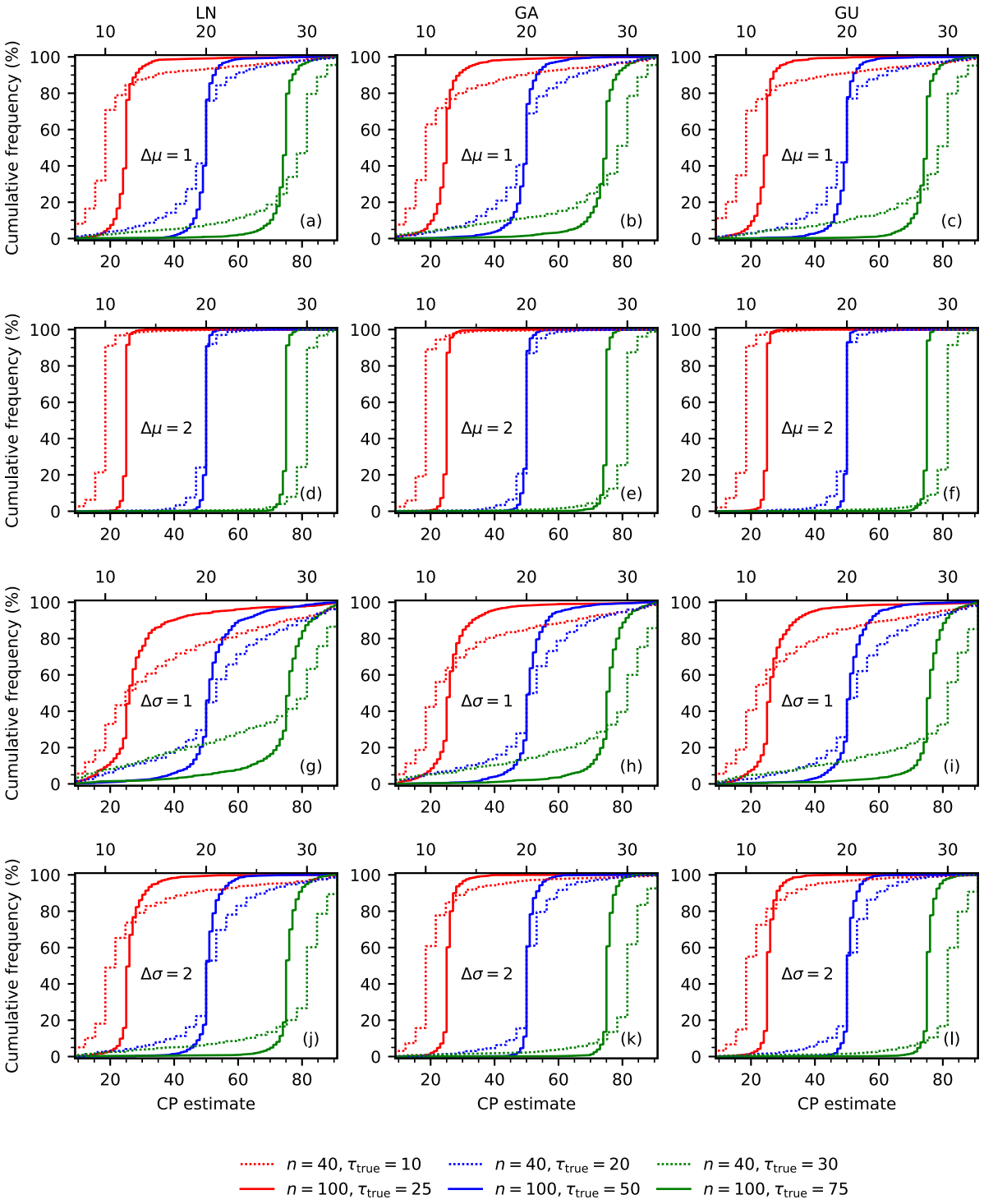

Fig. 3. The cumulative frequency distribution of CPs when there is a change in mean or standard deviation.

that CP. This in turn means that if the construction method for the curve makes very good use of the information in the sample, then it may result in confidence random sets that contain only one or two points, but have a very high probability of containing the actual CP. This implies that for low confidence levels the sets will be very conservative. This manifests itself in Fig. 4 where in (a) the actual coverage is always above 37% for $\Delta\mu = 1$; it always exceeds 65% for $\Delta\mu = 2$ in (b), and in (c) it is higher than 70% for $\Delta\mu = 2$. In Fig. 4(d–f) it can be seen that for changes in the standard deviation the actual coverage and nominal coverage start to coincide at lower values. It follows that in practice, the confidence sets with relatively high nominal confidence carry the best information. The sets at low confidence levels are much too conservative. The results for CML show that all distributions provide accurate actual coverage for nominal coverages above 90%. For changes in the mean, the results for CLMo (CMoM) are nearly identical to those for CML. Results for $n = 40$ were generated as well, but the impact from series length on actual coverage probability was small, so they have not been included here. Table 1 shows detailed information about actual versus nominal coverage for confidence curves constructed by CML and CLMo for confidence levels $\gamma = 0.90, 0.95, 0.99$.

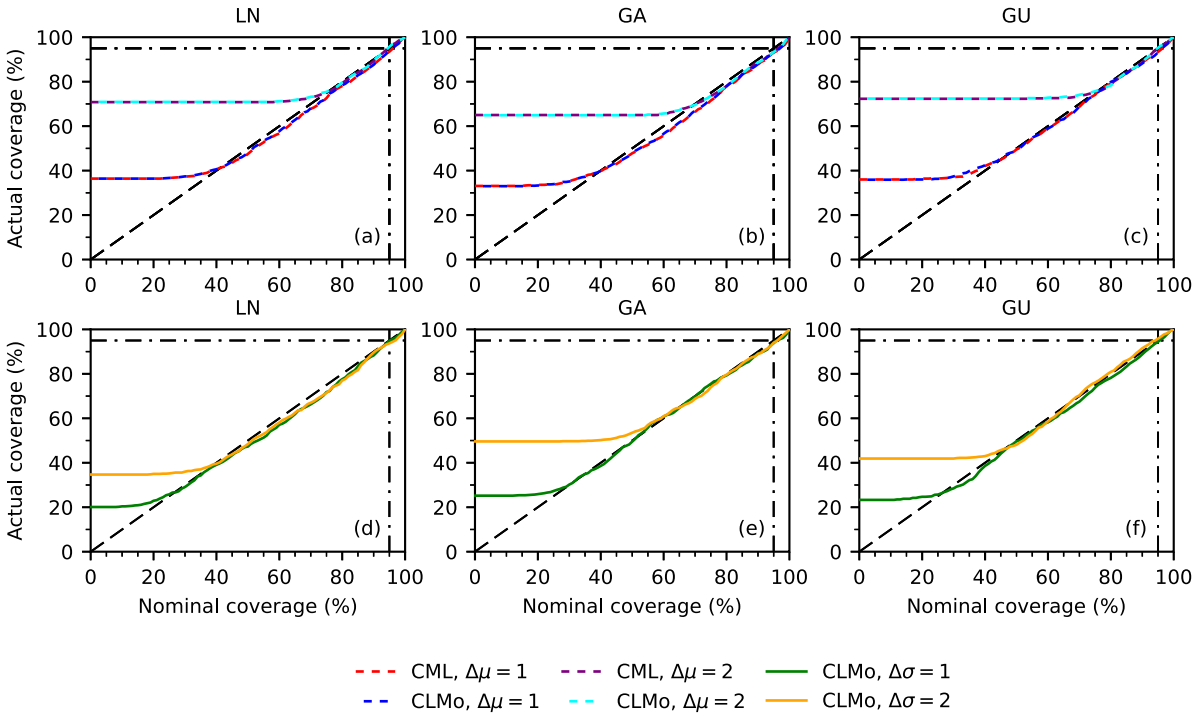

Fig. 4. Actual versus nominal coverage probability for a change in the mean or standard deviation when $\tau_{\text{true}} = n/2$ and series length $n = 100$.

Table 1
The actual coverage of confidence sets by CML and CLMo. Cells with conservative coverage are gray.

Confidence level			0.9		0.95		0.99	
Distribution	$\Delta\mu$	n	CML	CLMo	CML	CLMo	CML	CLMo
LN	1	40	0.855	0.860	0.919	0.922	0.980	0.981
		100	0.879	0.876	0.935	0.940	0.990	0.989
	2	40	0.880	0.885	0.927	0.934	0.974	0.976
		100	0.892	0.895	0.957	0.957	0.990	0.991
GA	1	40	0.859	0.854	0.905	0.911	0.982	0.984
		100	0.877	0.879	0.931	0.930	0.975	0.980
	2	40	0.889	0.896	0.945	0.945	0.988	0.989
		100	0.887	0.887	0.932	0.934	0.983	0.988
GU	1	40	0.864	0.868	0.925	0.929	0.985	0.984
		100	0.890	0.882	0.934	0.932	0.985	0.985
	2	40	0.881	0.882	0.937	0.943	0.985	0.985
		100	0.886	0.886	0.951	0.952	0.993	0.993

An indication of the spread in actual coverage can be provided as follows. If the actual coverage were equal to the nominal coverage, then the number k of M confidence sets $R_{\gamma,\text{m}}$, $m = 1, 2, \ldots, M$ that contained the true CP would be distributed according to a binomial distribution

$$\text{P}_\text{r}(k) = \left(\begin{array}{c} M \\ k \end{array} \right) \gamma^k (1-\gamma)^{M-k} \tag{17}$$

For the binomial distribution, the variance is $M\gamma(1-\gamma)$, so the standard deviation of k/M is $\sqrt{\gamma(1-\gamma)/M}$. For $M = 1000$, the standard deviation of the distribution of k for $\gamma = 0.90$ is 0.009; for $\gamma = 0.95$ it is 0.007, and for $\gamma = 0.99$ it is 0.003. When combining this information with Table 1, please recall that the location of the CP is a discrete RV, so for some confidence levels it might not be possible to define a confidence set with that exact coverage.

3.3. The uncertainty in the confidence curves

As mentioned in the introduction, it may be necessary to automatically split a set of confidence curves into groups for further analysis. Here Un is proposed as a tool to do so. One way to evaluate the suitability of Un is to compare its associated type I and type II errors to a classical hypothesis test for the null hypothesis that no CP is present. For this purpose, a comparison with the classical Pettitt test is performed. The value Un for a confidence curve is calculated according to (B.7). It summarizes the uncertainty of a confidence curve.

The bounds on Un depend on the distribution and the parameters, so in principle they should be determined by Monte Carlo experiments for each individual case. In practice splitting the series into three classes: definitely no CP, definitely a CP, and 'to be examined further' might allow rough bounds to be established that depend only weakly on the distribution parameters. See also Appendix C

3.3.1. The uncertainty in the confidence curves for the null hypothesis

The CML approach implicitly assumes that a CP is present, so it would seem that it should be preceded by a test for the presence of a CP. However, if Un is calculated for synthetic time series generated without a CP, then it turns out to be quite high in most cases, near one for 80% ($n = 40$) to 90% ($n = 100$) of all curves, as shown in Fig. 5. Examples of confidence curves for time series without a CP are shown in Fig. 1(a,g). As high Un in the presence of a CP means that the method supplies only a very limited amount of information on CP location, it is tempting to simply say that if Un exceeds a certain bound, then either there is no CP or the method cannot reliably detect the CP location. For example, Fig. 5b shows that for CML ($n = 100$), 95% of all Un values exceed

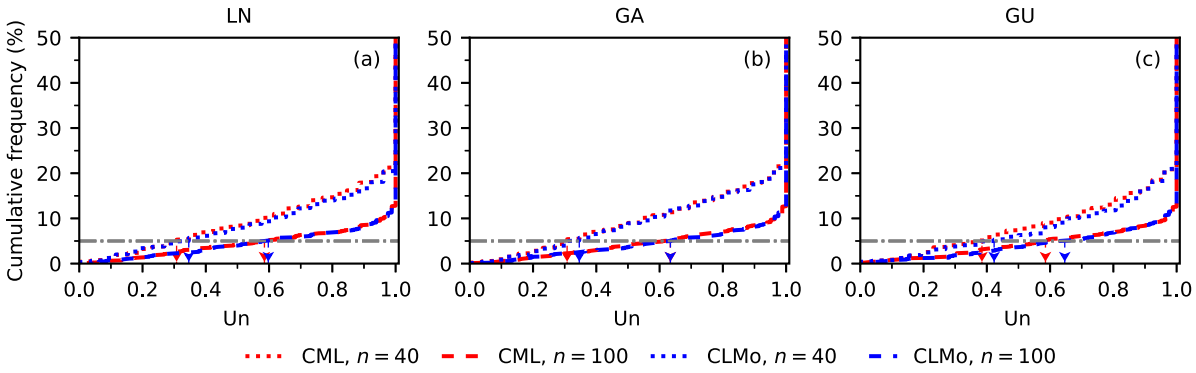

Fig. 5. Cumulative frequency of Un when H_0 holds (vertical scale truncated at 50%).

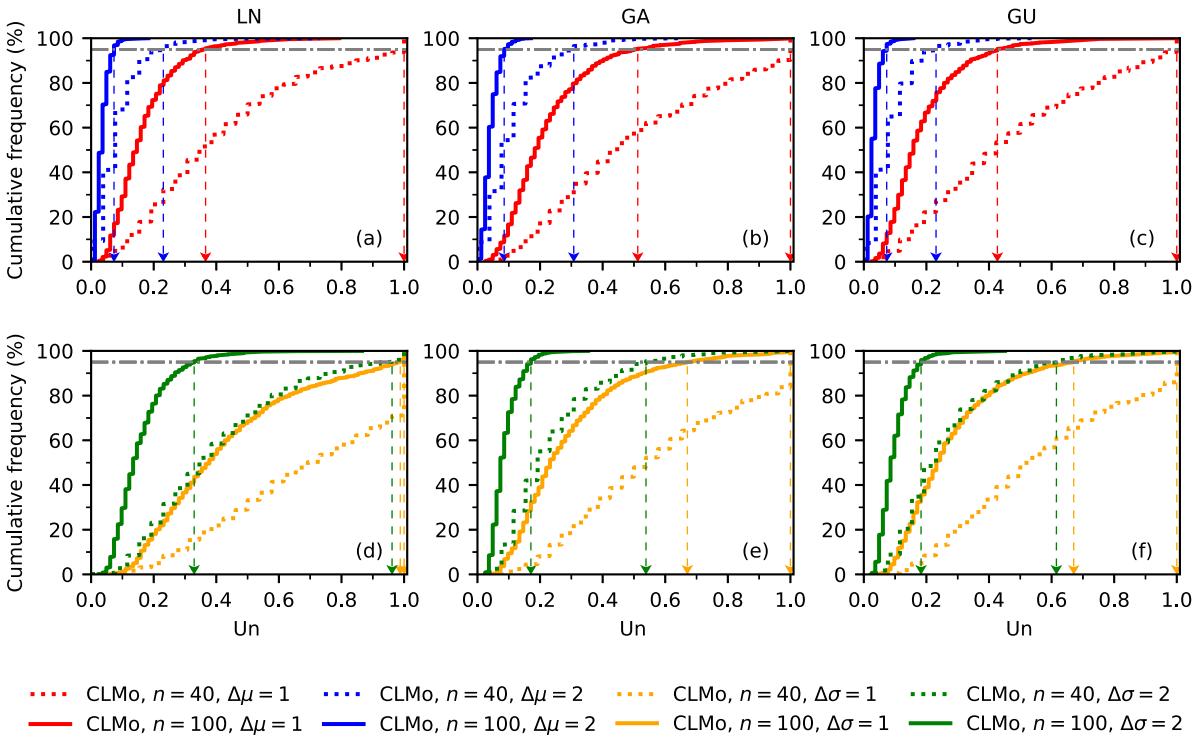

Fig. 6. Cumulative frequency of Un for a CP in the middle of the series and a change in the mean or standard deviation.

0.63. If that bound were used to reject the null hypothesis, then for this particular distribution and this particular parameter set, that choice would result in a 5% type I error. The viability of this approach depends on the distribution of Un in cases where the alternative hypothesis holds.

3.3.2. *The uncertainty of confidence curves for the alternative hypothesis*

Fig. 6 shows the frequency distribution of Un for CLMo when H_1 holds and the CP lies in the middle of the time series (the values for CML for a change in the mean are very similar). For $n = 100$, $\Delta\mu = 1$, 95% of the values lie below 0.4 for LN and for GA 95% of the Un values lie below 0.45 (Fig. 6b). For $\Delta\mu = 2$ these bounds less than 0.1 for all distributions. Fig. 6(d–f) show the frequency distribution for Un when H_1 holds and the CP lies in the middle of the time series. For $n = 100$, $\Delta\sigma = 2$ and the LN distribution, 95% of the values lie below 0.35. For GA the 95% of the Un values lie below 0.2. For GU 95% of the Un values lie below 0.2. It should be noted that the scale parameter of GA equals the variance divided by the mean, so for fixed mean it increases with the square of the standard deviation. For higher standard deviations this leads to a distribution that tends to produce many low values with a few very high values mixed in. Special care may be needed in the calculations for low means and high standard deviation.

3.3.3. *Uncertainty as a tool to select curves and data that need closer inspection*

There are two types of error that are of interest when testing a hypothesis. If H_0 is rejected when there is no CP, then that is a type I error and if H_0 is accepted when there is a CP, then that is a type II error. For example, for $\Delta\mu = 1$, $n = 100$ and distribution LN, Fig. 5(a) implies that rejection of H_0 for Un ≤ 0.2 would result in a very small type I error, while Fig. 6(a) implies that non-rejection of H_0 for Un ≥ 0.5 would result in a very small type II error. The subset of time series where $0.2 <$ Un < 0.5 would then need further study by visual inspection or additional tests. If none of the original time series actually has a CP, then that subset would contain less than 5% of the original set of time series, while if all series actually have a CP, then that subset would contain about 30% of the original set.

Fig. 7 is based on the frequency distribution of Un over the synthetic samples sets and shows how a particular choice of a Un value as a bound for acceptance of H_0 would translate into type I and type II

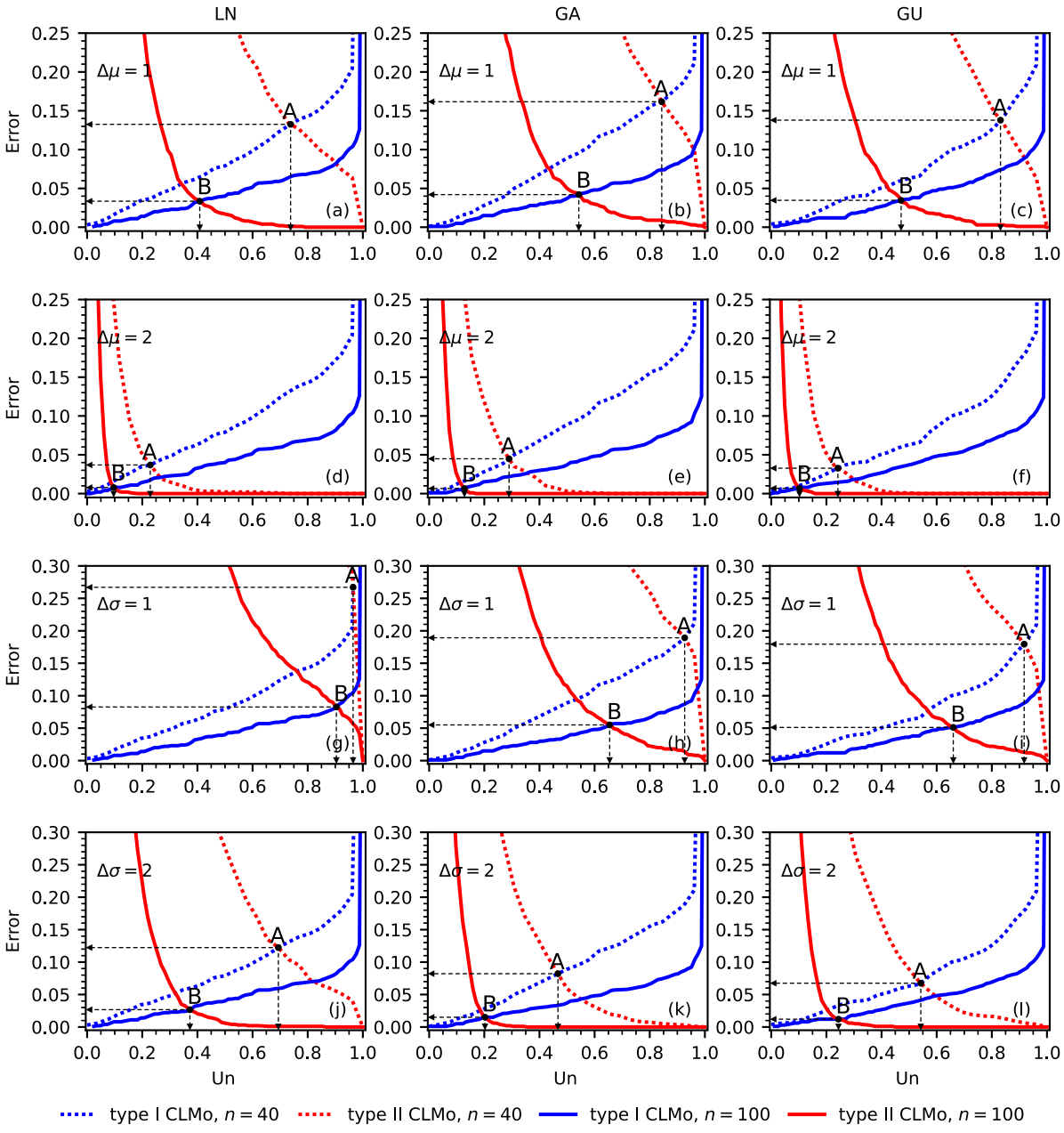

Fig. 7. Un versus type I and type II errors for a CP in the middle of the series and a change in the mean or the standard deviation.

errors for that set of samples. For different applications of the methods, the relative importance of the type I and type II error will differ. The point marked 'A' corresponds to the Un value for which the type I and type II errors are equal for $n = 40$. The point 'B' corresponds to the Un value for which the type I and type II errors are equal for $n = 100$. By plotting the value pairs of type I and type II errors associated with a particular value of Un over a range of Un values, it is possible to visualize the relation between the errors. For $n = 100$ and a change in the mean of 1 or 2 Fig. 7(a–f) shows that there are choices of Un boundary that, for these distributions and parameters, result in both type I and type II errors that are smaller than 5%. For $n = 100$ and $\Delta\sigma = 2$ similar results are obtained. For $n = 100$ and $\Delta\sigma = 1$ an upper bound of 5% on both errors can be achieved only for GU.

To see how a null hypothesis test based on Un would do when compared with the classical Pettitt test, the curve of error pairs is drawn for both tests for a CP in the middle of the time series in Fig. 8. The results show that in principle Un could even serve as the basis for a hypothesis test. For a change in the standard deviation the standard Pettitt test is much less effective (Fig. 8g–l). A modification of that test, specifically designed for the change to be detected, would be needed. Here the test based on Un is not the most effective for any particular change, but it is the one that will detect changes in all parameters.

3.4. The similarity index between confidence curves

To evaluate the similarity between confidence curves for the same synthetic time series, the similarity index $\tilde{J}$ was calculated by (16) for $\Delta\mu = 1, 2$, $\tau_{\text{true}} = n/2$, and $n = 40, 100$. Details on the calculation of $\tilde{J}$ and its properties can be found in Zhou et al. (2020, Appendix C). Fig. 9 shows the resulting cumulative frequency distributions of $\tilde{J}$. The confidence curves for synthetic data calculated by CML are very similar to those calculated by CLMo. The similarity increases with increasing $\Delta\mu$ and n. For the GU distribution similarity seems lower. To provide a point of reference for the similarity values, $\tilde{J}$ was calculated for 16000 random pairs of H_0 confidence curves. The result was that 95% of the

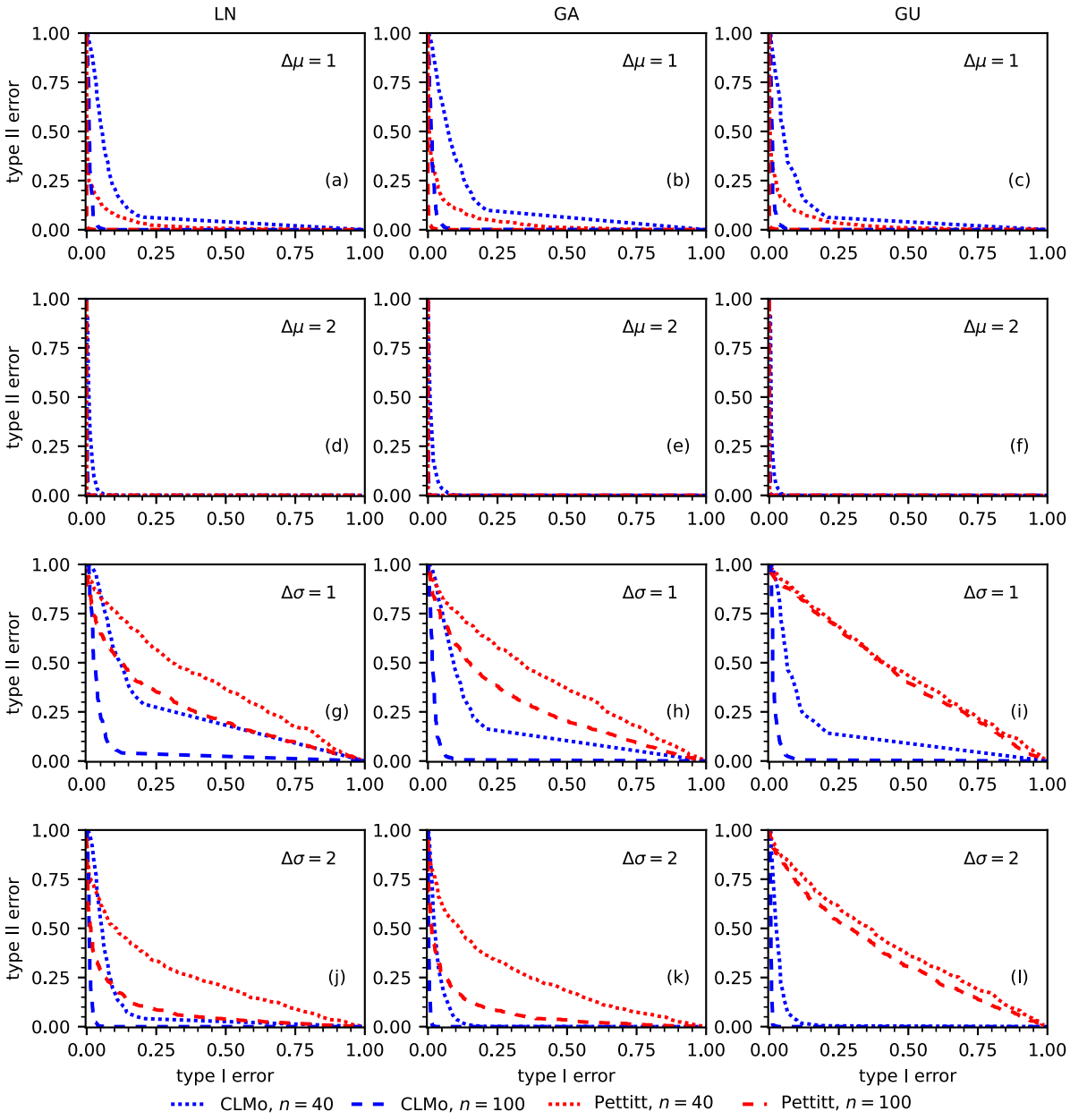

Fig. 8. Comparison of error rates for the two null hypothesis tests, where for H_1, the CP is in the middle of the series.

pairs had a similarity below 0.7, for all series lengths, distributions, and methods.

4. Change point detection and uncertainty in real hydrometeorological data

To examine the performance of the CML and CLMo methods on real world data, three time series of measurements were taken from previous publications: annual average naturalized flow at Itaipu (Conte et al., 2019) - case study 1, annual average temperature at Tuscaloosa (Reeves et al., 2007) - case study 2, and annual average rainfall at Tucumán (Jandhyala et al., 2010) - case study 3. The CPs found in the original studies are used as a reference. Both methods were used to construct confidence curves for CPs with each of the three distributions: LN, GA, and GU. The uncertainties for the confidence curves were determined as was the similarity between the CML and CLMo curve for each case. The confidence set at confidence level 95% is also shown. In Appendix F the methods are applied to four time series of annual maximum discharge on the Yangtze River in China that were also analyzed in Zhou et al. (2019) by a non-parametric method — case study 4.

4.1. Case study 1

Conte et al. (2019) found a significant CP in 1971 in the annual average naturalized flow at the Itaipu Hydroelectric Plant in Brazil from 1931 to 2015 by the bootstrap Pettitt test. In this case the value of $|\Delta\mu|/\sigma \approx 1.33$ suggests CML and CLMo should do reasonably well. And so they do: Un is low (0.03 to 0.07) and $\tilde{J}$ is high (≥ 0.96). All give a 95% confidence interval of about three years (see Fig. 10).

4.2. Case study 2

The annual average temperature time series from 1940 to 1986 in Tuscaloosa (USA) was selected because there was only one documented reason for a CP during this period. The time series was used in Reeves

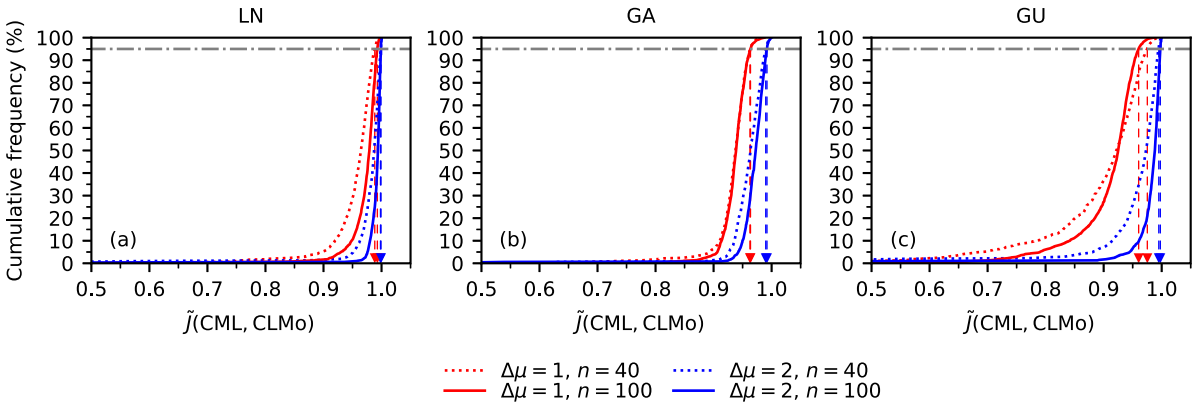

Fig. 9. The similarity index between confidence curves generated by the CML and CLMo methods.

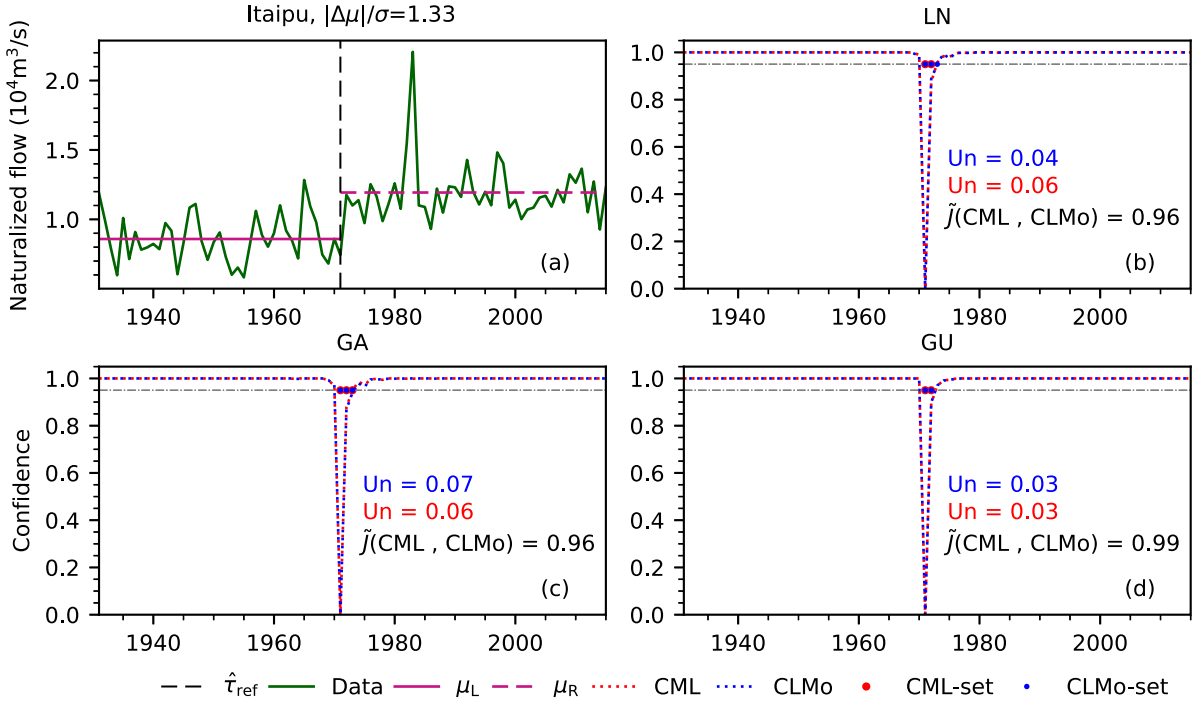

Fig. 10. Confidence curves for CP in annual average naturalized discharge time series of Itaipu.

et al. (2007), and a CP located at the year of 1957 was found by eight different methods. The value of $|\Delta\mu|/\sigma = 1.33$ again suggests the methods should do well. Both LN and GA based methods find a reasonably precise confidence curve for the CP with a 95% confidence interval [1955, 1959]. In this case GU is not doing as well as LN and GA. Moreover, GU combined with CLMo seems to be confused by the sudden drop from 1975 to 1976 (Fig. 11(b–d)).

A possible explanation is the difference in parameters for the Gumbel distribution found by the different methods. While one would hope that CML, CMoM, and CLMo would give similar results, all parameter estimates are RVs and their variance may be quite large for small samples. Given the very different formulas used to obtain the estimates, it should not be surprising that, without large samples to reduce the variance, very different results can be found. This in turn may lead to different points being selected as CP. Table 2 gives the estimated parameters and the corresponding values of the log-likelihood. It can be seen that a CP in 1975 results in a value for the profile log-likelihood that is close to the minimum and that the pseudo log-likelihood values are close to the profile log-likelihood value for 1975. For 1957, the CLMo and CMoM parameter approximations of the location are close to the CML value, but the scale parameter estimates are different, this results in a large deviation of the pseudo log-likelihood value from the profile log-likelihood value for 1957.

Table 2
Different Gumbel based parameter estimates for left and right subseries at Tuscaloosa with the CP is at a prescribed location.

CP	Method	left		right		log-likelihood
		loc	scale	loc	scale	
1957	CML	17.48	0.7091	16.73	0.4636	−40.0
	CMoM	17.53	0.4000	16.76	0.3417	−69.8
	CLMo	17.56	0.3502	16.75	0.3580	−100
1975	CML	17.07	0.5896	16.65	0.5552	−45.4
	CMoM	17.10	0.4662	16.68	0.4255	−49.8
	CLMo	17.08	0.4938	16.67	0.4480	−47.7

4.3. Case study 3

In Jandhyala et al. (2010), the annual average rainfall time series from 1884 to 1996 at Tucumán in Argentina was investigated and a

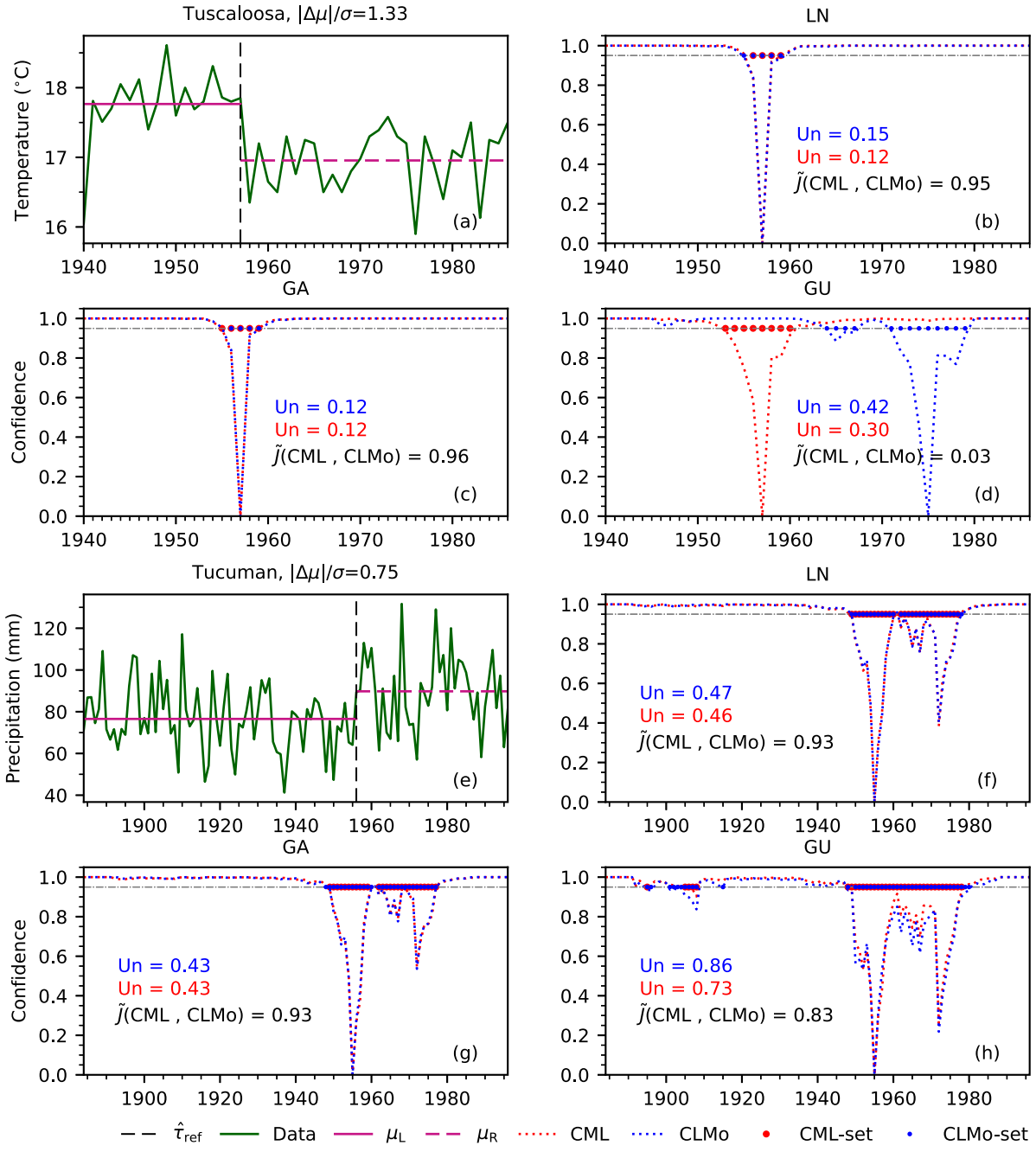

Fig. 11. Confidence curves for CP in time series of Tuscaloosa (a–d) and Tucumán (e–h).

CP was found in 1956 by a Bayesian method. The result was confirmed by Wu et al. (2001); they believed the change was caused by the construction of a dam in Tucumán from 1952 to 1962. Fig. 11(e) shows the data and the CP in 1956. In Fig. 11(f–h) the results of CML and CLMo with different distributions are shown. The value of $|\Delta\mu|/\sigma = 0.75$ suggests that the uncertainty here will be bigger. This is confirmed by the results. All three methods generate curves with several local minima, a global one at 1955 and a deep local one near 1972. This leads to relatively high uncertainties. For LN and GA the uncertainties (0.43 to 0.47) are still low enough to make the presence of a change point probable. Further investigation is needed to determine whether or not this type of curve means that a check for multiple change points should be made. Extension of the methods discussed here to multiple change points would involve the use of 2D confidence sets.

5. Conclusion and discussion

This study proposed new parametric methods to quantify the uncertainty in the location of a CP. The methods discussed here are intended for situations where it is of interest to look for the location of a CP and the statistical properties of the time series are sufficiently clear to determine expressions for the likelihood or pseudolikelihood. To keep the description of the methods readable, only the AMOC case with i.i.d. RVs in the (sub)series was presented. Extension to multiple CPs and/or series with short range dependence is possible in principle and will be investigated in future articles. Furthermore, when the methods as described in this study were applied to time series with short range dependence, they still worked, although the performance was reduced. The question of how to detect CPs in time series resulting from long-memory processes is a different topic (Berkes et al., 2006; Aue and

Horváth, 2012). Moreover, it turns out that distinguishing between long memory processes and short memory processes with shifting means is quite difficult (Rea et al., 2011).

The two new methods (CLMo, CMoM) were compared to a method (CML) described in Cunen et al. (2018). All methods were able to detect changes in the mean and in the standard deviation. They all involve a choice of a distribution family that is used to define a likelihood function for the CP. In this likelihood function, the parameters of the distribution are 'nuisance' parameters. It would be possible to add an additional nuisance parameter to represent dependence. The CML method deals with the nuisance parameters by using a profile likelihood; the CMoM and CLMo methods use a pseudolikelihood with parameter estimates based on moments and L-moments respectively. All methods define a deviance function based on the likelihood for the possible CP locations. An MC calculation is then used to assign approximate probabilities to the deviance function values. These approximate probabilities then define the confidence curve. The reason for the introduction of CMoM and CLMo is that CML, which uses ML parameter estimates, can be very costly in terms of computations and therefore in terms of time. Even for the gamma and Gumbel distributions, where the ML method is relatively cheap, the cost of CML was at least 8 times that of CMoM.

A statistical analysis of the results of a large number of synthetic data series of two lengths, 40 and 100, showed that CLMo, CMoM and CML performed equally well. Performance in terms of actual coverage of the associated confidence sets for high confidence levels was satisfactory and nearly independent of series length. Coverage for lower confidence levels was very conservative due to the discrete nature of the CP variable. For all distributions, the confidence curves produced by CLMo, CMoM, and CML were very close to each other, so using CLMo or CMoM instead of CML does not result in loss of quality.

In this article Un was introduced to provide a summary of the uncertainty about the CP location shown by a confidence curve. The amount of uncertainty about the CP location decreased with increasing series length and/or with increasing size of the change at the CP. Given that a longer series provides more data and that a larger change should be easier to distinguish from random noise, this was to be expected. Preliminary findings suggest that a test based on Un may perform on a par with the classical Pettitt test as long as the series are not too short and the change is large enough. This would combine in one method a null hypothesis test and confidence set estimates of CP location at multiple confidence levels.

When applied to measurement series from literature, all methods produced results compatible with the results reported in the literature. In fact, in all cases where the literature reported one or more CPs, the lowest point on the confidence curve coincided with one of those CPs. A somewhat surprising, but most welcome result was that the choice of distribution (Gumbel, gamma, or log-normal) used to calculate the likelihood had very little influence on the ability of the methods to recover information on a possible CP from the measurement series. To see if this holds more generally, more experiments with both synthetic data series and measurement time series are planned.

Both the experiments on synthetic data series and the results for measurement time series suggest that the series should have a length of about 100 points, and changes in the mean are detected if they exceed one standard deviation. For changes in the standard deviation, more experiments are needed to see how absolute or relative size of the change influences the method sensitivity.

The two new methods, CLMo and CMoM, introduced in this article complement the AED-BP method from Zhou et al. (2020). The AED-BP method has as advantage that it is non-parametric and relatively fast, but it tends to generate confidence curves with somewhat larger, and therefore less informative, confidence sets. Moreover, it needs an additional calculation to properly detect changes in standard deviation. A viable approach would be to start with AED-BP, apply CLMo when the results are not conclusive or a change in the mean is not expected, and use the more expensive CML method when the CLMo results still display large uncertainty.

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.

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.

Data availability

The authors do not have permission to share data

Acknowledgments

This work was partially developed within the framework of the Panta Rhei research initiative (Montanari et al., 2013; McMillan et al., 2016) 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'. 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 methods. The authors also thank A. Viglione (Politecnico di Turino) for his valuable comments on a draft of this article. Our special thanks to T. Iliopoulou and an anonymous reviewer for their comments, insightful questions and suggestions.

Funding

This work was supported by the China Scholarship Council under Grant number 201706710004.

Appendix A. Confidence curves

In the literature, the definition of confidence curves 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'. Schweder and Hjort (2016, Definition 4.3) gave a more general abstract definition of a confidence curve. A variation on that definition is given below. In this paper γ is used to denote a confidence level, as opposed to Cunen et al. (2018) and Zhou et al. (2020) where α is used. For more details on confidence curves, please consult Zhou et al. (2020, Appendix A).

Definition 1. If λ is a parameter of a random sample X, then a *confidence set* for λ with *confidence level* γ is a random set $R(X)$ such that

$$\Pr(\lambda \in R(X)) = \gamma \tag{A.1}$$

where $\Pr(E)$ denotes the probability of event E.

A confidence set is a generalized confidence interval. The standard concepts that hold for confidence intervals can therefore be extended to confidence sets. A confidence set with confidence level γ has a *nominal coverage probability* of γ; in other words, it is constructed to contain λ with a probability γ. However, the construction may use approximations. Therefore, the *actual coverage probability* is defined, which is the probability that λ is in $R(X)$ when the set is actually constructed for a sample taken from X. The actual coverage probability can be estimated by MC experiments. If the actual coverage probability is smaller than nominal coverage probability γ, the set is called *permissive*; when the actual coverage probability is larger than γ, the set is called *conservative*.

Definition 2. Suppose X is a random sample of size n, and λ is a parameter of the random sample 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
$$\min_{\lambda \in V} g(\lambda, x) = g\left(\hat{\lambda}(x), x\right) = 0 \tag{A.2}$$
for all realizations x of X.
2. For the true value λ_{true} of the property λ, the RV $g\left(\lambda_{\text{true}}, X\right)$ has the uniform distribution on the unit interval.

The estimate $\hat{\lambda}(x)$ is merely a reference point. It is the confidence curve as a whole that is meaningful. If $\text{cc}(\cdot,\cdot)$ is a confidence curve according to Definition 2, then for fixed λ_0 the function $\text{cc}\left(\lambda_0, X\right)$ is a RV. Hence, we can speak of the distribution of $\text{cc}\left(\lambda_0, X\right)$. If λ_{true} is the true value of λ, then, for a given confidence level $\gamma \in [0,1]$, $\text{cc}\left(\lambda_{\text{true}}, X\right)$ is uniformly distributed on $[0,1]$. Therefore,

$$\Pr\left(\text{cc}\left(\lambda_{\text{true}}, X\right) \le \gamma\right) = \gamma \tag{A.3}$$

Now define confidence sets $R_\gamma(X) = \{\lambda : \text{cc}(\lambda, X) \le \gamma\}$ with a confidence level γ. By definition $\lambda_{\text{true}} \in R_\gamma(X)$ if and only if $\text{cc}\left(\lambda_{\text{true}}, X\right) \le \gamma$, so

$$\Pr\left(\lambda_{\text{true}} \in R_\gamma(X)\right) = \Pr\left(\text{cc}(\lambda_{\text{true}}, X) \le \gamma\right) = \gamma \tag{A.4}$$

Note that the confidence sets $R_\gamma(X)$ are nested sets because they are derived from a confidence curve.

Appendix B. A summary of total uncertainty for a confidence curve

To have a reference for the size of confidence sets for CPs we introduce the following notation. For a set S with a finite number of elements, let $\#S$ denote the number of elements and let $\text{Choice}(k; S)$ denote a random set obtained by drawing k elements from S at random without replacement and with equal probability of selection for each element. Now for a given fixed element $s_0 \in S$, $n = \#S$ and a given value $0 \le \gamma \le 1$, the following equations hold

$$\Pr\left(s_0 \in \text{Choice}(k; S)\right) = \frac{k}{n} \tag{B.1}$$

$$\Pr\left(s_0 \in \text{Choice}(\lceil \gamma n \rceil; S)\right) = \frac{\lceil \gamma n \rceil}{n} \ge \gamma \tag{B.2}$$

where $\lceil r \rceil = \min\{k \in \mathbb{Z} : k \ge r\}$.

Visual inspection of a confidence curve cc can give a subjective impression of the location of the CP and its uncertainty, but a more objective measure would be needed for automated analysis of large sets of time series. In the case that the CP is restricted to L_{CP} defined in (1), it is easy to define random sets such that the probability that the true CP τ_{true} lies in the set is approximately γ, independently of the properties of the sample. Simply take

$$\mathcal{R}_\gamma = \text{Choice}\left(\left\lceil \gamma\left(n - 2n_{\min} + 1\right)\right\rceil; L_{\text{CP}}\right) \tag{B.3}$$

which has a coverage probability of

$$\Pr\left(\tau_{\text{true}} \in \mathcal{R}_\gamma(X)\right) = \frac{\left\lceil \gamma\left(n - 2n_{\min} + 1\right)\right\rceil}{n - 2n_{\min} + 1} \ge \gamma \tag{B.4}$$

For a realization R_γ of a confidence set with confidence level γ for a CP restricted to L_{CP}, we take as a *summary of relative uncertainty*

$$\text{Un}\left(R_\gamma\right) = \frac{\#R_\gamma - 1}{\gamma\left(n - 2n_{\min}\right)} \tag{B.5}$$

Now for large n, the value of $\text{Un}\left(R_\gamma\right)$ is zero for a one point set, approximately one for a realization of $\mathcal{R}_\gamma$ and larger than one for very uninformative sets.

This measure is useful for a set at a given confidence level, but for automated analysis of a cc a level needs to be selected. The highest confidence level for which a non-trivial $\mathcal{R}_\gamma$ can be constructed for which the equals sign holds is

$$\gamma_{\max} = \frac{n - 2n_{\min}}{n - 2n_{\min} + 1} \tag{B.6}$$

so that is the level that will be used. For Un applied to a confidence curve is defined as (leaving out the factor $1/\gamma_{\max}$)

$$\text{Un}(\text{cc}) = \frac{\left(\sum_{k=n_{\min}}^{n-n_{\min}} \mathbf{1}_{\text{cc}(k) \le \gamma_{\max}}\right) - 1}{n - 2n_{\min}} \tag{B.7}$$

Appendix C. The effect of shifting or scaling the time series on the confidence curve

C.1. Location-scale distribution families

If the pdf $f(x;\theta)$ with $\theta = (\xi, \varsigma)$ is of the form

$$f(x;\theta) = \frac{1}{\varsigma} g\left(\frac{x - \xi}{\varsigma}\right) \tag{C.1}$$

where ξ is the location and ς is the scale, then for the CML method it can be shown that the deviance function $D_{\text{prof}}(\tau, ay + b)$ is equal to $D_{\text{prof}}(\tau, y)$. For the CMoM and CLMo methods, a similar equality holds for D_{pseu} under the condition that the estimates of the parameters satisfy

$$\tilde{\xi}(ay + b) = a\tilde{\xi}(y) + b \tag{C.2}$$

$$\tilde{\varsigma}(ay + b) = a\tilde{\varsigma}(y) \tag{C.3}$$

If $D(\tau, ay + b) = D(\tau, y)$, then tests on synthetic time series with $\theta_L = (0, 1)$ while varying θ_R are representative for the performance of the method.

C.2. Distribution families with a scale parameter

If the pdf $f(x;\theta)$ with $\theta = (\varsigma, \eta)$ is of the form

$$f(x;\theta) = \frac{1}{\varsigma} g\left(\frac{x}{\varsigma}; \eta\right) \tag{C.4}$$

where ς is the scale and η is a shape parameter, then for the CML method it can be shown that $D_{\text{prof}}(\tau, ay) = D_{\text{prof}}(\tau, y)$. For the CMoM and CLMo methods a similar equality holds for D_{pseu} under the condition that the estimates of the parameters satisfy

$$\tilde{\varsigma}(ay) = a\tilde{\varsigma}(y) \tag{C.5}$$

$$\tilde{\eta}(ay) = \tilde{\eta}(y) \tag{C.6}$$

If $D(\tau, ay) = D(\tau, y)$, then tests on synthetic time series with $\theta_{\text{L},1} = \varsigma_{\text{L}} = 1$ while varying the other parameters are representative for the performance of the method.

Appendix D. Details on pdfs and parameter estimates

D.1. Location-scale distribution families

If the pdf $f(x;\theta)$ with $\theta = (\xi, \varsigma)$ is of the form

$$f(x;\theta) = \frac{1}{\varsigma} g\left(\frac{x - \xi}{\varsigma}\right) \tag{D.1}$$

where ξ is the location and ς is the scale, then for the CML method it can be shown that the deviance function $D_{\text{prof}}(\tau, ay + b)$ is equal to $D_{\text{prof}}(\tau, y)$. For the CMoM and CLMo methods, a similar equality holds for D_{pseu} under the condition that the estimates of the parameters satisfy

$$\tilde{\xi}(ay + b) = a\tilde{\xi}(y) + b \tag{D.2}$$

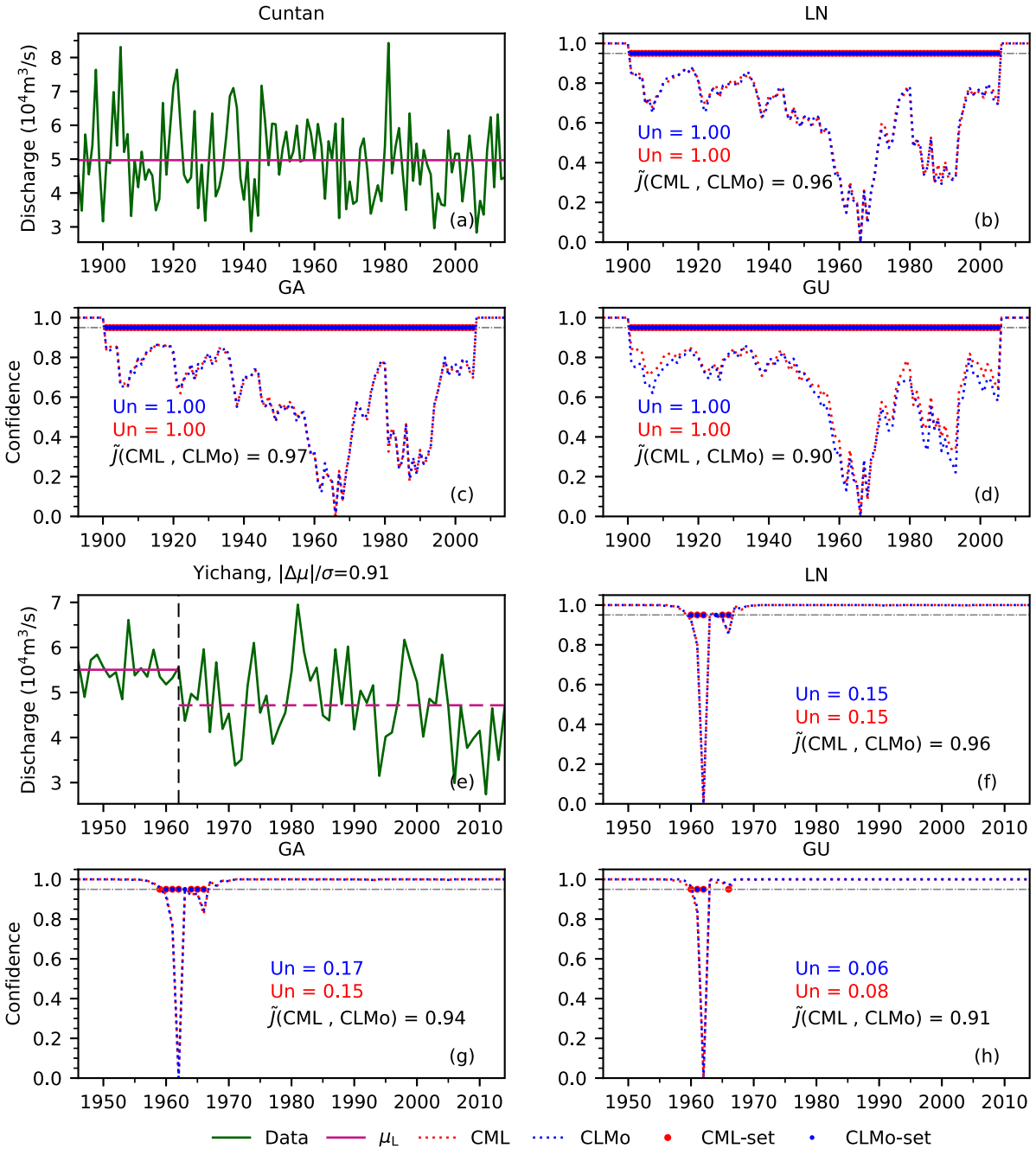

Fig. F.12. Confidence curves for CP in annual maximum discharge time series of Cuntan (a–d) and Yichang (e–h).

$$\tilde{\varsigma}(ay+b) = a\tilde{\varsigma}(y) \tag{D.3}$$

If $D(\tau, ay+b) = D(\tau, y)$, then tests on synthetic time series with $\theta_L = (0,1)$ while varying θ_R are representative for the performance of the method.

D.2. Distribution families with a scale parameter

If the pdf $f(x;\theta)$ with $\theta = (\varsigma, \eta)$ is of the form

$$f(x;\theta) = \frac{1}{\varsigma} g\left(\frac{x}{\varsigma};\eta\right) \tag{D.4}$$

where ς is the scale and η is a shape parameter, then for the CML method it can be shown that $D_{\text{prof}}(\tau, ay) = D_{\text{prof}}(\tau, y)$. For the CMoM and CLMo methods a similar equality holds for D_{pseu} under the condition that the estimates of the parameters satisfy

$$\tilde{\varsigma}(ay) = a\tilde{\varsigma}(y) \tag{D.5}$$

$$\tilde{\eta}(ay) = \tilde{\eta}(y) \tag{D.6}$$

If $D(\tau, ay) = D(\tau, y)$, then tests on synthetic time series with $\theta_{L,1} = \varsigma_L = 1$ while varying the other parameters are representative for the performance of the method.

Appendix E. The computational cost of the methods

For a time series of length n, with N MC runs for distribution approximation, all three methods (CML, CLMO, CMoM) need $(1+N)\left(n - 2n_{\min} + 1\right)$ deviance calculations. Each deviance calculation needs

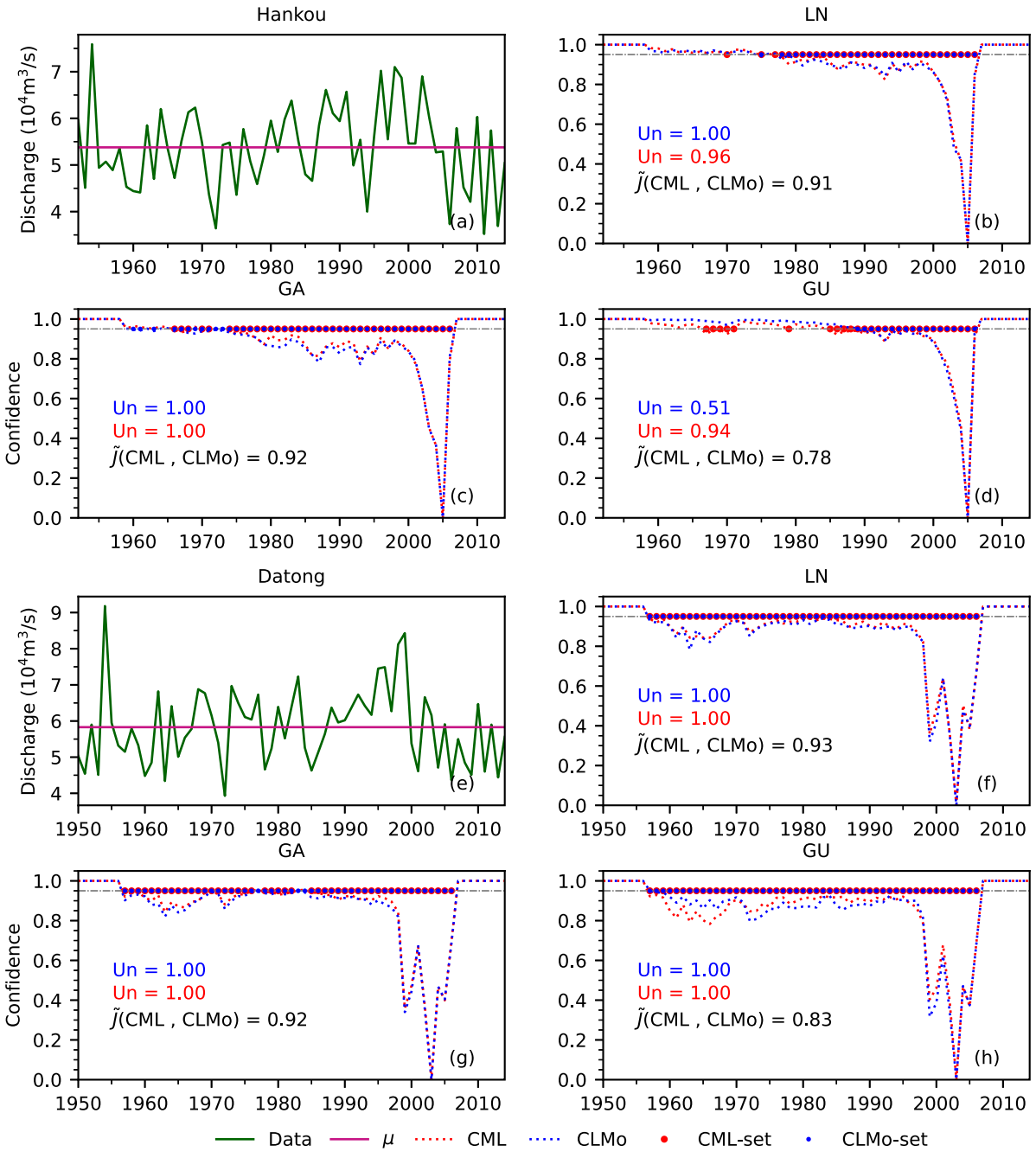

Fig. F.13. Confidence curves for CP in annual maximum discharge time series of Hankou (a–d) and Datong (e–h).

$2\left(n-2n_{\min}+1\right)$ parameter estimates and $\left(n-2n_{\min}+1\right)\times n$ calculations of the logarithm of the pdf. Each pair of ML parameter estimates will need at least n calculations of the logarithm of the pdf. The costs of a pair of MoM or LMo parameter estimates may be lower, but will still be on the order of n arithmetic operations. A relatively big difference in cost occurs for those distributions where ML needs to solve a minimization problem, while MoM and LMo provide explicit formulas. For all methods the total number of operations for one sample will be

$$O\left(\left(1+N\right)\left(n-2n_{\min}+1\right)^2 n\right) \tag{E.1}$$

where O stands for 'on the order of'. The difference in cost between the methods does not show up in the O notation, because it arises from multiplication factors that do not depend on n. A MoM or LMo parameter estimate involves on the order of n additions and multiplications plus a constant number of more complex operations. An ML estimate where the solution is not available in explicit form will involve solving a minimization problem; this in turn may involve between 5 and 20 evaluations of expressions derived from the log-likelihood. While these evaluations are order n in the operations count, they are likely to be more costly (perhaps a factor of 2 to 10) than the order n addition and multiplication operations needed by MoM and LMo. So, in theory ML may well take anywhere from 10 to 200 times as long. For GU and GA, where the ML problems correspond to a one dimensional search for the point where a nonlinear function is zero, in practice the cost of ML was between 8 and 11 times that of CMoM.

With $n = 100$ and $N = 1000$, the computational cost is not negligible. When one of these methods is itself analyzed statistically, for instance, by studying $M \geq 1000$ time series, this becomes a major problem. For $n_{\min} = 1$, $n = 100$ and $M = N = 1000$ the cost exceeds $O\left(10^{12}\right)$

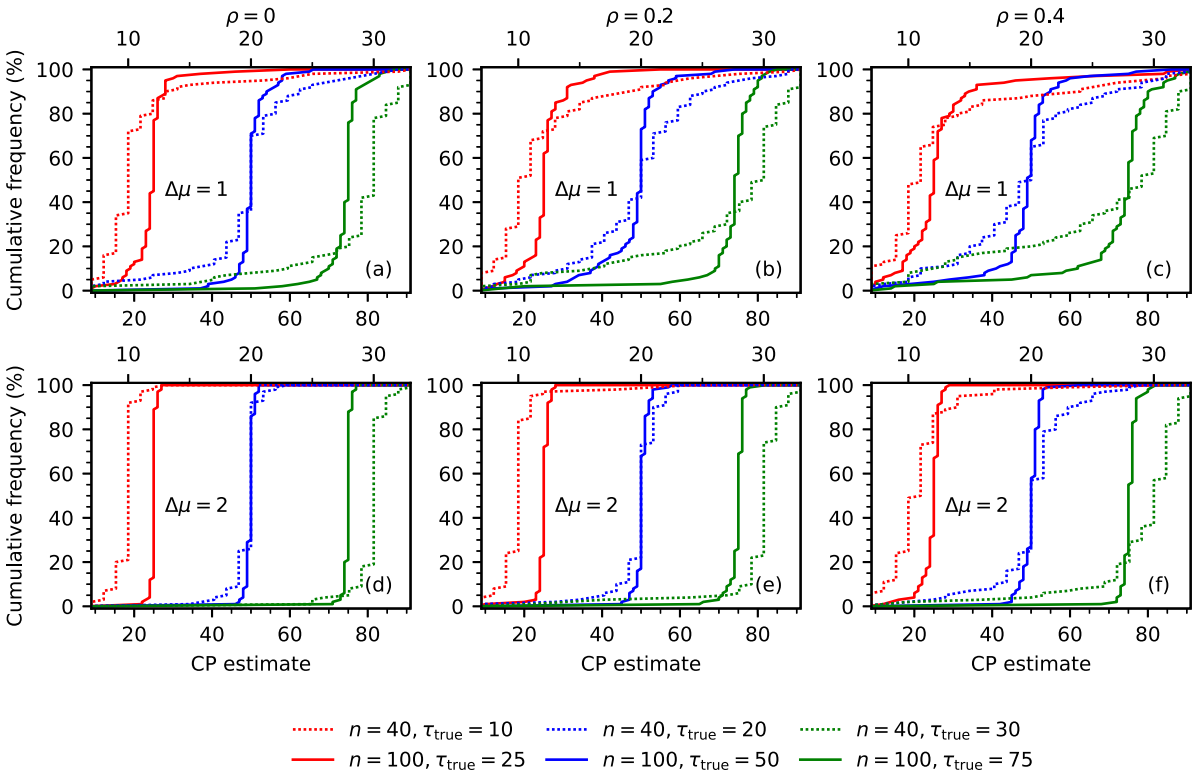

Fig. G.14. The cumulative frequency distribution of CPs when there is a change in mean in an AR(1) LN series with lag-one autocorrelation ρ.

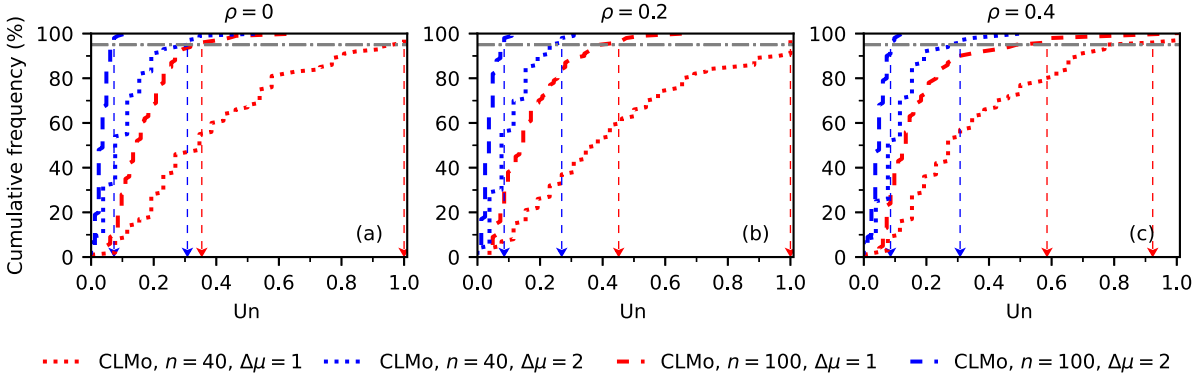

Fig. G.15. Cumulative frequency of Un for a change in the mean and a CP in the middle of an AR(1) LN series with lag-one autocorrelation ρ.

calculations of the logarithm of the pdf. In practice, for one series taken from the GA distribution with $n = 100$, $n_{\min} = 9$ and $N = 1000$, a CML curve for one series took 314 seconds and CMoM took 37 seconds. Counting flops is complicated by the presence of the log and Gamma functions. Calculating flop rates is difficult because the current implementation is in Matlab®, not in C or Fortran. Moreover, runs for different parameter sets were done in parallel. The calculations were performed on a six core Intel® Xeon® W-2133 at 3.60 GHz. A rough estimate of the code performance would be between 0.04 (CLMo) and 0.4 (CML) GFlops per core; Intel (2020) gives an Adjusted Peak Performance (APP) of 160 GFlops, so about 27 GFlops per core. In theory, there is room for improvement, but to verify this, an optimized implementation in a compiled language would be needed.

Appendix F. Case study 4

Four time series of annual maximum discharge on the Yangtze River in China were analyzed by classical methods (Pettitt, CUSUM, Cramér–von Mises) in Zhou et al. (2019) and by nonparametric confidence curves in Zhou et al. (2020). The stations Cuntan, Yichang, Hankou and Datong along the Yangtze River were selected to examine the impacts from the construction of the Three Georges dam. Construction officially started in 1994. There followed a series of interventions in the flow of the Yangtze River, first by partial damming, and then by the filling, in stages, of the reservoir. Construction was completed in 2009, but the reservoir was not yet completely filled at that point.

For Cuntan, which lies upstream of the Three Gorges dam, a time series of annual maximum flow from 1893 to 2014 was examined. Earlier studies did not find clear CPs. All confidence curves in Fig. F.12(b–d) show that there is no clear indication of a CP. All Un values are near one, this strongly suggests that there is no CP.

For Yichang, which lies about 40 km downstream of the Three Gorges dam, a time series of annual maximum flow from 1946 to 2014 was examined. An earlier study found a possible CP in 1962 (Xie et al., 2014). In Zhou et al. (2019) CUSUM found a CP in 1962, while Pettitt and Cramér–von Mises found a CP in 1966. All confidence curves in Fig. F.12(f–h) show that there is a clear CP near 1962. At the 95% confidence level the LN and GA based methods provide a set with

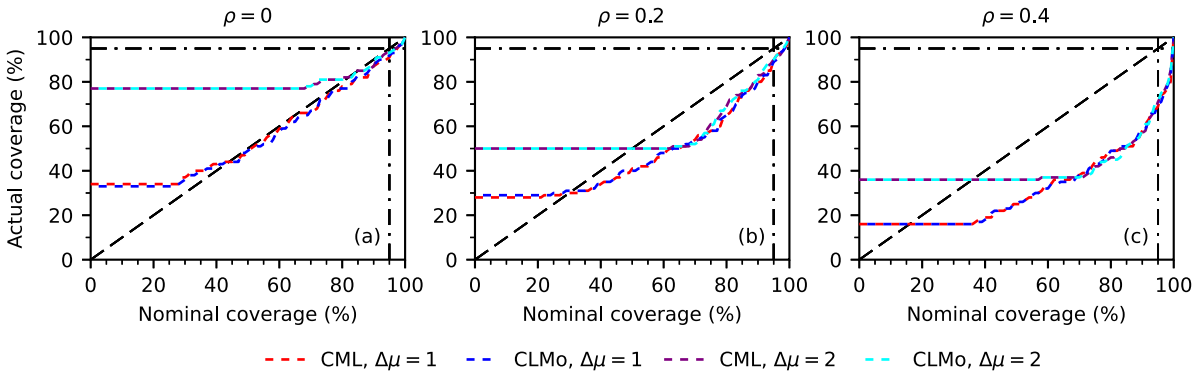

Fig. G.16. Actual versus nominal coverage probability for a change in the mean in an AR(1) LN series with lag-one autocorrelation ρ when $\tau_{\text{true}} = n/2$ and series length $n = 100$.

about 7 candidates, while for GU, CML selects 4 years and CLMo selects 2 years. The value of $|\Delta\mu|/\sigma$ at the CP in 1962 is near one.

For Hankou, approximately 700 km downstream of the Three Gorges dam, a time series of annual maximum flow from 1950 to 2014 was examined. Earlier studies did not find clear CPs. All methods, except for CLMo with GU, have Un close to one, see Fig. F.13(b–d). However, given the closeness of the lowest point on the confidence curve to the end of the series and the narrowness of the 80% confidence set, more data is needed to decide whether there is a CP near 2005 or not.

For Datong, about 1200 km downstream of the Three Gorges dam, a time series of annual maximum flow from 1952 to 2014 was examined. Earlier studies did not find clear CPs. Again, all methods have a Un that is nearly one, see Fig. F.13(f–h). The shape of the confidence curve suggests that more data is needed to decide whether or not there is a CP near 2003.

Appendix G. Effects of short range dependence

Autoregressive time series of order one (AR(1)) with an underlying LN distribution were generated according to the procedure used by Vogel et al. (1998) summarized below. First an AR(1) model is used to generate a time series Y'_k

$$Y'_k = \phi Y'_{k-1} + G_k \text{ for } k = 1, 2, \ldots \tag{G.1}$$

where Y'_0 is normally distributed with mean μ' and standard deviation σ', the G_k are i.i.d. according to a normal distribution with mean $\mu' = (1-\phi)\,\mu'$ and standard deviation $\sigma'\sqrt{1-\phi^2}$, and Y'_0 is independent of all G_k. Next a time series Y_k is obtained by setting

$$Y_k = \exp\left(Y'_k\right) \tag{G.2}$$

If

$$\mu' = \log \frac{\mu}{\sqrt{1+\frac{\sigma^2}{\mu^2}}} \tag{G.3}$$

$$\sigma' = \sqrt{\log\left(1+\frac{\sigma^2}{\mu^2}\right)} \tag{G.4}$$

$$\phi = \frac{\log\left(\rho\left[\exp\left(\left[\sigma'\right]^2\right)-1\right]+1\right)}{[\sigma']^2} \tag{G.5}$$

then Y_k will have a LN distribution with mean μ and standard deviation σ. The series Y_k will have lag-one autocorrelation ρ. Eq. (G.3) differs from Eq. 2 in Vogel et al. (1998) because of a typing error. When needed, a CP at k is introduced by starting a new series with initial value y_k, but μ' and σ' derived from μ_R and σ_R.

These are preliminary results for groups of 100 samples instead of the groups of 1000 samples used in the main paper. Samples were generated using the 'multFibonacci' random number generator from Matlab. Fig. G.14 shows that the confidence curves still have their minimum close to the actual CP. The algorithms used the 'multFibonacci' generator from Matlab to obtain random numbers. Figs. G.15 and G.16 show, that due to the difference between the assumed i.i.d. LN distribution and the actual AR(1) LN distribution, the actual coverage becomes increasingly permissive and the uncertainty about the CP increases.

References

Aue, A., Horváth, L., 2012. Structural breaks in time series. J. Time Series Anal. 34 (1), 1–16. http://dx.doi.org/10.1111/j.1467-9892.2012.00819.x.

Beaulieu, C., Chen, J., Sarmiento, J.L., 2012. Change-point analysis as a tool to detect abrupt climate variations. Phil. Trans. R. Soc. A 370 (1962), 1228–1249.

Belisle, P., Joseph, L., MacGibbon, B., Wolfson, D.B., du Berger, R., 1998. Change-point analysis of neuron spike train data. Biometrics 54 (1), 113. http://dx.doi.org/10.2307/2534000.

Berkes, I., Horváth, L., Kokoszka, P., Shao, Q.-M., 2006. On discriminating between long-range dependence and changes in mean. Ann. Statist. 34 (3), http://dx.doi.org/10.1214/009053606000000254.

Birnbaum, A., 1961. Confidence curves: An omnibus technique for estimation and testing statistical hypotheses. J. Amer. Statist. Assoc. 56 (294), 246–249.

Brodsky, E., Darkhovsky, B.S., 2013. Nonparametric Methods in Change Point Problems. Springer Science & Business Media.

Chen, J., Gupta, A.K., 2001. On change point detection and estimation. Comm. Statist. Simulation Comput. 30 (3), 665–697.

Chen, J., Gupta, A.K., 2011. Parametric Statistical Change Point Analysis: With Applications to Genetics, Medicine, and Finance. Springer Science & Business Media.

Chu, P.-S., Zhao, X., 2004. Bayesian change-point analysis of tropical cyclone activity: The Central North Pacific case. J. Clim. 17 (24), 4893–4901. http://dx.doi.org/10.1175/jcli-3248.1.

Cong, Z., Shahid, M., Zhang, D., Lei, H., Yang, D., 2017. Attribution of runoff change in the alpine basin: A case study of the Heihe Upstream Basin, China. Hydrol. Sci. J. 62 (6), 1013–1028. http://dx.doi.org/10.1080/02626667.2017.1283043.

Conte, L.C., Bayer, D.M., Bayer, F.M., 2019. Bootstrap Pettitt test for detecting change points in hydroclimatological data: Case study of Itaipu Hydroelectric Plant, Brazil. Hydrol. Sci. J. 64 (11), 1312–1326.

Cunen, C., Hermansen, G., Hjort, N.L., 2018. Confidence distributions for change-points and regime shifts. J. Statist. Plann. Inference 195, 14–34. http://dx.doi.org/10.1016/j.jspi.2017.09.009.

Delicado, P., Goria, M.N., 2008. A small sample comparison of maximum likelihood, moments and L-moments methods for the asymmetric exponential power distribution. Comput. Stat. Data Anal. 52 (3), 1661–1673. http://dx.doi.org/10.1016/j.csda.2007.05.021.

Eastwood, V.R., 1993. Some nonparametric methods for changepoint problems. Canad. J. Statist./ la Revue Canadienne de Statistique 21 (2), 209–222, URL: http://www.jstor.org/stable/3315813.

Gong, G., Samaniego, F.J., 1981. Pseudo maximum likelihood estimation: Theory and applications. Ann. Statist. 861–869.

Haktanir, T., 1991. Statistical modelling of annual maximum flows in Turkish rivers. Hydrol. Sci. J. 36 (4), 367–389. http://dx.doi.org/10.1080/02626669109492520.

Hamed, K., Rao, A.R., 2019. Flood Frequency Analysis. CRC Press.

Harrigan, S., Murphy, C., Hall, J., Wilby, R.L., Sweeney, J., 2014. Attribution of detected changes in streamflow using multiple working hypotheses. Hydrol. Earth Syst. Sci. 18 (5), 1935–1952. http://dx.doi.org/10.5194/hess-18-1935-2014.

Hušková, M., Kirch, C., 2008. Bootstrapping confidence intervals for the change-point of time series. J. Time Series Anal. 29 (6), 947–972. http://dx.doi.org/10.1111/j.1467-9892.2008.00589.x.

Intel, 2020. APP Metrics for Intel® Microprocessors: Intel® Xeon® Processor. Technical Report Rev. 3., Intel Corporation, Santa Clara, CA.

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

Karim, M.A., Chowdhury, J.U., 1995. A comparison of four distributions used in flood frequency analysis in Bangladesh. Hydrol. Sci. J. 40 (1), 55–66. http://dx.doi.org/10.1080/02626669509491390.

Kolokytha, E., Oishi, S., Teegavarapu, R.S.V. (Eds.), 2017. Sustainable Water Resources Planning and Management Under Climate Change. Springer Singapore, http://dx.doi.org/10.1007/978-981-10-2051-3.

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. Hydrol. Sci. J. 49 (1), 7–19.

Landwehr, J.M., Matalas, N.C., Wallis, J.R., 1979. Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles. Water Resour. Res. 15 (5), 1055–1064. http://dx.doi.org/10.1029/WR015i005p01055.

Lettenmaier, D.P., Burges, S.J., 1982. Gumbel's extreme value I distribution: A new look. J. Hydraul. Eng. 108 (4), 502–514.

McMillan, H., Montanari, A., Cudennec, C., Savenije, H., Kreibich, H., Krueger, T., Liu, J., Mejia, A., Loon, A.V., Aksoy, H., Di Baldassarre, G., 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., Veena, S., Toth, E., van Nooyen, R., Xia, J., 2016. Panta Rhei 2013–2015: Global perspectives on hydrology, society and change. Hydrol. Sci. J. 61 (7), 1174–1191. http://dx.doi.org/10.1080/02626667.2016.1159308.

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., Di Baldassarre, G., Yu, B., Hubert, P., Huang, Y., Schumann, A., Post, D.A., V., S., Harman, C., Thompson, S., Rogger, M., Viglione, A., McMillan, H., Characklis, G., Pang, Z., Belyaev, V., 2013. "Panta Rhei–Everything flows": Change in hydrology and society—The IAHS scientific decade 2013–2022. Hydrol. Sci. J. 58 (6), 1256–1275. http://dx.doi.org/10.1080/02626667.2013.809088.

Nuzzo, R., 2014. Scientific method: Statistical errors. Nature 506 (7487), 150–152. http://dx.doi.org/10.1038/506150a.

Perreault, L., Bernier, J., Bobée, B., Parent, E., 2000. Bayesian change-point analysis in hydrometeorological time series. Part 1. The normal model revisited. J. Hydrol. 235 (3), 221–241. http://dx.doi.org/10.1016/S0022-1694(00)00270-5.

Perreault, L., Haché, M., Slivitzky, M., Bobée, B., 1999. Detection of changes in precipitation and runoff over eastern Canada and U.S. using a Bayesian approach. Stoch. Environ. Res. Risk Assess. (SERRA) 13 (3), 201–216. http://dx.doi.org/10.1007/s004770050039.

Pettitt, A.N., 1979. A non-parametric approach to the change-point problem. J. R. Stat. Soc. Ser. C. Appl. Stat. 28 (2), 126–135.

Rea, W., Reale, M., Brown, J., Oxley, L., 2011. Long memory or shifting means in geophysical time series? Math. Comput. Simulation 81 (7), 1441–1453. http://dx.doi.org/10.1016/j.matcom.2010.06.007.

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

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. Cambridge University Press, http://dx.doi.org/10.1017/CBO9781139046671.

Sheskin, D.J., 2003. Handbook of Parametric and Nonparametric Statistical Procedures, third ed. Chapman and Hall/CRC, http://dx.doi.org/10.1201/9781420036268.

Tao, H., Gemmer, M., Bai, Y., Su, B., Mao, W., 2011. Trends of streamflow in the Tarim River Basin during the past 50 years: Human impact or climate change? J. Hydrol. 400 (1–2), 1–9.

Teegavarapu, R., 2018. Trends and Changes in Hydroclimatic Variables: Links to Climate Variability and Change. Elsevier.

Thompson, S.A., 2017. Hydrology for Water Management. CRC Press.

Vogel, R.M., Tsai, Y., Limbrunner, J.F., 1998. The regional persistence and variability of annual streamflow in the United States. Water Resour. Res. 34 (12), 3445–3459. http://dx.doi.org/10.1029/98wr02523.

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

Xie, H., Li, D., Xiong, L., 2014. Exploring the ability of the Pettitt method for detecting change point by Monte Carlo simulation. Stoch. Environ. Res. Risk Assess. 28 (7), 1643–1655.

Xiong, L., Guo, S., 2004. Trend test and change-point detection for the annual discharge series of the Yangtze River at the Yichang hydrological station / Test de tendance et détection de rupture appliqués aux séries de débit annuel du fleuve Yangtze à la station hydrologique de Yichang. Hydrol. Sci. J. 49 (1), 99–112. http://dx.doi.org/10.1623/hysj.49.1.99.53998.

Zhou, C., van Nooijen, R., Kolechkina, A., van de Giesen, N., 2020. Confidence curves for change points in hydrometerological time series. J. Hydrol. 590, 125503. http://dx.doi.org/10.1016/j.jhydrol.2020.125503.

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