Use of confidence curves to represent uncertainty in hydro-meteorological time series analysis

Zhou, C.

DOI
10.4233/uuid:f495093f-c5e1-4133-b136-76aae1ae2ac7

Publication date
2022

Document Version
Final published version

Citation (APA)
Zhou, C. (2022). *Use of confidence curves to represent uncertainty in hydro-meteorological time series analysis.* [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:f495093f-c5e1-4133-b136-76aae1ae2ac7

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

USE OF CONFIDENCE CURVES TO REPRESENT UNCERTAINTY IN HYDRO-METEOROLOGICAL TIME SERIES ANALYSIS

USE OF CONFIDENCE CURVES TO REPRESENT UNCERTAINTY IN HYDRO-METEOROLOGICAL TIME SERIES ANALYSIS

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology,
by the authority of the Rector Magnificus prof.dr.ir.T.H.J.J. van der Hagen,
chair of the Board for Doctorates
to be defended publicly on Thursday 13 October 2022 at 10:00 o'clock

by

Changrang ZHOU

Master of Engineering in Hydrology and Water Resources,
Hohai University, China,
born in Chuzhou, China.

This dissertation has been approved by the promotors

Promotor: Prof. dr. Ir. N. C. van de Giesen
Copromotor: Dr. R. R. P. van Nooijen

Composition of the doctoral committee:

Rector Magnificus,	Chairman
Prof. dr. Ir. N. C. van de Giesen,	Delft University of Technology
Dr. R. R. P. van Nooijen,	Delft University of Technology
Independent members:	
Prof. dr. Ir. M. Kok,	Delft University of Technology
Prof. rer. nat. Dr.-Ing A. Bárdossy,	University of Stuttgart
Prof. dr. Ir. P. H. A. J. M. van Gelder,	Delft University of Technology
Dr. hab. K. Kochanek,	Institute of Geophysics Polish Academy of Sciences
Dr. -Ir. G.H.W. Schoups,	Delft University of Technology

This research is funded by Delft University of Technology (TU Delft) and China Scholarship Council (CSC) under Grant 201706710004

keywords:	Confidence curves; Confidence sets; Uncertainty
Published & distributed by:	Changrang Zhou
Front & Back by:	Changrang Zhou

ISBN 000-00-0000-000-0

If I have seen further it is by standing on the shoulders of Giants.

Isaac Newton (1642-1727)

CONTENTS

SUMMARY

Climate change is incompatible with the assumption of stationarity. This has lead to a sharp increase in the detection and study of nonstationarity in hydro-meteorological processes. Most hydro-meteorological processes are still analyzed by studying time series of observations.

From the perspective of statistical characteristics, a stationary time series does not show significant changes. On the contrary, a nonstationarity time series often shows a slowly increasing/decreasing trend or a sudden change. A sudden change or a change point is a time point that a time series shows a great change in its statistical characteristics, for instance in the mean or the standard deviation.

For stationary cases, hydrologists have a large number of statistical tools to analyse these time series. These tools can not only help hydrologists to gain a deep insight into time series, but they can also analyse the corresponding uncertainty. For nonstationary cases, the detection of changes has drawn the majority of attention, however, the uncertainty associated with the detection has still been rarely studied. Therefore, this PhD research aims at bridging the gap between nonstationarity detection and the uncertainty of detection. To be more specific the main scope is rooted in analysing the uncertainty associated with detecting a change point in hydro-meteorological time series.

When it comes to representing uncertainties, a traditional choice is using a confidence interval with a certain confidence level. In this research instead, the uncertainty is represented by confidence curves because they are capable of capturing more information by including all confidence intervals at all confidence levels and they visualize uncertainty in a curve.

To verify the general applicability of a confidence curve in representing uncertainties, both a discrete parameter and a continuous parameter are considered in this research. The location of a change point is considered as a discrete parameter, and the dependence parameter in copula models will be considered as a continuous one. Additionally, in order to simplify the construction of a confidence curve, several new approaches have been presented in this research.

Based on results and findings, confidence curves have been proven to be more informative and theoretically they can represent uncertainties of all types of parameter of interest. With a confidence curve, hydrologists can easily read the uncertainty of the detected change point and this would also provide decision-makers a better insight into the nonstationarity of a time series of a hydro-meteorological observations.

SAMENVATTING

Klimaatverandering is onverenigbaar met de aanname van stationariteit. Dit heeft geleid tot een sterke toename van de detectie en bestudering van niet-stationariteit in hydro-meteorologische processen. Het grootste deel van de hydro-meteorologische processen wordt nog steeds geanalyseerd aan de hand van tijdreeksen van waarnemingen.

Vanuit het oogpunt van statistische karakteristieken vertoont een stationaire tijdreeks geen significante veranderingen. Een tijdreeks met niet-stationariteit daarentegen vertoont vaak een langzaam stijgende/dalende tendens of een plotselinge verandering. Een plotselinge verandering of een "change point" is een tijdstip waarop een tijdreeks een grote verandering in zijn statistische kenmerken vertoont, bijvoorbeeld in het gemiddelde of de standaardafwijking.

Voor stationaire gevallen beschikken hydrologen over een grote hoeveelheid statistische hulpmiddelen om deze tijdreeksen te analyseren. Deze hulpmiddelen kunnen hydrologen niet alleen helpen om een diep inzicht te krijgen in tijdreeksen, maar ook om de bijbehorende onzekerheid te analyseren. Voor niet-stationaire gevallen heeft de detectie van veranderingen de meeste aandacht getrokken, maar de onzekerheid die gepaard gaat met de detectie is nog steeds zelden bestudeerd. Daarom beoogt dit proefschrift een brug te slaan tussen de detectie van niet-stationariteit en de onzekerheid van de detectie. Meer specifiek gaat het om de analyse van de onzekerheid die gepaard gaat met de detectie van een veranderingspunt in hydro-meteorologische tijdreeksen.

Wanneer het erom gaat onzekerheden weer te geven, is een traditionele keuze het gebruik van een betrouwbaarheidsinterval met een bepaald betrouwbaarheidsniveau. In dit onderzoek wordt de onzekerheid daarentegen weergegeven door betrouwbaarheidscurven, omdat deze meer informatie kunnen weergeven door alle betrouwbaarheidsintervallen met elk betrouwbaarheidsniveau op te nemen en omdat zij de onzekerheid visualiseren in een curve.

Om de algemene toepasbaarheid van een vertrouwenscurve bij de weergave van onzekerheden te verifiëren, worden in dit onderzoek zowel een discrete parameter als een continue parameter in aanmerking genomen. De plaats van een veranderingspunt wordt beschouwd als een discrete parameter, en de afhankelijkheidsparameter in copula modellen wordt beschouwd als een continue parameter. Om de constructie van een vertrouwenscurve te vereenvoudigen, worden in dit onderzoek bovendien verschillende nieuwe benaderingen voorgesteld.

Op basis van resultaten en bevindingen is bewezen dat betrouwbaarheidscurven informatiever zijn en theoretisch kunnen zij onzekerheden van alle soorten parameters van belang weergeven. Met een betrouwbaarheidscurve kunnen hydrologen de onzekerheid van het gedetecteerde veranderingspunt gemakkelijk aflezen en dit zou besluitvormers ook een beter inzicht verschaffen in de niet-stationariteit van een tijdreeks van een hydro-meteorologische waarneming.

LIST OF SYMBOLS

LIST OF SYMBOLS

Symbol	Meaning
γ	Confidence level.
λ	A statistic of a random sample; λ_{true} is the true value of a statistic; $\hat{\lambda}$ is the estimate of λ.
$\hat{\lambda}$	The estimate of a statistic λ.
$Choice(k;S)$	A random set obtained by drawing k elements from S at random equally without replacement.
$R(\cdot)$	A random set obtained by drawing two elements from a random sample S. Let random sample be X, and $X = \{x_1, x_2, ..., x_m\}$. There will be $m(m-1)/2$ combinations of $R(X)$.
$R_\gamma(\cdot)$	The random set with a confidence level of γ of a random sample
cc$(\cdot)$	A confidence curve for the true value of statistic λ_{true}; when a confidence level is given, there will be at least one confidence set being generated $R_\gamma(\lambda)$; the confidence curve with a confidence level of γ at cc$(\lambda_{\text{true}}, Y)$ is approximate to confidence level γ, which is $\Pr(\text{cc}(\lambda_{\text{true}}, X) \leq \gamma) \approx \gamma$.
$F(\cdot;\theta,\zeta)$	A cumulative distribution with a location parameter θ and shape parameter ζ.
$n_{\min}$	The minimum length of resampling, $n_{\min} = 2\lfloor \log n \rfloor$, where $\lfloor\cdot\rfloor$is the running down towards the nearest integer.
$\Delta\mu$, $\Delta\sigma$	A change in sample mean of a hydrological time series.
$\Delta\sigma$	A change in standard deviance of a hydrological time series.
τ	The location of a change point.
ML	Maximum Likelihood estimate.
Pseudo ML	Estimates other than ML, for instance Method-of-Moments (MoM) and Linear-Moment method (LMo).
$\ell(\cdot)$	A log-likelihood function $\ell(\tau,\theta,\zeta;y)$, where τ is the parameter of interest, and θ,ζ are nuisance parameters. Traditionally, nuisance parameters are estimated by ML based on y, but in this study we will apply pseudo ML to estimate nuisance parameters.
$\ell_{\text{prof}}(\cdot)$	A profile loglikelihood function and in this study $\ell(\tau,\hat{\theta},\hat{\zeta};y) = \ell_{\text{prof}}(\tau,y)$.
$\ell_{\text{emp}}(\cdot)$	Empirical log-likelihood function ratio.
Λ	A likelihood ratio, and $-2\log\Lambda = 2\{\ell(\tau,y) - \ell(\hat{\tau},y)\}$. In this study, $-2\log\Lambda = \ell_{\text{prof}}(\tau,y) - \ell_{\text{prof}}(\hat{\tau},y)$.
Λ_{emp}	Empirical likelihood ratio, and $\ell_{\text{emp}}(\tau,y) = -2\log\Lambda_{\text{emp}}(\tau,y)$
$K_\tau(\cdot)$	An approximate distribution of τ based on a random sample.

$D(\tau,\cdot)$	A deviance function of τ, and it can be calculated by $D(\tau,\cdot) = 2(\ell(\hat{\tau},y) - \ell(\tau,y))$. In this study, $D(\tau,y) = 2(\ell_{\text{prof}}(\hat{\tau},y) - \ell_{\text{prof}}(\tau,y))$, and a confidence curve for τ can be approximated by $\text{cc}(\tau,y_{\text{obs}}) = K_\tau(D(\tau,y_{\text{obs}}))$.
$D_{\text{pseu}}(\tau,\cdot)$	A deviance function of τ, and the estimates of nuisance parameters are by pseudo ML.
$D_{\text{emp}}(\tau,\cdot)$	Empirical deviance function, and $D_{\text{emp}}(\tau,y) = 2(\ell_{\text{emp}}(\hat{\tau}_{\text{emp}},y) - \ell_{\text{emp}}(\tau_{\text{emp}},y))$ where $\hat{\tau}_{\text{emp}}$ the estimate of τ by ℓ_{emp}.
AED	A method to construct a confidence curve for τ by Approximate Empirical Deviance funcion, and bootstrap is used to resample from hydrological observations.
CML	Confidence curves based on maximum likelihood estimate to estimate nuisance parameters in a log-likelihood function, which uses Monte Carlo (MC) simulation to resample.
CLMo	Confidence curves based on LMo to estimate nuisance parameters in a log-likelihood function, which uses MC simutlation to resample.
$\tilde{J}$	Similarity index to measure the similarity between two curves and $\tilde{J} \in [0,1]$. It is one for idential curves, and smaller than one when curves that differ. In this study, $\tilde{J}$ is used to measure the similarity between two confidence curves generated by any two methods.
Un	Uncertainty measurements to measure the uncertainty in change point detection represent by a confidence curve, and $0 \leq \text{Un} \leq 1$. If Un is closer to 1, the uncertainty of a detected change point is higher so that we have confidence to believe $\hat{\tau}$ might not be a real change point. If Un is closer to 0, the uncertainty of a detected change point is lower. In this case the $\hat{\tau}$ could be a real change point.

1

Introduction

The book on confidence distributions
written by Tore Schweder and Nils Lid Hjort,
was given as a reference book to me.
Since then, it has been my go-to book all the time.

1

1.1. THE IMPORTANCE OF UNCERTAINTY ANALYSIS

Uncertainty analysis can provide an overview of the error in model estimation and the inaccuracies of measurements by using all available information [1]. The evaluation of uncertainty can be helpful in decision making and it has to be taken into account when interpreting the outputs of models [2]. Uncertainty analysis is also an aid for researchers to solve environmental problems and conduct associated risk assessment [3]. Water-related projects also rely on the estimation for the sake of safety, for instance hydrologists design dikes/reservoirs/dams by considering of the magnitude of the design flood with a specific return period. Government, insurance and real estate companies need to know the uncertainty and risk to make investment decisions.

The uncertainty is mainly comprised of deterministic and stochastic components, and there are many methods that can be used to estimate uncertainty, such as Monte Carlo simulations [4], bootstrap [5], frequentist [6], and Bayesian analysis [7].

The uncertainty of environmental problems becomes more complicated as a result of changes in land use, increasing urbanization and population and climate change [8–13]. Given the inherent uncertainty of the future, predictions inevitably involve statistics, and these statistics may or may not be influenced by changes in the environment. It is widely accepted that statistical models are often used as tools to solve practical problems, but uncertainties are inevitable in statistical models. Therefore, it is necessary to show the uncertainty in the outputs of statistical models.

1.2. TRADITIONAL TIME SERIES ANALYSIS BASED ON NULL HYPOTHESIS STATISTICAL TESTS

The traditional way to conduct time series analysis is based on statistical inference. Statistics developed for Null Hypothesis Statistical Testing (NHST) are very popular in a variety of sciences. The concept of statistical significance can be traced back to Edgeworth [14], and Fisher [15, first published in 1925] made it well known. It was originally a tool to indicate when a result warranted further scrutiny. Neyman and Pearson [16] used statistical significance to interpret the results of statistical inference, and hypothesis testing based on significance level became widely used.

The NHST depends on a pre-set $p-$value threshold, and it is used as an aid to report the significance of the statistical inference. However, the conclusions drawn from a NHST are often biased [17] and a $p-$value can only provide quite limited information about data and is very easily misinterpreted [18]. For instance, NHST only provides a 'Yes' of 'No' answer to whether to accept the null hypothesis or not, and it leaves no room for uncertainty analysis. An increasing number of statisticians suggested users to abandon the declaration of 'statistical significance' because the statistical significance based on $p < 0.05$ is not informative [18–20].

According to Davidian and Louis [21], significance of statistics is 'the science of learning from data, and of measuring controlling, and communicating uncertainty'. Wasserstein *et al.* [19] summarized that when it comes to statistical inference, it is recommended to 'Accept uncertainty. Be thoughtful, open and modest'. A confidence interval is a traditional way to show the uncertainty for a parameter of interest, but it only consists of two bounds of an interval at a given confidence level. Clearly, it is more informative to use

a distribution estimator than an interval estimator [22]. To better represent the uncertainty in statistical inference and to construct a distribution estimator for a parameter of interest, new statistical methods are needed. Statisticians encourage users to use confidence distributions to represent the uncertainty for a parameter of interest [22–30]. Ratnasingam and Ning [31] said "It (A confidence distribution) also can provide confidence intervals of all nominal levels for a parameter of interest through confidence curves."

In this PhD research, confidence curves will be employed to represent the uncertainty in statistical inference. A confidence curve is a variant of a confidence distribution, and it is based on an exact or approximate probability distribution of a deviance function based on a sample [32]. According to Schweder and Hjort [29, page 66], a confidence curve is a canonical graphical curve that can represent a confidence distribution, and

$$\mathrm{cc}(\psi) = \begin{cases} 1 - 2C(\psi), & \psi \leq \hat{\psi}_{0.5} \\ 2C(\psi) - 1, & \psi \geq \hat{\psi}_{0.5} \end{cases}$$

where ψ is a parameter of interest; $\mathrm{cc}(\psi)$ is a confidence curve for ψ; and $C(\psi)$ is a confidence distribution for ψ. With a confidence curve, the uncertainty can be represented graphically.

1.3. UNCERTAINTY ANALYSIS IN PARAMETER ESTIMATION

Uncertainty exists in almost all models and measurements, and in this research the focus will be on the uncertainty in estimating parameters. Generally, there are two types of parameters, discrete and continuous parameters. Therefore, this study will aim at constructing confidence curves for both discrete and continuous parameters.

1.3.1. CONFIDENCE CURVES FOR DISCRETE PARAMETERS

There are many discrete parameters in the real world, and in this study, the location of a change point in hydro-meteorological time series is taken as a parameter of interest. The location of change point is a moment in time where there is an abrupt change in one or more of the properties of the time series such as the mean, the median, or the standard deviation [33, 34].

The first study of finding change points was in product quality assessment in manufacturing [35]. With the development of change point detectors, finding change points has been widely applied in many fields, for instance, oceanography [36], economics, finance [37], biology [38] and meteorology [39, 40].

It is clear that climate change affects the hydrological cycle [41], and there is an increasing risk from extreme precipitation [42], floods [43], and heatwaves. When it comes to finding change points in hydro-meteorological time series, abundant data is often needed [44]. The number of change points can be single or multiple, but in this PhD thesis change point detection is conducted under the assumption of 'At Most One Change' (AMOC). It is noticeable that different detectors and different sample lengths might lead to different results. The traditional change point relies on NHST, but the uncertainty in the results still needs to be addressed and well represented.

For hydro-meteorological time series, the location of a change point is a discrete parameter which lies within the range of possible time. Currently, there is no theory that in-

dicates the probability distribution of a deviance function for a discrete parameter [32]. However, a probability distribution based on deviance function, or in our case, an approximate confidence curve for discrete parameters can be constructed. According to a previous study [32], it can be constructed by simulations to represent the uncertainty in finding the location of a sudden change. Therefore, confidence curves for the location of a change point constructed by Monte Carlo simulations will be considered.

1.3.2. CONFIDENCE CURVES FOR CONTINUOUS PARAMETERS

Continuous parameters play an important role in time series, for instance the design flood with a given return period, the frequency of the occurrence of a certain magnitude of rainfall. In this study, the dependence parameter in copulas is taken as a parameter of interest. A copula is a joint distribution of multiple variables, and according to Sklar [45], the joint behaviour of multiple random variables with continuous margins can be described uniquely by a copula function. In hydrology, the joint behaviour exists in many phenomena [46–49], for instance the duration and the intensity of precipitation, the streamflow concentrated in the lower stream and streamflow from the upper stream, surface drought and groundwater drought.

The uncertainty in continuous parameter estimation is often examined by Monte Carlo simulation or bootstrap and represented by confidence intervals. As we discussed in 1.2, compared to an interval estimator, a confidence curve is more informative and confidence intervals with different confidence levels can be easily extracted from it. According to Wilks' theorem, for continuous parameters, the probability distribution of a deviance function can be approximated by the $\chi^2_{(1)}$ distribution [32], and with a confidence curve confidence intervals at all confidence levels can be extracted from it [29]. Therefore, compared to discrete parameters, the construction of confidence curves for continuous parameters is simpler.

1.4. KNOWLEDGE GAP

Nowadays, uncertainty analysis is playing an increasingly important role in hydrological time series analysis, but it is rare to find studies using confidence curves to represent uncertainties. The concept of confidence curves has been known in statistics since Birnbaum [50], and it is getting increasing attention in statistics [22, 24, 27, 29, 51]. However, the application of it is rare. Therefore, this PhD research is an exploration of how to use confidence curves to conduct uncertainty analysis in hydrological time series analysis. Hydrological frequency analysis often assumes that observations are stationary [52–54], and change point detection plays an important role in determining the stationarity of hydrological observations. Cunen *et al.* [32] proposed a parametric method to construct confidence curves for the change point problem, but there are some limitations of the method. For instance, it is expensive computationally, and based on a parametric distribution. It would be interesting to apply the method to construct confidence curves for the location of a change point in hydrological time series. Modifications to the method will improve the efficiency of the existing method and bridge the gap between the change point detection in hydrological time series and the statistical development.

1.5. OBJECTIVES AND OUTLINES

This PhD research aims to use confidence curves to represent uncertainty in hydrological time series analysis. Two parameters are taken as parameters of interest, the location of a change point and the dependence parameter in copulas.

Chapter 2 introduces traditional methods based on NHST to find the location of change points, and three non-parametric change point detectors are studied: Pettitt's test, the Cramér von Mises (CvM) test and the CUSUM test. In addition, the properties of the aforementioned change-point detectors are analyzed and compared, a better method is suggested. Traditional NHST methods give not enough flexibility for uncertainty analysis, and a new technique to represent uncertainty in change point detection is called for.

To represent uncertainty by confidence curves for the location of a change point, the theoretical background of confidence curves is introduced in Chapter 3. In this chapter, we will present how to construct confidence curves by a profile log-likelihood function which is 'method B' proposed by Cunen *et al.* [32]. We then provide the steps needed to follow to construct confidence curves for the location of a change point. We then show how to examine the property of a confidence curve. Limitations of the method will also be presented.

The method proposed by Cunen *et al.* [32] uses a two-stage maximum likelihood estimator (mle) to estimate nuisance parameters and the parameter of interest, which is computationally expensive. The mle could be very difficult for many parametric distributions, for instance, Generalized Extreme Value (GEV) distributions [55, 56]. Therefore, Chapter 4 modifies the 'method B', and introduces how to use a pseudo maximum likelihood estimator (pmle) to construct confidence curves. We then provide the steps to follow when constructing confidence curves by the modified approach. We then show how to measure the similarity of two confidence curves by different methods and how to quantify uncertainty level by confidence curves.

Moreover, the 'method B' proposed by Cunen *et al.* [32] assumes that the real parametric distribution of observations is known, which does not hold most of the time in hydrology. Therefore, Chapter 5 will explore how to construct confidence curves for the location of a change point by a non-parametric approach based on the approximate empirical likelihood ratio [57–59] and bootstrapping. We then provide the steps to construct confidence curves by this non-parametric approach. We then show the properties of confidence curves, the similarity index between confidence curves, and the uncertainty level by confidence curves using parametric and non-parametric approaches.

In Chapter 6, a confidence curve for the dependence parameter in copulas is constructed. Since the dependence parameter is continuous, the confidence curve for it will be constructed according to Wilks' theorem.

Chapter 7 summarizes the key contribution of this PhD research, the knowledge generated and the limitations and perspectives for future study.

REFERENCES

[1] R. Coleman, *Calculus on normed vector spaces*, Universitext (Springer, New York, 2012) pp. xii+249.

[2] L. Uusitalo, A. Lehikoinen, I. Helle, and K. Myrberg, *An overview of methods to evaluate uncertainty of deterministic models in decision support,* Environmental Modelling & Software **63**, 24 (2015).

[3] E. P. Smith, *Uncertainty analysis,* Wiley StatsRef: Statistics Reference Online (2014).

[4] M. Dehghani, B. Saghafian, F. Nasiri Saleh, A. Farokhnia, and R. Noori, *Uncertainty analysis of streamflow drought forecast using artificial neural networks and Monte-Carlo simulation,* International Journal of Climatology **34**, 1169 (2014).

[5] Y.-M. Hu, Z.-M. Liang, B.-Q. Li, and Z.-B. Yu, *Uncertainty assessment of hydrological frequency analysis using bootstrap method,* Mathematical Problems in Engineering **2013** (2013).

[6] K. Inoue and S. E. Chick, *Comparison of Bayesian and frequentist assessments of uncertainty for selecting the best system,* in *1998 Winter Simulation Conference. Proceedings (Cat. No. 98CH36274)*, Vol. 1 (IEEE, 1998) pp. 727–734.

[7] D. Huard and A. Mailhot, *Calibration of hydrological model GR2M using Bayesian uncertainty analysis,* Water Resources Research **44** (2008).

[8] K. Arnbjerg-Nielsen, P. Willems, J. Olsson, S. Beecham, A. Pathirana, I. Bülow Gregersen, H. Madsen, and V.-T.-V. Nguyen, *Impacts of climate change on rainfall extremes and urban drainage systems: a review,* Water Science and Technology **68**, 16 (2013).

[9] L. Cheng, A. AghaKouchak, E. Gilleland, and R. W. Katz, *Non-stationary extreme value analysis in a changing climate,* Climatic Change (2014), 10.1007/s10584-014-1254-5.

[10] E. M. Fischer and R. Knutti, *Anthropogenic contribution to global occurrence of heavy-precipitation and high-temperature extremes,* Nature Climate Change **5**, 560 (2015).

[11] I. Giuntoli, J.-P. Vidal, C. Prudhomme, and D. M. Hannah, *Future hydrological extremes: The uncertainty from multiple global climate and global hydrological models,* Earth Syst. Dynam. (2015).

[12] I. Haddeland, J. Heinke, H. Biemans, S. Eisner, M. Flörke, N. Hanasaki, M. Konzmann, F. Ludwig, Y. Masaki, J. Schewe, *et al.*, *Global water resources affected by human interventions and climate change,* Proceedings of the National Academy of Sciences **111**, 3251 (2014).

[13] T. A. Solaiman, *Uncertainty estimation of extreme precipitations under climate change: A non-parametric approach*, Ph.D. thesis, University of Western Ontario, London, Ontario, Canada (2011).

[14] F. Edgeworth, *The mathematical method of statistics,* Journal of the Statistical Society of London , 649 (1886).

[15] R. A. Fisher, *Statistical methods for research workers,* in *Breakthroughs in statistics* (Springer, 1992) pp. 66–70.

[16] J. Neyman and E. S. Pearson, *On the use and interpretation of certain test criteria for purposes of statistical inference: Part I,* Biometrika **20A**, pp. 175 (1928).

[17] S. Greenland, *Valid p-values behave exactly as they should: Some misleading criticisms of p-values and their resolution with s-values,* The American Statistician **73**, 106 (2019).

[18] L. G. Halsey, *The reign of the p-value is over: What alternative analyses could we employ to fill the power vacuum?* Biology Letters **15**, 20190174 (2019).

[19] R. L. Wasserstein, A. L. Schirm, and N. A. Lazar, *Moving to a world beyond "p < 0.05",* The American Statistician **73**, 1 (2019).

[20] S. T. Ziliak, *How large are your G-Values? Try Gosset's Guinnessometrics when a little "p" is not enough,* The American Statistician **73**, 281 (2019).

[21] M. Davidian and T. A. Louis, *Why statistics?* (2012).

[22] K. Singh, M. Xie, W. E. Strawderman, *et al.*, *Confidence distribution (CD)–distribution estimator of a parameter,* in *Complex datasets and inverse problems* (Institute of Mathematical Statistics, 2007) pp. 132–150.

[23] B. Efron, *RA Fisher in the 21st century,* Statistical Science , 95 (1998).

[24] T. Schweder and N. L. Hjort, *Confidence and likelihood,* Scandinavian Journal of Statistics **29**, 309 (2002).

[25] K. Singh, M. Xie, W. E. Strawderman, *et al.*, *Combining information from independent sources through confidence distributions,* The Annals of Statistics **33**, 159 (2005).

[26] M. Xie, K. Singh, and W. E. Strawderman, *Confidence distributions and a unifying framework for meta-analysis,* Journal of the American Statistical Association **106**, 320 (2011).

[27] M.-g. Xie and K. Singh, *Confidence distribution, the frequentist distribution estimator of a parameter: A review,* Int. Stat. Rev. **81**, 3 (2013).

[28] S. Nadarajah, S. Bityukov, and N. Krasnikov, *Confidence distributions: A review,* Statistical Methodology **22**, 23 (2015).

[29] T. Schweder and N. L. Hjort, *Confidence, likelihood, probability. Statistical inference with confidence distributions.* (Cambridge: Cambridge University Press, 2016) pp. xx + 500.

[30] D. R. Bickel, *Confidence intervals, significance values, maximum likelihood estimates, etc. sharpened into Occam's razors,* Communications in Statistics-Theory and Methods **49**, 2703 (2020).

[31] S. Ratnasingam and W. Ning, *Confidence distributions for skew normal change-point model based on modified information criterion,* Journal of Statistical Theory and Practice **14**, 1 (2020).

[32] C. Cunen, G. Hermansen, and N. L. Hjort, *Confidence distributions for change-points and regime shifts,* Journal of Statistical Planning and Inference **195**, 14 (2018).

[33] P. Gao, X. Zhang, X. Mu, F. Wang, R. Li, and X. Zhang, *Trend and change-point analyses of streamflow and sediment discharge in the Yellow River during 1950–2005,* Hydrological Sciences Journal–Journal des Sciences Hydrologiques **55**, 275 (2010).

[34] C. Zhou, R. van Nooijen, A. Kolechkina, and M. Hrachowitz, *Comparative analysis of nonparametric change-point detectors commonly used in hydrology,* Hydrological Sciences Journal **64**, 1690 (2019).

[35] E. S. Page, *Continuous inspection schemes,* Biometrika **41**, 100 (1954).

[36] R. Killick, I. A. Eckley, K. Ewans, and P. Jonathan, *Detection of changes in variance of oceanographic time-series using changepoint analysis,* Ocean Engineering **37**, 1120 (2010).

[37] J. Chen and A. K. Gupta, *Parametric statistical change point analysis: With applications to genetics, medicine, and finance*, 2nd ed. (Birkhäuser/Springer, New York, 2012) pp. xiv+273.

[38] E. Brodsky and B. S. Darkhovsky, *Nonparametric methods in change point problems*, Vol. 243 (Springer Science & Business Media, 2013).

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

[40] M. Gocic and S. Trajkovic, *Analysis of changes in meteorological variables using Mann-Kendall and Sen's slope estimator statistical tests in Serbia,* Global and Planetary Change **100**, 172 (2013).

[41] M. G. Donat, A. L. Lowry, L. V. Alexander, P. A. O'Gorman, and N. Maher, *More extreme precipitation in the world's dry and wet regions,* Nature Climate Change **6**, 508 (2016).

[42] J. Lehmann, D. Coumou, and K. Frieler, *Increased record-breaking precipitation events under global warming,* Climatic Change **132**, 501 (2015).

[43] Y. Hirabayashi, R. Mahendran, S. Koirala, L. Konoshima, D. Yamazaki, S. Watanabe, H. Kim, and S. Kanae, *Global flood risk under climate change,* Nature Climate Change **3**, 816 (2013).

[44] H. McMillan, A. Montanari, C. Cudennec, H. Savenije, H. Kreibich, T. Krueger, J. Liu, A. Mejia, A. Van Loon, H. Aksoy, *et al.*, *Panta Rhei 2013–2015: Global perspectives on hydrology, society and change,* Hydrological Sciences Journal **61**, 1174 (2016).

[45] A. Sklar, *Fonctions de répartition à n dimensions et leurs marges,* Publ. Inst. Statist. Univ. Paris **8**, 229 (1959).

[46] G. Salvadori and C. De Michele, *Frequency analysis via copulas: Theoretical aspects and applications to hydrological events,* Water Resources Research **40** (2004).

[47] G. Salvadori and C. De Michele, *On the use of copulas in hydrology: Theory and practice,* Journal of Hydrologic Engineering **12**, 369 (2007).

[48] R. B. Nelsen, *An introduction to copulas* (Springer Science & Business Media, 2007).

[49] C. Genest and A.-C. Favre, *Everything you always wanted to know about copula modeling but were afraid to ask,* Journal of Hydrologic Engineering **12**, 347 (2007).

[50] A. Birnbaum, *A unified theory of estimation. I.* Ann. Math. Stat. **32**, 112 (1961).

[51] T. Schweder, *Confidence is epistemic probability for empirical science,* Journal of Statistical Planning and Inference **195**, 116 (2018).

[52] M. J. Machado, B. A. Botero, J. López, F. Francés, A. Díez-Herrero, and G. Benito, *Flood frequency analysis of historical flood data under stationary and non-stationary modelling,* Hydrology and Earth System Sciences **19**, 2561 (2015).

[53] T. B. M. J. Ouarda and S. El-Adlouni, *Bayesian nonstationary frequency analysis of hydrological variables 1,* JAWRA Journal of the American Water Resources Association **47**, 496 (2011).

[54] B. Renard, X. Sun, and M. Lang, *Bayesian methods for non-stationary extreme value analysis,* in *Extremes in a Changing Climate* (Springer, 2013) pp. 39–95.

[55] P. Prescott and A. Walden, *Maximum likelihood estimation of the parameters of the generalized extreme-value distribution,* Biometrika **67**, 723 (1980).

[56] J. R. Hosking, *Algorithm AS 215: Maximum-likelihood estimation of the parameters of the generalized extreme-value distribution,* Journal of the Royal Statistical Society. Series C (Applied Statistics) **34**, 301 (1985).

[57] M. Csörgö and L. Horváth, *Limit theorems in change-point analysis,* Vol. 18 (John Wiley & Sons Inc, 1997).

[58] C. Zou, Y. Liu, P. Qin, and Z. Wang, *Empirical likelihood ratio test for the change-point problem,* Statistics & Probability Letters **77**, 374 (2007).

[59] A. B. Owen, *Empirical likelihood ratio confidence intervals for a single functional,* Biometrika **75**, 237 (1988).

2

CHANGE POINT DETECTION BY NULL HYPOTHESIS STATISTICAL TESTS

The story of Null Hypothesis Statistical Tests starts with Ronald A. Fisher (1890-1962),
gets extended by Jerzy S. Neyman (1894-1981),
and it has become an indispensable tool in all scientific studies.

Parts of this chapter have been published in"Zhou, C., van Nooijen, R., Kolechkina, A., and Hrachowitz, M. Comparative analysis of nonparametric change-point detectors commonly used in hydrology, Hydrological Sciences Journal, Page: 1690-1710, 64(14), 2019. "

2.1. INTRODUCTION

Today environmental scientists are well aware of the changes that affect the systems they study. Changes in land use, increasing urbanization and climate change combine to complicate the process of predicting the future behaviour of these systems [1–3]. These predictions are needed to answer practical questions like "How high should this dam be to be functional for 50 years?" or "Can we safely develop this coastal area?". Given the inherent uncertainty about the future, predictions inevitably involve statistics, for instance, the probability of certain amounts of precipitation or runoff. These statistics may or may not be influenced by changes in the environment.

One type of change one may look for is a change point [4, 5] a moment in time where there is an abrupt change in one or more of the properties of the time series such as the mean, the median, or the standard deviation.

The art of finding change points was studied first to detect changes in product quality in manufacturing [6, 7]. One of the earliest papers that addressed this question by developing and using a formal statistical test in a hydrological context was written by McGilchrist and Woodyer [8]. They looked for change points in an 88-year-long series of yearly rainfall at Walgett, New South Wales, Australia.

Change point analysis was initially restricted to univariate time series of independent variables under the assumption of 'At Most One Change' (AMOC). It was extended to series with multiple change points [9, 10] and to multivariate time series [11]. New methods were developed to consider dependence within a series, or high-dimensional multivariate time series [12–19]. Detecting change points in a series with trend was studied by analysing a two-phase regression model, see for example Lund *et al.* [14], Wang [20] and Beaulieu *et al.* [21].

Hydrological processes are widely thought to have changing properties [22–24]. Many types of human intervention may result in change points in hydrological time series, for instance, construction of dams, changes in instrumentation or measurement protocol and relocation of measurement stations.

Sometimes the potential cause of a change point in a time series is known, for example, the relocation of a measurement station. These are referred to as 'documented change points', where detected change points can be examined in context. But on other occasions, there are no explicitly documented potential causes for change points and only the outcome of the statistical change point analysis can be used to judge the reliability of the result [25–28].

As in other areas of statistics, there are parametric and non-parametric (distribution free) methods for change point detection. Parametric methods assume that observations are from a known parameterized family of distributions. A number of classical parametric methods have been developed, see for example Chernoff and Zacks [29], Kander *et al.* [30], Hawkins [31] or Gurevich and Vexler [32]. In practice, there is often not enough information on the type of distribution of a hydrological sample to make an informed choice for the distribution family and subsequently perform a parametric change point detection analysis. Therefore, only nonparametric tests are studied in this paper.

Previous studies have analysed the Pettitt's method in terms of its ability to detect the correct time of change for different distributions [33] and sensitivity for the gamma

distribution [34], but comparative studies of multiple methods are rare.

Time series analysis of hydrological data is a complex topic due to dependence in the time series and the complexities of multivariate data. This study considers only one specific context: under ideal circumstances and for a time series containing only one variable, can change point analysis be used for exploratory data analysis and what are its limitations? Questions to be answered are:

- Can the probability of incorrectly signaling a change point be predicted?
- What is the probability of correctly detecting a change point?
- How close are the estimates to the correct location?
- What is the effect of time series length?
- Is there a relationship between the size of the change and the answers to the above questions?
- Does it matter when our series starts or ends? In other words: is it safe to look at parts of a time series that contain a given range of potential change points, but have different start or end years?

The following change point detection methods are considered: the method described in Pettitt [4], which we refer to as 'Pet-CP', a method based on the two-sample Cramér von Mises test statistic, which we refer to as 'CvM-CP' [35, 36], and a method based on CUSUM median statistics, which we refer to as 'CUSUM-CP' [8, 37, 38]. Xiong *et al.* [36] used CvM-CP to detect the change point in multivariate time series, but this study applies CvM-CP in the univariate situation.

2.2. METHODOLOGY AND DATA

This study contains two groups of experiments. The first uses synthetic data series to examine how well the methods perform. The second takes four time series of the annual maximum runoff and uses the methods to look for change points in the full series and subseries with different start and/or end years. From a statistical point of view, a time series of hydrological measurements of length n can be seen as a vector of n observations $(x_1, x_2, \ldots, x_n)$ corresponding to one sample of a random vector $(X_1, X_2, \ldots, X_n)$. The vector components may or may not be independent, and they may or may not have the same marginal distribution. The methods for change point analysis used in this study have three components:

- a test statistic;
- an exact (or approximate) distribution of the test statistic under the null hypothesis;
- an estimator $\hat{\tau}$ for the point in time τ where the change occurs (the change point).

For these tests the null hypothesis is: There is no change point. To apply one of these methods, first a significance level is set, next the statistic is calculated and, finally, if the null hypothesis is rejected, then the estimator $\hat{\tau}$ is applied and the resulting change point location is reported. All tests given here are described in a form suitable for independent vector components and the presence of at most one change point, so either the n vector components have the same distribution, or the first τ are from one distribution and the remaining $(n-\tau)$ are from a second distribution. If the vector components are not independent, then either adjustment of the distribution of the test statistic, or pre-processing of the time series is indicated [39], and if there are multiple change points, then the tests need to be extended; both are outside the scope of this chapter. Background information on change detection can be found in Kundzewicz and Robson [39, 40].

2.2.1. CHANGE POINT DETECTION METHODS

(1) PET-CP METHOD

The Pettitt's test was specifically designed to detect a single change point [4]. The following two-sample test statistic is defined as

$$U_\tau = \sum_{i=1}^{\tau} \sum_{j=\tau+1}^{n} \operatorname{sgn}(X_i - X_j) \tag{2.1}$$

where $\operatorname{sgn}(\cdot)$ is a sign function (see A.2). The Pettitt's test statistic itself is given by

$$K_n = \max_{1\le\tau<n} |U_\tau| \tag{2.2}$$

If the null hypothesis does not hold, then the estimator for the change point location is:

$$\hat{\tau} = \min\left(\arg\max_{1\le\tau\le n} |U_\tau|\right) \tag{2.3}$$

According to Pettitt [4], the limit distribution of K_n for large n is given by:

$$\Pr\left(K_n\sqrt{\frac{3}{n^2+n^3}} \le \alpha\right) = 1 + 2\sum_{j=1}^{\infty}(-1)^j \exp\left(-2j^2\alpha^2\right) \tag{2.4}$$

where the right-hand side represents the cdf of the Kolmogorov distribution and α is the significance probability associated with $K_n\sqrt{\frac{3}{n^2+n^3}}$. Most papers that apply this test use this limit distribution (see [41], [34]), so it will be used here as well.

(2) CVM-CP TEST

The original Cramér von Mises (CvM) test was intended to determine whether all observations in a sample of n independent observations were drawn from a given probability distribution [42]. A modification can be used to test whether or not two samples were drawn from the same distribution [43]. Holmes *et al.* [35] developed a method on the basis of the two-sample CvM test statistic to detect the change point within the multivariate series. This was a further development of the approach proposed by Gombay and Horváth [44]. According to Bücher *et al.* [45], the method developed by Holmes *et al.*

[35] performs much better than that based on the two-sample Kolmogorov-Smirnov test statistic. Moreover, it is not only useful in detecting the change point within a univariate time series, but can also be applied to get a change point in a multivariate hydrological time series, such as multivariate series based on copla models [36]. The notation from Xiong *et al.* [36] is used to describe the CvM-CP detection method. We refer to A.1, the indicator function is $[\![\cdot]\!]$, so the empirical distribution function is:

$$F_\tau(X_k) = \frac{1}{\tau}\sum_{i=1}^{\tau}[\![X_i \leq X_k]\!] \tag{2.5}$$

For the part of the sample up to a potential change point, and the empirical distribution function for the part of a sample after the potential change point:

$$F^*_{n-\tau}(X_k) = \frac{1}{n-\tau}\sum_{i=\tau+1}^{n}[\![X_i \leq X_k]\!] \tag{2.6}$$

For a time series of one variable, the CvM-CP test statistic is defined in terms of $(n-1)$ two-sample statistics:

$$S_\tau = \frac{1}{n}\sum_{k=1}^{n}(D(\tau, X_k))^2 \tag{2.7}$$

$$D(\tau, X_k) = \frac{\tau(n-\tau)}{n^{\frac{3}{2}}}(F_\tau(X_k) - F^*_{n-\tau}(X_k)) \tag{2.8}$$

The CvM-CP statistic is given by:

$$S = \max_{1\leq\tau<n} S_\tau \tag{2.9}$$

The distribution for this value under the null hypothesis is not known exactly and an asymptotic distribution is not available. It was approximated empirically from a sample of size 10 000 taken from the standard uniform distribution, as in Holmes *et al.* [35]. If the null hypothesis does not hold, then the estimator for the change point location is:

$$\hat{\tau} = \min\left(\arg\max_{1\leq\tau<n} S_\tau\right) \tag{2.10}$$

The general approach of choosing the lowest index τ if there are multiple equal maxima was proposed in Antoch *et al.* [46].

(3) CUSUM-CP METHOD

Page [7] was the first to suggest the use of a cumulative sum to find changes in a parameter of interest. McGilchrist and Woodyer [8] used it to detect a change point for even sample lengths; this is the variant used in this study. Chiew and McMahon [37] used this method to detect change in annual flow of Australian rivers.

The test is defined in terms of a one-sample test statistic for each potential change point

$$V_\tau = \sum_{j=1}^{\tau}(2[\![K \leq X_j]\!] - 1) \tag{2.11}$$

In (2.11), K is a random variable corresponding to one of several quantities. We follow McGilchrist and Woodyer [8], who used the sample median. The test statistic is

$$T_n = \frac{2}{n} \max_{1 \le \tau \le n} |V_\tau| \tag{2.12}$$

and the estimator for the change point location is

$$\hat{\tau} = \min \left(\arg\max_{1 \le \tau < n} |V_\tau| \right) \tag{2.13}$$

According to McGilchrist and Woodyer [8], under the null hypothesis the limit distribution of T_n for large n is the same as that of the Kolmogorov-Smirnov test statistic. It follows that

$$\Pr\left(T_n \sqrt{\frac{n}{4}} < x \right) = 1 + 2 \sum_{j=1}^{\infty} (-1)^j \exp\left(-2 j^2 x^2\right) \tag{2.14}$$

where the right-hand side represents the cdf of the Kolmogorov distribution. Most papers that apply this test use this limit distribution, so it will be used here as well.

2.2.2. CRITERIA USED TO EVALUATE THE PERFORMANCE OF THE TESTS

The first property to be checked is the empirical type I error probability. For a significance level of 5% the test should reject the null hypothesis, H_0, 'There is no change point', for 5% of the synthetic time series without change point. To see how well the tests do when detecting change points, we want to approximate the power of the test, which is defined as the probability that a test correctly rejects H_0 without considering the accuracy of the estimate of the change point [47]. If, for a set of N samples with a change point, the test rejects N_{rej}, then the empirical probability of correct rejection is:

$$\text{power} \approx \frac{N_{\text{rej}}}{N} \tag{2.15}$$

While high power is desirable, it is also important that the estimate of the point in time where the change takes place is accurate. A very strict measure of this is the ability of a change point detection test. This is defined as the empirical probability that the test will correctly reject the null hypothesis and correctly identify the location of the change point [33]. If for N_{cor} out of N samples the null hypothesis is rejected and the change point correctly identified, then this is given by:

$$\text{ability} \approx \frac{N_{\text{cor}}}{N} \tag{2.16}$$

2.2.3. DATA SOURCES: SYNTHETIC AND OBSERVATIONAL

(1) GENERATION OF THE SYNTHETIC TIME SERIES

Each synthetic time series consisted of n observations of independent random variables where $n = 10, 20, \ldots, 100, 200, 500, 1000$. Homogeneous synthetic series were generated by sampling M times from the same distribution and used to determine the rejection rate of the null hypothesis 'there is no change point'. Time series with exactly one change

point τ, with $\tau = \frac{n}{10}, \frac{2n}{10}, \ldots, \frac{9n}{10}$, were generated by sampling from a given distribution type with mean μ_L, and the standard deviation σ_L for the left-hand part of the series up to and including X_τ and and mean μ_R and standard deviation σ_R for the right-hand part of the series. The following notation is used:

$$\Delta\mu = \mu_R - \mu_L; \Delta\sigma = \sigma_R - \sigma_L$$

To study the sensitivity to a change in the mean, series were generated with $\mu_L = 0, \sigma_L = \sigma_R = 1$ and $\mu_R = 0.5, 1, 2, 4, 8$. To study the sensitivity to a change in the standard deviation, series were generated with $\mu_L = \mu_R = 0, \sigma_L = 1$ and $\mu_L = \mu_R = 0, \sigma_R = 0.5, 2, 4, 8$.

To allow statistical analysis of the results for each specific combination of type of distribution, $\Delta\mu$, $\Delta\sigma$, change point location τ, and series length n we generated M synthetic time series. For most combinations, M was equal to 10000, except for CvM-CP in the case of series of length 200 and 500, where $M = 1000$ was used, and sample length of 1000, where $M = 5000$ was used, as CvM-CP turned out to be much more expensive to calculate for long series than the other tests.

(2) TYPE OF DISTRIBUTION

Genralized Extreme Value disrtbutions (GEV) have been employed in many hydrological time series analysis, for instance in [48] and [49], therefore in this study GEV distributions will also be used. The following shows the four distribution types to be considered:

- normal distribution;
- generalized extreme value (GEV) distribution with shape −0.15, which corresponds to the three-parameter reverse Weibull distribution with shape 20/3;
- GEV distribution with shape 0, which corresponds to the Gumbel distribution; and
- GEV distribution with shape 0.15, which corresponds to the three-parameter Fréchet distribution with shape 20/3. The value 0.15 was chosen as representative for thick-tailed GEV distributions [50].

Formulas for the GEV can be found in, for instance, van Nooijen and Kolechkina [51]. B.1 provides arguments to limit the number of different parameter combinations in case of location–scale distribution families such as those given above.

(3) SOURCE OF THE REAL-WORLD DATA

For a given location, the first and last year of a period for which suitable data is available may depend on pre-processing, willingness to allow for missing data and access to recent data. This raises the question whether or not change point detection results depend on the choice of first and last year. To examine this in the context of real data, measurements from the Yangtze River in China were used. The methods were applied to annual maximum runoff (AMR) observations from four gauge stations: Cuntan (1893–2014), Yichang (1946–2014), Hankou (1952–2014) and Datong (1950–2014) collected by the Ministry of Water Resources of the People's Republic of China in corresponding periods. The locations of the measurement stations are shown in Fig. 2.1 and the four AMR time series used are shown in Fig. 2.2. Over the last 70 years, the Yangtze River basin has been sub-

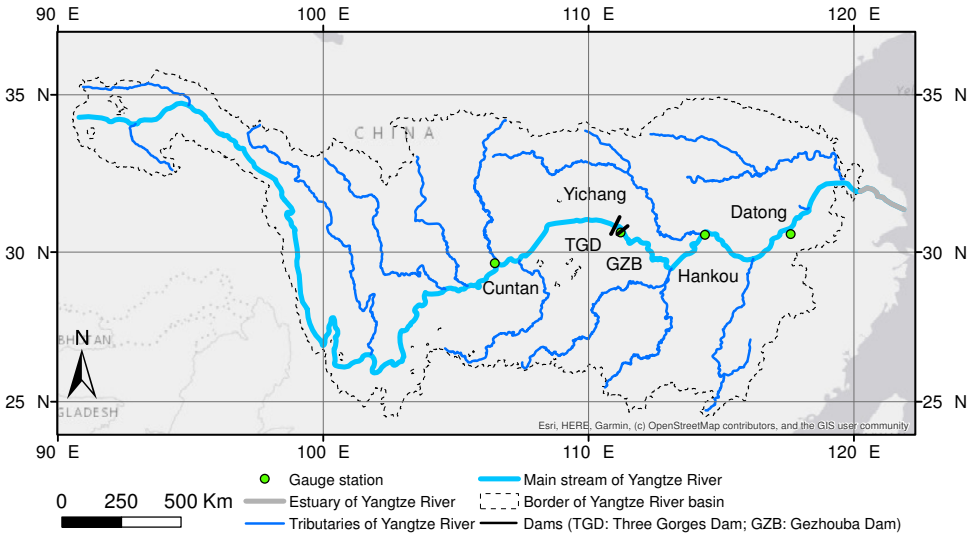


Figure 2.1: Location of four gauge stations and three dams on Yangtze River.

Table 2.1: Details on some of the dams on the Yangtze River and its tributary.

Dam name	Danjiangkou	Gezhouba	Three Gorges
Orientation	111 ° 29'17"E 32 ° 33'22"N	111 ° 16'20"E 30 ° 44'23"N	111 ° 00'12"E 30 ° 49'23"N
Construction time	1958–1973	1970–1988	1993–2009
Capacity	900MW	2715MW	22500MW
Reservoir capacity	17.45 km^3	1.58 km^3	39.3 km^3
Location	In Hanjiang, upstream of Hankou	38 km upstream of Yichang	44 km upstream of Yichang

ject to large-scale human intervention [52]. Reservoir construction has resulted in the building of over 10 000 dams since the 1960s [53]. Information on the largest two dams in the Yangtze and one in its Hanjiang tributary is given in Table 2.1 (locations are shown in Fig. 2.1).

For the Yichang, Hankou and Datong series, previous investigations suggest the series can be treated as uncorrelated [54, 55] at the 5% significance level. Zhang *et al.* [56] used detrended fluctuation analysis to find the long-range correlation of three datasets from the Yangtze River and concluded that the daily streamflow (1893-–2009) from Cuntan station showed no significant correlation.

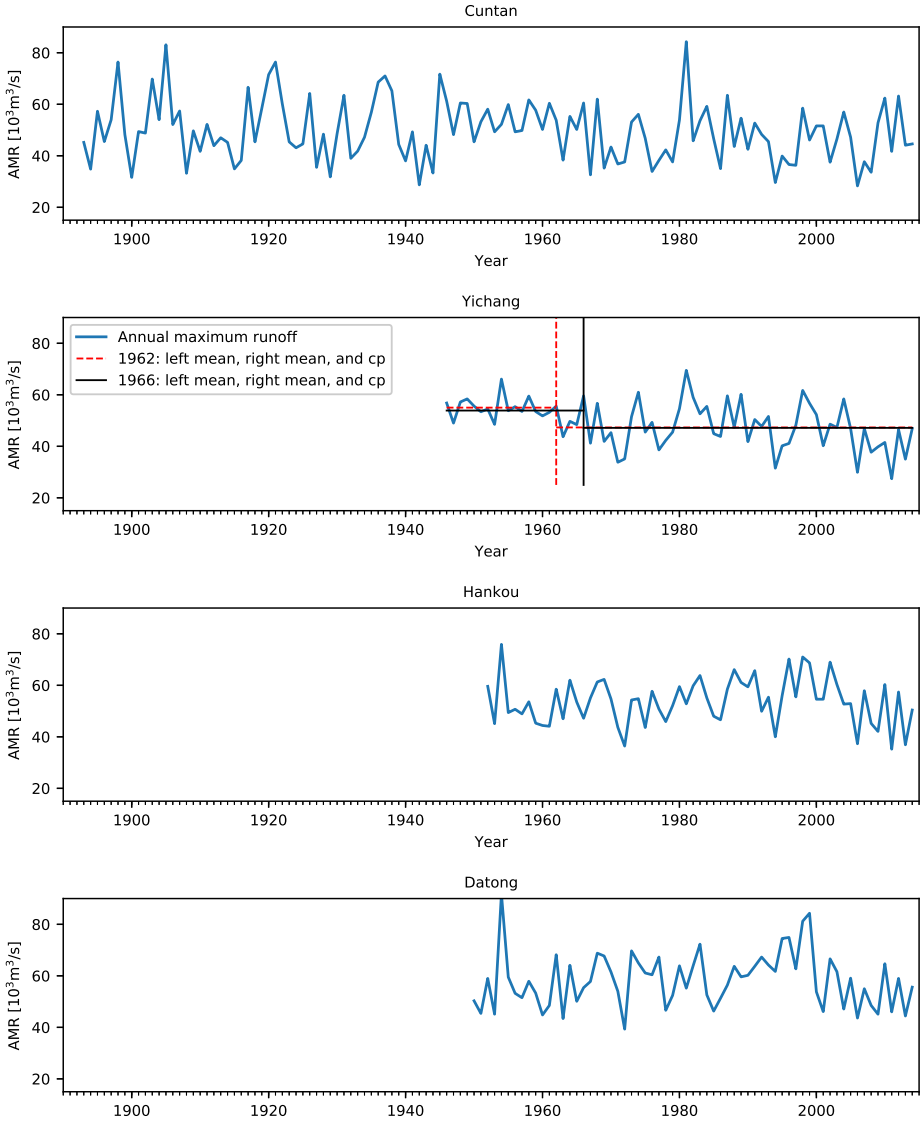


Figure 2.2: Annual Maximum Runoff of the four hydrological stations in Yangtze River.

2.3. ANALYSIS OF THE PERFORMANCE OF THE TESTS FOR DIFFERENT INPUT DATA

The results of the experiments with synthetic data are followed by the results of the experiments on the time series of observed AMR.

2.3.1. SYNTHETIC EXPERIMENT

For all tests the significance level was set to 0.05. In other words, it is allowed to incorrectly assume the existence of a change point in 5% of all applications of the test. If the real rejection rate of the null hypothesis 'there is no change point' is higher than this value, then change points will appear more likely than they are in reality, possibly leading to unnecessary efforts to allow for non-existent change. If the real rejection rate of the null hypothesis is lower than this value, then change points will appear less likely than they are in reality, possibly leading to a failure to allow for real change.

Figure 2.3 shows the rejection rates for the different methods and distributions as a function of sample size.

Rejection rate of H_0 as a function of sample size for each of the tests (significance level $\alpha = 0.05$). For sample lengths of 1000 and 5000, Monte Carlo simulations are applied for the CvM test.

We can see that Pet-CP and CUSUM-CP start well below the expected rejection rate, while CvM-CP stays close to the chosen significance level. Given that the CvM-CP rejection rate was determined from an empirical distribution, it is not surprising that it does so well; for the other tests we used a limit distribution to approximate the quantile. It is clear that for small samples ($n \leq 100$) the limit distributions are not sufficiently accurate, and use of either the exact distribution or an empirical distribution would be preferable. The traditional statistical remedy "use a larger sample" is not an option for time series of extreme values where longer series are simply not available. An alternative traditional remedy for this problem, "use an improved approximation of the distribution", is simple in theory, but complicated in practice because calculation of the exact distribution, or alternatively the generation of an approximate distribution by Monte Carlo methods can be quite expensive.

2.3.2. ONE CHANGE POINT PRESENTS

(1) SENSITIVITY TO A CHANGE IN THE MEAN

The power and ability to correctly identify the change point are shown in Fig. 2.4 and 2.5, respectively.

We can see that for all tests both power and ability increase considerably with an increase in the magnitude of the change $\Delta\mu$ in the mean. The plots of power vs the location of the actual change point τ are nearly symmetrical with respect to a vertical line at $\tau = n/2$. For Pet-CP and CvM-CP the power is higher than for CUSUM-CP when $\Delta\mu \leq 1$, except for GEV with $k = 0.15$ (see the bottom row in Fig. 2.4). For $\Delta\mu \geq 2$, , all tests have 100% power for $\tau = 20, 30, \ldots, 80$. If we look at the ability as a function of the location of the change point, then for Pet-CP and CvM-CP the function is nearly symmetrical with respect to a vertical line at $\tau = n/2$, and the highest abilities are reached when the actual change point is near $n/2$. From Figures 2.4 and 2.5, it is clear that the power and ability

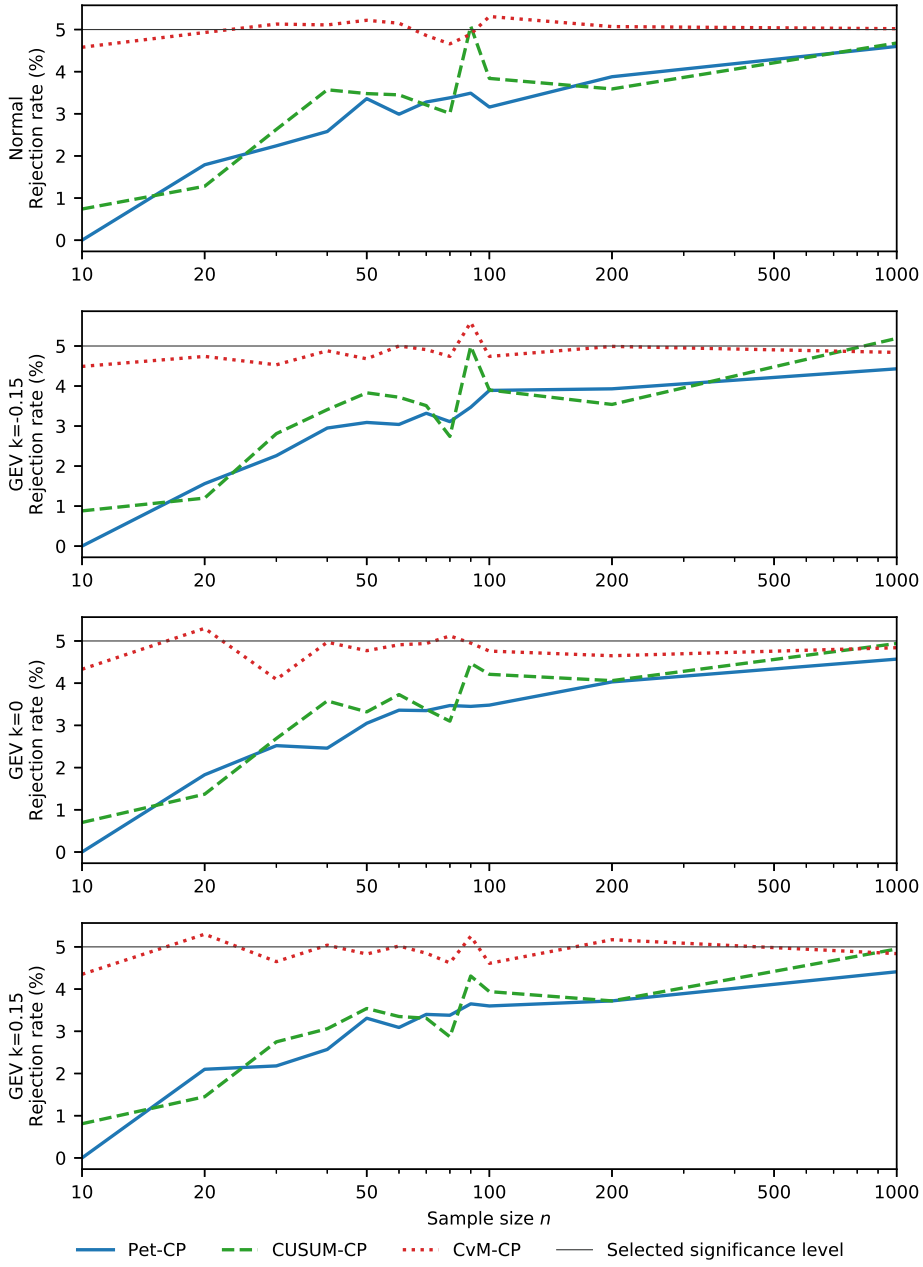


Figure 2.3: Rejection rate of H0 as a function of sample size for each of the tests (significance level $\alpha = 0.05$). For sample lengths of 1000 and 5000 (not shown here), Monte Carlo simulations are applied for the CvM-CP test.

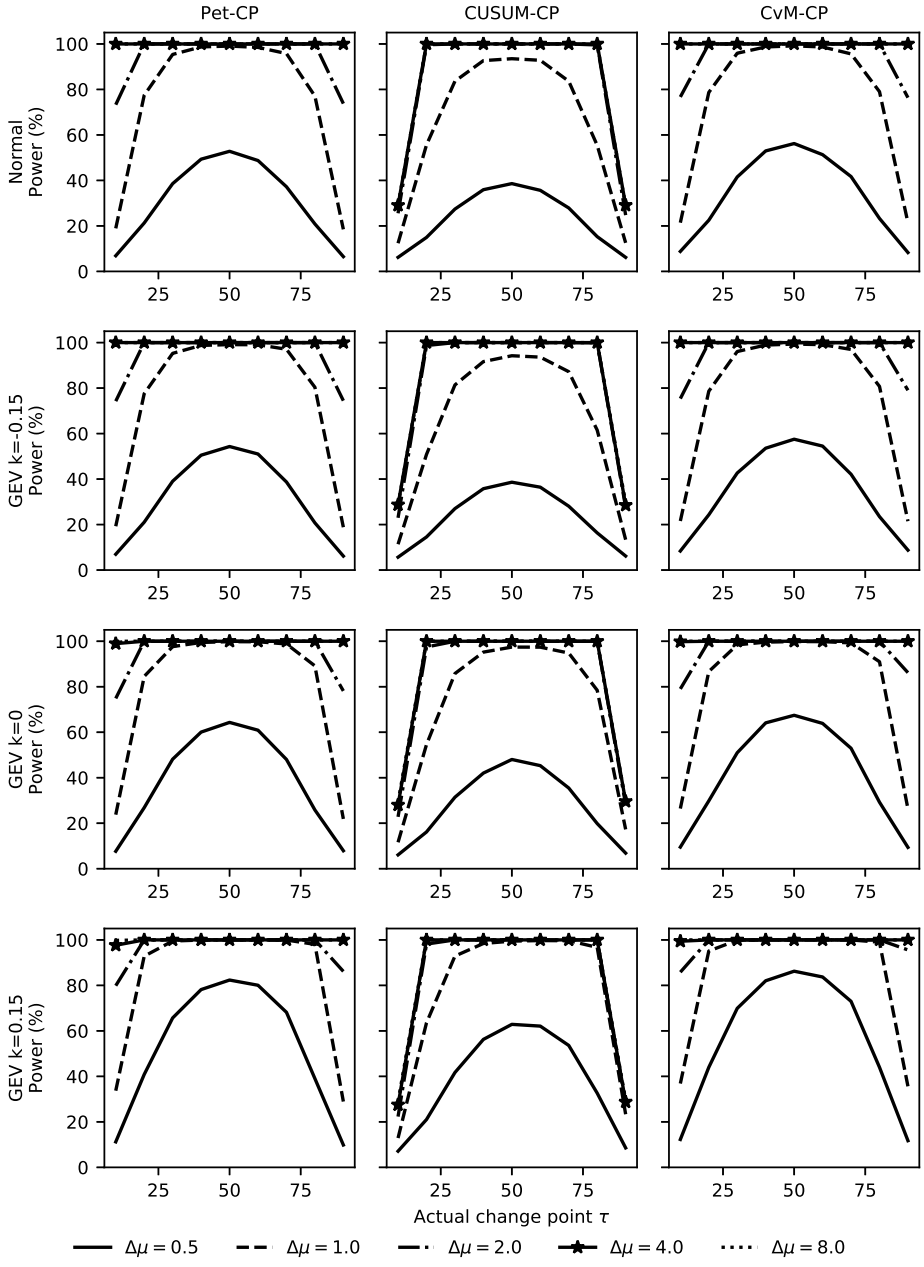


Figure 2.4: Power of all tests for a change in the mean ($n = 100$).

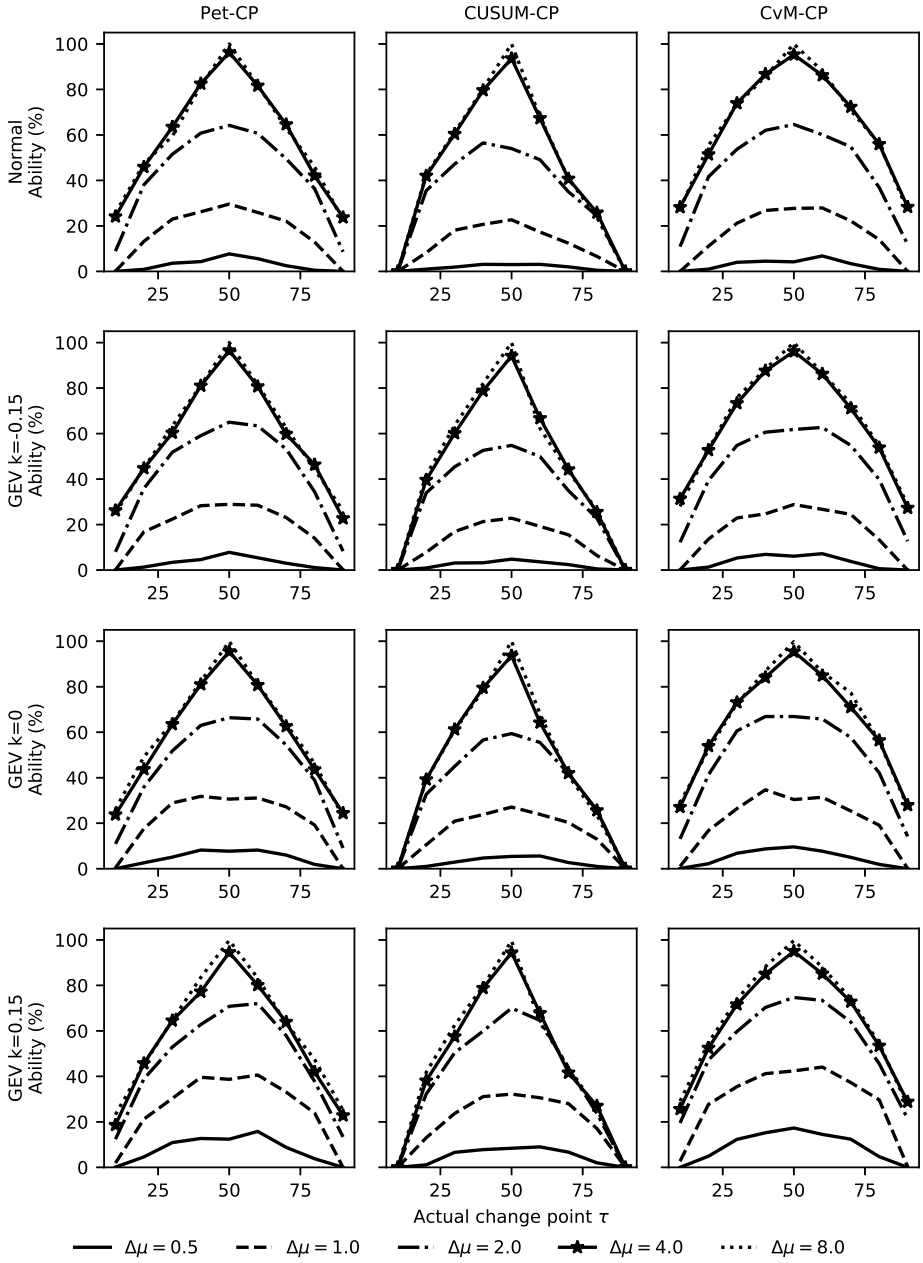


Figure 2.5: Ability of all the tests for a change in the mean ($n = 100$).

vary with location for each test; the ability tends to be more sensitive to the magnitude of the change and the location of the change point. For instance, for Pet-CP, when the magnitude of change is the same, the ability (Fig. 2.5, row 1, column 1) varies much more than the power (Fig. 2.4, row 1, column 1). The differences in shape indicates the ability of Pet-CP is much more sensitive to location of a change point than the power.

For all three methods, the abilities increase as $|\Delta\mu|$ increases and stabilize for $|\Delta\mu \geq 4|$. When τ is near the middle of the series, the ability increases from less than 10% to nearly 100% for increasing $|\Delta\mu|$. When τ is near the ends of the series, the abilities stay well below 100%. For a series of length 100, detecting a change in the first or last 20 elements of the series, there is a low probability of it being estimated correctly, regardless of the size of the change.

(2) SENSITIVITY TO A CHANGE IN THE STANDARD DEVIATION

The results for power (Fig. 2.6) and ability (Fig. 2.7) show that Pet-CP and CUSUM-CP cannot detect a change in the standard deviation.

While CvM-CP can detect a change in the standard deviation, its ability to do so is much lower than in the case of a change in the mean. For a change of a factor of two in the standard deviation, the power is low as well (see the first two columns in both Figs 2.6 and 2.7). The power and ability plots of CvM-CP are nearly symmetrical with respect to a vertical line at $\tau = n/2$, and they reach their highest point when the actual change point is located near $n/2$. From the first two columns in Fig. 2.7, the abilities of Pet-CP and CUSUM-CP stay below 1%. The CvM-CP method shows similar abilities for change points at locations τ and $(n-\tau)$. For $\tau = 10$ and $\tau = 90$, its ability is near zero (see the last column in Fig. 2.7). It seems that only for very large changes in standard deviation ($\Delta\sigma \geq 6$) and only for the change points $\tau = 40 \sim 60$ near the midpoint of the series does the ability rise above 50% (Fig. 2.7).

For Pet-CP, the lower sensitivity to a change in σ seems to be known [57], but the reasoning behind this is difficult to find. One possible line of reasoning is given in B.2. For CUSUM-CP, the original source states that it is intended for detection of changes in the mean, so its failure for the standard deviation was perhaps to be expected.

(3) UNCERTAINTY OF THE ESTIMATORS FOR A CHANGE IN THE MEAN

The ability gives the empirical probability that the estimated change point coincides with the actual change point. In cases where there is a large difference between power and ability, additional information may be needed. The main question in that case is whether the correctly detected, but incorrectly placed change points are clustered near the correct value or not. Results for the normal distribution are presented in Fig. 2.8. For all tests, the boxplots for change point estimates when the actual change point is at τ or $(n-\tau)$ show very similar uncertainty.

For $\Delta\mu = 0.5$, the systematic error (bias) near the ends of the series and the spread in the estimate are both too large for practical use. Take CvM-CP for example, and $\Delta\mu = 0.5$ (Fig. 2.8, row 3, column 1): for synthetic series of length 100 with a change point at position 10, the boxplot of the estimates has median near 42 and interquantile range of about 22. For a change point at position 20, the boxplot of the estimates shows a median near 32 with an interquartile range of about 18. Similar, but negative, biases occur for

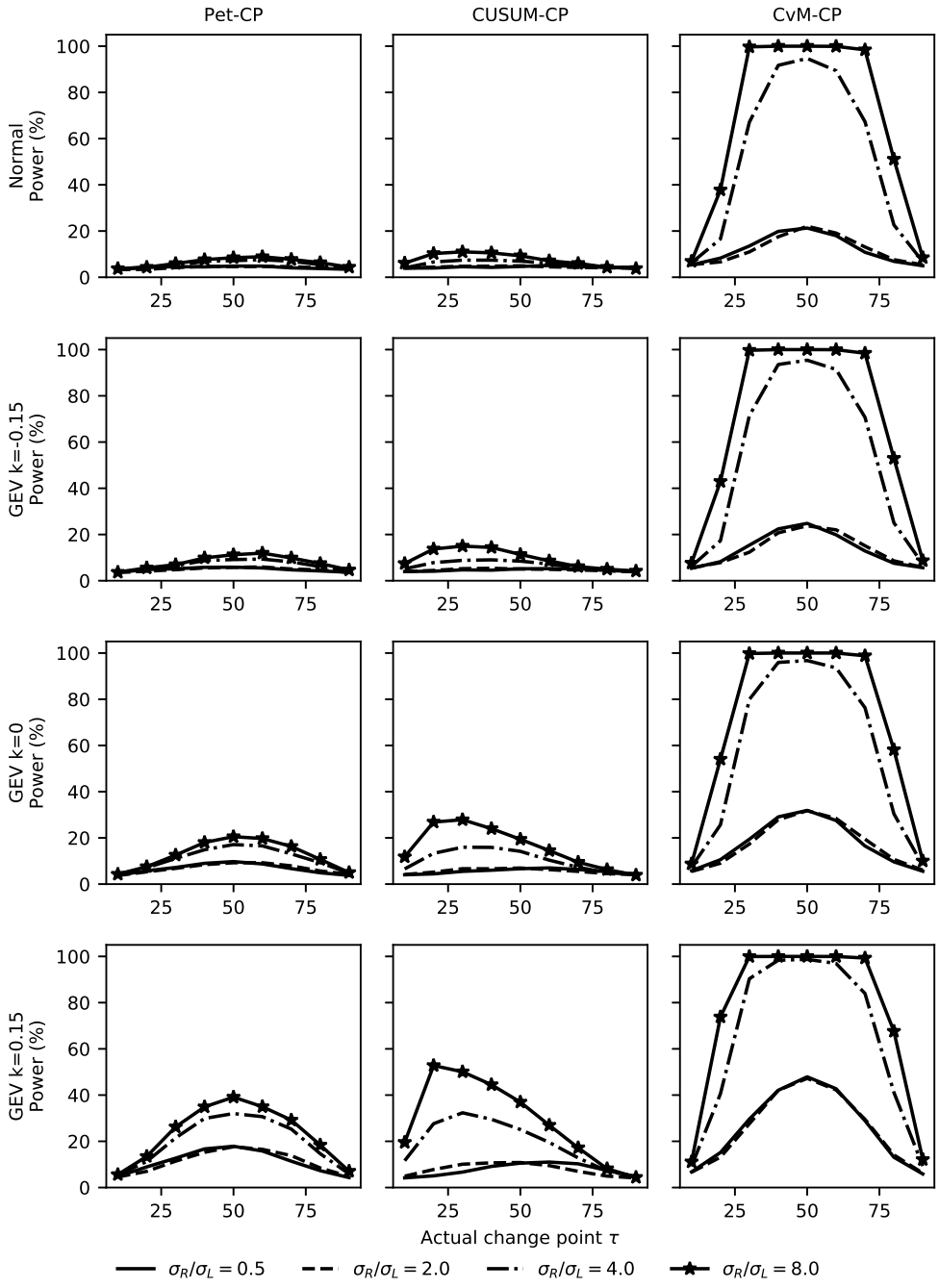


Figure 2.6: Power of all the tests for a change in the standard deviation ($n = 100$).

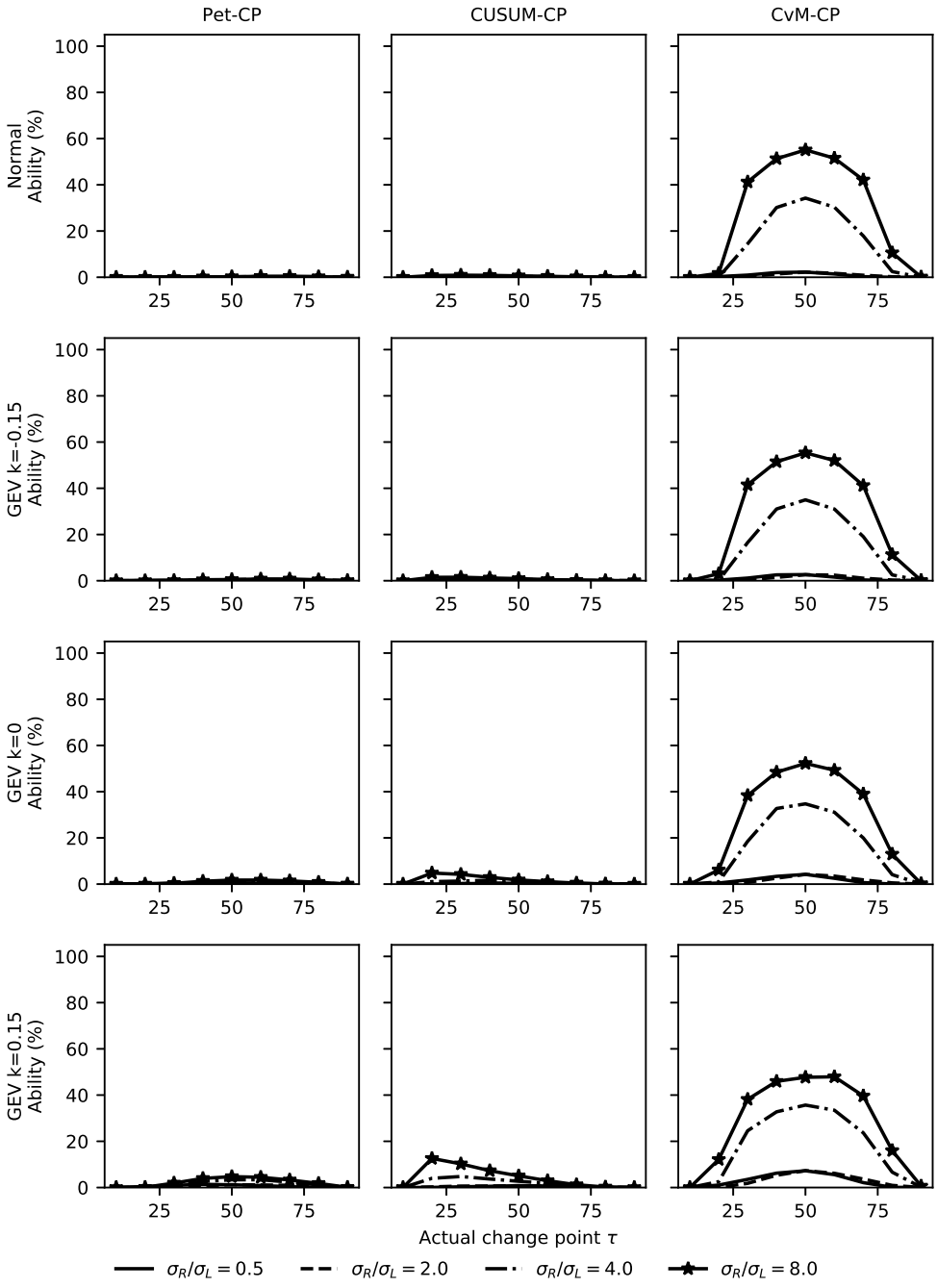


Figure 2.7: Ability of all the tests for a change in the standard deviation ($n = 100$).

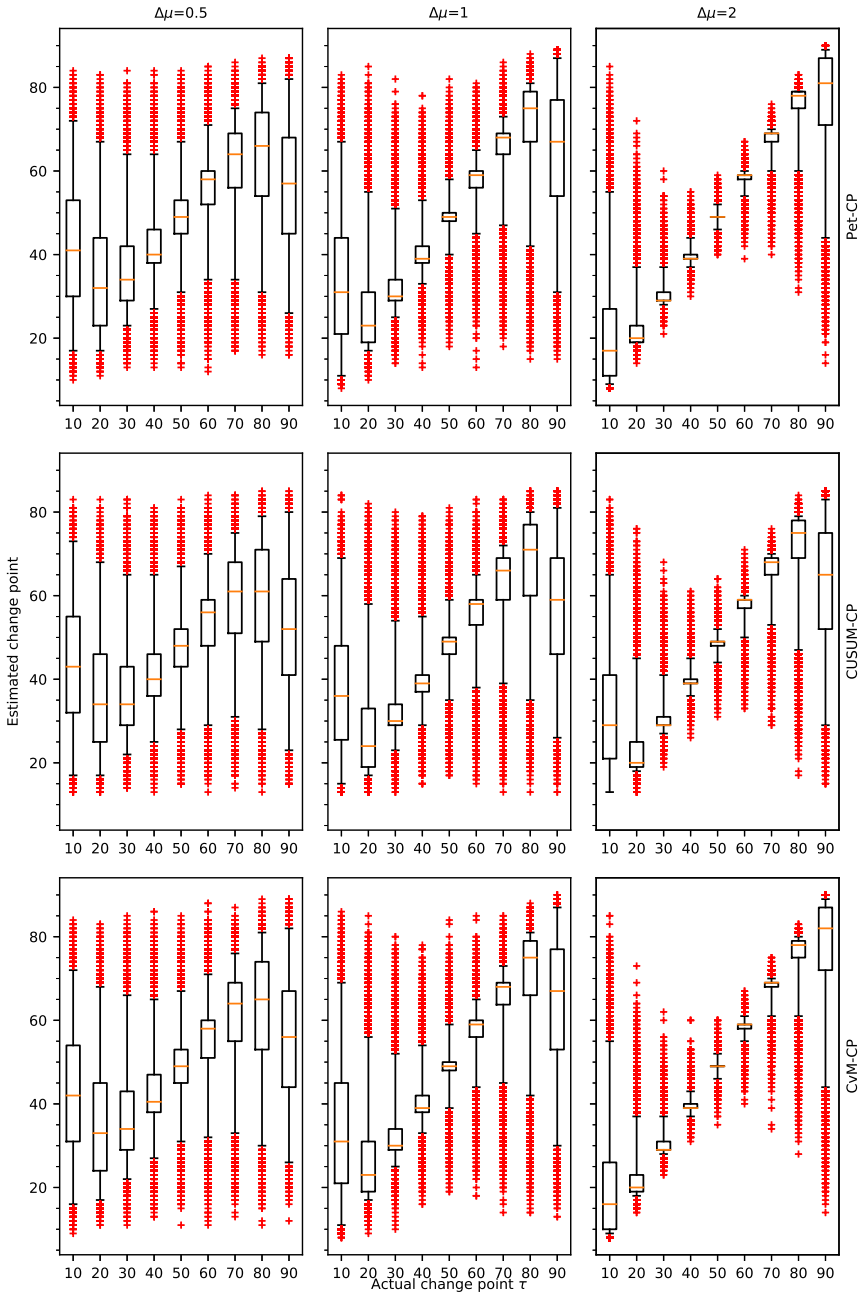


Figure 2.8: Boxplots of the error in the change point estimates based on 50 000 samples for a change in the mean. The whiskers are at 2.5% and 97.5%; the crosses show the estimates outside that range.

change points near the end of the series. Similar bias and spread occur for the other methods at $\Delta\mu = 0.5$.

For $\Delta\mu = 1$, the systematic error near the ends is still large. Moreover, the 95% confidence interval is large even for the centre point of the series. For $\Delta\mu = 2$, there are still problems with the systematic error near the end of the series, but in the case of CvM-CP (see the last plot in the last row of Fig. 2.8, points between position 20 and position 80), the distribution of the spread in the estimates approaches reasonable values.

The results presented here imply that change points near the end of the series, if detected, will almost always result in a relatively large error in the estimated change point.

(4) UNCERTAINTY OF THE ESTIMATORS FOR A CHANGE IN THE STANDARD DEVIATION

Results for the normal distribution are presented in Fig. 2.9. For all tests, the boxplots for change point locations k and $n-k$ show very similar uncertainty. Take for example the row of boxplots for $\hat{\tau}$ as found by Pet-CP in Fig. 2.9: when τ_{true} is located at k and $(n-k)$, the boxplots for Pet-CP have similar widths and the interquartile distances are close to 20. The wide interquartile ranges indicate considerable uncertainty for the location of changes in the standard deviation.

For both Pet-CP and CUSUM-CP, it is clear from the systematic error and the 95% confidence interval that the methods cannot be used to detect a change in standard deviation. The plots in the last row of Fig. 2.9 show that, for CvM-CP, the results improve with increasing size of the change, but only reach usable levels for the changes $\Delta\sigma = 2$. The spread and bias in the estimated change point locations are illustrated by the boxplot. Only for CvM-CP, $\Delta\sigma \geq 2$ and $\tau = 40 \sim 60$ is there any hope of getting a reliable answer.

(5) INFLUENCE OF THE SAMPLE SIZE ON ABILITY

For the mean, the ability of the detectors first increases as sample size n increases from 10 to 100 (Fig. 2.10). When sample size exceeds 100, the ability of the detectors becomes nearly constant, and the ability for $n = 1000$ is nearly the same as for $n = 100$. From the first plot in the first row of Fig. 2.10, for all magnitudes of change, the ability of Pet-CP equals 0 when the sample size is 10. Therefore, when the sample size is 10, Pet-CP is not capable of finding a change point and it is visibly outperformed by CUSUM-CP and CvM-CP.

Based on the first two plots in the bottom row of Fig. 2.10, the ability of both Pet-CP and CUSUM-CP stays at very low levels. Accordingly, in the case of Pet-CP and CUSUM-CP, a detection of a shift in the standard deviation is not possible, and the magnitude of $\Delta\sigma$ has no significant influence on their ability. For CvM-CP, the ability to detect a change in standard deviation increases considerably as the sample size increases from 30 to 100 (Fig. 2.10, last row, third column). The ability found for length $n = 1000$ suggests this increase continues more slowly between $n = 100$ and $n = 1000$. Therefore, compared to Pet-CP and CUSUM-CP, CvM-CP is superior in finding a change point in the standard deviation. Considering that the performance of CvM-CP is comparable to that of Pet-CP and CUSUM-CP in detecting a change point in the mean, its better performance in finding a change point in the standard deviation makes CvM-CP much more attractive in change point detection.

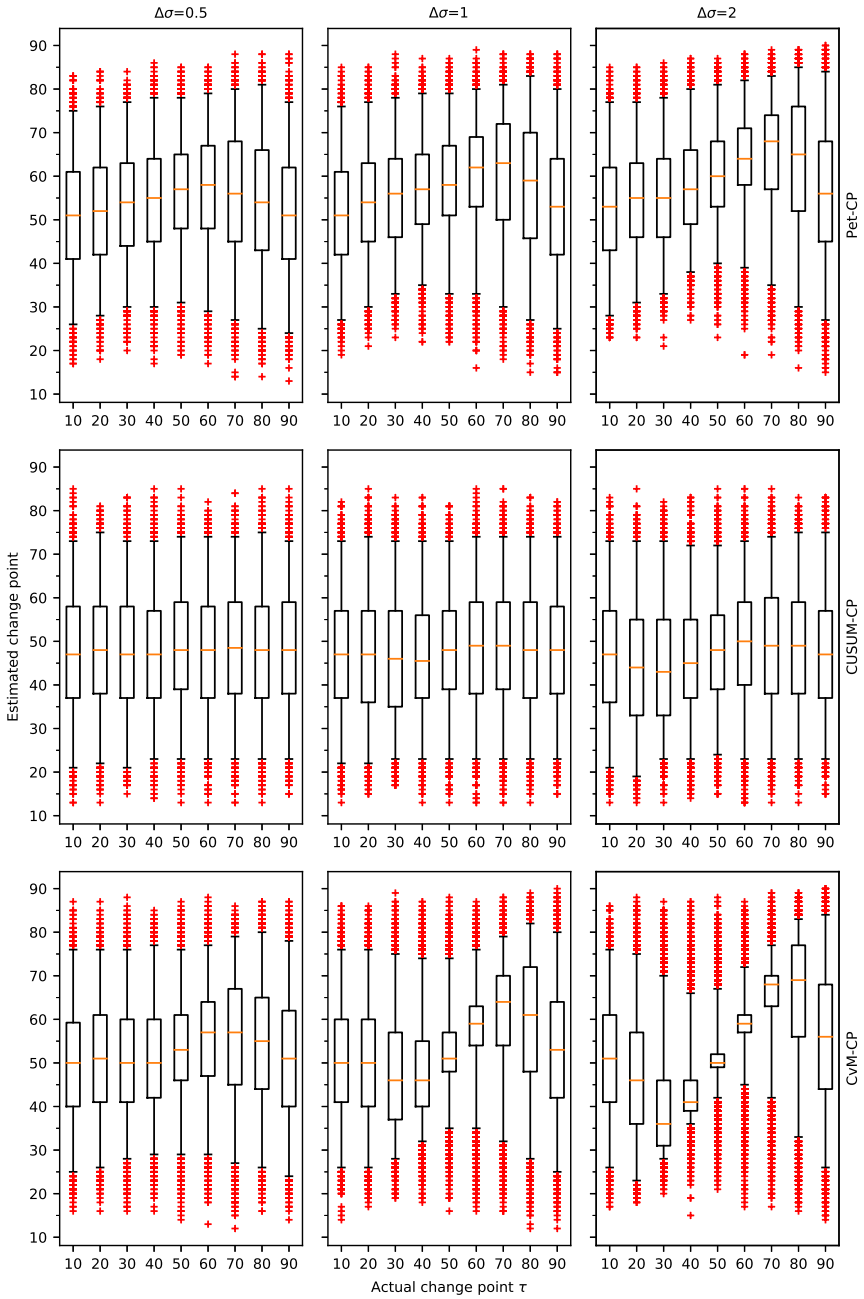


Figure 2.9: Boxplots of the error in the change point estimates based on 50 000 samples for a change in the standard deviation. The whiskers are at 2.5% and 97.5%; the crosses show the estimates outside that range.

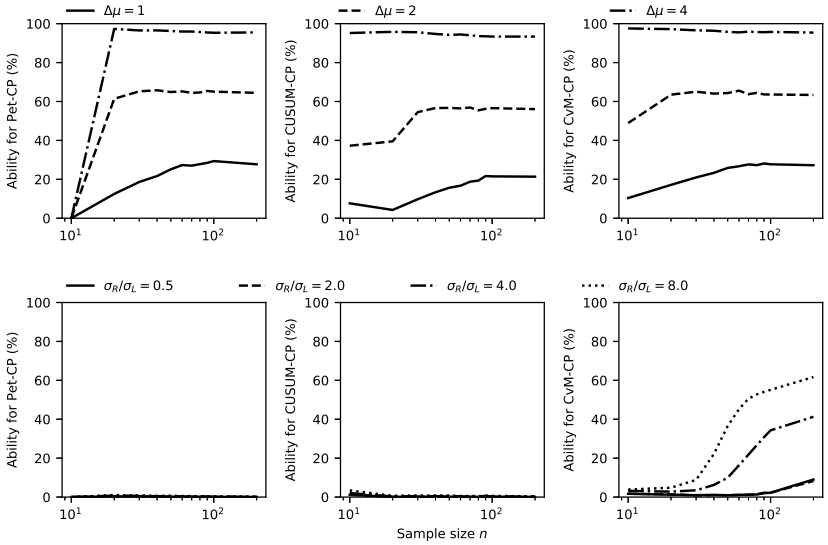


Figure 2.10: Ability of the different tests for a change in the mean and standard deviation at the midpoint of the series as a function of sample size n.

For change points near the start (or end) of the series, both power (Fig. 2.11 and ability (Fig. 2.12) decrease with increasing series length. From the power and ability of Pet-CP and CvM-CP shown in the first and third columns of Fig. 2.11 and 2.12, their performance in finding a change point located near the start (or end) is very similar and it stays constant till sample length 150; after that their performance decreases rapidly to a relatively low level. But for CUSUM-CP, its power and ability start decreasing when the sample length exceeds 20. For instance, in the middle column of Fig. 2.11, the power of CUSUM-CP decreases from 100% to 40% when the sample size changes from 20 to 30 for $\Delta\mu = 8$. From the experiments, we have observed that ability and power for similar relative change point locations, for instance $2n/10$, have similar values for different sample sizes. In brief: adding points at the end of a series makes detection of change points at the start of the series less likely. At the same time it makes detection of change points that were near the end before the addition of points at the end more likely.

2.4. APPLICATION OF THE TESTS TO HISTORICAL DATA FOR THE YANGTZE RIVER

2.4.1. EFFECT OF THE START AND END POINT OF THE SERIES

To investigate the influence of the time series length in practice, we took the longest time series corresponding to Cuntan station (1893-2014) and looked for change points in subseries. The starting year was varied from 1893 to 1957 and the end year from 1964 to 2014. The results are presented in Fig. 2.13, where a marker at a given pair of years indicates whether or not a change point was found.

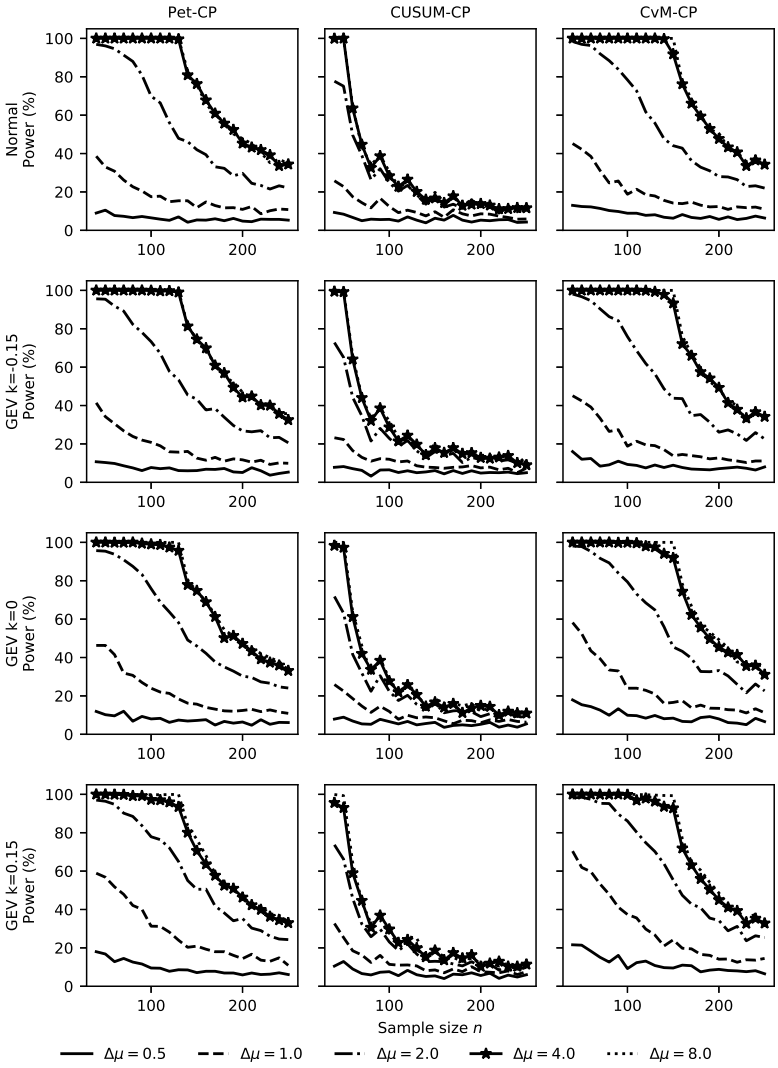


Figure 2.11: Power of the different tests for a change in mean at location 10 as a function of sample size n (number of samples $M = 1000$).

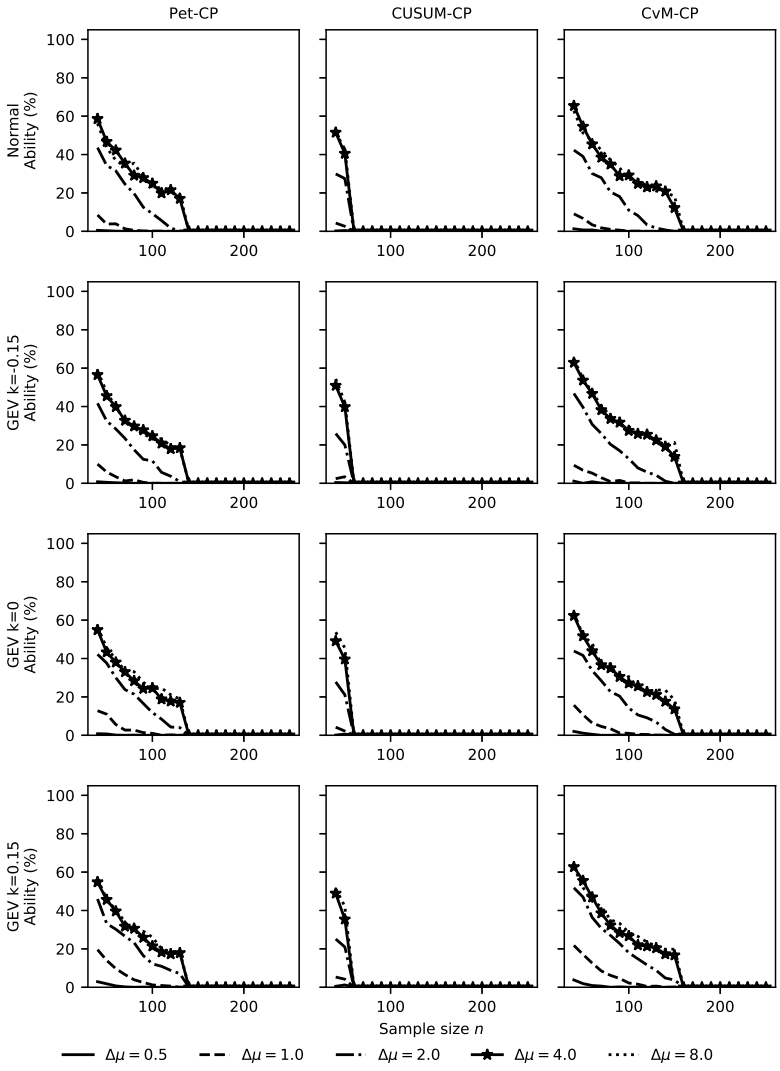


Figure 2.12: Ability of the different tests for a change in mean at location 10 as a function of sample size n (number of samples $M = 1000$).

Table 2.2: Change in the mean (μ) and standard deviation (σ) at each detected change point – Yichang station.

Change point	μ	σ	μ_L	μ_R	σ_L	σ_R	$\Delta\mu/\sigma$	$\Delta\sigma/\sigma$
	m^3/s						-	-
1962	49104	8642	55047	47101	4065	8876	-0.91	0.56
1966			54152	46896	4847	9038	-0.84	0.48

In Fig. 2.13, the different coloured points denote the different years of significant change for Cuntan station for subseries of years with different start and end years. The bottom plot shows that, depending on which subseries is used, CvM-CP may find three different change points. Comparison of the top and bottom rows shows a similar pattern of detection for subseries ending after 1995 for CvM-CP and Pet-CP. For series ending in 1980, Pet-CP detects 1966 as a change point for more starting years than the other two methods.

For time series with different combinations of start/end year, 1944 and 1966 are found as change points in some subseries by all three methods, but subseries with a significant change point located at 1968 are only found by CvM-CP. It is clear that for all methods the detection and location of a change point depend on the choice of subseries. In other words, different combinations of start/end year will lead to different change point detection results. The other time series showed similar effects.

As start and end year change, the change point appears, disappears and reappears, possibly in a different year. This is a cause for concern. If two researchers have access to datasets with different start and end points, then they may come to different conclusions about the presence and location of change points. This is particularly unfortunate if, for example, a design decision taken in 2020 on the basis of the absence of a change point in a time series turns out to be invalid in 2030, when the time series – now extended with data for the intervening years – shows a change point in 2010 that invalidates the analysis made in 2020.

Time series of yearly maxima increase in length by one year each year. If this can lead to the appearance or disappearance of change points far from the end of the series, it calls into question the reliability of the results.

2.4.2. CHANGE POINT DETECTION

The results of the application of the methods to the entire AMR time series of four gauge stations are as follows: Yichang station is the only one where change points are detected at the 5% significance level (see Table 2.2). For that station Pet-CP and CvM-CP find a change point in 1966 and CUSUM-CP finds one in 1962. The relative changes in mean and standard deviation for the change points are given in Table 2.2.

Other studies have also looked for change points in various types of hydrological series in the Yangtze River basin. For example, Xie *et al.* [33] applied the Pettitt's method and found a change in 1962 in the series of annual maxima at Yichang station for the period 1882–2010, with a p value of 0.0183. They also found a change in 1979 in the series of annual maxima series for 1952–2000 at Hankou station, with a p value of 0.2131. Xiong and Guo [55] studied the time series of mean annual flows at Yichang station and found a

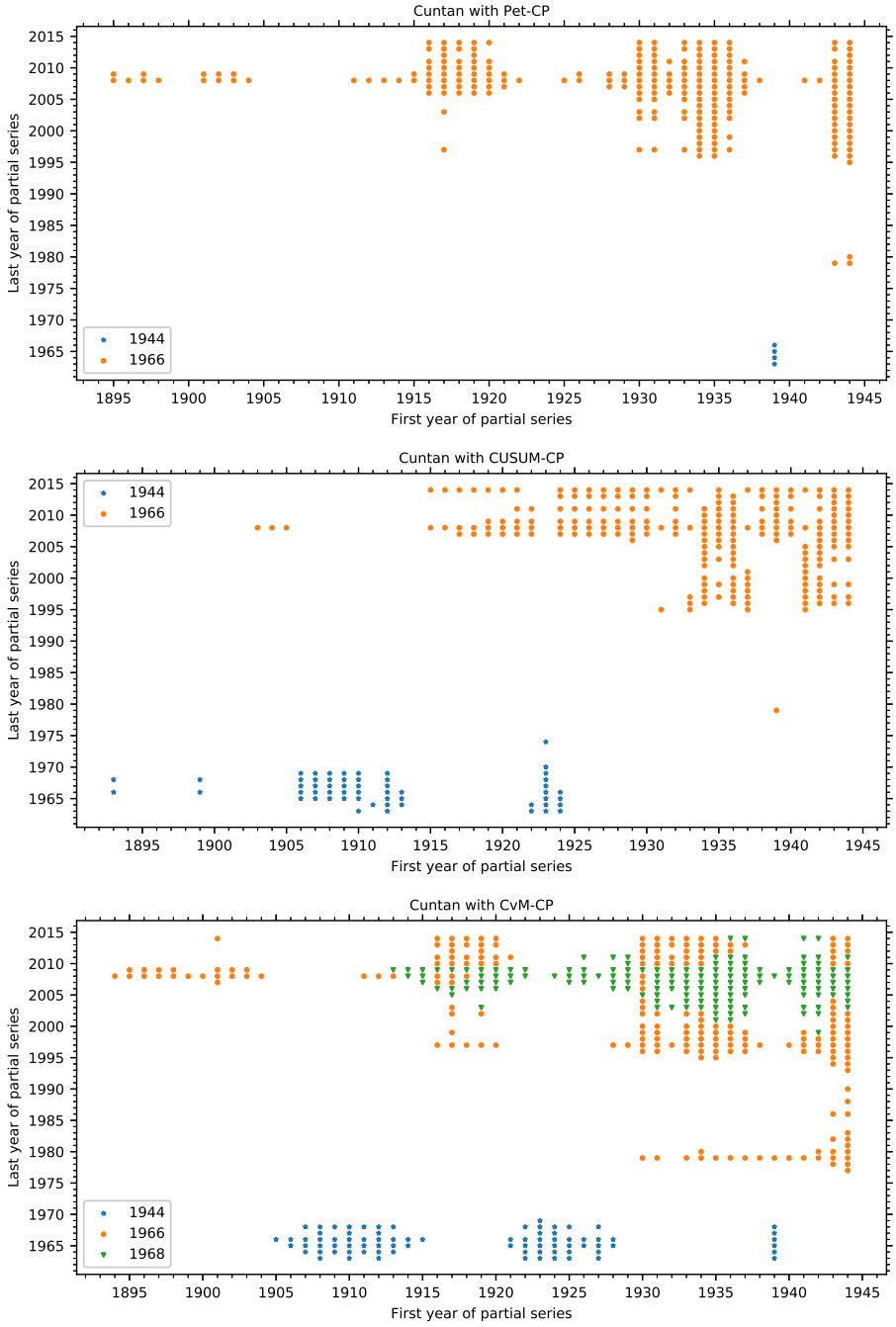


Figure 2.13: Plot of change points found in subseries of the Cuntan data by the three methods. A marker at a given coordinate pair (x, y) indicates whether or not a change point was found for a series starting in year x and ending in year y.

peak in the posterior distribution for the change point in 1968, close to the points found in this study.

None of the methods found a significant change point at a measurement station in the construction period of the dams upstream of that station. For the Three Gorges Dam (TGD) project the non-detection of a change point after the start of construction is in line with the analysis of the Yichang series of annual mean flows for the period 1882–2001 by Xiong and Guo [55], who found a peak only in the posterior distribution for the change point in 1968. However, this does not necessarily mean there is no change, Xiong and Guo [55] wrote:

> "As the change points for both the annual minimum and the annual mean series occurred before 1993 (the year in which the Three Gorges Project commenced), one can state that, since the construction of the Three Gorges Project there have not been any significant changes in the annual minimum or the annual mean series. However, it is very possible that the above conclusions might change with time, as the Three Gorges Project will definitely exert some influences on the flow regime of the Yangtze River at the Yichang hydrological station. Any change in the characteristics of the hydrological time series of Yichang station in the future could be a reason for modifying the initial construction and operation plan for the Three Gorges Project."

Our results for Yichang are consistent with those of earlier studies. To our knowledge, no study has yet found definite physical causes for a change point near 1966. It would be tempting to conclude that, between 1946 and 2014, the construction of the TGD project has not had a significant influence on Yichang station, but filling of the reservoir started only in 2003, so any change point resulting from dam operation would be very near the end of the gauge station time series and therefore much less likely to be detected by the methods used here.

2.5. CONCLUSIONS

The performance of several methods to detect an abrupt change in the statistical properties of synthetic and real times series was examined. The methods studied were Pettitt's test (Pet-CP), a CUSUM-based test (CUSUM-CP) and a test based on the Cramér von Mises two-sample test (CvM-CP). Based on experiments with synthetic data series from four distribution families: normal, generalized extreme value (GEV) with shape $k = -0.15$ (reverse Weibull), GEV with shape $k = 0$ (Gumbel) and GEV with shape $k = 0.15$ (Fréchet), it was found that the CvM-CP method had the best overall performance. However, all three methods have a serious short-coming: not only do they have great difficulty in detecting changes near the start or end of the time series, but they also tend to make large systematic errors in estimating the location of such changes.

The methods Pet-CP and CUSUM-CP could not detect a change in standard deviation for any of the distributions. For CvM-CP, the probability of correctly signalling a change in the standard deviation was much lower than for a change in the mean. The tests showed that, for a change in the mean, test ability did not differ much for samples from the different distributions.

For Pet-CP, CvM-CP and CUSUM-CP the power and ability to detect change points

plotted as a function of the change point are roughly symmetrical relative to a vertical line at $n/2$.

For the initial application of the tests to the annual maximum runoff time series from four gauge stations on the Yangtze River, the methods found change points only in the Yichang station series. Moreover, no change points were found after 1993, the start of the Three Gorges Dam (TGD) project. This is in line with findings by Xiong and Guo [55] for the period up to 2001, but the findings presented in this study on detection of change points near the end of a time series suggest that this cannot be considered as evidence that the TGD project did not cause an abrupt change in statistical properties of annual maximum runoff.

With respect to the questions posed in at the start of this study we found the following answers:

For the probability of incorrectly signaling a change point, it was found that, for CvM-CP, where an empirical distribution of the test statistic was used, the false positive rate was correct. For Pet-CP and CUSUM-CP, where a limit distribution of the test statistic was used, this turned out not to be fully justified even when the total time series length reached 100. For short series (less than 100 points) the asymptotic estimates of distribution quantiles for Pet-CP and CUSUM-CP were too high, and the resulting null hypothesis rejection rates were too low. We would recommend to either use special small sample approximations of the distribution, or generate an empirical distribution by a Monte Carlo method and use that as the test statistic distribution.

The probability of correctly detecting a change point for a change in the mean near the start and end of a time series was low (less than 10% for a change in the mean corresponding to one times the standard deviation, 1SD, of the signal). For a change in the standard deviation, only CvM-CP showed reasonable power.

When we considered all estimated change point locations, we found that estimates of change points near the start and end of a time series have a large bias (97.5% of all location estimates of a change at location 10 was beyond location 20 for a series with a change in the mean corresponding to 1SD of the signal) and a large uncertainty in the location estimate.

The effect of the length of the time series was twofold. For a change in the mean and a change point located in the middle of the series, it seems that the detection rate improves until a length of about 70 is reached. However, for a change point location at a fixed distance from the end of the series, the ability and power will decrease as the series length increases. This is particularly dramatic in case of a change point close to the start of the series, say at year 10. For a change in the standard deviation and a change point located in the middle of the series, only CvM-CP detects anything; and here detection keeps improving up to at least series length 200.

As was to be expected, larger changes result in better detection results. However, it is clear that relatively large changes are needed to get acceptable results.

Moreover, it mattered what start or end year was chosen for a time series. In other words: it was not safe to look at parts of a time series that contain a given range of potential change points, but had different start or end years. Application of the tests to real data series showed that when different start and end years were used, different results were indeed obtained. These experiments with detection of change points in subseries

of annual maxima demonstrated that change points may seem to appear and disappear when the end points of the series are shifted.

In summary, we found that, even under ideal circumstances of independent variables, no trend and, at most, one change point, the results of these methods need to be interpreted with great care: a few years of additional data or missing data may change the outcome of the detection experiment and change points near the start or the end of the time series are likely to be either missed or reported in the wrong location. NHST based method can only provide a 'Yes/No' result based on a pre-set significance level and leaves no room for uncertainty analysis. A new way to conduct uncertainty analysis to the change point detection should be proposed properly.

REFERENCES

[1] H. McMillan, A. Montanari, C. Cudennec, H. Savenije, H. Kreibich, T. Krueger, J. Liu, A. Mejia, A. Van Loon, H. Aksoy, *et al.*, *Panta Rhei 2013–2015: Global perspectives on hydrology, society and change,* Hydrological Sciences Journal **61**, 1174 (2016).

[2] A. Montanari, G. Young, H. Savenije, D. Hughes, T. Wagener, L. Ren, D. Koutsoyiannis, C. Cudennec, E. Toth, S. Grimaldi, *et al.*, *"Panta Rhei—everything flows": Change in hydrology and society–the IAHS scientific decade 2013–2022,* Hydrological Sciences Journal **58**, 1256 (2013).

[3] Z. W. Kundzewicz, *Nonstationarity in water resources–central European perspective 1,* JAWRA Journal of the American Water Resources Association **47**, 550 (2011).

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

[5] P. Gao, X. Zhang, X. Mu, F. Wang, R. Li, and X. Zhang, *Trend and change-point analyses of streamflow and sediment discharge in the Yellow River during 1950–2005,* Hydrological Sciences Journal–Journal des Sciences Hydrologiques **55**, 275 (2010).

[6] B. P. Dudding and W. J. Jennett, *Quality Control Charts: Being Part 1 of a Revision of B. S. 600: 1935, The Application of Statistical Methods to Industrial Standardisation and Quality Control* (British Standards Institution, 1942).

[7] E. S. Page, *Continuous inspection schemes,* Biometrika **41**, 100 (1954).

[8] C. A. McGilchrist and K. D. Woodyer, *Note on a distribution-free CUSUM technique,* Technometrics **17**, 321 (1975).

[9] É. Lebarbier, *Detecting multiple change-points in the mean of Gaussian process by model selection,* Signal Processing **85**, 717 (2005).

[10] M. Lavielle and G. Teyssiere, *Detection of multiple change-points in multivariate time series,* Lithuanian Mathematical Journal **46**, 287 (2006).

[11] D. S. Matteson and N. A. James, *A nonparametric approach for multiple change point analysis of multivariate data,* Journal of the American Statistical Association **109**, 334 (2014).

[12] B. K. Ray and R. S. Tsay, *Bayesian methods for change-point detection in long-range dependent processes,* Journal of Time Series Analysis **23**, 687 (2002).

[13] I. Berkes, L. Horváth, P. Kokoszka, Q.-M. Shao, *et al.*, *On discriminating between long-range dependence and changes in mean,* The Annals of Statistics **34**, 1140 (2006).

[14] R. Lund, X. L. Wang, Q. Q. Lu, J. Reeves, C. Gallagher, and Y. Feng, *Changepoint detection in periodic and autocorrelated time series,* Journal of Climate **20**, 5178 (2007).

[15] E. Gombay, *Change detection in autoregressive time series,* Journal of Multivariate Analysis **99**, 451 (2008).

[16] H. Cho and P. Fryzlewicz, *Multiple-change-point detection for high dimensional time series via sparsified binary segmentation,* Journal of the Royal Statistical Society: Series B: Statistical Methodology , 475 (2015).

[17] T. Zhang and L. Lavitas, *Unsupervised self-normalized change-point testing for time series,* Journal of the American Statistical Association **113**, 637 (2018).

[18] X. Shao and X. Zhang, *Testing for change points in time series,* Journal of the American Statistical Association **105**, 1228 (2010).

[19] Y. Xie, J. Huang, and R. Willett, *Change-point detection for high-dimensional time series with missing data,* IEEE Journal of Selected Topics in Signal Processing **7**, 12 (2012).

[20] X. L. Wang, *Comments on "Detection of undocumented changepoints: A revision of the two-phase regression model",* Journal of Climate **16**, 3383 (2003).

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

[22] G. Thirel, V. Andréassian, C. Perrin, J.-N. Audouy, L. Berthet, P. Edwards, N. Folton, C. Furusho, A. Kuentz, J. Lerat, *et al.*, *Hydrology under change: An evaluation protocol to investigate how hydrological models deal with changing catchments,* Hydrological Sciences Journal **60**, 1184 (2015).

[23] E. Hajani, A. Rahman, and E. Ishak, *Trends in extreme rainfall in the state of New South Wales, Australia,* Hydrological Sciences Journal **62**, 2160 (2017).

[24] Z. Sa'adi, S. Shahid, T. Ismail, E.-S. Chung, and X.-J. Wang, *Trends analysis of rainfall and rainfall extremes in Sarawak, Malaysia using modified Mann–Kendall test,* Meteorology and Atmospheric Physics **131**, 263 (2019).

[25] R. Lund and J. Reeves, *Detection of undocumented changepoints: A revision of the two-phase regression model,* Journal of Climate **15**, 2547 (2002).

[26] M. J. Menne and C. N. Williams Jr, *Detection of undocumented changepoints using multiple test statistics and composite reference series,* Journal of Climate **18**, 4271 (2005).

[27] X. L. Wang, *Penalized maximal F test for detecting undocumented mean shift without trend change,* Journal of Atmospheric and Oceanic Technology **25**, 368 (2008).

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

[29] H. Chernoff and S. Zacks, *Estimating the current mean of a normal distribution which is subjected to changes in time,* The Annals of Mathematical Statistics **35**, 999 (1964).

[30] Z. Kander, S. Zacks, *et al.*, *Test procedures for possible changes in parameters of statistical distributions occurring at unknown time points,* The Annals of Mathematical Statistics **37**, 1196 (1966).

[31] D. M. Hawkins, *Self-starting CUSUM charts for location and scale,* Journal of the Royal Statistical Society: Series D (The Statistician) **36**, 299 (1987).

[32] G. Gurevich and A. Vexler, *Retrospective change point detection: From parametric to distribution free policies,* Communications in Statistics—Simulation and Computation® **39**, 899 (2010).

[33] H. Xie, D. Li, and L. Xiong, *Exploring the ability of the Pettitt method for detecting change point by Monte Carlo simulation,* Stochastic Environmental Research and Risk Assessment **28**, 1643 (2014).

[34] I. Mallakpour and G. Villarini, *A simulation study to examine the sensitivity of the Pettitt test to detect abrupt changes in mean,* Hydrological Sciences Journal **61**, 245 (2016).

[35] M. Holmes, I. Kojadinovic, and J.-F. Quessy, *Nonparametric tests for change-point detection à la Gombay and Horváth,* Journal of Multivariate Analysis **115**, 16 (2013).

[36] L. Xiong, C. Jiang, C.-Y. Xu, K.-x. Yu, and S. Guo, *A framework of change-point detection for multivariate hydrological series,* Water Resources Research **51**, 8198 (2015).

[37] F. Chiew and T. McMahon, *Detection of trend or change in annual flow of Australian rivers,* International Journal of Climatology **13**, 643 (1993).

[38] A. Abdul Rahman, S. S. Syed Yahaya, and A. M. A. Atta, *The effect of median based estimators on CUSUM chart,* Journal of Telecommunication, Electronic and Computer Engineering **10**, 49 (2018).

[39] Z. Kundzewicz and A. Robson, *Detecting trend and other changes in hydrological data* (World Meteorological Organization, 2000).

[40] Z. W. Kundzewicz and A. J. Robson, *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**, 7 (2004).

[41] K. R. Ryberg, G. A. Hodgkins, and R. W. Dudley, *Change points in annual peak streamflows: Method comparisons and historical change points in the united states,* Journal of Hydrology **583**, 124307 (2020).

[42] T. W. Anderson and D. A. Darling, *A test of goodness of fit,* Journal of the American Statistical Association **49**, 765 (1954).

[43] T. W. Anderson, *On the distribution of the two-sample Cramer-von Mises criterion,* The Annals of Mathematical Statistics , 1148 (1962).

[44] E. Gombay and L. Horváth, *Change-points and bootstrap,* Environmetrics: The official journal of the International Environmetrics Society **10**, 725 (1999).

[45] A. Bücher, I. Kojadinovic, T. Rohmer, and J. Segers, *Detecting changes in cross-sectional dependence in multivariate time series,* Journal of Multivariate Analysis **132**, 111 (2014).

[46] J. Antoch, M. Hušková, and Z. Prášková, *Effect of dependence on statistics for determination of change,* Journal of Statistical Planning and Inference **60**, 291 (1997).

[47] N. G. Reich, J. A. Myers, D. Obeng, A. M. Milstone, and T. M. Perl, *Empirical power and sample size calculations for cluster-randomized and cluster-randomized crossover studies,* PLoS One **7**, e35564 (2012).

[48] Hosking, *Estimation of the generalized extreme-value distribution by the method of probability-weighted moments,* Technometrics **27**, 251 (1985).

[49] J. R. Hosking, *Algorithm AS 215: Maximum-likelihood estimation of the parameters of the generalized extreme-value distribution,* Journal of the Royal Statistical Society. Series C (Applied Statistics) **34**, 301 (1985).

[50] D. Koutsoyiannis, *Statistics of extremes and estimation of extreme rainfall: II. Empirical investigation of long rainfall records/Statistiques de valeurs extrêmes et estimation de précipitations extrêmes: II. Recherche empirique sur de longues séries de précipitations,* Hydrological Sciences Journal **49** (2004).

[51] R. van Nooijen and A. Kolechkina, *Estimates of extremes in the best of all possible worlds,* in *3rd STAHY international workshop on statistical methods for hydrology and water resources management* (2012) pp. 1–2.

[52] Y. Wang, Y. Ding, B. Ye, F. Liu, and J. Wang, *Contributions of climate and human activities to changes in runoff of the Yellow and Yangtze rivers from 1950 to 2008,* Science China Earth Sciences **56**, 1398 (2013).

[53] S. Yang, I. Belkin, A. Belkina, Q. Zhao, J. Zhu, and P. Ding, *Delta response to decline in sediment supply from the Yangtze River: Evidence of the recent four decades and expectations for the next half-century,* Estuarine, Coastal and Shelf Science **57**, 689 (2003).

[54] Q. Zhang, C. Liu, C.-y. Xu, Y. Xu, and T. Jiang, *Observed trends of annual maximum water level and streamflow during past 130 years in the Yangtze River basin, China,* Journal of Hydrology **324**, 255 (2006).

[55] L. Xiong and S. Guo, *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,* Hydrological Sciences Journal **49**, 99 (2004).

[56] Q. Zhang, Y. Zhou, V. P. Singh, and X. Chen, *The influence of dam and lakes on the Yangtze River streamflow: Long-range correlation and complexity analyses,* Hydrological Processes **26**, 436 (2012).

[57] P. P. Talwar and J. E. Gentle, *Detecting a scale shift in a random sequence at an unknown time point,* Applied Statistiss , 301 (1981).

3

BACKGROUND OF CONFIDENCE CURVES

"... but here is a safe prediction for the 21st century:
statisticians will be asked to solve bigger and more complicated problems.
I believe there is a good chance that objective Bayes methods
will be developed for such problems, and that something like
fiducial inference will play an important role in this development.
Maybe Fisher's biggest blunder will become a big hit in the 21st century!"
Bradley Efron (Caltech; Stanford University)

3.1. DEFINITIONS OF CONFIDENCE CURVES

In the literature confidence curves have been defined in several different ways. The following definition provides a starting point, γ is used to denote a confidence level.

Definition 1. A confidence interval with confidence level (also known as confidence coefficient) γ 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

$$\Pr(u(X) \leq \lambda_0 \leq v(X)) = \gamma$$

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 γ for a parameter λ is a random set $R(X)$ such that

$$\Pr(\lambda \in R(X)) = \gamma$$

Here γ is the nominal coverage probability of the set. In the calculation of γ 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 γ, then the true value lies in the set with probability greater than γ. 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 γ, 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 and Molchanov [1]. 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 [2] 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 [3, 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

$$\min_{\lambda \in V} g(\lambda, x) = g\left(\hat{\lambda}(x), x\right) = 0 \tag{3.1}$$

   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

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

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

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

For the case with unknown σ, see, for example, Schweder and Hjort [3, page 73].

3.2. COVERAGE PROBABILITY OF CONFIDENCE SETS

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 $\gamma \in [0,1]$ and a given property value λ, it is possible to determine

$$\Pr\left(\text{cc}(\lambda, X) \le \gamma\right) \tag{3.2}$$

Next, define the sets

$$R_\gamma(x) = \left\{\lambda : \text{cc}(\lambda, x) \le \gamma\right\} \tag{3.3}$$

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

$$\Pr\left(\text{cc}(\lambda_{\text{true}}, X) \le \gamma\right) = \gamma \tag{3.4}$$

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

$$\Pr\left(\lambda_{\text{true}} \in R_\gamma(X)\right) = \Pr\left(\text{cc}(\lambda_{\text{true}}, X) \leq \gamma\right) \tag{3.5}$$

Combined with the fact that $\text{cc}(\lambda_{\text{true}}, X)$ is uniformly distributed on $[0,1]$, it now follows that the $R_\gamma(X)$ is a confidence set with confidence level γ. 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

$$\frac{1}{m}\sum_{j=1}^{m}\left[\!\left[\text{cc}\left(\lambda_0, x^{(j)}\right) \leq \gamma\right]\!\right] - \gamma \tag{3.6}$$

goes to zero as m increases, where $[\![\cdot]\!]$ is the indicator function (A.1).

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 $R_\gamma(x)$. Moreover, the sets $R_\gamma(x)$ derived from a confidence curve are nested, which further limits the number of available sets. As each subset will correspond to one confidence level γ, and there are infinitely many confidence levels, the best we can hope to achieve is $\Pr\left(\text{cc}(\lambda_{\text{true}}, X) \leq \gamma\right) \approx \gamma$, so

$$\Pr\left(\text{cc}(\lambda_{\text{true}}, X) \leq \gamma\right) - \gamma \tag{3.7}$$

cannot be zero for all γ, and therefore (3.6) cannot go to zero for all γ, 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 $\gamma(n-1)$ points is selected randomly from $\{1,2,\ldots,n-1\}$, then the probability that the actual change point τ_{true} lies in that set is approximately γ. This suggests that a set $R_\gamma(x)$ that contains more points than $\gamma(n-1)$ indicates large uncertainty at that confidence level, while sets $R_\gamma(x)$ that are much smaller correspond to low uncertainty at that confidence level. In such a way the size of the sets $R_\gamma(x_{\text{obs}})$ for the observed sample x_{obs} can be linked to the uncertainty in the location of the change.

3.3. CONFIDENCE CURVES FOR THE LOCATION OF A CHANGE POINT BASED ON PARAMETRIC LOG-LIKELIHOOD AND DEVIANCE FUNCTIONS

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

3.3.1. DESCRIPTION OF CHANGE POINT DETECTION PROBLEM

The formulation of 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 $\tau \in \{1, 2, ..., n-1\}$ such that the original random vector is split into two sub-series: a sub-series with i.i.d. random variables $\{Y_1, Y_2, ..., Y_\tau\}$ and a sub-series with i.i.d. random variables $\{Y_{\tau+1}, Y_{\tau+2}, ..., Y_n\}$, but the distributions of the variables in the two sub-series 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 $F(\cdot;\theta,\zeta)$ and the pdf as $f(\cdot;\theta,\zeta)$ where θ and ζ are vectors. Together the parameters in θ and ζ fully determine the distribution. In the model only the parameters in θ change at the CP. The parameter vectors for the left and the right sub-series will be θ_L and θ_R respectively. The null hypothesis can now be expressed as $\theta_L = \theta_R$. 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. In the remainder of the thesis y will represent a realization of Y, and y_{obs} will represent the actual observed time series.

Both the parametric and the non-parametric methods to construct confidence curves for the location of CP 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. Intuitively it is clear that parameter estimation for very small samples will be difficult. Some studies considering this are Lettenmaier and Burges [4], Delicado and Goria [5]. These suggest that for short sub-series the results may vary considerably for sample to sample.

A minimum sub-series length $n_{\min}$ will therefore be used. As a result only a subset of CP locations, given by

$$L_{\text{CP}} = \{n_{\min}, n_{\min+1}, ..., n - n_{\min}\} \tag{3.8}$$

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

$$n_{\min} = \lfloor 2 \log n \rfloor \tag{3.9}$$

where the notation $\lfloor \cdot \rfloor$ 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 [6]. This would seem undesirable when looking for parameter changes.

Moreover, our non-parametric method which uses an approximate empirical likelihood is related to the empirical likelihood method discussed in Zou *et al.* [7] 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 approxima-

tion we use for the empirical log-likelihood does not hold everywhere, but it does hold for change points at $n_{\min}, n_{\min+1}, \ldots, n - n_{\min}$.

3.3.2. CONFIDENCE CURVES BASED ON THE PROFILE LOG-LIKELIHOOD

Cunen *et al.* [8] presented a method to construct confidence curves (Definition 3) based on the log-likelihood function ℓ. In the case of change point detection ℓ is given by

$$\ell(\tau, \theta_L, \theta_R, \zeta; y) = \sum_{i=1}^{\tau} \log f(y_i; \theta_L, \zeta) + \sum_{i=\tau+1}^{n} \log f(y_i; \theta_R, \zeta) \tag{3.10}$$

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 [9], 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 (3.10) 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

$$\ell_{\text{prof}}(\tau; y) = \sup_{\theta_L, \theta_R, \zeta} \ell(\tau, \theta_L, \theta_R, \zeta; y) \tag{3.11}$$

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

$$\ell_{\text{prof}}(\tau; y) = \ell(\tau, \hat{\theta}_L(\tau; y), \hat{\theta}_R(\tau; y), \hat{\zeta}(\tau; y); y) \tag{3.12}$$

The value of τ for which ℓ_{prof} is maximal will be denoted by $\hat{\tau}(y)$. The values $\hat{\theta}_L(\hat{\tau}(y); y)$, $\hat{\theta}_R(\hat{\tau}(y); y)$, 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

$$D(\tau, y) = 2\left(\ell_{\text{prof}}(\hat{\tau}(y); y) - \ell_{\text{prof}}(\tau; y)\right) \tag{3.13}$$

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

$$\begin{aligned} &\forall r \in \mathbb{R} : K_\tau(r) = \\ &\Pr\left(D(\tau, Y) < r \mid \tau, \theta_L = \hat{\theta}_L(\hat{\tau}(y); y),\right. \\ &\left.\theta_R = \hat{\theta}_R(\hat{\tau}(y); y), \zeta = \hat{\zeta}(\hat{\tau}(y); y)\right) \end{aligned} \tag{3.14}$$

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 $\mathrm{cc}(\tau_{\text{true}}; Y)$ is uniformly distributed on $[0,1]$ when τ_{true} is the true value of τ. Nevertheless, as in Cunen *et al.* [8], D will be used to define the function that will be referred to as a confidence curve

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

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 $k \in \{n_{\min}, n_{\min}+1, \ldots, n-n_{\min}\}$ 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 $\theta = \hat{\theta}_{\text{R}}(\hat{\tau}(y_{\text{obs}}); y_{\text{obs}})$, $\zeta = \hat{\zeta}(\hat{\tau}(y_{\text{obs}}); y_{\text{obs}})$; $n_{\min}$ is used to avoid calculation of profile log-likelihoods based on a handful of points.
3. Approximate the curve $\mathrm{cc}(\tau; y_{\text{obs}}) = K_\tau\left(D(\tau, y_{\text{obs}})\right)$ by

$$K_{\tau,N}\left(D(\tau, y_{\text{obs}})\right) = \frac{1}{N}\sum_{j=1}^{N}\left[\!\left[D\left(\tau, y^{(j,\tau)}\right) < D(\tau, y_{\text{obs}})\right]\!\right] \tag{3.15}$$

If the parameter of interest is continuous (λ), according to Wilks' theorem, the probability distribution of deviance function $D(\cdot)$, $K(D(\cdot))$ is approximately a χ^2_1 [8], similar suggestions can be found in Schweder and Hjort [3, Section 1.6, 2.4]. For continuous variable for instance the dependence parameter in copulas, $K(D(\cdot))$ is approximately the distribution function of a χ^2_1.

3.4. CONCLUSIONS AND REMARKS

From the 'method B' of Cunen *et al.* [8], if a parametric distribution of observations is known, a confidence curve for a parameter of interest can be constructed based on a parametric profile log-likelihood function. It provides a way to construct confidence curves for a discrete parameter, for instance the location of change point τ, by deviance function and Monte Carlo simulation. In the parametric likelihood function, only the location of a change point is the parameter of interest, and parameters in parametric distributions for real data are taken as nuisance. Therefore, it is a multi-parametric problem.

However, Cunen *et al.* [8] used a maximum likelihood estimator (mle) to estimate all parameters in two steps. All nuisance parameters are estimated in the first step by the mle. The estimated nuisance parameters are substituted into a profile log-likelihood function and the location of a change point is estimated by the mle method. Given the fact that estimating nuisance parameters by mle in the first step is often computationally expensive, a more efficient way to estimate all parameters in the framework

of Cunen *et al.* [8] is using a pseudo maximum likelihood estimator (pmle) to estimate nuisance parameters, then estimate a parameter of interest by maximizing the profile log-likelihood function. Therefore, in Chapter 4, we compared confidence curves based on two parametric methods. One is the method by Cunen *et al.* [8] and the other is the newly proposed method based on pmle (including Method of Moments: MoM or Linear-Moments: LMo).

Furthermore, the 'method B' 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. Clearly, a new method that depends on empirical frequency of samples is necessary. Therefore, in Chapter 5 we proposed a new non-parametric method and compared it with the parametric method by Cunen *et al.* [8]. One is the method by Cunen *et al.* [8], and the other is a method based on an approximate empirical likelihood function.

To be clear, the method proposed by Cunen *et al.* [8] is called Confidence curve based on profile likelihood, Maximum Likelihood and deviance function (CML) in Chapter 4 and 5. In Chapter 4, the confidence curve based on pmle (MoM or LMo) is called Confidence curve by Method of moments (CMoM) or Confidence curve by Linear Moments (CMLo). In Chapter 5, the non-parametric method is called confidence curve based on Approximate Empirical likelihood and Deviance function and bootstrapping (AED).

REFERENCES

[1] I. Molchanov and I. S. Molchanov, *Theory of random sets* (Springer, 2005).

[2] A. Birnbaum, *Confidence curves: An omnibus technique for estimation and testing statistical hypotheses,* Journal of the American Statistical Association **56**, 246 (1961).

[3] T. Schweder and N. L. Hjort, *Confidence, likelihood, probability. Statistical inference with confidence distributions.* (Cambridge: Cambridge University Press, 2016) pp. xx + 500.

[4] D. P. Lettenmaier and S. J. Burges, *Gumbel's extreme value I distribution: A new look,* Journal of Hydraulic Engineering **108**, 502 (1982).

[5] P. Delicado and M. N. Goria, *A small sample comparison of maximum likelihood, moments and L-moments methods for the asymmetric exponential power distribution,* Comput. Statist. Data Anal. **52**, 1661 (2008).

[6] J. M. Landwehr, N. Matalas, and J. Wallis, *Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles,* Water Resources Research **15**, 1055 (1979).

[7] C. Zou, Y. Liu, P. Qin, and Z. Wang, *Empirical likelihood ratio test for the change-point problem,* Statistics & Probability Letters **77**, 374 (2007).

[8] C. Cunen, G. Hermansen, and N. L. Hjort, *Confidence distributions for change-points and regime shifts,* Journal of Statistical Planning and Inference **195**, 14 (2018).

[9] S. A. Murphy and A. W. Van der Vaart, *On profile likelihood,* Journal of the American Statistical Association **95**, 449 (2000).

4

CONFIDENCE CURVES BASED ON THE PSEUDO MAXIMUM LIKELIHOOD METHOD

Pseudo likelihood, which is also called Quasi likelihood was first introduced by Robert W.M. Wedderburn (1947-1975). Now, it has been widely used, for instance fitting generalized linear models.

Parts of this chapter have been submitted as "Zhou, C., van Nooijen, R., and Kolechkina, A., Capturing the uncertainty about a sudden change in the properties of time series with confidence curves, Journal of Hydrology, under review, 2021.".

4.1. INTRODUCTION

Today, the need to take into account climate variability and the results of human interventions in water management and hydrology seems clear [1, 2]. 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 behaviour 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 [3]. 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 [3, 4], or to see whether it is necessary to split a series into two parts for further analysis [5]. Beaulieu *et al.* [6] 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 [7].

In this chapter 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 [6, 8–10]. Theoretical work on CP detection in general was done, for example, by Pettitt [11], Chen and Gupta [12], Chen and Gupta [13], or Brodsky and Darkhovsky [14].

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 [11], 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. Strictly speaking, the methods discussed in this chapter 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' [15]. This is of interest in situations where a large set of time series is studied and it would not be feasible to analyse 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.

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 [16]. 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 [17] 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. It therefore makes sense to examine both types of CP tests.

An example of a nonparametric detection method using confidence curves can be found in Zhou *et al.* [18]. The current chapter 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 pseudo likelihood.

A method based on the first approach can be found in Cunen *et al.* [19] 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 estimates* (CML). A potential problem with this method is that it is very computationally intensive (refer to G for detailed computational costs.). Even in the case of just one CP, two optimizations of a log-likelihood need to be done for each possible CP location 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 G, 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 sample length. Run-times on a desktop workstation by spftwares for instance MATLAB may take hundreds of seconds for a single sample. Removal of the maximum likelihood estimator (mle) 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 a pseudo maximum likelihood estimator (pmle). More information on pseudo likelihood can be found in, for example, Gong and Samaniego [20]. 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 Linear Moments* (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 nec-

essary 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 remainder of this chapter 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 analysed. 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.

4.2. METHODOLOGY

All time series will be modelled as a vector Y of n independent random variables $Y_1, Y_2, ..., Y_n$. In the remainder of the chapter, 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 *independent identically distributed* (i.i.d.) random variables. 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. random variables $\{Y_1, Y_2, ..., Y_\tau\}$ and a subseries with i.i.d. random variables $\{Y_{\tau+1}, Y_{\tau+2}, ..., Y_n\}$, 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 $\theta_{\text{L}} = \theta_{\text{R}}$.

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 [21], Delicado and Goria [22]. These suggest that for short sub-series the results may vary considerably for sample to sample. A minimum sub-sample length $n_{\min}$ by (3.9) will therefore be considered. Therefore, candidates of CP location will be within the subset given by (3.8).

4.2.1. 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

$$\ell(\tau, \theta_{\text{L}}, \theta_{\text{R}}; y) = \sum_{i=1}^{\tau} \log f(y_i; \theta_{\text{L}}) + \sum_{j=\tau+1}^{n} \log f(y_j; \theta_{\text{R}}); \tau \in \{1, 2, ..., n-1\} \tag{4.1}$$

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 (4.1) over all θ_L, θ_R; the resulting function is called the *profile log-likelihood*

$$\ell_{\text{prof}}(\tau; y) = \sup_{\theta_L,\theta_R} \ell(\tau, \theta_L, \theta_R; y) \tag{4.2}$$

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 (4.2) is equivalent to

$$\ell_{\text{prof}}(\tau; y) = \ell\left(\tau, \hat{\theta}_L(\tau, y), \hat{\theta}_R(\tau, y); y\right) \tag{4.3}$$

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

$$\hat{\tau}(y) = \min\left(\text{argmax}_{\tau \in L_{CP}} \ell\left(\tau, \hat{\theta}_L(\tau, y), \hat{\theta}_R(\tau, y); y\right)\right) \tag{4.4}$$

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 (4.1)

$$\ell_{\text{pseu}}(\tau; y) = \ell\left(\tau, \tilde{\theta}_L(\tau, y), \tilde{\theta}_R(\tau, y); y\right) \tag{4.5}$$

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

$$\tilde{\tau}(y) = \min\left(\text{argmax}_{\tau \in L_{CP}} \ell\left(\tau, \tilde{\theta}_L(\tau, y), \tilde{\theta}_R(\tau, y); y\right)\right) \tag{4.6}$$

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 $\hat{\tau}(y)$

$$D_{\text{prof}}(\tau, y) = 2\{\ell_{\text{prof}}(\hat{\tau}(y); y) - \ell_{\text{prof}}(\tau; y)\} \tag{4.7}$$

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

$$D_{\text{pseu}}(\tau, y) = 2\{\ell_{\text{pseu}}(\tilde{\tau}(y); y) - \ell_{\text{pseu}}(\tau; y)\} \tag{4.8}$$

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

$$\begin{gathered}\forall r \in \mathbb{R} : K_{\tau,\text{prof}}(r) = \\ \Pr\left(D_{\text{prof}}(\tau, y) < r | \tau, \theta_L = \hat{\theta}_L(\hat{\tau}(y), y), \theta_R = \hat{\theta}_R(\hat{\tau}(y), y); y\right)\end{gathered} \tag{4.9}$$

and for CLMo (CMoM) from

$$\begin{gathered}\forall r \in \mathbb{R} : K_{\tau,\text{pseu}}(r) = \\ \Pr\left(D_{\text{pseu}}(\tau, y) < r | \tau, \theta_L = \tilde{\theta}_L(\tilde{\tau}(y), y), \theta_R = \tilde{\theta}_R(\tilde{\tau}(y), y); y\right)\end{gathered} \tag{4.10}$$

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.* [19] and in this paper, a *confidence curve* is defined using the distribution of the deviance function

$$\text{cc}(\tau, y_{\text{obs}}) = K_\tau(D(\tau, y_{\text{obs}})) \tag{4.11}$$

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 $\theta^*_{\text{L}}(\tau^*(y_{\text{obs}}), y_{\text{obs}})$ and $\theta^*_{\text{R}}(\tau^*(y_{\text{obs}}), y_{\text{obs}})$.

2. For each possible location $\tau \in \{n_{\min}, n_{\min+1}, ..., n - n_{\min}\}$, and $j = 1, 2, ..., N$, draw a new sample $y^{(j;k)}$ where the components $y_i^{(j,k)}$ $(k = 1, 2, ..., \tau)$ are distributed according to the distribution $F(\cdot;\theta)$ with $\theta = \theta^*_{\text{L}}(\tau^*(y_{\text{obs}}), y_{\text{obs}})$ and the components $y_i^{(j;\tau)}$ $(k = \tau+1, \tau+2, ..., n)$ are distributed according to the distribution $F(\cdot;\theta)$ with $\theta = \theta^*_{\text{R}}(\tau^*(y_{\text{obs}}), y_{\text{obs}})$.

3. Approximate the confidence curve $\text{cc}(\tau, y_{\text{obs}}) = K_\tau(D(\tau, y_{\text{obs}}))$ by

$$K_{\tau,N}\left(D(\tau, y_{\text{obs}})\right) = \frac{1}{N}\sum_{j=1}^{N} [\![D(\tau, y^{(j,\tau)}) < D(\tau, y_{\text{obs}})]\!] \tag{4.12}$$

where $[\![\cdot]\!]$ is the indicator function; 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).

4.2.2. PROPERTIES OF CONFIDENCE CURVES FOR THE LOCATION OF A CHANGE POINT

The performance of CML and CLMo (CMoM) methods will be examined and compared by exploring some properties [18] 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.

- 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 3.3.

- A summary of the *uncertainty* about the CP associated with a confidence curve cc, which is as follows

$$\mathrm{Un}(\mathrm{cc}) = \frac{\left(\sum_{k=n_{\min}}^{n-n_{\min}} [\![\mathrm{cc}(k) \leq \gamma_{\max}]\!]\right) - 1}{n - 2n_{\min}} \tag{4.13}$$

where

$$\gamma_{\max} = \frac{n - 2n_{\min}}{n - 2n_{\min} + 1} \tag{4.14}$$

as defined in Appendix E.

- The *similarity index* is used to measure the similarity of two confidence curves. In D.2 a derivation is given for

$$\tilde{J}(\mathrm{cc}, \mathrm{cc}') = \frac{\sum_{k=n_{\min}}^{n-n_{\min}} \min\left(1 - \mathrm{cc}(k), 1 - \mathrm{cc}'(k)\right)}{\sum_{k=n_{\min}}^{n-n_{\min}} \max\left(1 - \mathrm{cc}(k), 1 - \mathrm{cc}'(k)\right)} \tag{4.15}$$

where cc and cc′ are a pair of confidence curves. This index was proposed in Zhou *et al.* [18] and resembles the Ružička index [23]. It is one for identical curves and smaller than one for curves that differ.

4.2.3. A CONFIDENCE CURVE BASED NULL HYPOTHESIS TEST

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's test is performed.

4.2.4. 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 [24–27]. The pdfs and the relations between the parameters and moments and L-moments are given in Appendix F. 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 sample 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 sub-series up to τ was $\mu_{\mathrm{L}} = 2$, and the mean for the sub-series 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. 4.1(a-f).

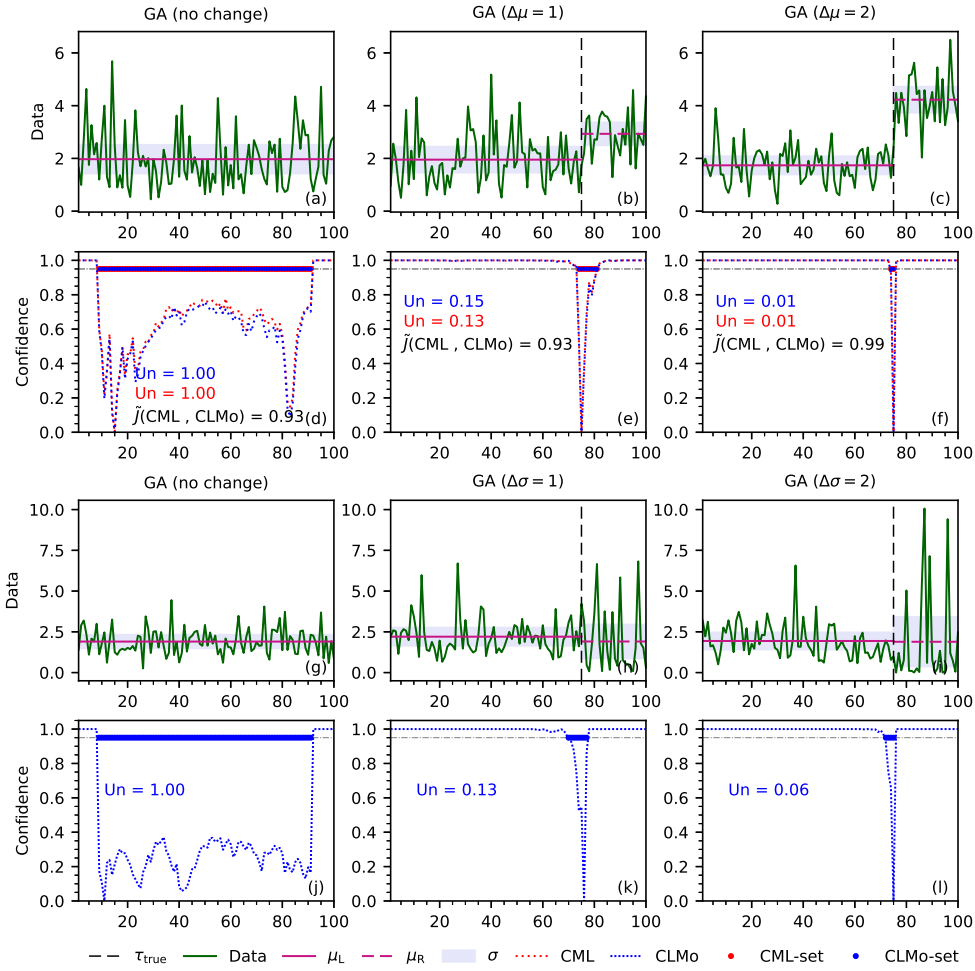


Figure 4.1: Synthetic GA distributed data sets with a change in the mean $\Delta\mu$ or the standard deviation $\Delta\sigma$, and the corresponding confidence curves.

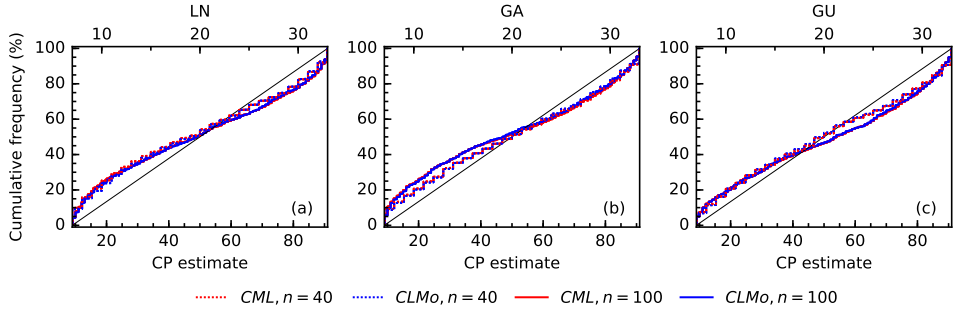


Figure 4.2: The cumulative frequency distribution of CPs for $n = 40, 100$ when H_0 holds .

The standard deviation for the left and right sub-series will be referred to as σ_L and σ_R respectively. Additional experiments were done for CLMo with $\mu_R = \mu_L$, $\sigma_R = \sigma_L + \Delta\sigma$, where $\Delta\sigma = 1, 2, 3$. Examples of data series synthetic data sets with $\Delta\sigma = 0, 1, 2$ are given in Fig. 4.1(g-l). The effect of shifting (different μ_L) or scaling (different σ_L) a time series is discussed in B.3.

4.3. RESULTS FOR SYNTHETIC DATA WITH A CHANGE IN THE MEAN

The performance of the confidence curves produced by CML and CLMo as represented by the properties listed in Section 4.2.2 are examined. For all methods, $N = 1000$ MC simulations were used to generate the approximate confidence curve.

4.3.1. THE CUMULATIVE FREQUENCY DISTRIBUTION OF THE CHANGE POINT ESTIMATE

(1) THE CUMULATIVE FREQUENCY DISTRIBUTION OF THE CHANGE POINT ESTIMATE WHEN THE NULL HYPOTHESIS HOLDS

Figure 4.2 shows the cumulative frequency distribution of the CP estimates found by CML and CLMo when the null hypothesis holds (no CP). In this case, estimating a CP becomes an event of randomly picking a point from all possible candidates. The possible candidates are the elements of the set L_{CP} defined in (3.8). The black lines in Fig. 4.2 show the corresponding uniform frequency distribution. The experimental results do not match this exactly, but do approximate it. For LN, GA, and GU the methods CML and CLMo give similar results.

(2) THE CUMULATIVE FREQUENCY DISTRIBUTION OF THE CHANGE POINT ESTIMATE WHEN THE ALTERNATIVE HYPOTHESIS HOLDS

Figure 4.3 shows the frequency distribution of detected CPs by the two methods when the alternative hypothesis holds for $\Delta\mu = 1, 2$, $\tau_{true} = n/4, n/2, 3n/4$, and $n = 40, 100$. For CML and CLMo, the points where the confidence curve is zero are spread around the true CP. The spread decreases with increasing $\Delta\mu$ and n. For example, for $\Delta\mu = 1$ and $n = 100$,

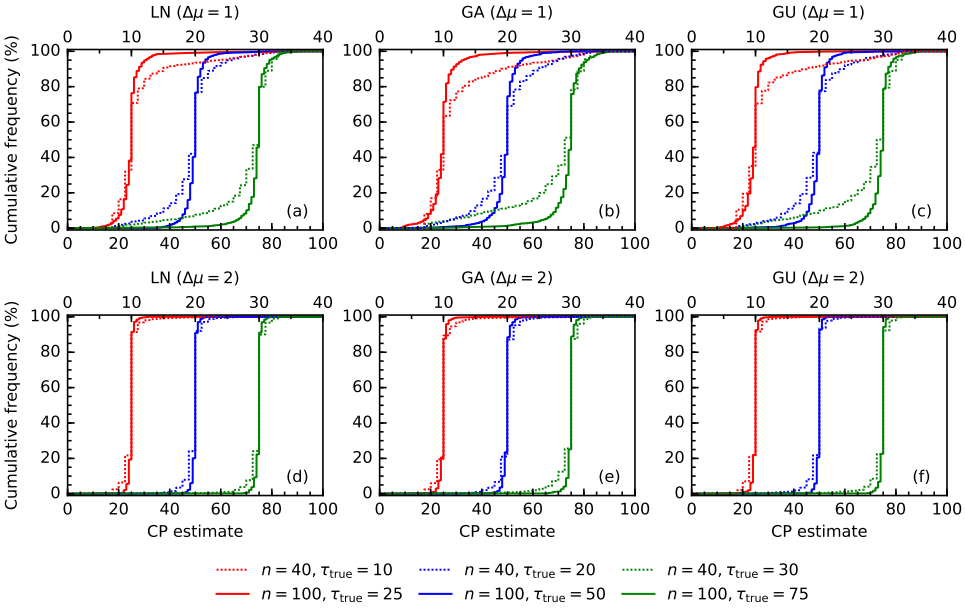


Figure 4.3: The cumulative frequency distribution of CPs for $n = 40; 100$ when there is a change in the mean.

about 90% of the estimates lie within ± 10 points of the actual CP. For $\Delta\mu = 2$, the spread reduces to ± 5 points. For $\Delta\mu = 4$ the plot is not shown, but the spread was reduced to about ± 2 points. The frequency distributions found by all methods for synthetic time series drawn from the three distributions are very similar.

4.3.2. 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* (see also Section 3.1). 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 $m = 1, 2, ..., M$ 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. 4.4, plots of the actual versus nominal coverage are shown.

When interpreting Fig. 4.4, it is important to recall that if a CP is present, then there is only a finite number of possible locations for 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

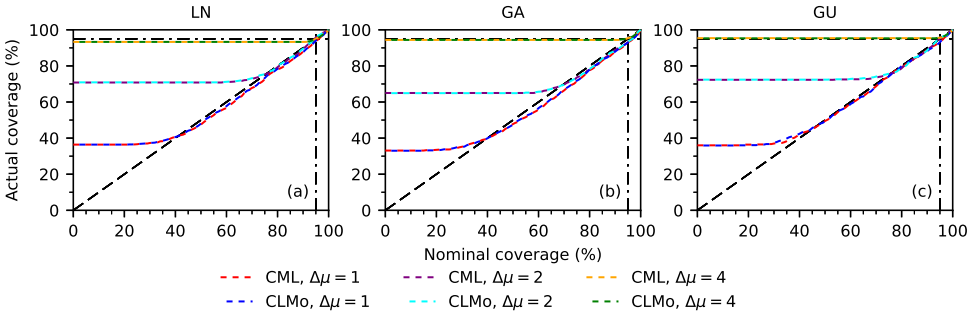


Figure 4.4: Actual versus nominal coverage probability for a change in the mean when $\tau_{\text{true}} = n/2$ and sample length $n = 100$.

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. 4.4 where in (a) the actual coverage is always above 37% for $\Delta\mu = 1$; it always exceeds 70% for $\Delta\mu = 2$ in (b), and in (c) it is higher than 90% for $\Delta\mu = 4$. It follows that in practice, only the confidence sets with relatively high nominal confidence are of interest. The sets at low confidence levels are much too conservative. The results for CML show that, all distributions provide accurate actual coverage for nominal coverage above 90%. At $\Delta\mu = 4$ the method seems almost certain of the CP. The results for CLMo (CMoM) are nearly identical to those for CML. Results for $n = 40$ were generated as well, but the impact from sample length on actual coverage probability was small, so they have not been included here.

Table 4.1 shows details about actual versus nominal coverage for confidence curves constructed by CML and CMoM/CLMo for confidence levels $\gamma = 0.90, 0.95, 0.99$.

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 $R_{\gamma,m}$, $m = 1, 2, ..., M$, that contained the true change point would be distributed according to a binomial distribution

$$\Pr(k) = \left(\begin{array}{c} M \\ k \end{array} \right) \gamma^k (1-\gamma)^{M-k} \tag{4.16}$$

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 4.1, please recall that the location of the CP is a discrete random variable, so for some confidence levels it might not be possible to define a confidence set with that exact coverage.

4.3.3. THE UNCERTAINTY IN THE CONFIDENCE CURVES

The value of Un for a confidence curve is calculated according to Appendix E. It summarizes the uncertainty of a confidence curve.

Table 4.1: The actual coverage of confidence sets by CML and CLMo. Cells with conservative coverage are grey.

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
	4	40	0.954	0.953	0.960	0.962	0.982	0.981
		100	0.933	0.934	0.951	0.951	0.984	0.986
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
	4	40	0.944	0.945	0.960	0.959	0.982	0.984
		100	0.944	0.944	0.954	0.951	0.987	0.983
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
	4	40	0.959	0.958	0.961	0.961	0.979	0.983
		100	0.954	0.953	0.956	0.957	0.985	0.985

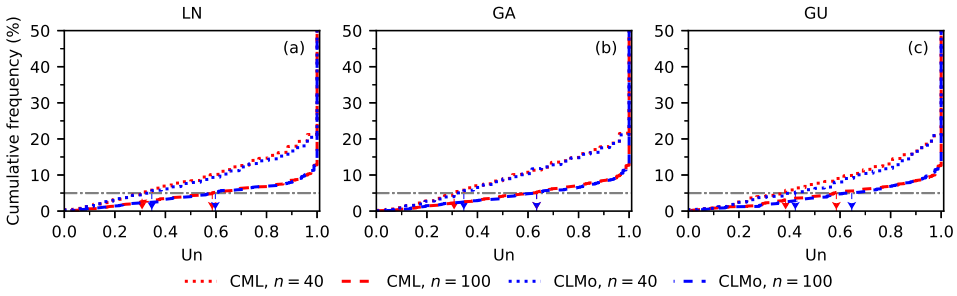


Figure 4.5: Cumulative frequency of Un when H_0 holds.

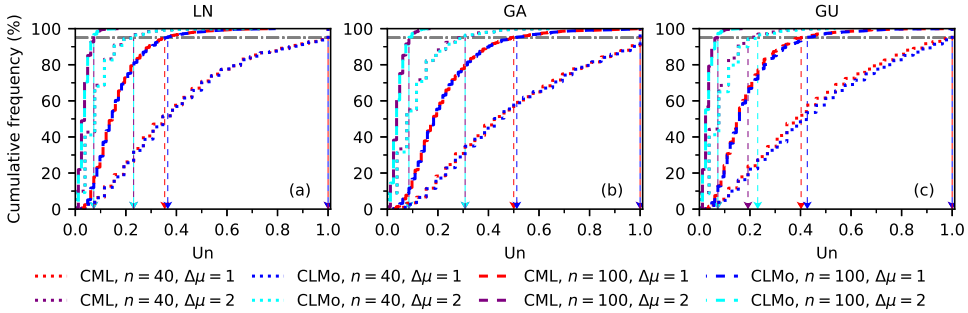


Figure 4.6: Cumulative frequency of Un for a CP in the middle of the series and a change in the mean.

(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. 4.5. Examples of confidence curves for time series without a CP are shown in Fig. 4.1(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. The viability of this approach depends on the distribution of Un in cases where the alternative hypothesis holds.

(2) THE UNCERTAINTY OF CONFIDENCE CURVES FOR THE ALTERNATIVE HYPOTHESIS

Figure 4.6 shows the frequency distribution for Un when H_1 holds and the CP lies in the middle of the time series. The values of Un are very similar for CML and CLMo. If $n = 100$ and $\Delta\mu = 1$, then 95% of the values lie below 0.4 for LN and GU, while for GA the 95% of the Un values lie below 0.5. For $\Delta\mu = 2$ these values are halved, while for $\Delta\mu = 4$ (not shown), the Un is nearly zero. For $n = 40$ the distribution of Un is much more spread out.

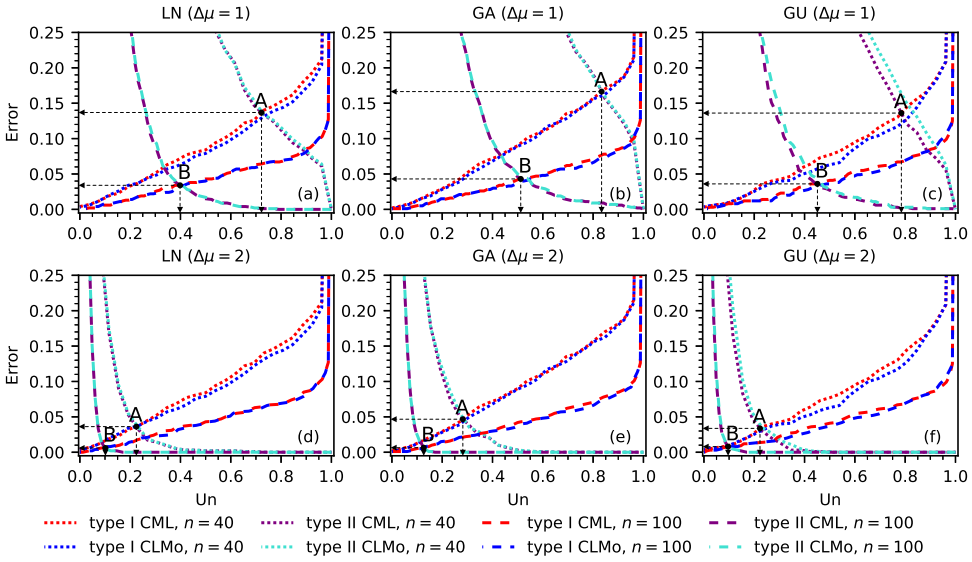


Figure 4.7: Un versus type I and type II errors for a CP in the middle of the series and a change in the mean.

(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. The rejection of H_0 when there is no CP (type I error) and the acceptance of H_0 when there is a CP (type II error). For example, for $\Delta\mu = 1$, $n = 100$, and distribution LN, Fig. 4.5(a) implies that rejection of H_0 for Un ≤ 0.2 would result in a very small type I error, while Fig. 4.6(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 the subset would be less than 5% of the original set of time series, while if all series actually have a CP, then the subset would be about 30% of the original set.

Figure 4.7 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 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 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. To see how a null hypothesis test based on Un would do when compared with the classical Pettitt's test, the curve of error pairs is drawn for both tests in Fig. 4.8. The results show that in principle Un could serve as the basis for a hypothesis test.

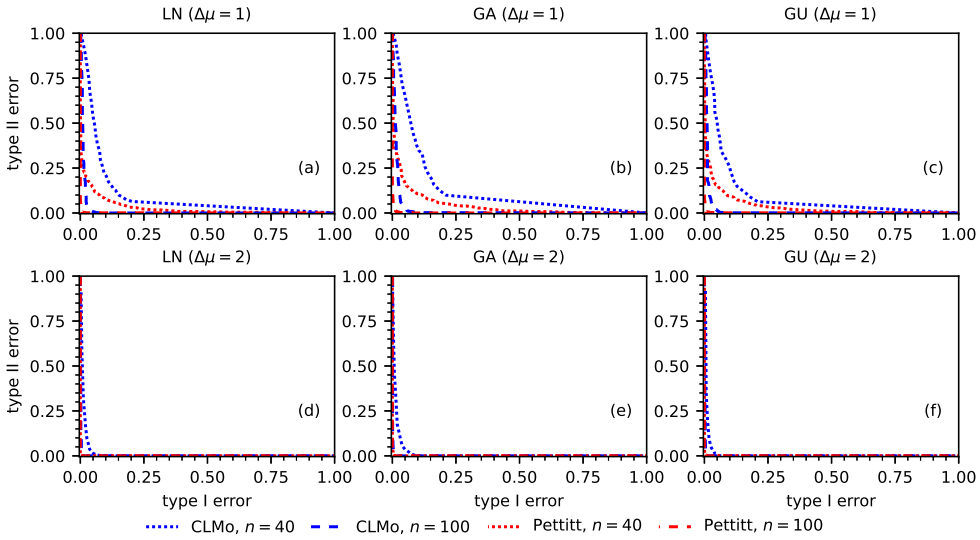


Figure 4.8: Comparison of the two null hypothesis tests where for H_1 the CP is in the middle of the series.

4.3.4. 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 (4.15) for $\Delta\mu = 1, 2, 4$, $\tau_{\text{true}} = n/2$, and $n = 40, 100$. Details on the calculation of $\tilde{J}$ and its properties can be found in Appendix D. Figure 4.9 shows the resulting cumulative frequency distributions of $\tilde{J}$. The confidence curves for synthetic data calculated as by CML are very similar to those calculated by CLMo (CMoM). 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 pairs had a similarity below 0.7, for all sample lengths, distributions, and methods.

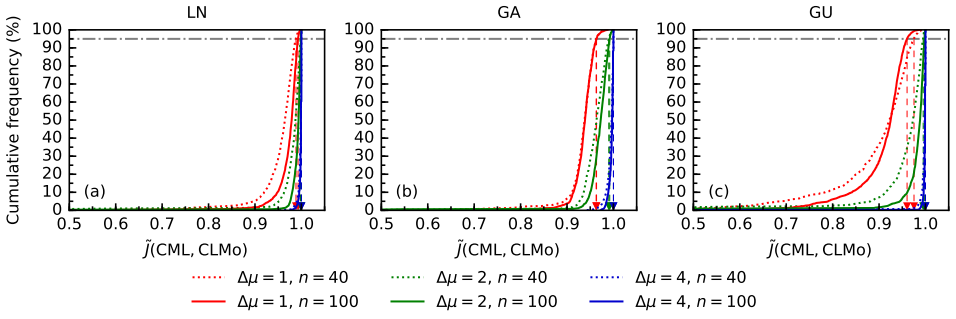


Figure 4.9: The similarity index between confidence curves generated by the CML and CLMo methods.

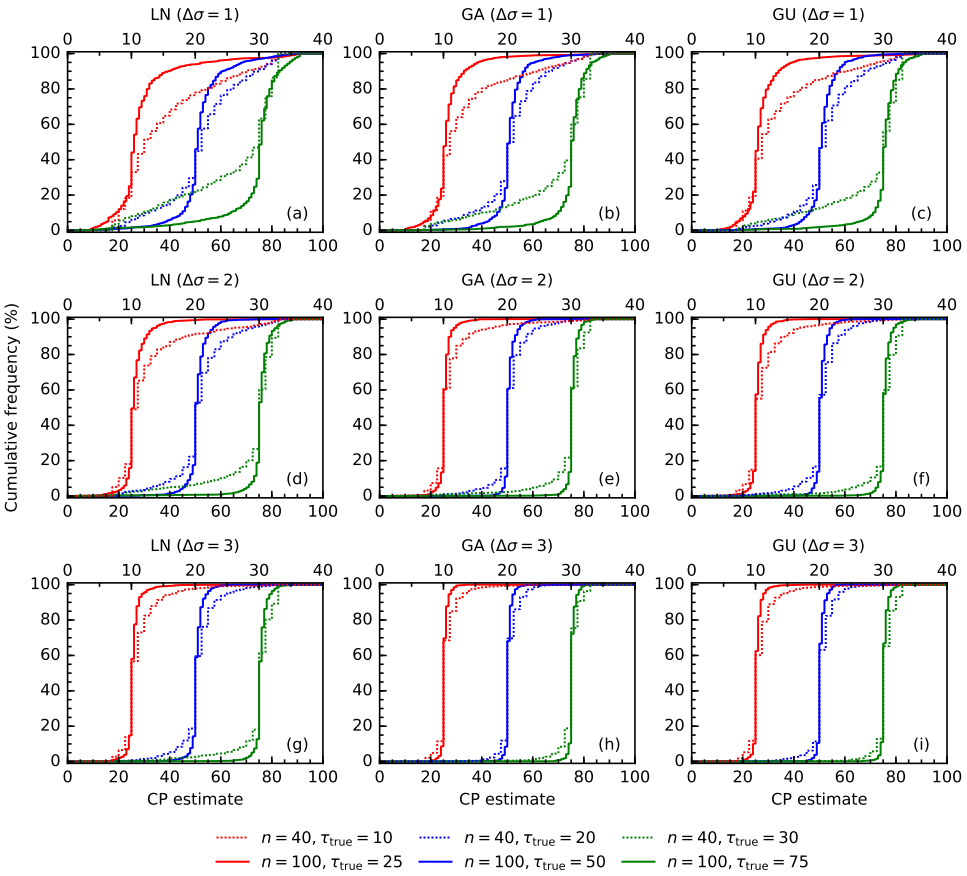


Figure 4.10: The cumulative frequency distribution of CPs for $n = 40, 100$ when there is a change in the standard deviation.

4.4. RESULTS FOR CLMO FOR SYNTHETIC DATA WITH A CHANGE IN THE STANDARD DEVIATION

One advantage of a parametric method is that it looks for changes in all parameters at the same time. To examine this further limited experiments were performed for synthetic series with a change in the standard deviation.

4.4.1. THE CUMULATIVE FREQUENCY DISTRIBUTION OF THE CP ESTIMATE WHEN THE ALTERNATIVE HYPOTHESIS HOLDS

Figure 4.10 shows the frequency distribution of detected CPs by CLMo when the alternative hypothesis holds for $\Delta\sigma = 1, 2, 3$, $\tau_{\text{true}} = n/4, n/2, 3n/4$, and $n = 40, 100$. The spread decreases with increasing size of the change.

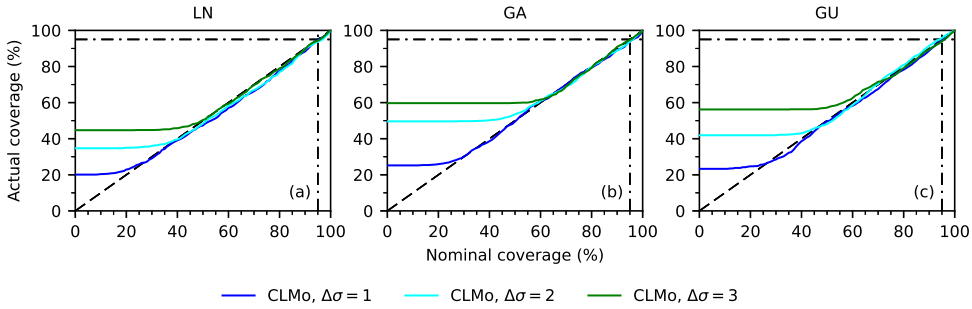


Figure 4.11: Actual versus nominal coverage probability for a change in the standard deviation when $\tau_{\text{true}} = n/2$ and sample length $n = 100$.

4.4.2. ACTUAL VERSUS NOMINAL COVERAGE PROBABILITY

When interpreting Fig. 4.11, it is important to recall the earlier remark that, if a CP is present, then there is only a finite number of possible locations for that CP. This implies that for low confidence levels the sets will be very conservative. This manifests itself in Fig. 4.11(a) where for LN, the actual coverage is always above 20% for $\Delta\sigma = 1$; it always exceeds 35% for $\Delta\sigma = 2$, and it is higher than 45% for $\Delta\sigma = 3$. For GA and GU the lower bounds on the actual coverage are even higher. This in turn means that in practice, only the confidence sets with relatively high nominal coverage, for example above 60%, are of interest. The sets at low confidence levels are much too conservative.

4.4.3. UNCERTAINTY AS A BASIS FOR A NULL HYPOTHESIS TEST

Figure 4.12 shows 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. For $\Delta\sigma = 3$ these values are almost halved for LN, GA, and GU. 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. If the standard Pettitt's test is used for a change in the standard deviation, then it is much less effective than for a change in the mean (Fig. 4.14). A different test, specifically designed for the change to be detected, would be needed. Here the test based on Un is not the most effective for a particular change, but it is the one that will detect all parameter changes.

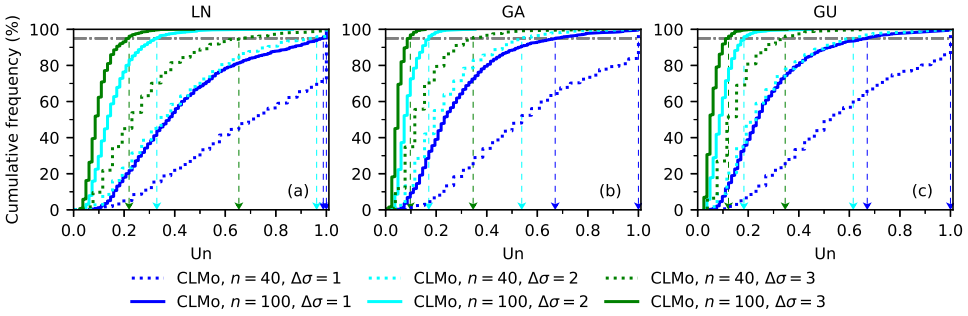


Figure 4.12: The cumulative frequency of Un for a CP in the middle of the series and a change in the standard deviation.

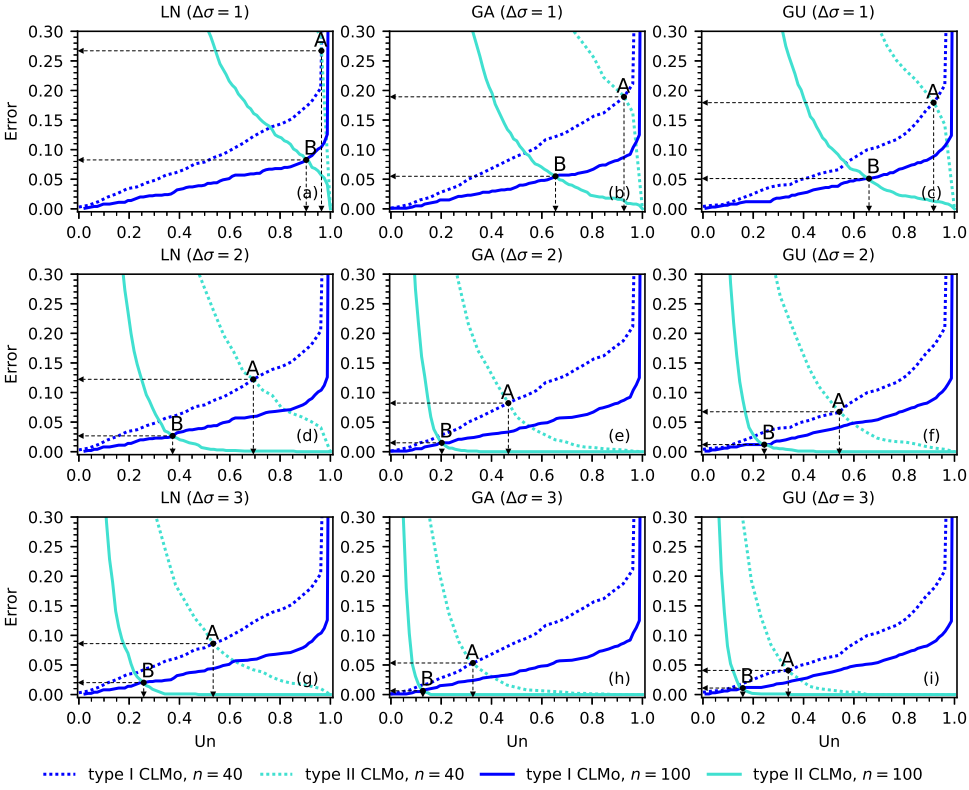


Figure 4.13: Un versus type I and type II errors for a CP in the middle of the series and a change in the standard deviation.

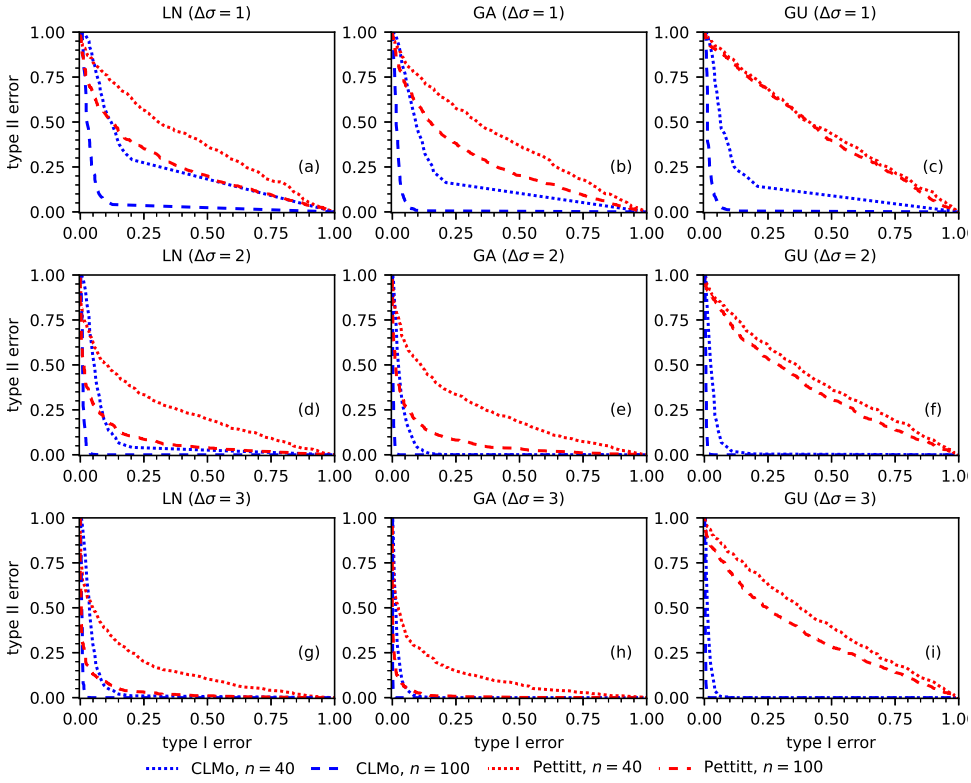


Figure 4.14: Comparison of the two null hypothesis tests for a change in the standard deviation when the CP is in the middle of the series.

4.5. CHANGE POINT DETECTION AND UNCERTAINTY IN REAL HYDROMETEOROLOGICAL DATA

To examine the performance of the CML and CLMo methods on real world data, seven time series of measurements were taken from previous publications by Conte *et al.* [8] - case study 1, Zhou *et al.* [28] - case study 2, Jandhyala *et al.* [29] - case study 3, Reeves *et al.* [30] - case study 4. The CPs found in the original studies are used as a reference, see Table 4.2. Both methods were used to construct confidence curves for CPs with each

Table 4.2: Change points found in the hydrometeorological series and statistical properties of the series.

Time series	from	to	τ_{ref}	σ	$\frac{\Delta\mu}{\sigma}$
Tucumán	1884	1996	1956	18mm	0.76
Tuscaloosa	1940	1986	1957	0.61°C	−1.3
Itaipu	1931	2015	1971	$2.5 \times 10^3 \text{m}^3/\text{s}$	1.3
Cuntan	1893	2014	not found	$12 \times 10^3 \text{m}^3/\text{s}$	−0.45
Yichang	1946	2014	1962, 1966	$8.6 \times 10^3 \text{m}^3/\text{s}$	−1.3
Hankou	1952	2014	not found	$8.9 \times 10^3 \text{m}^3/\text{s}$	−0.89
Datong	1950	2014	not found	$11 \times 10^3 \text{m}^3/\text{s}$	−0.76

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.

The details of the time series used for analysis are:

- Conte *et al.* [8] used the bootstrap Pettitt's 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.* [28]. 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. In Zhou *et al.* [28] of the four stations only Yichang station yielded change points. The paper applied three methods to this series: the Pettitt's 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 $\Delta\mu/\sigma = -0.84$.
- 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.* [29], a change point in the time series was found near 1956 by a Bayesian method. Wu *et al.* [31] 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 resulting from equipment changes or station relocation, namely in November 1957 [30]. All eight methods used in that study found a change point in the year of 1957.

4.5.1. CASE STUDY 1

Conte *et al.* [8] found a significant CP in 1971 in the annual average naturalized discharge of the Itaipu Hydroelectric Plant in Brazil from 1931 to 2015 by the bootstrap Pettitt's test. In this case the value of $|\Delta\mu|/\sigma$ suggests the methods should do reasonably well. And so they do: Un is low and $\tilde{J}$ is high. All give a 95% confidence interval of about three years (Fig. 4.15).

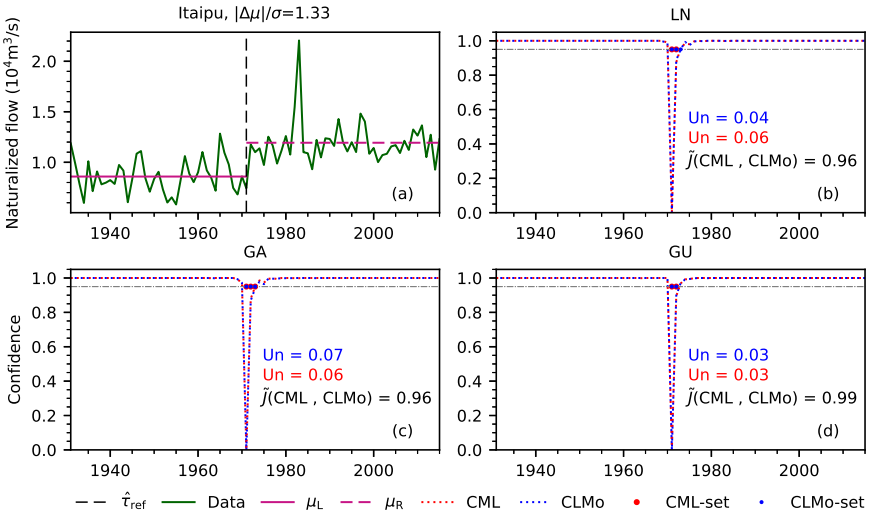


Figure 4.15: Confidence curves for CP in annual average naturalized discharge time series of Itaipu.

4.5.2. CASE STUDY 2

Four time series of annual maximum discharge on the Yangtze River in China were analysed in Zhou *et al.* [28]. 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.

(1) CUNTAN

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. 4.16(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.

(2) YICHANG

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 [9]. In Zhou *et al.* [28] CUSUM found a CP in 1962, while Pettitt's and Cramér-von Mises found a CP in 1966. All confidence curves in Fig. 4.16(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 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.

(3) HANKOU

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. fig:Confidence-curves-for Hankou(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.

(4) DATONG

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. 4.17(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.

4.5.3. CASE STUDY 3

In Jandhyala *et al.* [29] the annual average rainfall time series from 1884 to 1996 at Tucumán] in Argentina was investigated, and a CP was found in 1956 by a Bayesian method. The result was confirmed by Wu *et al.* [31]; they believed the change was caused by the construction of a dam in Tucumán] from 1952 to 1962. Figure 4.18(a) shows the data and the CP in 1956. In Fig. 4.18(b-d) the results of CML and CMoM/CLMo with different distributions are shown.

The large Un value make it difficult to decide whether or not there is a CP. The confidence curves suggest that there could well be a CP near 1956, but there is considerable uncertainty about its precise location. Given the results on synthetic data series for relatively small values of $|\Delta\mu|/\sigma$, this is not surprising.

4.5.4. CASE STUDY 4

The annual average temperature time series from 1940 to 1986 in Tuscaloosa of USA was selected because there was only one documented reason for a CP during this period. The time series was used in Reeves *et al.* [30], and a CP located at the year of 1957 was found by eight different methods. Here the value of $|\Delta\mu|/\sigma$ offers more hope of finding a CP. Both LN and GA based methods find a reasonably precise confidence curve for the CP

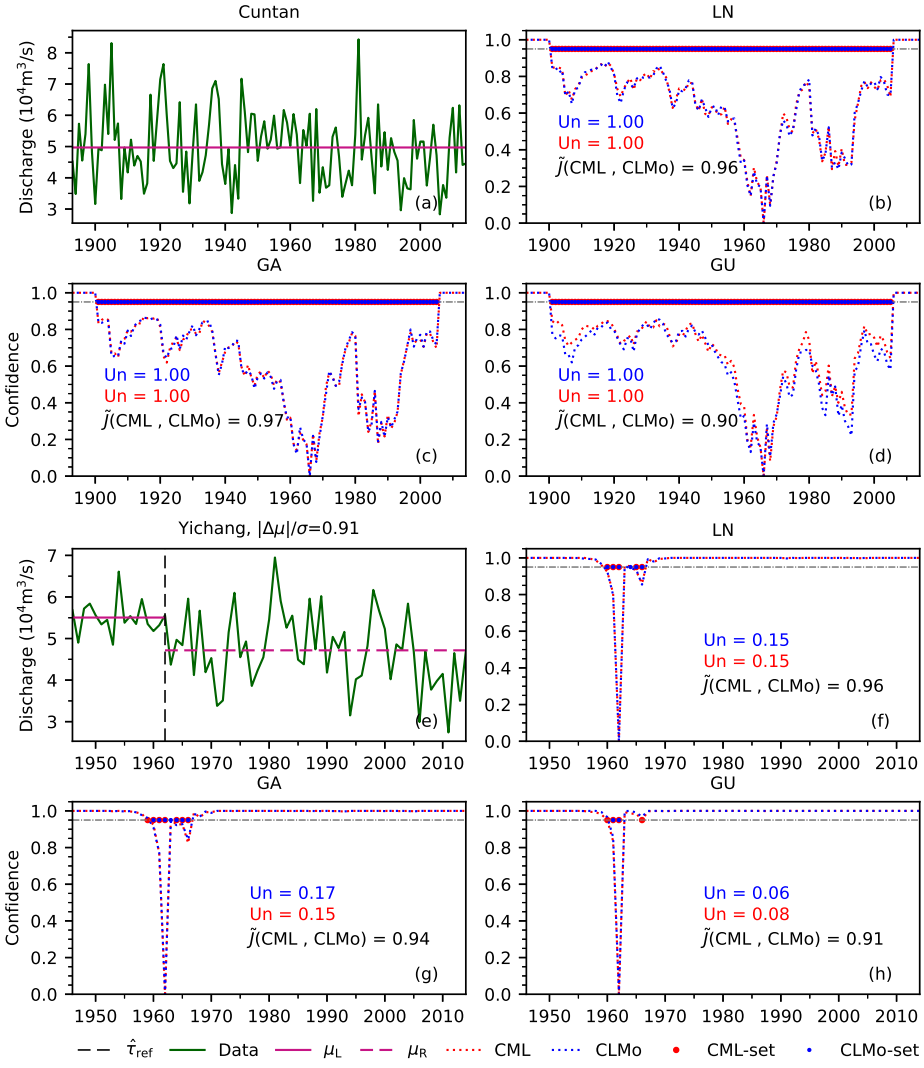


Figure 4.16: Confidence curves for CP in annual maximum discharge time series of Cuntan and Yichang.

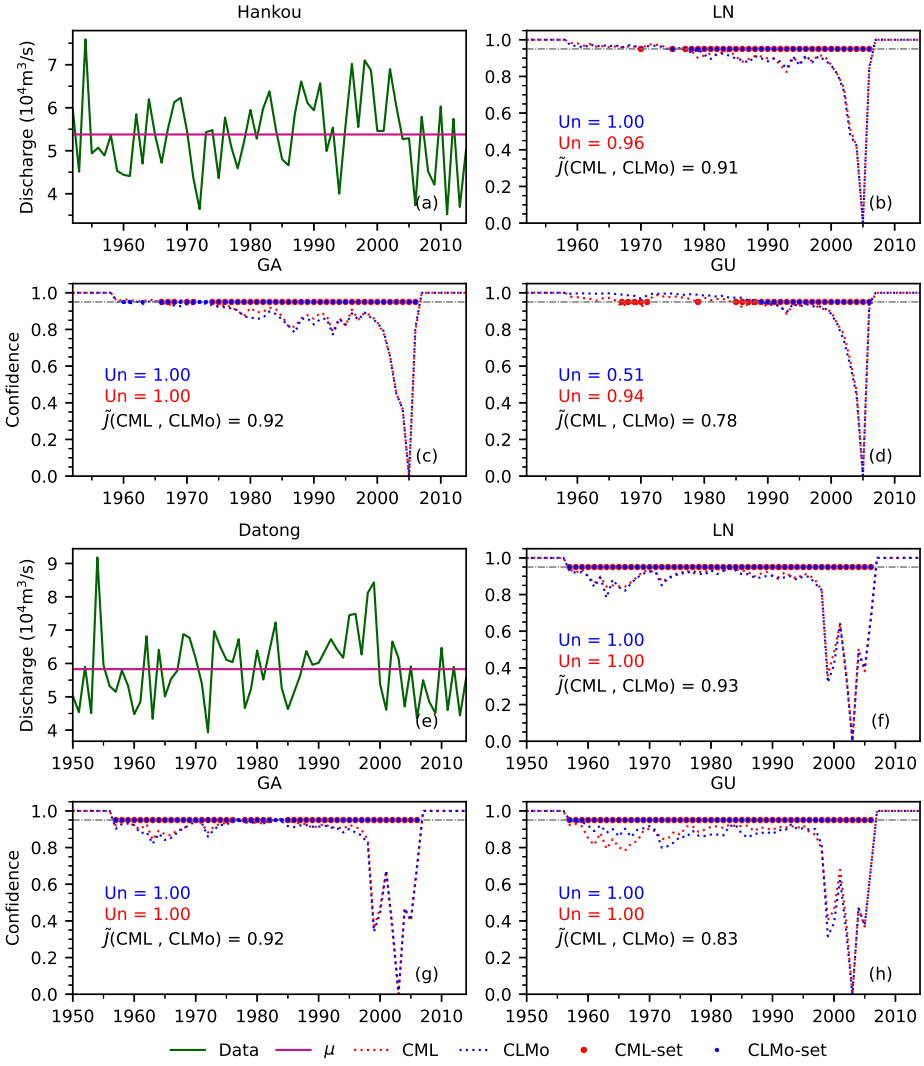


Figure 4.17: Confidence curves for CP in annual maximum discharge time series of Hankou and Datong.

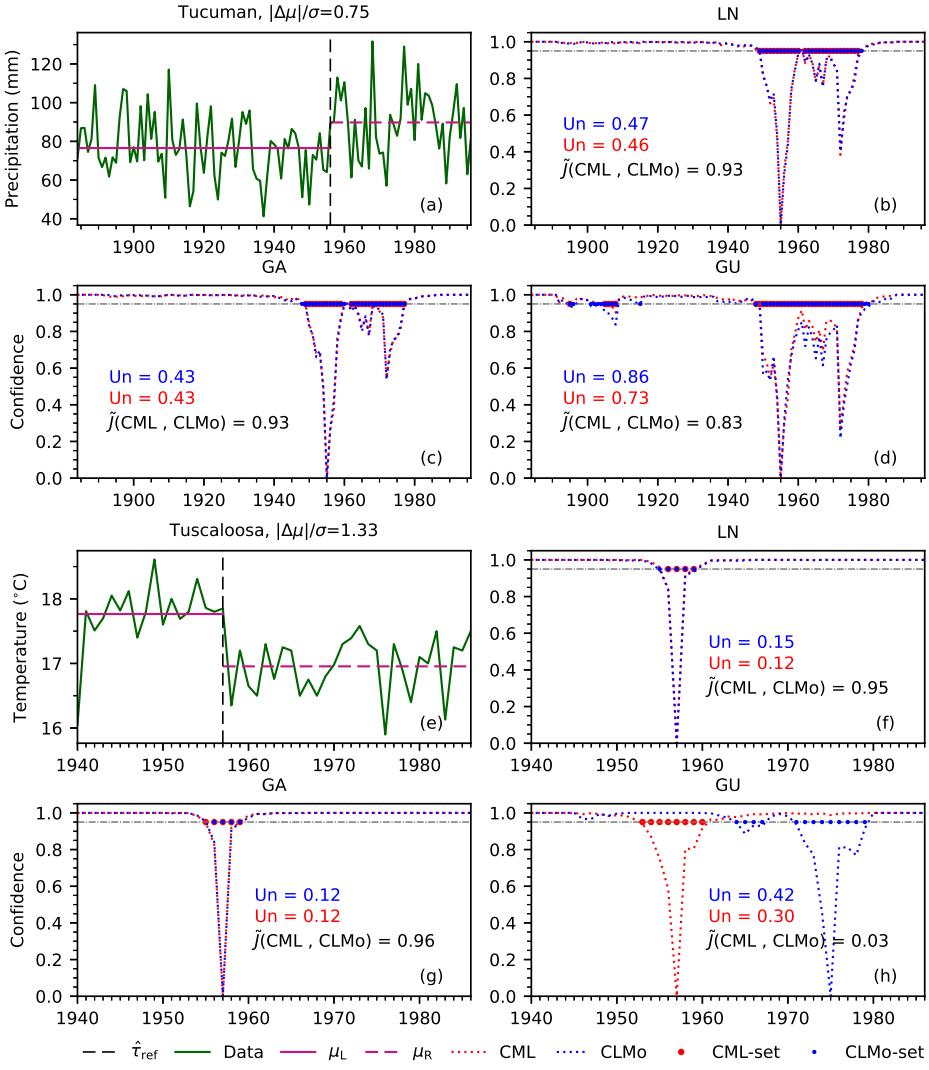


Figure 4.18: Confidence curves for CP in time series of Tucumán and Tuscaloosa.

Table 4.3: Different Gumbel based parameter estimates for left and right subseries at Tuscaloosa

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

with a 95% confidence interval of about 5 years. 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 in 1976 (Fig. 4.18(f-h)).

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 random variables 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 4.3 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.

4.6. CONCLUSION AND DISCUSSION

This study examined three parametric methods to construct confidence curves for change points (CPs) in time series. The methods are able to detect changes in the mean and the standard deviation. One method (CML) was based on Cunen *et al.* [19] and two faster variations on that method (CLMo, CMoM), which were proposed in the current study. All methods 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. The CML method deals with the nuisance parameters by using a profile likelihood; the CMoM and CLMo methods use a pseudo likelihood 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, see Appendix G.

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 CP detection equally well. Performance in terms of actual coverage of the associated confidence sets for high confidence levels was satisfactory and nearly independent of sample 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.

The uncertainty in the information produced about the CP decreased with increasing sample length and/or with increasing size of the change at the CP. In this chapter Un, a summary of the uncertainty shown by a curve, was defined to serve as the basis for a test for the null hypothesis of the absence of a CP. Preliminary findings suggest that a test based on this measure may perform on a par with the classical Pettitt's 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; changes in the mean are detected if they exceed one standard deviation. For changes in the standard deviation more experiments are needed to see whether absolute or relative size of the change determines the method sensitivity.

The two new methods, CLMo and CMoM, introduced in this article complement the AED method from Zhou *et al.* [18]. The AED 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, apply CLMo when the results are not conclusive or a change in the mean is not expected, and finally use CML when the CLMo result still displays large uncertainty.

REFERENCES

[1] E. Kolokytha, S. Oishi, and R. S. Teegavarapu, *Sustainable water resources planning and management under climate change* (Springer Singapore, 2017).

[2] R. Teegavarapu, ed., *Trends and changes in hydroclimatic variables* (Elsevier, 2018)

pp. 275–304.

[3] H. Tao, M. Gemmer, Y. Bai, B. Su, and W. Mao, *Trends of streamflow in the Tarim River Basin during the past 50 years: Human impact or climate change?* Journal of Hydrology **400**, 1 (2011).

[4] S. Harrigan, C. Murphy, J. Hall, R. Wilby, and J. Sweeney, *Attribution of detected changes in streamflow using multiple working hypotheses,* Hydrology and Earth System Sciences **18**, 1935 (2014).

[5] Z. Cong, M. Shahid, D. Zhang, H. Lei, and D. Yang, *Attribution of runoff change in the alpine basin: A case study of the Heihe Upstream Basin, China,* Hydrological Sciences Journal **62**, 1013 (2017).

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

[7] Z. W. Kundzewicz and A. J. Robson, *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**, 7 (2004).

[8] L. C. Conte, D. M. Bayer, and F. M. Bayer, *Bootstrap Pettitt test for detecting change points in hydroclimatological data: Case study of Itaipu Hydroelectric Plant, Brazil,* Hydrological Sciences Journal **64**, 1312 (2019).

[9] H. Xie, D. Li, and L. Xiong, *Exploring the ability of the Pettitt method for detecting change point by Monte Carlo simulation,* Stochastic Environmental Research and Risk Assessment **28**, 1643 (2014).

[10] L. Xiong and S. Guo, *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,* Hydrological Sciences Journal **49**, 99 (2004).

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

[12] J. Chen and A. K. Gupta, *On change point detection and estimation,* Communications in Statistics-Simulation and Computation **30**, 665 (2001).

[13] J. Chen and A. K. Gupta, *Parametric statistical change point analysis: With applications to genetics, medicine, and finance* (Springer Science & Business Media, 2011).

[14] E. Brodsky and B. S. Darkhovsky, *Nonparametric methods in change point problems,* Vol. 243 (Springer Science & Business Media, 2013).

[15] R. Nuzzo, *Scientific method: Statistical errors,* Nature News **506**, 150 (2014).

[16] V. R. Eastwood, *Some nonparametric methods for changepoint problems,* Canadian Journal of Statistics **21**, 209 (1993).

[17] D. J. Sheskin, *Handbook of parametric and nonparametric statistical procedures*, pp. 97-98 (Chapman and Hall/CRC, 2003).

[18] C. Zhou, R. van Nooijen, A. Kolechkina, and N. van de Giesen, *Confidence curves for change points in hydrometeorological time series,* Journal of Hydrology **590**, 125503 (2020).

[19] C. Cunen, G. Hermansen, and N. L. Hjort, *Confidence distributions for change-points and regime shifts,* Journal of Statistical Planning and Inference **195**, 14 (2018).

[20] G. Gong and F. J. Samaniego, *Pseudo maximum likelihood estimation: Theory and applications,* The Annals of Statistics , 861 (1981).

[21] D. P. Lettenmaier and S. J. Burges, *Gumbel's extreme value I distribution: A new look,* Journal of Hydraulic Engineering **108**, 502 (1982).

[22] P. Delicado and M. N. Goria, *A small sample comparison of maximum likelihood, moments and L-moments methods for the asymmetric exponential power distribution,* Comput. Statist. Data Anal. **52**, 1661 (2008).

[23] A. Schubert and A. Telcs, *A note on the Jaccardized Czekanowski similarity index,* Scientometrics **98**, 1397 (2014).

[24] K. Hamed and A. R. Rao, *Flood frequency analysis* (CRC press, 2019).

[25] S. A. Thompson, *Hydrology for water management* (CRC Press, 2017).

[26] T. Haktanir, *Statistical modelling of annual maximum flows in Turkish rivers,* Hydrological Sciences Journal **36**, 367 (1991).

[27] M. A. Karim and J. U. Chowdhury, *A comparison of four distributions used in flood frequency analysis in Bangladesh,* Hydrological Sciences Journal **40**, 55 (1995).

[28] C. Zhou, R. van Nooijen, A. Kolechkina, and M. Hrachowitz, *Comparative analysis of nonparametric change-point detectors commonly used in hydrology,* Hydrological Sciences Journal **64**, 1690 (2019).

[29] V. K. Jandhyala, S. B. Fotopoulos, and J. You, *Change-point analysis of mean annual rainfall data from Tucumán, Argentina,* Environmetrics **21**, 687 (2010).

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

[31] W. B. Wu, M. Woodroofe, and G. Mentz, *Isotonic regression: Another look at the changepoint problem,* Biometrika **88**, 793 (2001).

5

CONFIDENCE CURVES BASED ON EMPIRICAL METHOD

Thanks to Prof. Art B. Owen(Stanford University).
Now we can use an empirical and distribution-free
likelihood ratio to estimate our parameters of interest

Parts of this chapter have been published in"Zhou, C., van Nooijen, R., Kolechkina, A. and van de Giesen, N., Confidence curves for change points in hydrometeorological time series, Journal of Hydrology, Page: 1-19, 590, 2020."

5.1. INTRODUCTION

While it is clear that climate change affects the hydrological cycle [1] and that there is an increased risk of extremes in precipitation [2], discharge [3] and temperature, the effects on a regional scale may vary considerably [4]. 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 [5]. 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 [6]. The need for more hydrological data that is mentioned in McMillan *et al.* [7] 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 focuses 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 (CP) when the statistical characteristics of the series before and after the CP show a significant difference. Finding CPs has attracted attention from many fields, for instance, in oceanography [8], economics, finance [9], biology [10], and meteorology [11, 12]. In hydrology, CP detection plays an indispensable role in homogeneity tests for hydrological observations [13].

A number of methods have been developed to find change points; some require a parametric description of the probability distribution of the data points in the time series [9], others do not [14–17]. Traditional CP 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 [18].

In some cases, CP 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 CP is detected in the 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 [19], see Chapter 2 for more detailed information. 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 CP 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.* [20]. Their 'method B', which constructs a confidence curve for the location of a CP 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 (see CML in Chapter 3 and 4). 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 samples from the original sample. The new method is called Confidence curve based on Approximate Empirical likelihood ratio, Deviance function and bootstrapping (AED)

There are alternative approaches that can be used to represent the uncertainty in the CP 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 [21, 22] because of the non-repeatability of hydrological observations. An example of a Bayesian CP analysis method can be found in Perreault *et al.* [23]. 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.* [20] 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.

5.2. METHODOLOGY OF THE EMPIRICAL LOG-LIKELIHOOD RATIO METHOD

A change point detection problem has been introduced in Chapter 3, and in that chapter CML method proposed by Cunen *et al.* [20] is also introduced in details. To show the relations between CML and the method proposed in this study, it is necessary to make a few intermediate steps. In Appendix C, the detailed intermediate steps from confidence curves based on parametric likelihood to confidence curves based on approximate em-

pirical likelihood are shown.

The methodology of 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 (5.1).

$$\ell_{\text{apn}}(\tau; y) = \frac{\tau(n-\tau)}{n} \frac{\left(\frac{1}{\tau}\sum_{i=1}^{\tau} y_i - \frac{1}{n-\tau}\sum_{i=\tau+1}^{n} y_i\right)^2}{\frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \frac{1}{n}\sum_{j=1}^{n} y_j\right)^2} \tag{5.1}$$

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

$$D_{\text{apn}}(\tau; y) = 2\left(\ell_{\text{apn}}(\hat{\tau}_{\text{apn}}(y); y) - \ell_{\text{apn}}(\tau; y)\right) \tag{5.2}$$

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

$$K_{\text{apn},\tau}(r) = \Pr\left(D_{\text{apn}}(\tau; Y) < r\right) \tag{5.3}$$

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 $\tau \in \{n_{\min}, n_{\min}+1, \ldots, n-n_{\min}\}$, use bootstrapping to resample y_{obs} and get N new samples $y_{\text{res}}^{(j)}$ $(j = 1, 2, ..., N)$. 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} [24].
3. Approximate the curve $\text{cc}(\tau; y_{\text{obs}}) = K_{\text{apn},\tau}\left(D_{\text{apn}}(\tau, y_{\text{obs}})\right)$ by

$$\frac{1}{N}\sum_{j=1}^{N}\left[\!\left[D_{\text{apn}}\left(\tau, y_{\text{res}}^{(j)}\right) < D_{\text{apn}}(\tau, y_{\text{obs}})\right]\!\right] \tag{5.4}$$

Here $n_{\min}$ which has been used in (3.9), is used to avoid calculation of approximate empirical likelihoods based on a handful of points.

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

5.2.1. 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 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 $\tau = 25$ 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 point can be represented by $\Delta\mu$.

5.2.2. AN EXAMPLE OF CONFIDENCE CURVES FOR THE LOCATION OF CHANGE POINTS

Figure 5.1 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 are shown in Fig. 5.1(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. 5.1 shows information for just four data sets, so it cannot be used to draw conclusions about the relative performances of the methods. Figure 5.1(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. 5.1(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 CP in less than half of all experiments.

5.3. COMPARISON OF CONFIDENCE CURVES BY PARAMETRIC AND EMPIRICAL METHODS FOR SYNTHETIC DATA

In order to evaluate the performance of CML and AED, 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 CML will be considered. One using the log-normal pdf (LN-CML), one using the gamma pdf (GA-CML), and one using the Fréchet pdf (F-CML). The parametric distribution functions of the three distributions can be found in F. Notice that in this study, the shape parameter of the Fréchet distribution is fixed to avoid over-fitting to hydrometeorological data.

The properties of confidence curves and similarity for confidence curves from the two different methods are analyzed according to Chapter 4.3. The results for confidence curves by the two methods are shown as follows.

5.3.1. ACTUAL VERSUS NOMINAL COVERAGE PROBABILITY OF CONFIDENCE CURVES

For synthetic data, the actual coverage probability will be examined as well as the distribution of the estimate of the CP both when the null hypothesis H_0 holds and when it does not hold. The uncertainty measure Un (see E) of the confidence curves for the null and the alternative hypothesis will be examined respectively as well. These properties will be used to determine the relative merits of the methods. Finally, the similarity of the curves generated by CML and AED will be examined.

Figure 5.2 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

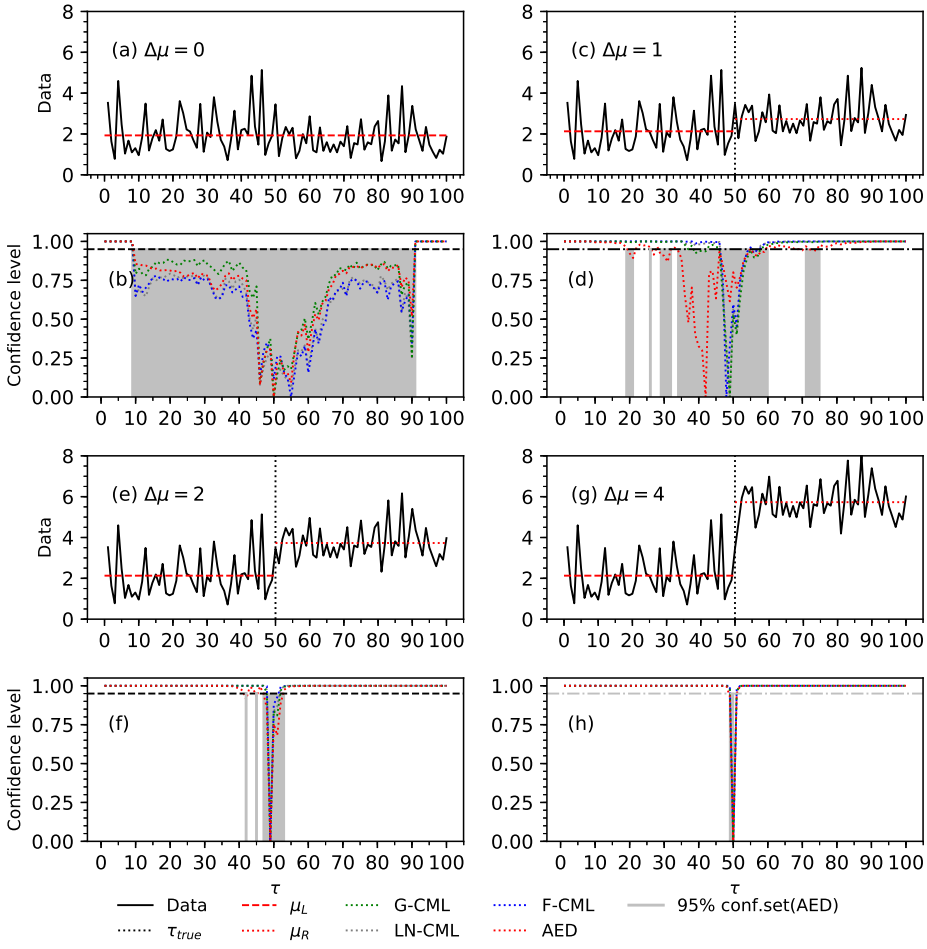


Figure 5.1: Confidence curves for CP location in synthetic data from a log-normal distribution.

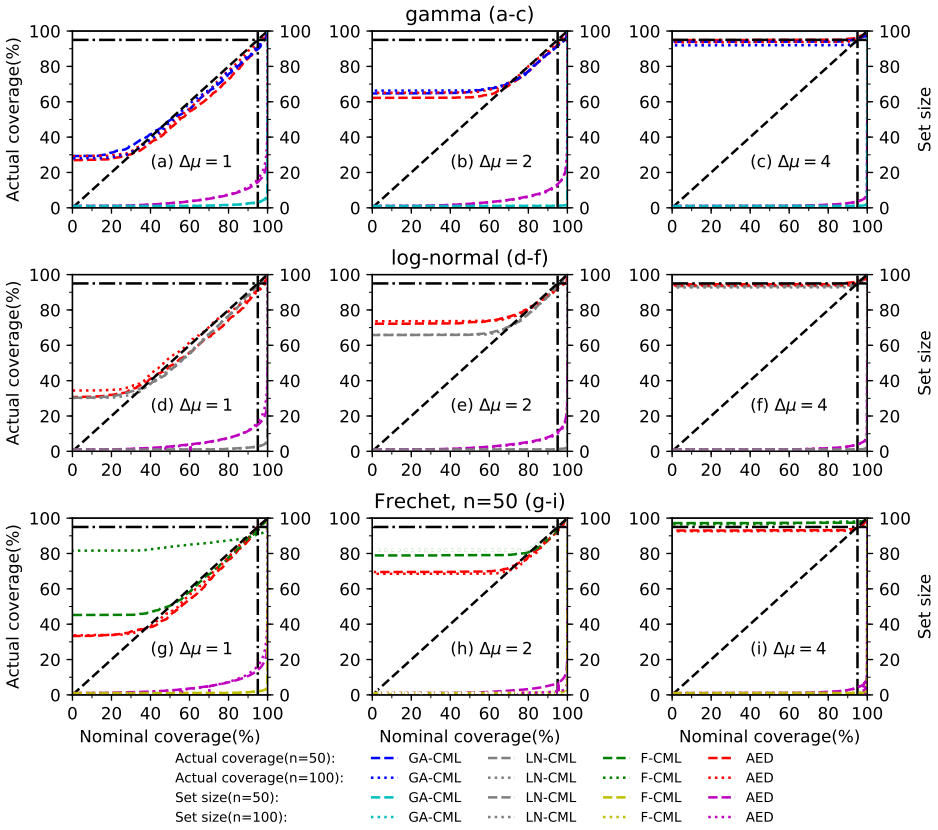


Figure 5.2: Actual coverage probabilities and confidence set size as a function of nominal coverage probabilities for synthetic data.

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 CML and AED deliver reasonable actual coverage probabilities.

For the nominal coverage probabilities $\gamma = 0.90, 0.95, 0.99$, estimates of the actual coverage probability of the confidence curves constructed by the CML variants and AED are listed in Table 5.1. For $\Delta\mu = 1$, the actual coverage probabilities are somewhat lower than the nominal coverage probabilities, so the sets are permissive. For $\Delta\mu = 4$, the actual coverage for $\gamma = 0.90$ and $\gamma = 0.95$ is often higher than the nominal coverage, the corresponding sets are conservative (see Chapter 3). Actual coverage tends to be closer to the nominal value for longer time series.

When combining this information with Table 5.1, please keep in mind that the location of the CP is a discrete random variable, so for some confidence levels it might not be possible to define a confidence set with that exact coverage.

Table 5.1: Actual coverage probability for given confidence levels (conservative coverage is marked by a grey background).

Confidence level			0.9		0.95		0.99	
Distribution	$\Delta\mu$	n	CML	AED	CML	AED	CML	AED
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
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
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

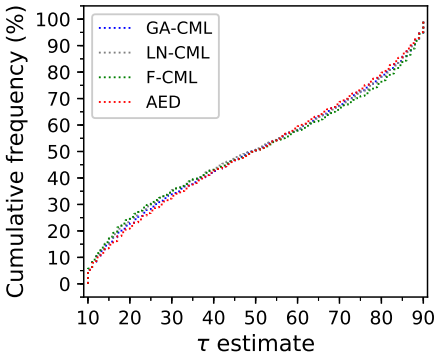


Figure 5.3: Frequency distribution of the CP estimate for the different methods applied to log-normal samples when H_0 holds.

5.3.2 THE FREQUENCY DISTRIBUTION OF THE ESTIMATED CHANGE POINTS WHEN THE NULL HYPOTHESIS HOLDS

Figure 5.3 shows the frequency distribution of the CP 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 CML and AED, 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 CML method has little or no effect on the outcome for the distributions considered here.

5.3.2. THE FREQUENCY DISTRIBUTION OF THE ESTIMATED CHANGE POINTS WHEN THE ALTERNATIVE HYPOTHESIS HOLDS

Figures 5.4 and 5.5 show the frequency distribution of the CP 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 are very close to $\hat{\tau}(y)$ for the CML variant based on the distribution that matches the data source.

When we compare Figures 5.4 and 5.5, it is clear that LN-CML, GA-CML, and AED perform very well for $\Delta\mu = 1$ and $\Delta\mu = 4$ on all data, while F-CML struggles with data from the log-normal and the gamma distributions 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 distributions, and one distribution with a parameter dependent support, the Fréchet distribution.

5.3.3. SIMILARITY OF CONFIDENCE CURVES

The similarity of CML and AED confidence curves was measured by the similarity index $\tilde{J}$ (see D). Figure 5.6 shows examples of the resulting frequency distribution. The sample length has very limited influence on the similarity index, but the size of the change

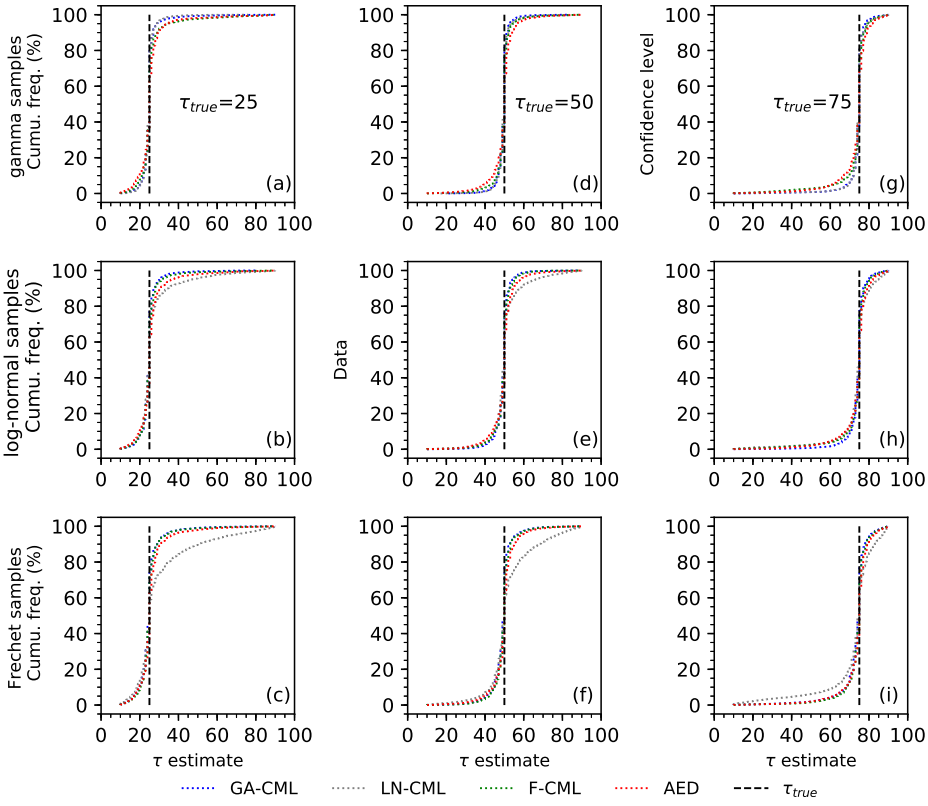


Figure 5.4: Frequency distribution of the CP estimate for the different methods under the alternative hypothesis H_1 with $\tau = 25, 50, 75$ and with magnitudes of change $\Delta\mu = 1$.

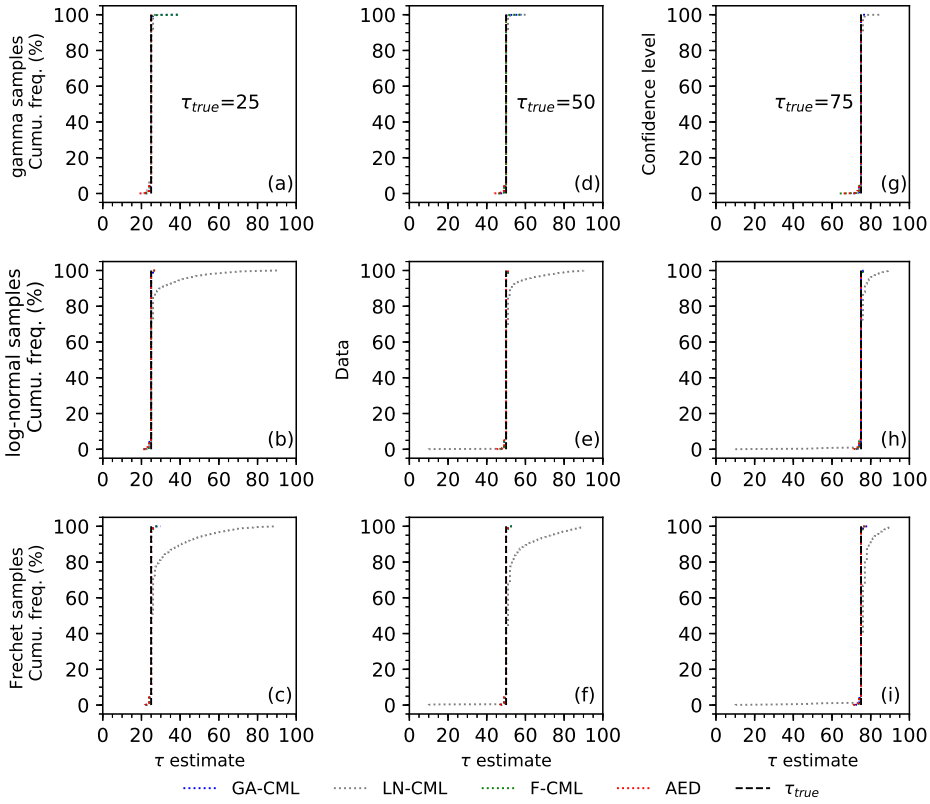


Figure 5.5: Frequency distribution of the CP estimate for the different methods under the alternative hypothesis H_1 with $\tau = 25, 50, 75$ and with magnitudes of change $\Delta\mu = 4$.

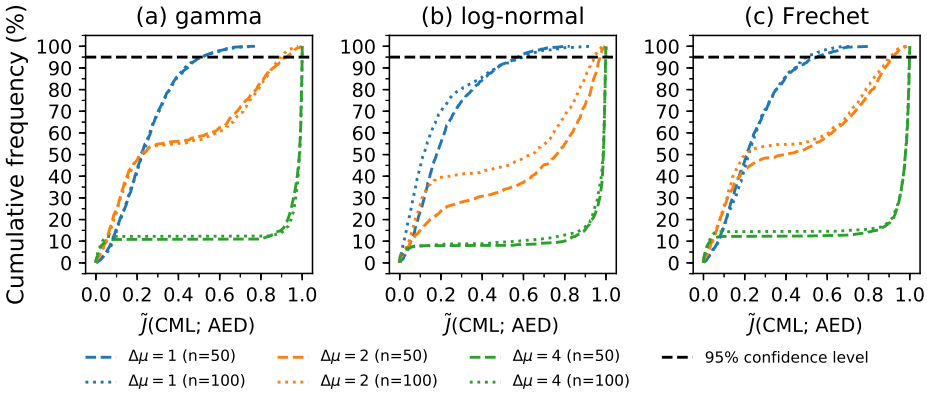


Figure 5.6: Similarity index $\tilde{J}$ between confidence curves constructed by CML and AED.

5

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. 5.6, 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.

5.3.4. THE UNCERTAINTY IN THE CONFIDENCE CURVES

The uncertainty measure Un for a confidence curve is calculated according to Appendix E. It reflects the relative uncertainty of a confidence curve.

(1) THE UNCERTAINTY OF CONFIDENCE CURVES FOR THE NULL HYPOTHESIS

The method based on approximate empirical ratio assumes that there is a CP. This it has in common with the CML and CMoM/CLMo methods mentioned in Chapter 4. Therefore, it is necessary to know the characteristics of the uncertainty measure when the null hypothesis holds for synthetic data. From Fig. 5.7, when synthetic data are generated without a CP, then the value of Un turns out to be quite high. The cumulative frequency distribution for Un have medians nearly one. For 80% ($n = 50$) and 90% ($n = 100$) of all curves. As an example of confidence curves for synthetic data without a CP is shown in Fig. 5.1(b) a high Un value often means a method could provide limited information about the location of a CP, therefore, if Un exceeds a certain bound, then either there is no CP or the method cannot reliably detect the CP location. The Un for synthetic data when there does exist a CP can give a better understanding about the viability of the methods.

(2) THE UNCERTAINTY OF CONFIDENCE CURVES FOR ALTERNATIVE HYPOTHESIS

Figure 5.8 shows the cumulative frequency for Un when the alternative hypothesis holds, and there exists a CP lies in the middle of a time series. The value of Un by the two methods decreases explicitly when H_1 holds, but it shows some differences between Un by the two methods. For the same sample size, CML method tends to provide lower uncer-

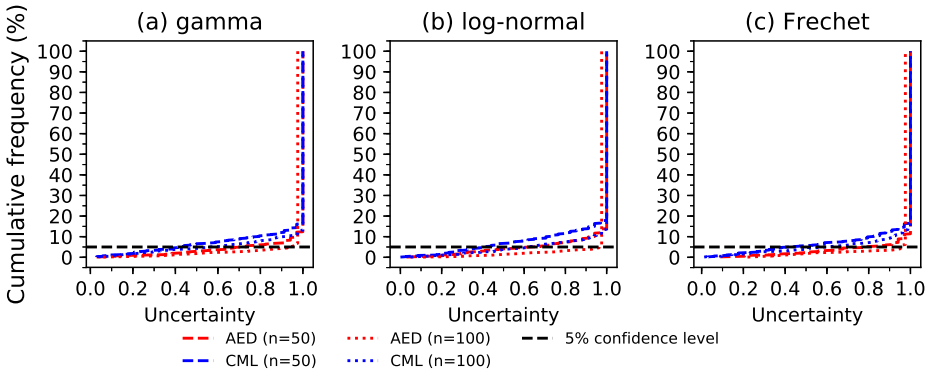


Figure 5.7: Uncertainty measure when null hypothesis holds.

tainty than AED method. This is as expected since CML is a method based on parametric distributions, and AED is a non-parametric one. The sample length influences the value of Un, and with the increase of n, the Un drops significantly. Besides, the impacts of size of change $\Delta\mu$ is negligible. For $n = 100$ and $\Delta\mu = 1$, nearly 60% of the uncertainty lies below 0.5 for GA-CML and AED. For $n = 100$ and $\Delta\mu = 2$, the value of uncertainty halved and for $n = 100$ and $\Delta\mu = 4$, the uncertainty is nearly zero.

5.4. ANALYSIS RESULTS FOR HYDROMETEOROLOGICAL DATA

The results of AED 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.

5.4.1. DATA SOURCE

To examine the performance of the AED and CML methods on real world data, seven time series of measurements were taken from previous publications by Jandhyala *et al.* [25] (case study 1), Reeves *et al.* [26] (case study 2), Conte *et al.* [27] (case study 3) and Zhou *et al.* [19] (case study 4). The detailed information of each time series can be found in Chapter 4.5. The CPs found in the original studies are used as a reference. The change in the mean at the reported change is given in terms of $\Delta\mu/\sigma$, see B.3. Both methods were used to construct confidence curves for CPs with each of the distributions.

5.4.2. ANALYSIS RESULTS

For the Tucumán, Tuscaloosa, and Itaipu, time series shown in Fig. 5.9(a, b, e), the confidence curves generated by AED are shown in Fig. 5.9(c, d, g). These curves show that, for each of time series, the uncertainty measure for each confidence curve Un = 0.32, 0.16, 0.07, which is low. Table 5.2 lists the CPs τ_{ref} found in the references, $\hat{\tau}_{\text{apn}}(y_{\text{obs}})$for AED, and the uncertainty Un of the confidence curve found by AED. 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.

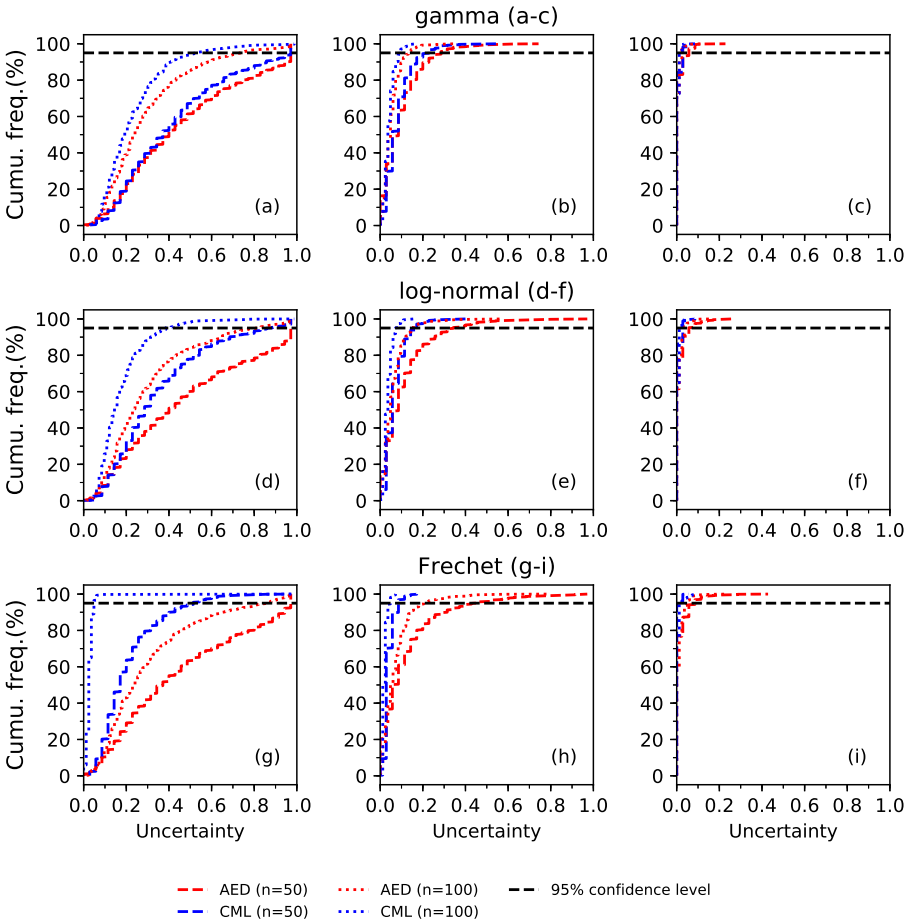


Figure 5.8: Uncertainty measure when the alternative holds.

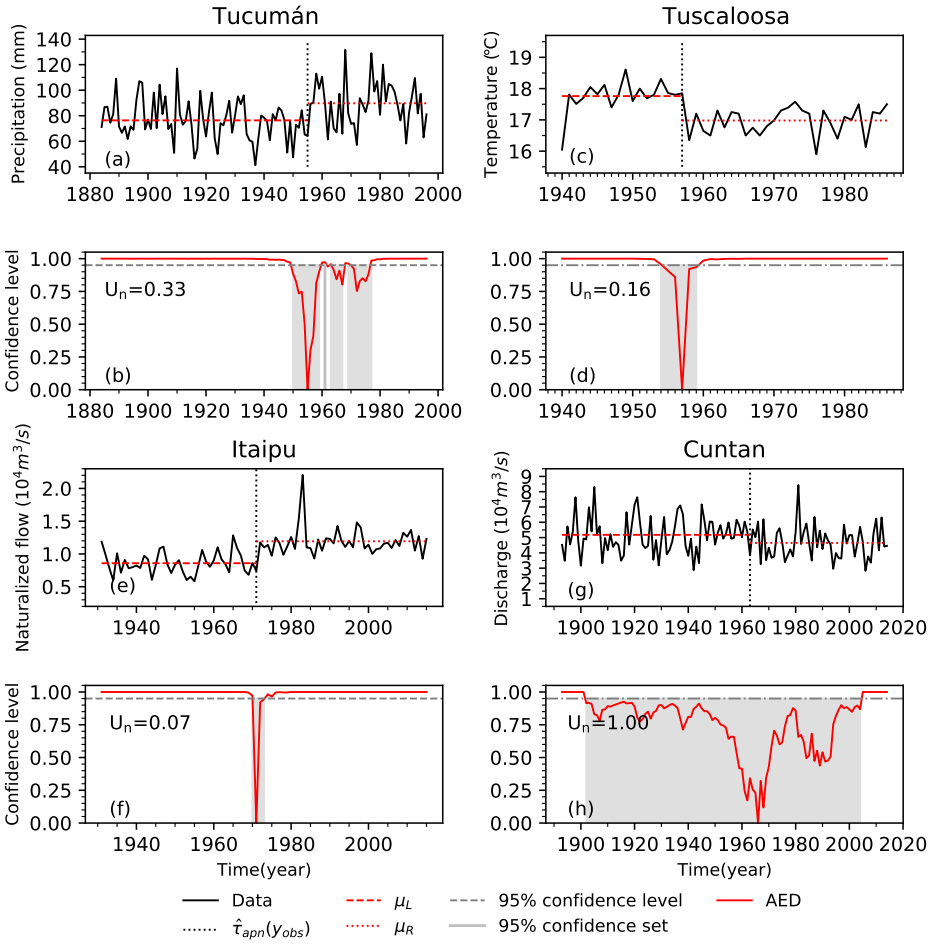


Figure 5.9: Change point analysis of hydrometeorological time series from published papers.

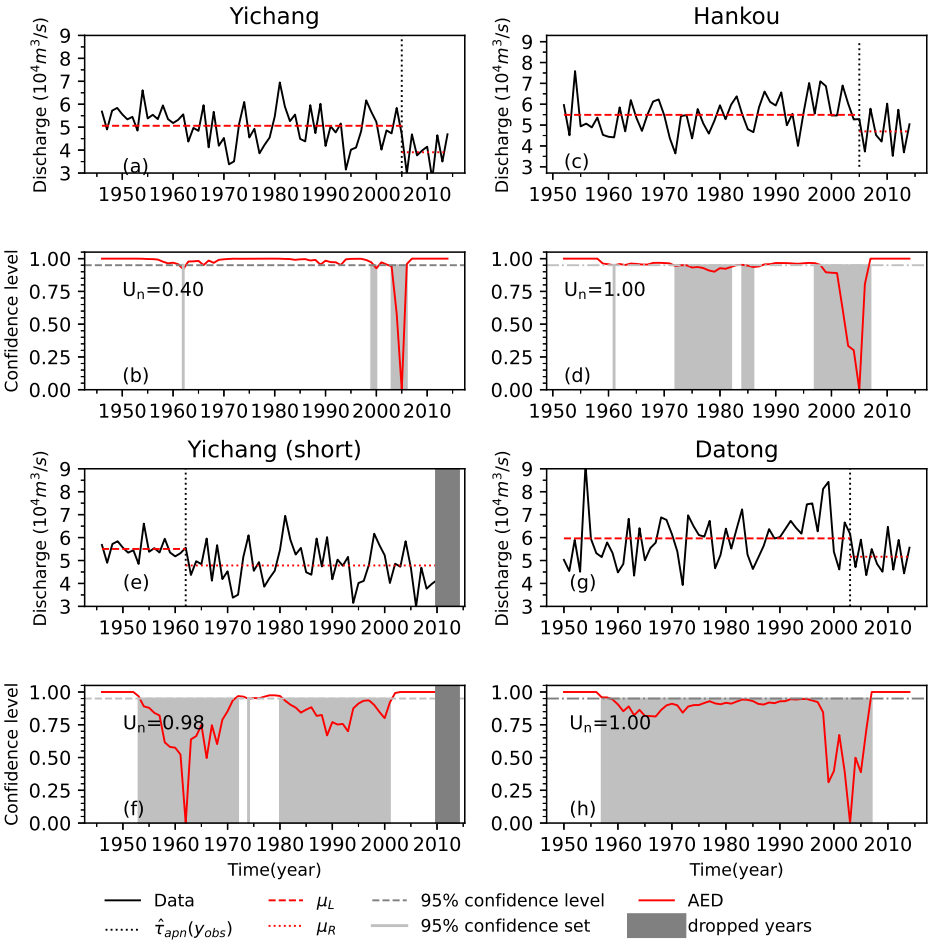


Figure 5.10: Change point analysis of annual maximum discharge time series in Yangtze River.

Table 5.2: Change points found in the hydrometeorological series and statistical properties of the series by AED.

Time series	from	to	τ_{ref}	σ	$\frac{\Delta\mu}{\sigma}$	Un	$\hat{\tau}_{\text{apn}}(y_{\text{obs}})$
Tucumán	1884	1996	1956	18mm	0.76	0.33	1955
Tuscaloosa	1940	1986	1957	0.61°C	−1.3	0.16	1957
Itaipu	1931	2015	1971	$2.5 \times 10^3 \text{m}^3/\text{s}$	1.3	0.07	1971
Cuntan	1893	2014	not found	$12 \times 10^3 \text{m}^3/\text{s}$	−0.45	1.00	n/a
Yichang	1946	2014	1962, 1966	$8.6 \times 10^3 \text{m}^3/\text{s}$	−1.3	0.40	2005
Yichang 'short'	1946	2010	n/a	$8.3 \times 10^3 \text{m}^3/\text{s}$	−0.87	0.98	n/a
Hankou	1952	2014	not found	$8.9 \times 10^3 \text{m}^3/\text{s}$	−0.89	1.00	n/a
Datong	1950	2014	not found	$11 \times 10^3 \text{m}^3/\text{s}$	−0.76	1.00	n/a

For Cuntan, Zhou *et al.* [19] did not find a significant CP. This agrees with the results of AED: the shape of the confidence curve in Fig. 5.9(h) and the Un = 1 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 5.2.

For Yichang station, AED strongly suggests that there is a CP near 2005, see Fig. 5.10(a). For this station, the Un is 0.40, which is acceptable, according to Fig. 5.10(c) or Table 5.2, 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. Figure 5.10(g) shows that for the time series from 1946 to 2010 (Yichang 'short') AED found a CP in 1962, but the uncertainty measure Un of confidence curve is 1, which indicates there is no CP in the 'short' Yichang time series. The size of the change is also smaller. To see whether there might be a second CP, years were successively dropped from the series. It turned out that 2005 was selected until it was masked by n_{min} (which was 8 in this case). Table 5.2 shows that when 2005 was masked, 1962 was found, but with a much wider 95% confidence set and therefore uncertainty is very high.

For Hankou, downstream of Yichang, Fig. 5.10(d) suggests there may be a CP in 2005. The uncertainty measure Un = 1, which indicates there is no CP in the time series. The putative change is a bit smaller, which may explain part of the additional uncertainty (Table 5.2). The methods used in the reference did not find a significant CP at this station.

Further downstream lies Datong station, but, while there is a drop in the confidence curve around 2003, the confidence curve with uncertainty measure Un = 1 (Table 5.2) is completely uninformative, see also Fig. 5.10(h). The methods used in the reference did not find a significant CP at this station.

From Table 5.2 and Fig. 5.10(d, g, 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.

5.5. CONCLUSION

This study provides a distribution-free way to construct confidence curves for CPs. 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 approximate probability distribution of the deviance function (AED). The method introduced by Cunen *et al.* [20] is used as an alternate source of confidence curves. It combines a parametric likelihood function with a deviance function and Monte Carlo simulation (CML). Both methods intrinsically provide confidence sets at all confidence levels that quantify the uncertainty in the results of CP detection. This is an advantage over classical CP detection methods (see Chapter 2 for details) that do not have this feature. Bayesian methods do provide a representation of uncertainty, but they need a prior distribution. The advantage of AED over CML 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 CP detection. Moreover, the performance of the AED is similar to that of CML. The similarity between confidence curves constructed by AED and CML 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 CP 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 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, see Chapter 4 for detailed results. The AED results for the AMR series for the stations Yichang and Hankou along the Yangtze river are among the first that show a possible CP 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 CP problem in a future study.

REFERENCES

[1] M. G. Donat, A. L. Lowry, L. V. Alexander, P. A. O'Gorman, and N. Maher, *More extreme precipitation in the world's dry and wet regions,* Nature Climate Change **6**, 508 (2016).

[2] J. Lehmann, D. Coumou, and K. Frieler, *Increased record-breaking precipitation events under global warming,* Climatic Change **132**, 501 (2015).

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

[4] K. Tamaddun, A. Kalra, and S. Ahmad, *Identification of streamflow changes across the continental United States using variable record lengths,* Hydrology **3**, 24 (2016).

[5] I. Haddeland, J. Heinke, H. Biemans, S. Eisner, M. Flörke, N. Hanasaki, M. Konzmann, F. Ludwig, Y. Masaki, J. Schewe, *et al.*, *Global water resources affected by human interventions and climate change,* Proceedings of the National Academy of Sciences **111**, 3251 (2014).

[6] G. Blöschl, M. F. Bierkens, A. Chambel, C. Cudennec, G. Destouni, A. Fiori, J. W. Kirchner, J. J. McDonnell, H. H. Savenije, M. Sivapalan, *et al.*, *Twenty-three unsolved problems in hydrology (UPH)–a community perspective,* Hydrological Sciences Journal **64**, 1141 (2019).

[7] H. McMillan, A. Montanari, C. Cudennec, H. Savenije, H. Kreibich, T. Krueger, J. Liu, A. Mejia, A. Van Loon, H. Aksoy, *et al.*, *Panta Rhei 2013–2015: Global perspectives on hydrology, society and change,* Hydrological Sciences Journal **61**, 1174 (2016).

[8] R. Killick, I. A. Eckley, K. Ewans, and P. Jonathan, *Detection of changes in variance of oceanographic time-series using changepoint analysis,* Ocean Engineering **37**, 1120 (2010).

[9] J. Chen and A. K. Gupta, *Parametric statistical change point analysis: With applications to genetics, medicine, and finance* (Springer Science & Business Media, 2011).

[10] E. Brodsky and B. S. Darkhovsky, *Nonparametric methods in change point problems*, Vol. 243 (Springer Science & Business Media, 2013).

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

[12] M. Gocic and S. Trajkovic, *Analysis of changes in meteorological variables using Mann-Kendall and Sen's slope estimator statistical tests in Serbia,* Global and Planetary Change **100**, 172 (2013).

[13] Z. W. Kundzewicz and A. J. Robson, *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**, 7 (2004).

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

[15] S. Lee, J. Ha, O. Na, and S. Na, *The CUSUM test for parameter change in time series models,* Scandinavian Journal of Statistics **30**, 781 (2003).

[16] D. M. Hawkins and K. Zamba, *A change-point model for a shift in variance,* Journal of Quality Technology **37**, 21 (2005).

[17] G. Gurevich and A. Vexler, *Retrospective change point detection: From parametric to distribution free policies,* Communications in Statistics—Simulation and Computation® **39**, 899 (2010).

[18] R. L. Wasserstein, A. L. Schirm, and N. A. Lazar, *Moving to a world beyond "$p < 0.05$"*, The American Statistician **73**, 1 (2019).

[19] C. Zhou, R. van Nooijen, A. Kolechkina, and M. Hrachowitz, *Comparative analysis of nonparametric change-point detectors commonly used in hydrology,* Hydrological Sciences Journal **64**, 1690 (2019).

[20] C. Cunen, G. Hermansen, and N. L. Hjort, *Confidence distributions for change-points and regime shifts,* Journal of Statistical Planning and Inference **195**, 14 (2018).

[21] S. G. Coles and J. A. Tawn, *A Bayesian analysis of extreme rainfall data,* Journal of the Royal Statistical Society: Series C (Applied Statistics) **45**, 463 (1996).

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

[23] L. Perreault, J. Bernier, B. Bobée, and E. Parent, *Bayesian change-point analysis in hydrometeorological time series. part 1. The normal model revisited,* Journal of Hydrology **235**, 221 (2000).

[24] P. Hall and M. Martin, *On the bootstrap and two-sample problems,* Australian Journal of Statistics **30**, 179 (1988).

[25] V. K. Jandhyala, S. B. Fotopoulos, and J. You, *Change-point analysis of mean annual rainfall data from Tucumán, Argentina,* Environmetrics **21**, 687 (2010).

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

[27] L. C. Conte, D. M. Bayer, and F. M. Bayer, *Bootstrap Pettitt test for detecting change points in hydroclimatological data: Case study of Itaipu Hydroelectric Plant, Brazil,* Hydrological Sciences Journal **64**, 1312 (2019).

6

CONFIDENCE CURVES FOR THE DEPENDENCE PARAMETER IN COPULAS

The term of 'copulas' was first introduced by Abe Sklar (1925-2020).
Now copulas have been a wide spread tool in multivariate analysis.

Parts of this chapter have been submitted as "Zhou, C., van Nooijen, R., Kolechkina, A., Gargouri-Ellouze, E., and van de Giesen, N., Using confidence curves to capture the uncertainty in dependence structure in copula models, Hydrological Sciences Journal, under review, 2021."

6.1. INTRODUCTION

The modelling of the interdependence of hydrological time series and the communication of the uncertainty of the results of hydrological studies are two major challenges that face modern hydrology. Conveying information on uncertainty is even mentioned as one of a list of twenty-three unsolved problems in hydrology [1]. An important tool to address the first challenge is the copula concept, introduced by Sklar [2]. This simplifies working with multivariate distributions by decomposing them into marginals and a dependence structure given by a copula. Until the year of 2000, hydrology and water management studies mostly used multivariate versions of univariate distributions. Examples are articles, that model the relation between the intensity and duration of a storm when investigating the probabilistic structure of runoff [3], model extreme rainfall [4], study floods [5], or model low flow events [6].

In fact, most of the mathematical literature was limited to such distributions until about 1980. The reason for this is the high complexity of multivariate distributions. This problem was at the same time illustrated and partially solved by Sklar [2]. He showed that any n-dimensional multivariate distribution can be constructed by providing the *cumulative distribution function* (cdf) for each of its n marginals and then combining these using a copula, a function from the unit hypercube $[0,1]^n$ to the unit interval $[0,1]$ that satisfies a specific set of conditions (see Appendix I). The generality of these conditions is such that the number of possible multivariate distributions is overwhelming even without taking into account all the possible combinations of marginal distributions, but at the same time copulas make it possible to fit and study the marginal distributions and the dependence structure separately. For a modern overview and hydrological examples see, for instance, Salvadori and De Michele [7, 8], Favre *et al.* [9], or Genest and Favre [10]. As a result, copulas are now used in many hydrological studies. In De Michele and Salvadori [11], Zhang and Singh [12], and Gargouri-Ellouze [13] bivariate copulas are used as a model for the joint distribution of rainfall parameters. Shiau [14], Kwon and Lall [15] model the joint distribution of drought duration and severity with copulas. Gartsman *et al.* [16] consider flood risk at different sites that are linked through shared processes and use copulas to model the joint risk. Bárdossy [17] uses copulas to model dependence between different ground water parameters. Debele *et al.* [18] studied the depednence between seasonal peak flood and annual maxima design quantiles in San River basin by copula. In Grimaldi and Serinaldi [19], bivariate and trivariate copulas are used to study drought properties.

In this chapter, a new combination of fitting method and uncertainty representation is proposed, studied and applied to hydrological data. The fitting method used estimates the copula parameter directly without fitting the marginals Genest *et al.* [20], and uncertainty in the copula parameter is then represented by a confidence curve. While Ko and Hjort [21] also use copulas and confidence curves, the fitting method used there is different. They compare two options: maximum likelihood to the fit marginals and the copula at the same time, or fitting the marginals first and then fitting the copula. Apart from the difference in fitting methods, the main difference between this chapter and Ko and Hjort [21] is that they emphasize the study of the fitting method and in particular the effects of fitting the wrong model, whereas this chapter emphasizes the study of the properties of the confidence curve and its possible applications in hydrology.

With regards to the second challenge mentioned earlier, confidence curves for parameters, as originally proposed by Birnbaum [22] and later amended and extended by Schweder and Hjort [23], may well offer a way of not only communicating, but also studying uncertainty in statistical results. A confidence curve provides a collection of confidence intervals for a given parameter at all confidence levels. Such curves provide a useful overview of the tradeoff between confidence and confidence interval size, but they are not yet in common use in hydrology. Once a confidence curve is constructed for the copula parameter, that curve can be used to construct confidence curves for quantities derived from the copula such as Kendall's τ, or, perhaps even more important, for the relation between specific return periods for the marginal distributions and the corresponding bivariate return periods [24].

In hydrology, the time series under study are often relatively short, so the number of parameters that can realistically be estimated is limited. It is therefore not surprising that in hydrology one parameter copula families are often used. Of these the Clayton, Frank and Gumbel-Hougaard copula families were selected for use in this chapter.

This chapter is organized as follows. First the method is presented. Next, some criteria are selected to evaluate the performance of the method. Then the results of the application of the method to synthetic time series are analysed. Once the validity of the method has been established, it is applied to two hydrological problems. The dependence between time series of annual maximum daily discharge of the Rhine and its tributaries was examined. The aim was to see to what extent the different subcatchments are likely to have extreme events in the same year. Next rainfall-runoff from a karst region in Tunisia was analysed to determine the probable delay between precipitation and runoff. Finally, we discuss the results and present our conclusions. Notations, definitions, and some background on copulas are provided in the appendices.

6.2. METHODOLOGY OF CONSTRUCTING CONFIDENCE CURVES FOR THE DEPENDENCE PARAMETER IN COPULAS

In this chapter, copulas will be fitted to time series of two dimensional random vectors $\boldsymbol{Z} = \boldsymbol{Z_1, Z_2, ..., Z_n}$. Each series $\boldsymbol{Z}$ has two associated time series of random variables that correspond to the components of the random vectors $X_i = Z_{i,1}$ and $Y_i = Z_{i,2}$. It is assumed that the X_i are *independent identically distributed* (i.i.d.) *random variables* (RVs) X_i and Y_i are not necessarily independent and that their joint distribution does not depend on the value of i. The joint cdf will be denoted by H. In the remainder of the chapter $\boldsymbol{X}$ will stand for the random sample $X_1, X_2, ..., X_n$; $\boldsymbol{x}$ will represent a realization $x_1, x_2, ..., x_n$ of that sample, and $\boldsymbol{x}_{\text{obs}}$ will stand for a specific series of observations. The same holds for $\boldsymbol{Y}, \boldsymbol{y}$ and $\boldsymbol{y}_{\text{obs}}$.

6.2.1. THE COPULAS

According to Sklar [2], the joint distribution of n random variables is fully specified by the combination of an $n-$dimensional copula and n one dimensional marginal distributions. Therefore, if X, Y are RVs, then their joint cdf H can be described by the marginal cdfs F and G for X and Y respectively and a copula C

$$H(x, y) = C(F(x), G(y)) \tag{6.1}$$

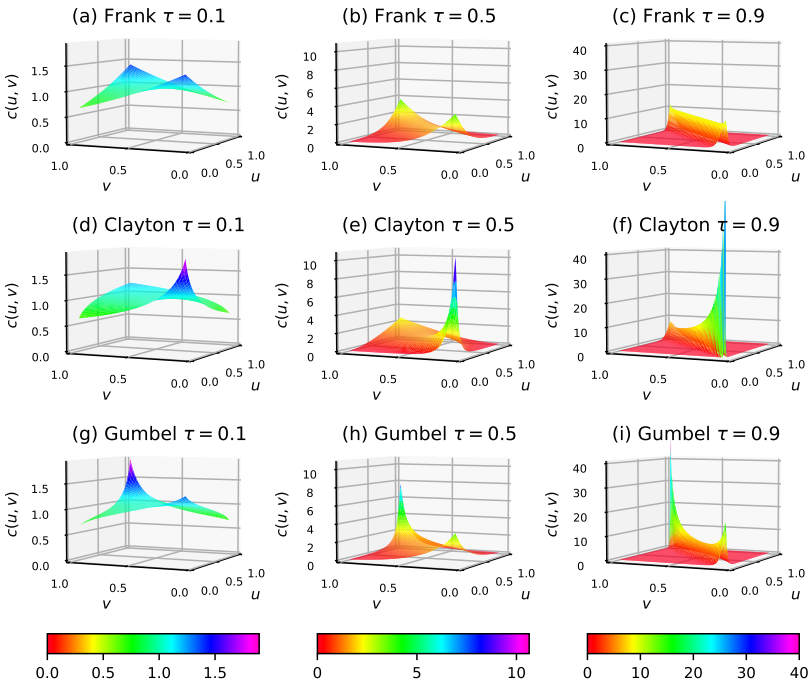


Figure 6.1: The shape of the pdf of different copulas for τ=0.1,0.5,0.9.

Table 6.1: Parameter ranges and the relation between θ and Kendall's τ, where $D_1(\theta)$ is the first Debye function [25].

Copula family	Parameter range	Relation between θ and Kendall's τ
Frank	$\theta \neq 0$	$\tau = 1 - \frac{4(1-D_1(\theta))}{\theta}$
Gumbel	$1 \leq \theta \leq \infty$	$\tau = 1 - \frac{1}{\theta}$
Clayton	$-1 \leq \theta \leq \infty$	$\tau = \frac{\theta}{\theta+2}$

Details about the Frank, Clayton, and Gumbel copulas are given in Appendix I.

For the three copulas, the parameter range and the relation between the copula parameter θ and Kendall's τ is given in Table 6.1. The relation provides a way to examine the uncertainty about the parameter on a common scale. Clayton and Frank copulas can model positive and negative correlation without limitations. However, for a Gumbel copula $\tau \geq 0$, which implies that a Gumbel copula can only model positive correlation. The shape of the pdfs of the copulas for different values of τ can be seen in Fig. 6.1. For $\tau = 0.1$ (Fig. 6.1(a,d,g)) the copula are relatively close to the pdf of a uniform 2D distribution; this can be seen most clearly in Fig. 6.1(a). All three copulas are symmetric around the $u = v$ axis. The Frank copula (6.1(a-c)) has an additional symmetry relative to the line $u = 1 - v$, while Clayton (6.1(d-f)) has a peak at (0,0) and Gumbel (6.1(g-i)) has a peak at (1,1).

6.2.2. COPULA PARAMETER ESTIMATION

In most cases, the parametric distribution of the observations is unknown; this complicates parameter estimation. Therefore, Genest *et al.* [20] use a pseudo log-likelihood approach. A rescaled version of the *empirical cumulative distribution function* (ecdf) is used for the marginals. The rescaled ecdfs for $\boldsymbol{X}$ and $\boldsymbol{Y}$ are

$$\hat{U}(x) = \tfrac{n}{n+1}\tfrac{1}{n}\textstyle\sum_{i=1}^{n}[\![X_i \le x]\!] \qquad \hat{V}(y) = \tfrac{n}{n+1}\tfrac{1}{n}\textstyle\sum_{i=1}^{n}[\![Y_i \le y]\!] \tag{6.2}$$

As pseudo log-likelihood Genest *et al.* [20] take

$$\hat{\ell}(\theta) = \sum_{i=1}^{n} \log\left(c\left(\hat{U}(X_i), \hat{V}(Y_i); \theta\right)\right) \tag{6.3}$$

They then introduce

$$\hat{\theta} = \text{argmax}_\theta \hat{\ell}(\theta) \tag{6.4}$$

the pseudo maximum likelihood estimator (pmle) for θ and show that it is a consistent and asymptotically normal estimator. Chen and Fan [26] show this even holds under model misspecification. For a given set of observations, the calculation is performed as follows. The value of τ corresponding to the pmle $\hat{\theta}$ is denoted by $\hat{\tau}$. For a given set of observations, the calculation is performed as follows

$$u_i = \frac{1}{n+1}\sum_{i=1}^{n}[\![x_{\text{obs},i} \le x_i]\!]; v_i = \frac{1}{n+1}\sum_{i=1}^{n}[\![y_{\text{obs},i} \le y_i]\!] \tag{6.5}$$

$$\ell(\theta) = \sum_{i=1}^{n} \log(c(u_i, v_i; \theta)) \tag{6.6}$$

which results in the estimate

$$\hat{\theta} = \text{argmax}_\theta \ell(\theta) \tag{6.7}$$

for the copula parameter θ.

Note that the use of u_i and v_i instead of x_i and y_i implies that we can apply any strictly increasing function w_x or w_y to the x_i or the y_i respectively without changing the u_i and v_i. In fact, we could even replace x_i by its rank in the sorted sequence of the x_i. This implies, for instance, that for the marginals of the copula, $u = 0.25$ corresponds to the bound of the first quartile of the distribution of X.

6.2.3. THE CONSTRUCTION OF APPROXIMATE CONFIDENCE CURVES

The definition of a confidence curve is given in Chapter 3. Construction of an exact confidence curve is quite difficult, because, like the construction of confidence distributions, it is not (yet) a question of applying a simple standard approach. However, there is a standard method to construct an approximate confidence curve for a parameter θ [23]. It assumes that a log-likelihood function is available. Here the pseudo log-likelihood defined earlier will be used instead. The construction of the approximate cc uses a function D, defined by

$$D(\theta) = 2\left(\ell(\hat{\theta}) - \ell(\theta)\right) \tag{6.8}$$

and which they refer to as the deviance function. For $\theta = \hat{\theta}$, this function assumes its minimum value, which is zero. The cdf for $D(\theta)$ is

$$K_\theta(\lambda) = \Pr\{D(\theta) \leq \lambda)\} \tag{6.9}$$

If ℓ were a true likelihood, then according to Wilks' theorem, the deviance function $D(\theta)$ is approximate χ^2_1 distributed. As Chen and Fan [26] have shown that the limit distribution for the estimator $\hat{\theta}$ is approximate normal, one might hope a version of Wilks' theorem could be proved for the current deviance, see also Schweder and Hjort [23]. To avoid the additional computation time needed for an MC approximation of K_θ, it will be approximated by a χ^2_1 distribution.

6.2.4. PROPERTIES OF CONFIDENCE CURVES

To allow for proper comparison between copulas, the confidence curves for θ have been translated into confidence curves for τ using Table 6.1. The following properties of these confidence curves will be examined:

- Actual versus nominal coverage probability of confidence intervals: As introduced in Chapter 3, the actual coverage probability of a confidence curve should be close to the nominal coverage probability. If the actual coverage probability is lower than the nominal one, then a confidence curve has a *permissive coverage*; it is too optimistic about finding the parameter in the interval. If the actual coverage probability is higher than the nominal one, then the confidence curve has *conservative coverage*, so it is unnecessarily pessimistic about finding the parameter in the interval.

- The width of the confidence interval. For a given confidence level, a confidence interval can be extracted from a confidence curve. The width of a confidence interval at a given confidence level shows the uncertainty level of the estimate of τ. If a confidence interval is small, then the estimate has low uncertainty. Since a confidence curve is comprised of confidence intervals at all confidence levels, the shape of it will be helpful for understanding the uncertainty in $\hat{\tau}$, and the narrower confidence intervals at high confidence levels, the less uncertain $\hat{\tau}$ is. Note that if a confidence interval for τ corresponding to confidence level λ and contained in [-1,1] is wider than 2λ, then does not provide any information.

- The difference between $\hat{\tau}$ and τ_{true}, where τ_{true} is the Kendall's τ for the copula from which the synthetic time series was drawn.

6.3. EVALUATION OF THE METHOD WITH SYNTHETIC DATA

In order to evaluate the method, synthetic data sets from the three copulas (Frank, Gumbel, Clayton) were generated. As the aim is to study and compare the properties of the confidence curves for all three copulas, only samples from copulas with positive τ were used. Because the different copulas have different parameter ranges and because those parameter ranges all extend to positive infinity, it would be difficult to directly compare results for confidence intervals for different copulas. Fortunately, for all three copulas

Table 6.2: The dependence structure described by Kendall's $\tau \in [0,1]$.

Copulas	Kendall's τ	0.1	0.3	0.5	0.7	0.9
Frank		0.91	2.9	5.7	11.4	38
Gumbel	θ	1.11	1.43	2	3.33	10
Clayton		0.22	0.86	2	4.67	18

there is a strictly increasing function that maps the copula parameter to a Kendall's τ value (Table 6.1). This allows display of the results for the copula in terms of τ.

6.3.1. SYNTHETIC TIME SERIES GENERATION

To determine the statistical properties of the method, synthetic time series of length $n = 50, 100, 200$ were generated by drawing from Clayton, Frank, and Gumbel copulas for $\tau = 0.1, 0.3, 0.5, 0.7, 0.9$. For each combination of copula family, length, and τ a set of $N = 1000$ time series was generated. For each set, the resulting 1000 confidence curves were used to analyse the coverage of the associated confidence intervals, the frequency distribution of $\hat{\tau}$, and the width of the confidence intervals at a given level. The corresponding value of θ for each copula was calculated using the formula from Table 6.1 and is listed in Table 6.2. The parameter value that was used to generate a specific series will be denoted by θ_{true} and the corresponding Kendall's τ by τ_{true}.

6.3.2. EXAMPLES OF SYNTHETIC DATA

Some examples of synthetic samples are shown in Fig. 6.2. When $\tau = 0.9$ (Fig. 6.2(c,f,i)), the correlation between u and v is clearly visible. For $\tau = 0.1$ (Fig. 6.2(a,d,g)), a plot of the (u_i, v_i) pairs does not show a clear pattern. From Fig. 6.2b, the Frank copula ($\tau = 0.5$) shows correlation over the whole range, but with peak of equal height in the lower left and the upper right corner. For Gumbel with $\tau = 0.5$, the plot in Fig. 6.2(e) displays some correlation over the whole range, somewhat stronger in the lower left corner, and very strong correlation in the upper right corner. For Clayton with $\tau = 0.5$, the role of the corners is reversed (Fig. 6.2(h)). See also Fig. 6.1.

The approximate confidence curves for τ for the bivariate copula samples in Fig. 6.2 are shown in Fig. 6.3. A confidence curve reaches its lower point, cc(τ) lines in (0,1]. A traditional 95% confidence level is presented in a dashed line in each plot, and one can extract a nominal confidence interval for τ_{true} with a confidence level of 95% from each confidence curve.

Figure 6.3 shows that in principle, a confidence curve for τ contains more information than one confidence interval for a single confidence level. It also shows that, as τ_{true} increases, the widths of the confidence intervals that make up the curve decrease. However, Fig. 6.5 shows that the actual coverage for high τ_{true} is permissive, so part of the decrease may be due to an underestimation of the interval width. Figure 6.4 shows that some of the decrease is real.

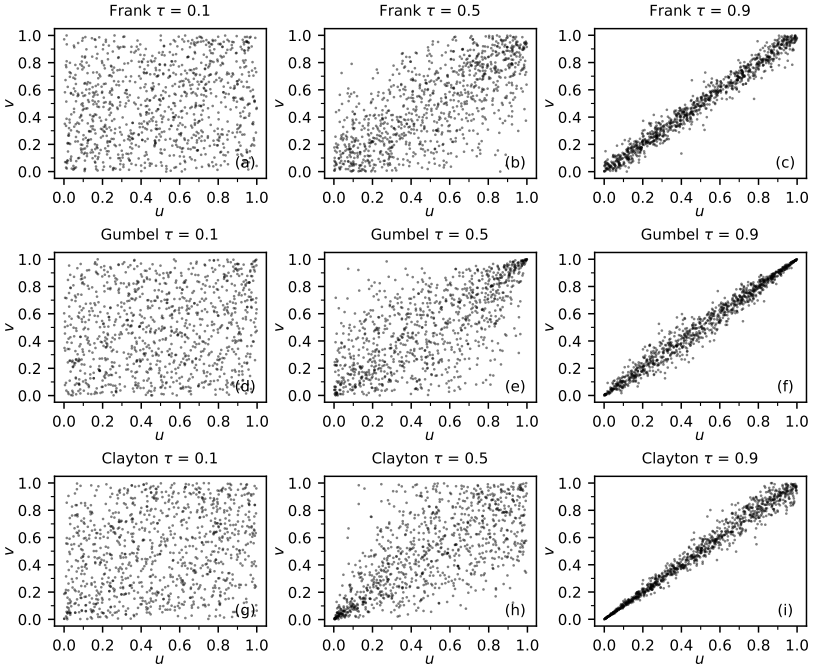


Figure 6.2: Scatter plots of a sample from a copula with $\tau = 0.1, 0.5, 0.9$ and $n = 100$.

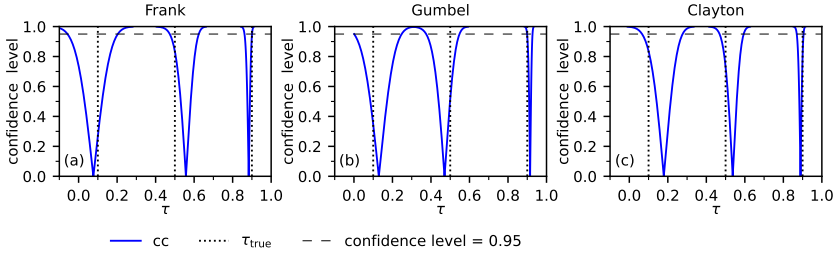


Figure 6.3: Example of confidence curves for τ for one of the synthetic samples for each copula with $\tau_{\text{true}} = 0.1, 0.5, 0.9$ and $n = 100$.

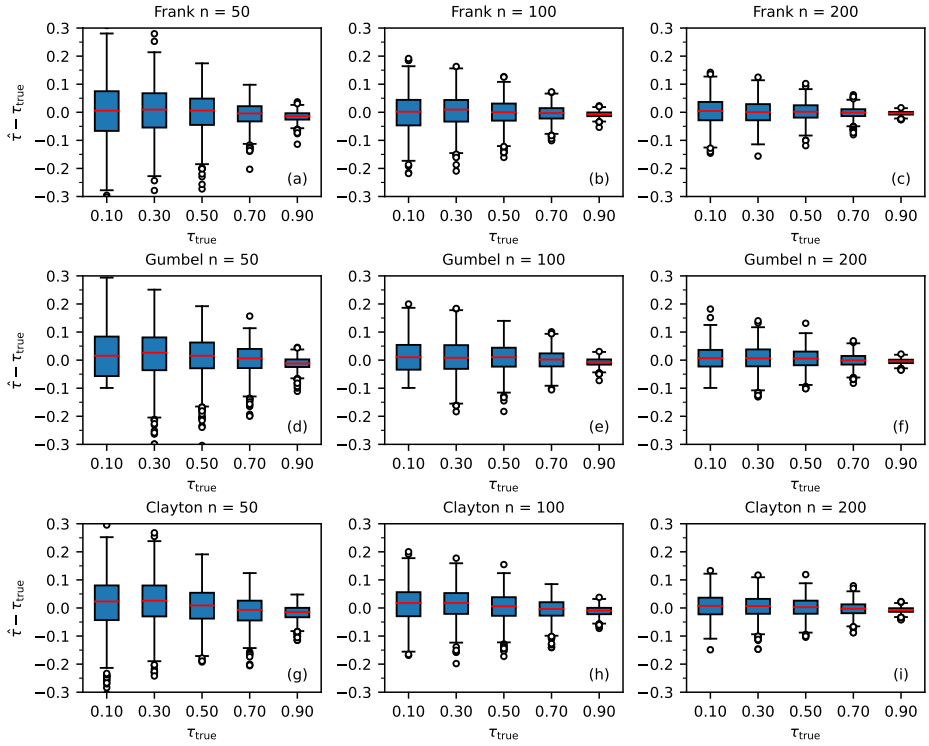


Figure 6.4: Boxplots of $\hat{\tau} - \tau_{\text{true}}$ for different copulas and different sample sizes.

6.3.3. SPREAD IN THE ESTIMATED KENDALL'S τ

The box plots in Fig. 6.4 show that for $\hat{\tau} - \tau_{\text{true}}$ both the spread of the outliers and the interquartile distance decrease with increasing sample length for all copulas. The spread of the outliers and the interquartile distance for $\hat{\tau} - \tau_{\text{true}}$ also decrease with increasing τ.

6.3.4. ACTUAL COVERAGE PROBABILITY FOR KENDALL'S τ

The actual coverage of the confidence interval associated with the confidence curve was examined for all confidence levels. The actual versus nominal coverage probability for random copula samples is shown in Fig. 6.5, and the coverage at the 95% confidence level is listed in Table 6.3. According to the results shown in Fig. 6.5, sample length has little effect on the actual coverage probability up to $\tau = 0.5$ and the actual coverage probability does not change much when n increases from 50 to 200. For $\tau = 0.9$, the sample length has a visible effect on the actual coverage probability for the Frank and Gumbel copulas but not for the Clayton copula (Fig. 6.5(g-i)).

The dependence level strongly influences the coverage. If the bivariate samples are weakly dependent, for instance $\tau = 0.1$ (Fig. 6.5(a,d,g)), the actual coverage probability is close to the nominal. For samples with high dependence, for instance $\tau = 0.9$ (Fig. 6.5(c,f,i)) the actual coverage probability is lower than the nominal. For bivariate sam-

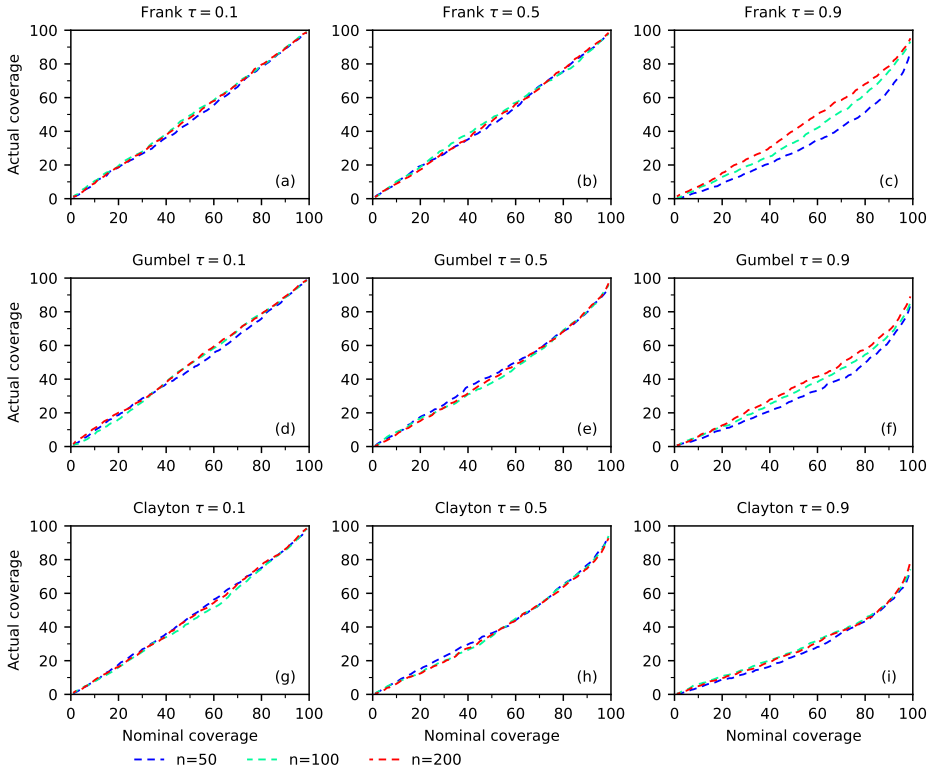


Figure 6.5: Actual coverage probability versus the nominal one for τ in copulas.

Table 6.3: Actual coverage probability of a confidence curve with a nominal coverage probability of 95%.

Bivariate random copulas		Actual coverage Nominal coverage: 0.95				
Kendall's τ		0.1	0.3	0.5	0.7	0.9
Frank	$n = 50$	94.2	92.6	93.1	91.5	73.0
	$n = 100$	95.0	93.2	92.6	92.0	84.6
	$n = 200$	94.4	94.6	93.3	93.6	86.5
Gumbel	$n = 50$	94.3	90.6	87.2	84.1	72.1
	$n = 100$	94.0	90.2	87.4	85.5	74.6
	$n = 200$	94.7	89.8	88.0	84.6	78.1
Clayton	$n = 50$	92.0	86.5	85.1	78.5	62.3
	$n = 100$	92.1	86.2	83.7	81.5	63.4
	$n = 200$	93.1	88.1	82.6	79.4	64.1

ples from a Frank copula with $\tau = 0.9$ and nominal coverage probability of 95%, the actual coverage probability is only 73% for $n = 50$, 84.6% for $n = 100$, and 86.5% for $n = 200$. Results are similar for the other copulas (Table 6.3).

The statistics for $\hat{\tau} - \tau_{\text{true}}$ in Fig. 6.4 show that the error in the estimate of τ decreases with increasing τ_{true}. The results in Fig. 6.4 are in line with this, but Fig. 6.5 shows that the interval widths for high τ are overly optimistic. So, while strong independence is associated with lower uncertainty, the approximate confidence curves calculated by the current version of our code are too optimistic for values of τ close to 1.

6.3.5. THE WIDTH OF 95% CONFIDENCE INTERVALS FOR THE DEPENDENCE PARAMETER IN COPULAS

Information on the distribution of the widths of confidence intervals for the 95% confidence level is shown in Fig. 6.6. The figure shows that the effect of the sample length n are significant and that the interval width decreases as n gets larger. Therefore, the uncertainty about $\hat{\tau}$ decreases as the sample length gets larger. These results are in qualitative agreement with those shown in Fig. 6.4.

6.3.6. EFFECTS OF THE MISSPECIFICATION OF COPULA

If we use the Clayton or Gumbel copula to fit a sample from one of the other copula families and calculate τ, then this results in a biased estimate. The Frank copula does much better in this respect. Box plots of the difference between the true value and the estimate of τ in the synthetic experiments are given in Fig. 6.7.

6.4. TWO EXAMPLES OF THE USE OF THE METHOD ON OBSERVED HYDROLOGICAL TIME SERIES

In practice, the method can be used to examine the uncertainty about dependence between time series and the effects of this uncertainty on an analysis based on that dependence. Two examples are given. The first example considers the correlation between

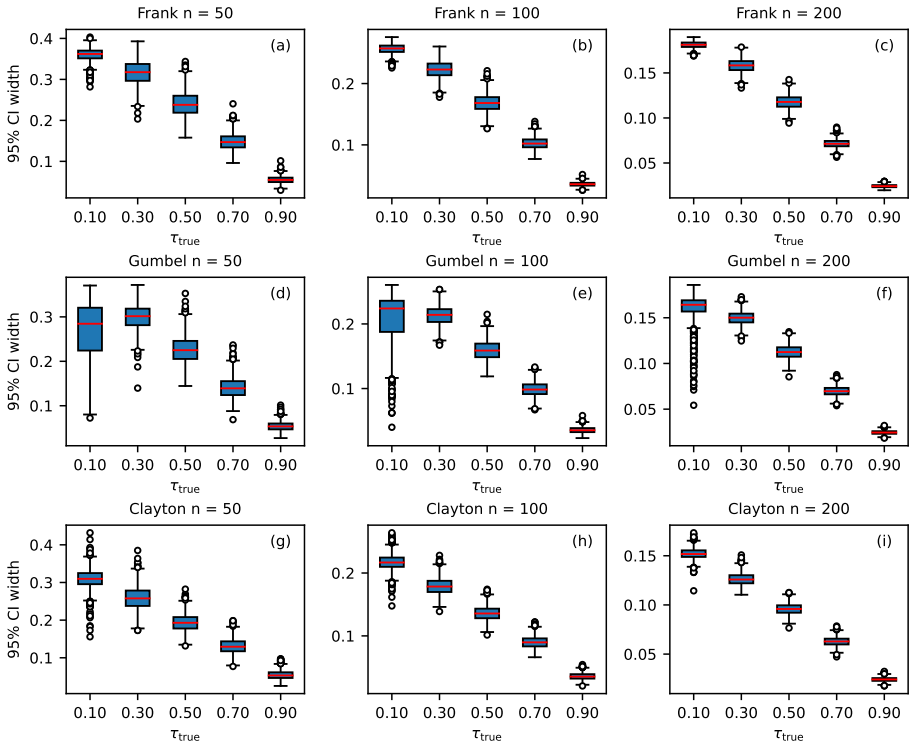


Figure 6.6: Box plots of the width of confidence intervals for a 95% confidence level.

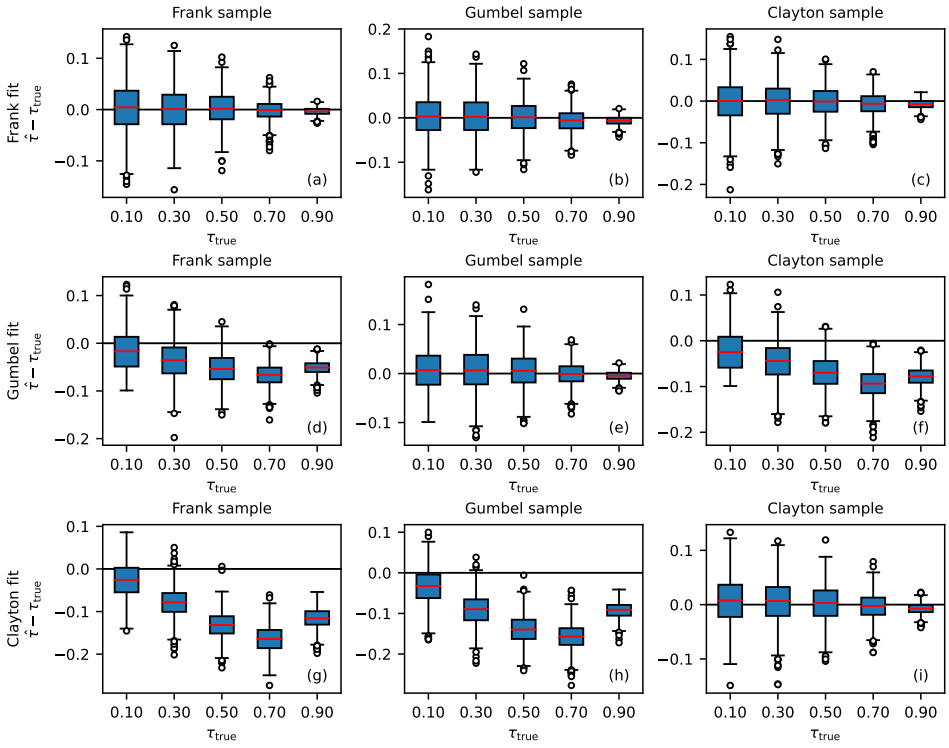


Figure 6.7: Box plots of the difference between the true value and the estimate of τ in the synthetic experiments with $n = 200$.

yearly extremes for several measurement stations on the Rhine and its tributaries. The second example investigates estimation of the lag between rainfall and runoff for a karst area and the uncertainty in that estimate.

6.4.1. DEPENDENCE STRUCTURE FOR EXTREME FLOWS IN TRIBUTARIES OF THE RIVER RHINE

In a river with many tributaries, such as the Rhine, the timing of high discharges in the main river and the different tributaries determines the resulting discharge in the main river. The risk of a flood in the main river would therefore depend strongly on the timing of high discharges in the tributaries. If timing of the peak discharges is such that they arrive at the same point in the main river at the same time, then they will reinforce each other. One way to assess this risk is to consider time series of known periods of high flows and check for cross correlation. However, travel times will depend on flood wave size and the state of the river bed, so future events may not display the behaviour seen in past events. At the same time, the tributaries of the catchments might change more slowly and therefore results linked to the catchment might be less variable in time. Now, for flood waves to cause a problem, a necessary condition is that for different tributaries flood waves with high return periods tend to occur in the same year. As the return periods are linked to the tributary catchment as a whole, this might be something that varies less than flood wave travel times. This results in the following question: do discharges with the same return period tend to occur in the same year? This question can be answered by determining a copula for the time series of return periods for the different tributaries.

METHODOLOGY

As was noted in 6.2.2, the preparatory step for the fitting method used in this chapter means that first mapping discharges to return periods will not change the result of the fitting process. This implies that using the fitting method to determine the copula parameters for the three different copulas and the associated uncertainty will tell us something about the dependence structure of the return periods. More specifically, are high return periods correlated and if yes, then how? For the corresponding copula, this would mean that there should be a peak in the pdf in the upper right hand corner of the (u, v) plane. The uncertainty in the dependence structure can be examined by looking at the copula corresponding to the estimated parameter and the difference between that copula and copulas corresponding to the lower or upper bound of a confidence interval for a given confidence level.

As the time series are series of annual maxima, a copula package by Hofert *et al.* [27] was used to extend the collection of one parameter copulas applied to the series with additional copulas specifically suited for extreme values: Galambos and Huesler-Reiss (see I.4). Of the other extreme value copulas, the Tawn copula was not considered because its Kendall's τ cannot exceed 0.418, and the t-EV copula was not considered because it has two parameters. To allow for high correlation both for low and for high return periods, the copulas that have different correlation structure for low and high parameters (Clayton, Gumbel, Galambos, Huesler- Reiss) were tried both in their standard orientation and after a 180 degree rotation, the rotated copula will be denoted by, for instance, Gum-

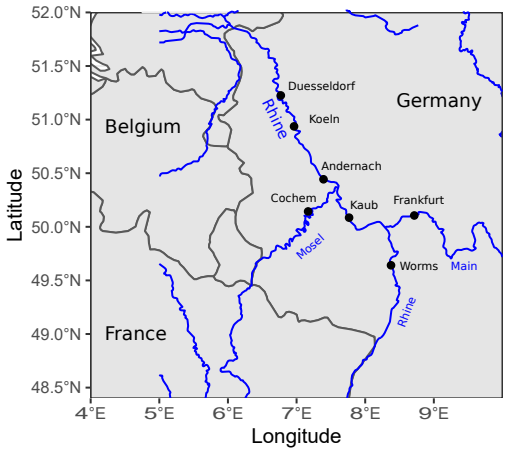


Figure 6.8: Locations of station on the Rhine River.

bel 180°. As in the rest of this chapter, the uncertainty in the parameter is represented by a confidence curve for the Kendall's τ . For a given τ, the shape of the Galambos and Huesler-Reiss copulas is very close to that of the Gumbel copula.

Time series of annual daily maximum flows for several stations for the Rhine and one station each for the Mosel and the Main were obtained from from the Global Runoff Data Center [28]. The stations are shown in Fig. 6.8.

6

RESULTS AND DISCUSSION

To illustrate the type of results that would be obtained, four pairs of measurement stations were selected that were expected to have different dependency structures. For each pair the discharge at a station downstream of the confluence point was determined. The pair Andernach and Koeln serves as a test case. For these stations a near perfect correlation was expected because no major tributaries enter the river between the stations. Figure 6.9(a) confirms this. Figure 6.9(b) shows that high discharges at the downstream station tend to be correlated as well. Values of the parameter θ and Kendall's τ values can be found in Tables 6.4 and 6.5 respectively. The other station pairs combine a station on a tributary with a station on the Rhine River upstream of the confluence. All pairs show definite correlation as the bounds of the confidence intervals up to 99% are well away from zero (Fig. 6.9(b-d)). For all pairs the best fit was obtained with the 180°rotated version version of Gumbel, Galambos or Huesler-Reiss. Given the shape of the pdf of these copulas, with a peak in the lower left quadrant, this could suggest stronger correlation for short return periods than for long return periods. However, the scarcity of points in the upper right corner could also have caused this preference for the rotated version (Fig. 6.9(f-h)). In the scatter plots (Fig. 6.9(e-h)) colour is used to show the discharge at a station downstream of the confluence of tributary and main river. An illustration of the variation in the shape of the pdf of a copula over a 95% confidence interval can be found in Fig. 6.10 for the pair Cochem and Kaub. For example, for the Frank copula Fig. 6.10(a) shows the pdf for $\hat{\theta}$, Fig. 6.10(f) shows the difference between the pdf at $\hat{\theta}$ and the

Table 6.4: Dependence parameters between time series and width of the 95% confidence intervals for the estimate of dependence parameter.

Time series pair	Frank		Clayton		Gumbel 180°		Galambos 180°		Huesler-Reiss 180°	
	τ	width	τ	width	τ	width	τ	width	τ	width
Andernach & Koeln	59.0	21.7	20.3	7.8	14.7	5.1	14.0	5.1	16.4	5.0
Cochem & Kaub	8.0	3.8	2.3	1.3	2.5	0.9	1.8	0.9	2.2	0.8
Cochem & Worms	5.4	3.2	1.5	1.0	1.9	0.7	1.2	0.6	1.7	0.7
Worms & Frankfurt	4.6	3.8	1.2	1.1	1.7	0.7	1.0	0.7	1.5	0.8

Table 6.5: Kendall's τ values between time series and width of the 95% confidence intervals.

Time series pair	Frank		Clayton		Gumbel 180°		Galambos 180°		Huesler-Reiss 180°	
	τ	width	τ	width	τ	width	τ	width	τ	width
Andernach & Koeln	0.93	0.02	0.91	0.03	0.93	0.02	0.93	0.02	0.93	0.02
Cochem & Kaub	0.60	0.14	0.53	0.14	0.60	0.14	0.60	0.14	0.58	0.13
Cochem & Worms	0.48	0.19	0.43	0.16	0.48	0.18	0.48	0.18	0.47	0.16
Worms & Frankfurt	0.43	0.26	0.37	0.22	0.42	0.25	0.42	0.24	0.42	0.23

pdf at the lower bound of the confidence interval, and Fig. 6.10(k) shows the difference between the pdf at $\hat{\theta}$ and the pdf at the upper bound of the confidence interval. Similar plots are shown for the other copulas.

The confidence curves can serve to explore the variation in a copula over a given confidence interval and therefore give a better insight into the effect in the area of interest. While the whole upper quadrant ($[0.5, 1] \times [0.5, 1]$) would be of interest, limitations deriving from viewing 3D information in 2D usually lead to examination of exceedance frequencies or, equivalently, return periods. The relation between return periods and copulas is discussed in Salvadori and De Michele [8]. For instance, the probability that the discharges in both rivers are in the top 10% of return periods corresponds with the integral of the pdf of the copula over the rectangle $[0.9, 1] \times [0.9, 1]$, which corresponds to the value $\overline{C}(1-0.9, 1-0.9)$ where

$$\overline{C}(u, v) = u + v - 1 + C(1-u, 1-v)) \tag{6.10}$$

which relates to a return period by $\mu_T / \overline{C}(1-0.9, 1-0.9)$ where μ_T is the the mean inter-arrival time, one year for annual maxima. We can now relate the copula to return periods for combined $\mu_T/(1-u)$, $\mu_T/(1-v)$ floods. The confidence curve allows us to pick a copula parameter range associated with a given confidence level. As a result we get a confidence interval for the return period of the combined floods.

6.4.2. DEPENDENCE STRUCTURE BETWEEN RAINFALL AND DISCHARGE FOR A KARST AREA

Rainfall onto and runoff out of a catchment are clearly related, but the relation is rarely purely causal. Unknown factors can and do cause variations in the relation. The simplest possible model would be a one in which the runoff is a shifted and scaled version of the rainfall. A slightly more complex model would assume that the joint distribution

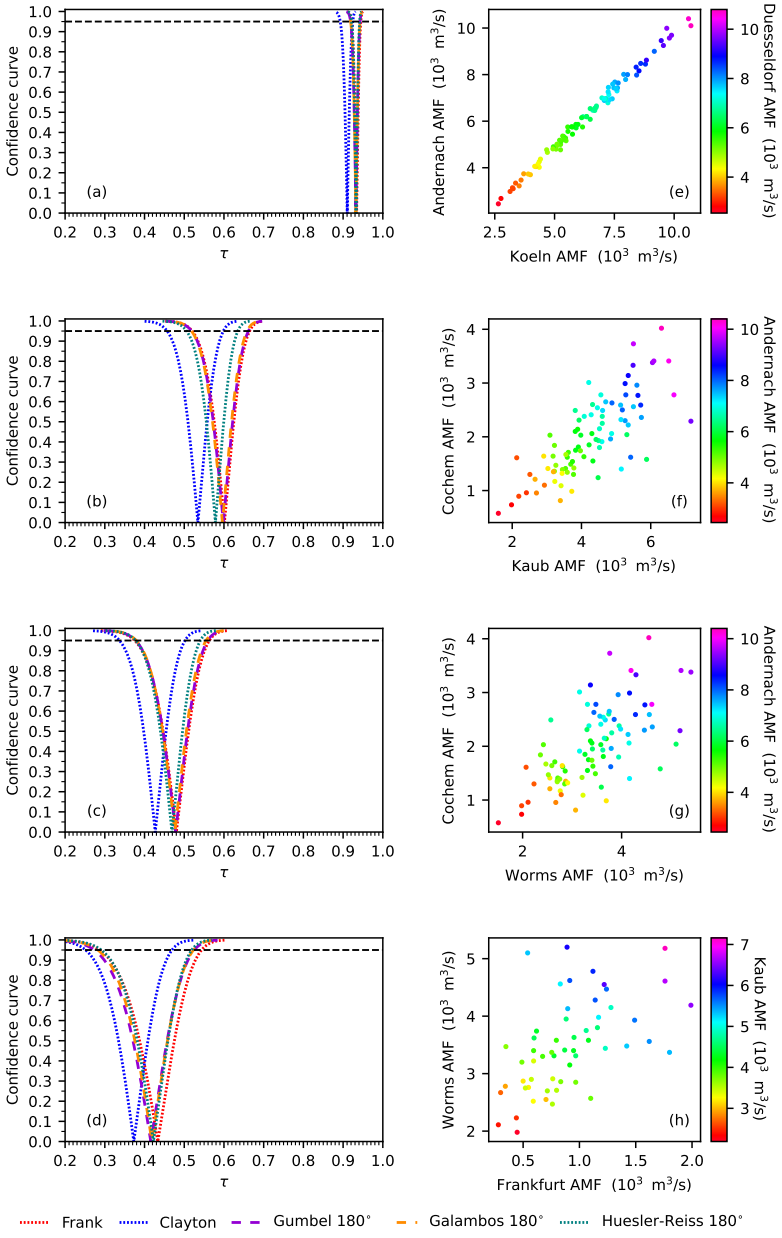


Figure 6.9: Confidence curves and scatter plots for pairs of stations, discharge at the downstream station is indicated by the colour of the dots in the scatter plots.

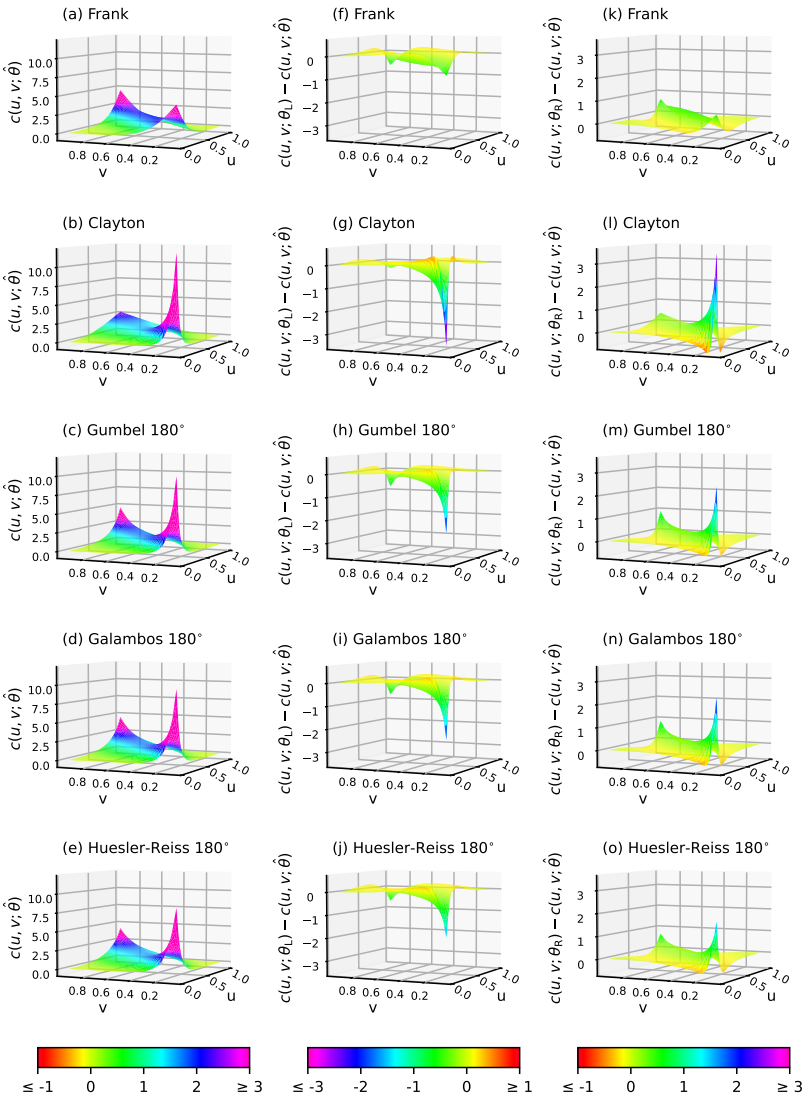
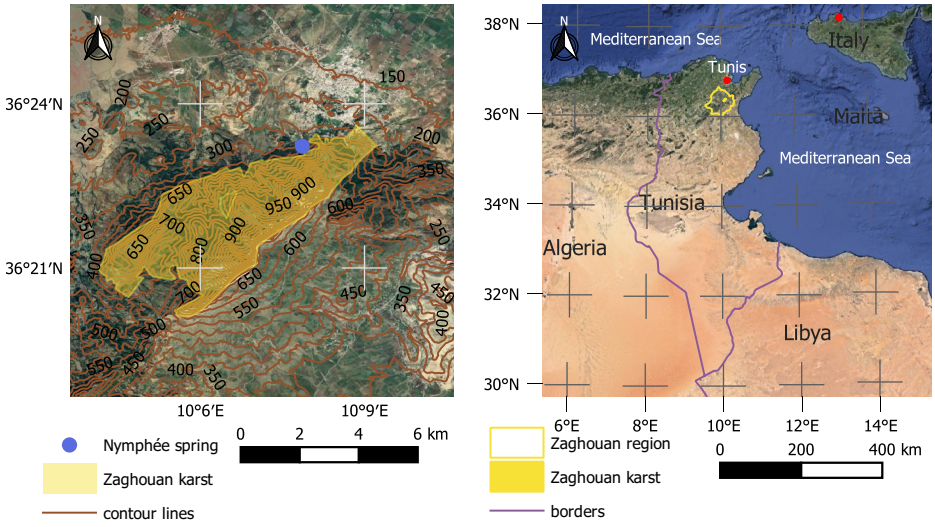


Figure 6.10: The pdfs for copulas for the station pair Cochem and Kaub.



(a) The Zaghouan district

(b) The Zaghouan karst area

Figure 6.11: The Djebel Zaghouan region

6

of a shifted version of the runoff and the rainfall has a non-trivial dependence structure. The dependence would be strongest for a shift that best corresponds to the mean delay between rainfall and runoff. To see whether such an approach can deliver useful information, it was be applied to data from the Djebel Zaghouan region in Tunisia.

GEOLOGICAL CONTEXT

The Djebel Zaghouan is the most important Jurassic formation of the Zaghouan massif and it is located at about fifty kilometers from Tunis (Tunisia). This massif is constituted from monoclinals of limestone overlapping each other. It is also contains marls of the Cretaceous and Eocene [29]. The Djebel Zaghouan is characterized by the presence of southern and transverse faults that have created individualized blocks [30]. These faults allow the infiltration. The Zaghouan karst aquifer covers an area of approximately $19.6 km^2$ (see Fig. 6.11). It has an eastern part where conditions are favourable for the storage of seepage water, whereas in its western part marl deposits strongly decrease the storage coefficient [31].

CLIMATE, HYDROLOGY, AND DATA

The Zaghouan region is located between two kinds of climate: upper semi-arid and sub-humid. It is characterized by an average annual rainfall of 467 mm with a heterogenous spatial and temporal distribution of values ranging from 245 to 625 mm. The discharge series used, which was recorded from 1915 to 1943, falls within the natural flow period. It was measured at Nymphée spring. The time series contains two years of unusual flows: a very wet one during the hydrological year running from September 1920 to August 1921 and a relatively dry one during 1926-1927. These resulted in total volumes of $6.5 \times 10^6 m^3$

and of $1.9 \times 10^6 m^3$ respectively flowing from the springs. These observations are consistent with the natural flow of the resurgences during this period. The original discharge series were recorded in graphical form with irregular time steps. To extract information, we proceeded by graphics digitalization. This digitalization allowed us to obtain a complete continuous regular discharge series on a weekly scale. The discharge on a daily and monthly scale were obtained by linear interpolation. Today, the aquifer is fully exploited to provide drinking water to the city of Zaghouan, and this overexploitation has prevented the natural resurgence of springs for decades.

METHODOLOGY

The lag between rainfall and runoff was estimated by fitting copulas to the rainfall time series and a version of the runoff series that was shifted by m steps for m ranging from 0 up to $m_{\max}$. A fixed window size was chosen equal to the length of the runoff series minus $m_{\max}$ to avoid artefacts due to different time series lengths for different lags. Different copulas were fitted to see if this made an appreciable difference in the results. The estimate of the copula parameter for lag m will be denoted by $\hat{\tau}(m)$ and the corresponding τ by $\hat{\tau}(m)$. The lag was estimated as

$$\hat{m} = \text{argmax}_m \hat{\theta}(m) \tag{6.11}$$

The pmle is asymptotically normal [20]. An estimate of the standard deviation of this normal distribution was obtained by taking the 95% confidence interval and translating this into a standard deviation for the normal distribution. For selected values this was checked against calculations using the R copula package, which gave very similar results, but took much longer to produce results and sometimes ran into problems when the procedure did not converge. This estimate can be used to get bounds on the lag as follows. First determine an interval $[m_0, m_1]$ around $\hat{m}$, for instance, the lags around $\hat{m}$ for which $\hat{\theta}(m)$ is strictly positive. Next, for all $m \in [m_0, m_1]$ draw n_R random values from the normal distribution $N(\hat{\theta}(m), \sigma(m))$ with mean $\hat{\theta}(m)$ and standard deviation $\sigma(m)$. This resulted in $n_R \times (m_1 - m_0 + 1))$ values $\theta_{i,j}$ with corresponding values $\tau_{i,j}$. For each i, the $\tau_{i,j}$ were then considered as alternatives to $\hat{\tau}(j)$ for $j = m_0, m_0 + 1, ..., m_1$; for each i, the lag corresponding to the maximum value of $\tau_{i,j}$ was labelled m_i. The cumulative frequency distribution of the m_i was then used to determine the bounds of a confidence interval for the lag. This procedure ignores the constraint that the curve should have exactly one local maximum in $[m_0, m_1]$ and is therefore probably on the conservative side.

RESULTS AND DISCUSSION

Figure 6.12 shows daily rainfall and runoff for the karst area. In addition a line is plotted representing the runoff shifted backward in time over a number of days corresponding to the estimated lag. The lag was estimated based on the Frank copula because of the results discussed in 6.3.6. The confidence curve for $\hat{\tau}(\hat{m})$ is shown in Fig. 6.13(a,c,e) for time steps of one day, one week, and one month respectively. The curve of $\hat{\tau}(m)$ versus the lag m is shown in Fig. 6.13(b,d,f) for time steps of one day, one week, and one month respectively. Please note that the Gumbel copula can only model positive correlations, so a result of zero was returned when the dependence would result in a negative τ . The

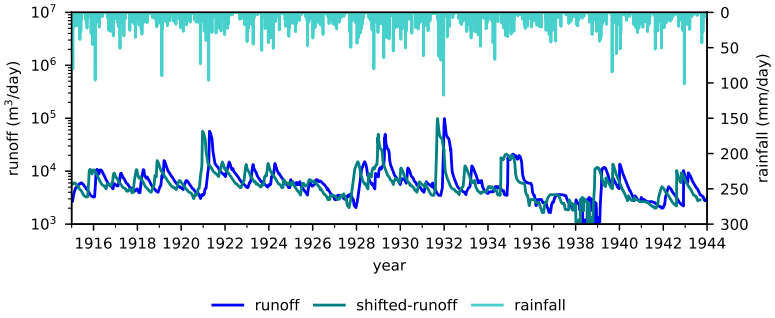


Figure 6.12: Daily rainfall, runoff, and shifted runoff.

Table 6.6: Table of lags and confidence intervals.

copula	time step	$\hat{\tau}(\hat{m})$	$\hat{m}$	90% interval	95% interval
Frank	day	0.114	115	[86.0, 131.0]	[83.0, 133.0]
	week	0.149	18	[13.0, 20.0]	[12.0, 21.0]
	month	0.233	3	[3.0, 5.0]	[2.0, 5.0]
Clayton	day	0.081	116	[85.0, 136.0]	[82.0, 141.0]
	week	0.122	18	[13.0, 21.0]	[12.0, 22.0]
	month	0.156	4	[3.0, 5.0]	[3.0, 6.0]
Gumbel	day	0.074	87	[79.0, 125.0]	[75.0, 128.0]
	week	0.140	17	[11.0, 19.0]	[11.0, 20.0]
	month	0.229	3	[2.0, 4.0]	[2.0, 5.0]

Frank copula consistently delivered the highest τ values. Table 6.6 provides values for the lags for different copulas and time steps. The lags found confirm and support the results found in rainfall-runoff modelling of the KARMA project, both with the conceptual KarstMod model [32] and with neural networks.

6.5. CONCLUSIONS AND DISCUSSION

Multivariate distributions and therefore copulas are an essential tool in modern hydrology. Given that the importance of taking into account uncertainty in univariate distribution parameters is clear, it is therefore logical to assume that the same applies to copula parameters. In this chapter, a new method was developed that uses confidence curves as a means to represent the uncertainty in the estimate of the copula parameter. Three Archimedean copulas were considered: Clayton, Frank and Gumbel. A pseudo maximum likelihood estimator (pmle) was used to estimate θ for synthetic and real data. For the copulas used here, there is a one-to-one correspondence between the copula parameter θ and Kendall's τ that respects ordering, therefore it is possible to transform parameter θ of each copula into the corresponding τ. This allows better comparison between results of different copulas and a more convenient interpretation, the parameter θ of each copula was transformed into the corresponding τ.

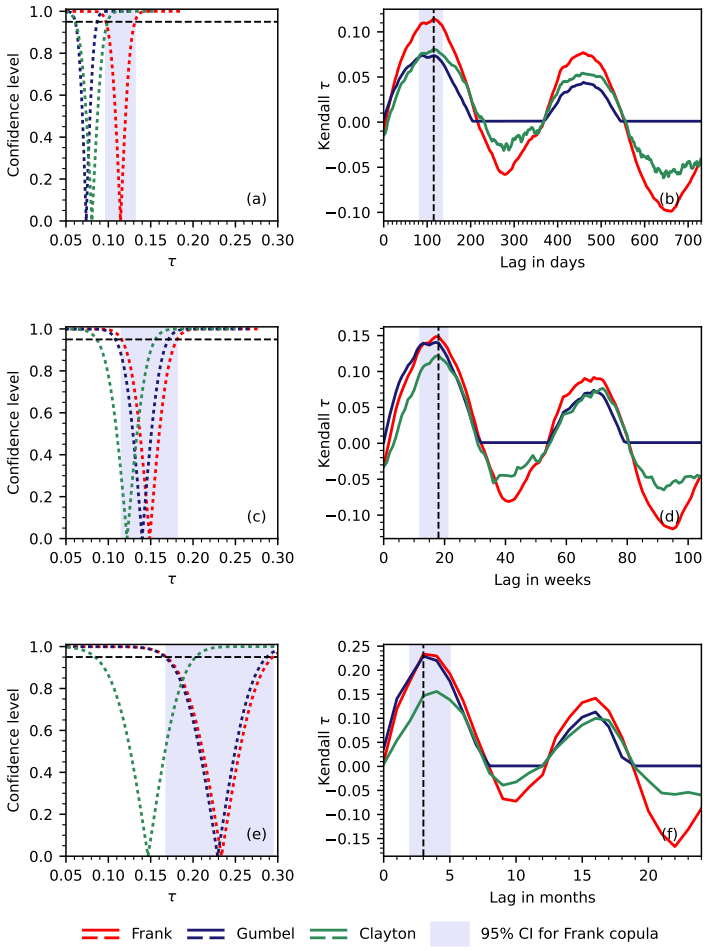


Figure 6.13: Kendall's τ for different lags and confidence curve for selected lag (CI = confidence interval).

Confidence curves have the advantage that they offer much more information than just one confidence interval. In fact they offer about the same amount of information as a posterior Bayes distribution, but without the need of first finding the correct prior. In hydrology, such information is especially important because the decisions based on hydrological analyses usually have a large impact, and the questions posed can rarely be adequately answered with a simple yes or no. A confidence curve allows an exploration of a range of answers based on different levels of confidence. A large number of experiments with synthetic time series were performed. The results showed that the pmle gave good parameter estimates and that confidence curves provided good confidence intervals for copulas with low positive Kendall's τ. For the Frank copula coverage was good up to $\tau = 0.5$; results of Clayton and Gumbel up to $\tau = 0.5$ showed increasing permissiveness of the coverage. For $\tau = 0.9$ the confidence curves for all three copula types were permissive. This may be due to the use of the chi-square distribution as an approximation to the distribution of the deviance. The actual coverage probability did not change significantly with the increase of sample length. The accuracy of the pmle of Kendall's τ improved and the spread decreased both with increasing sample length and with increasing Kendall's τ. The results for synthetic time series also showed that fitting the Frank copula to a time series generated from a Clayton or Gumbel copula gave good results for τ, while fitting Clayton or Gumbel to time series generated by another copula resulted in a biased estimate.

Two examples of applications of the method were given:

- An examination of the dependence structure of annual maximum daily flows for several pairs of measurement stations located in the German part of the Rhine River basin. That dependence structure provides joint return times for pairs of single station return times. The confidence curve then provides the uncertainty in those return times.

- An examination of the lag between rainfall and runoff for a Tunisian karst region. The confidence curves formed the basis for an estimate of a confidence interval for the lag for a given confidence level.

For the Rhine River, two results stand out. Firstly, all extreme value copulas fit best when rotated 180°, and secondly, the parameter estimate and the confidence curve for Kendall's τ delivered by the Frank copula are very close to the estimates and the confidence curves corresponding to the extreme value copulas. This suggests that the effect observed for synthetic data, namely that Frank seemed to give good results for Kendall's τ even when the time series was drawn from another copula, may well extend to real time series.

For the Tunisian karst region the pmle was mapped to a τ value and this was used to estimate the lag between rainfall and runoff. The confidence curve for the τ corresponding to the chosen lag served as an initial check on the relevance of the correlation. The confidence curve was also used to approximate the distribution of the points on the curve of the estimated $\hat{\tau}(m)$ versus the lag m. This in turn allowed generation of alternative curves and an estimate of a confidence interval for the lag. The lag found was in accordance with results of earlier research. When calculating τ for the different lags, the

Frank copula gave values that were larger in magnitude than the Clayton and Gumbel copulas. As the Gumbel copula cannot model negative correlations, this copula gave no results when the actual correlation was negative.

In both cases, the confidence curves for the copula parameter allowed simple propagation of the uncertainty in the parameter to quantities of direct hydrological significance, and in both cases, the Frank copula gave the highest estimate for Kendall's τ. All results show that confidence curves for copula parameters are a valuable addition to the hydrological tool set and can be used in a wide variety of hydrological settings. Earlier work already showed the value of confidence curves for change point analysis [33]. In certain cases, confidence curves may provide an alternative to Bayesian methods, for instance, in Cunen *et al.* [34], "it is apparent that confidence distribution by Method B based on the deviance and its distribution does a much better job than the Bayesian apparatus when it comes to delivering confidence intervals with correct coverage". Further research is planned to see whether the coverage can be corrected either by a correction factor as suggested for a more general case in Schweder and Hjort [23] or through the use of Monte Carlo simulation to generate an approximate probability distribution for the deviance.

REFERENCES

[1] G. Blöschl, M. F. Bierkens, A. Chambel, C. Cudennec, G. Destouni, A. Fiori, J. W. Kirchner, J. J. McDonnell, H. H. Savenije, M. Sivapalan, *et al.*, *Twenty-three unsolved problems in hydrology (UPH)–a community perspective,* Hydrological Sciences Journal **64**, 1141 (2019).

[2] A. Sklar, *Fonctions de répartition à n dimensions et leurs marges,* Publ. Inst. Statist. Univ. Paris **8**, 229 (1959).

[3] J. R. Córdova and I. Rodríguez-Iturbe, *On the probabilistic structure of storm surface runoff,* Water Resources Research **21**, 755 (1985).

[4] B. Bacchi, G. Becciu, and N. T. Kottegoda, *Bivariate exponential model applied to intensities and durations of extreme rainfall,* Journal of Hydrology **155**, 225 (1994).

[5] S. Yue, T. B. Ouarda, B. Bobée, P. Legendre, and P. Bruneau, *The gumbel mixed model for flood frequency analysis,* Journal of hydrology **226**, 88 (1999).

[6] F. Ashkar, N. El Jabi, and M. Issa, *A bivariate analysis of the volume and duration of low-flow events,* Stochastic Hydrology and Hydraulics **12**, 97 (1998).

[7] G. Salvadori and C. De Michele, *Frequency analysis via copulas: Theoretical aspects and applications to hydrological events,* Water Resources Research **40** (2004).

[8] G. Salvadori and C. De Michele, *On the use of copulas in hydrology: Theory and practice,* Journal of Hydrologic Engineering **12**, 369 (2007).

[9] A.-C. Favre, S. El Adlouni, L. Perreault, N. Thiémonge, and B. Bobée, *Multivariate hydrological frequency analysis using copulas,* Water Resources Research **40** (2004).

[10] C. Genest and A.-C. Favre, *Everything you always wanted to know about copula modeling but were afraid to ask,* Journal of Hydrologic Engineering **12**, 347 (2007).

[11] C. De Michele and G. Salvadori, *A Generalized Pareto intensity-duration model of storm rainfall exploiting 2-copulas,* Journal of Geophysical Research: Atmospheres **108**, n/a (2003).

[12] L. Zhang and V. P. Singh, *Gumbel–Hougaard copula for trivariate rainfall frequency analysis,* Journal of Hydrologic Engineering **12**, 409 (2007).

[13] A. Gargouri-Ellouze, Emnaand Chebchoub, *Modelling the dependence structure of rainfall depth and duration by Gumbel's copula,* Hydrological Sciences Journal **53**, 802 (2008).

[14] J. Shiau, *Fitting drought duration and severity with two-dimensional copulas,* Water Resources Management **20**, 795 (2006).

[15] H.-H. Kwon and U. Lall, *A copula-based nonstationary frequency analysis for the 2012–2015 drought in California,* Water Resources Research **52**, 5662 (2016).

[16] B. Gartsman, R. van Nooyen, and A. Kolechkina, *Implementation issues for total risk calculation for groups of sites,* Physics and Chemistry of the Earth, Parts A/B/C **34**, 619 (2009).

[17] A. Bárdossy, *Copula-based geostatistical models for groundwater quality parameters,* Water Resources Research **42** (2006).

[18] S. Debele, E. Bogdanowicz, and W. Strupczewski, *The impact of seasonal flood peak dependence on annual maxima design quantiles,* Hydrological Sciences Journal **62**, 1603 (2017).

[19] S. Grimaldi and F. Serinaldi, *Asymmetric copula in multivariate flood frequency analysis,* Advances in Water Resources **29**, 1155 (2006).

[20] C. Genest, K. Ghoudi, and L.-P. Rivest, *A semiparametric estimation procedure of dependence parameters in multivariate families of distributions,* Biometrika **82**, 543 (1995).

[21] V. Ko and N. L. Hjort, *Model robust inference with two-stage maximum likelihood estimation for copulas,* Journal of Multivariate Analysis **171**, 362 (2019).

[22] A. Birnbaum, *Confidence curves: An omnibus technique for estimation and testing statistical hypotheses,* Journal of the American Statistical Association **56**, 246 (1961).

[23] T. Schweder and N. L. Hjort, *Confidence, likelihood, probability. Statistical inference with confidence distributions.* (Cambridge: Cambridge University Press, 2016) pp. xx + 500.

[24] R. van Nooijen, C. Zhou, and A. Kolechkina, *A method to quantify the uncertainty in copula parameters when studying dependence structures of time series,* in *Proceedings of the 39th IAHR World Congress, 19-24 June 2022, Granada, Spain. In press.* (2022).

[25] M. Abramowitz, I. A. Stegun, and R. H. Romer, *Handbook of mathematical functions with formulas, graphs, and mathematical tables,* (1988).

[26] X. Chen and Y. Fan, *Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification,* Journal of econometrics **135**, 125 (2006).

[27] M. Hofert, I. Kojadinovic, M. Maechler, and J. Yan, *copula: Multivariate dependence with copulas.* R package version 1.0-1. (2020.).

[28] GRDC, *Long-term statistics and annual characteristics of grdc timeseries data / online provided by the Global Runoff Data Centre of WMO,* Koblenz: Federal Institute of Hydrology (bfg) (2021).

[29] G. Castany, *Geological study of the Tunisian eastern Atlas,* Besançon. Imp de l'IEST (1951).

[30] A. H. Ferjani, R. Guellala, S. Gannouni, and M. Inoubli, *Enhanced characterization of water resource potential in Zaghouan region, Northeast Tunisia,* Natural Resources Research **29**, 3253 (2020).

[31] M. Djebbi, M. Besbes, J. Sagna, and M. Rekaya, *Les sources karstiques de Zaghouan: Recherche d'un opérateur Pluie-Débit,* Sciences et techniques de l'environnement. Mémoire hors-série , 125 (2001).

[32] N. Mazzilli, V. Guinot, H. Jourde, N. Lecoq, D. Labat, B. Arfib, C. Baudement, C. Danquigny, L. Dal Soglio, and D. Bertin, *Karstmod: A modelling platform for rainfall-discharge analysis and modelling dedicated to karst systems,* Environmental Modelling & Software **122**, 103927 (2019).

[33] C. Zhou, R. van Nooijen, A. Kolechkina, and N. van de Giesen, *Confidence curves for change points in hydrometeorological time series,* Journal of Hydrology **590**, 125503 (2020).

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

7

Conclusions

Quiet people have the loudest minds.
Stephen Hawking (1942 2018)

7.1. KNOWLEDGE GENERATED

To conduct uncertainty analysis by constructing confidence curves for the parameter of interest, in this PhD research traditional time series analysis based on NHST methods have been explored.

7.1.1. TRADITIONAL NHST BASED CHANGE POINT DETECTORS

Traditional time series analysis depends on NHST methods, and the results of change point detection are often a 'Yes' or 'No' answer to accept the null hypothesis at a given significance level. There are two types of change point (CP) detectors, parametric and non-parametric detectors. Compared to parametric detectors, nonparametric ones do not need the assumption that the parametric distributions of hydrological observations are known. Therefore, three non-parametric detectors were considered: Pettitt's, Cramér von Mises (CvM) and CUSUM tests, and their performances were analyzed and compared according to the properties of each detector. From the power and ability of the three detectors we conclude that the CvM test outperforms its two counterparts.

However, when different start or end year of a time series is considered, traditional detectors tend to give different locations of a CP. Clearly, NHST methods leave no room for further uncertainty analysis. This calls for methods to represent the uncertainty in finding the location of a CP.

7.1.2. CONSTRUCTING CONFIDENCE CURVES FOR THE LOCATION OF A CHANGE POINT

To represent uncertainty in finding a change point, confidence curves should be constructed. Previous studies provided a parametric method to construct Confidence curves based on profile likelihood, deviance function and Monte Carlo simulations. All parameters are estimated by Maximum Likelihood estimate (CML) to construct confidence curves. In this research, two new methods were developed by using pseudo maximum likelihood estimator (pmle), for instance Linear Moments method (LMo) and Method of Moments (MoM), instead of standard maximum likelihood estimator (mle) for estimating the nuisance parameters of the log-likelihood function (CLMo/CMoM). With the profile log-likelihood function, a parameter of interest can be estimated and a confidence curve for the parameter can be constructed based on a deviance function and Monte Carlo simulation. The second one is a confidence curve based on the Approximate Empirical log-likelihood ratio, Deviance function and bootstrap (AED).

From the results of two new methods (CLMo/CMoM and AED) and the CML, the confidence curves constructed by the three methods have comparable performances. The two new methods are simpler and more efficient than CML. With confidence curves, not only the potential location of a CP can be seen, but also confidence sets at each confidence level can be extracted. The methods were applied to real hydro-meteorological time series, and the uncertainty in finding the location of a CP could graphically be represented.

The methods proposed in this research could provide some new insights into CP detection and uncertainty representation for discrete parameters.

7.1.3. CONSTRUCTING CONFIDENCE CURVES FOR THE DEPENDENCE PARAMETER IN COPULAS

Also, confidence curves for continuous parameters were constructed in this research. For this, the dependence parameter in copulas were considered. Copulas are widely used to describe multivariate phenomena, that can be commonly found in hydrological processes. According to Wilks' theorem, the probability distribution of the deviance function based on observations for continuous parameters can be approximated by a chi-square (χ_1^2) distribution. Therefore, the confidence curve for the dependence parameter can be constructed by calculating the probability distribution of deviance function with a χ_1^2 distribution.

The use of confidence curves for the dependence parameter in copulas can be helpful for studies that concentrate on the uncertainty analysis of continuous parameter estimation.

7.1.4. ANALYSING PROPERTY OF A CONFIDENCE CURVE AND SIMILARITY BETWEEN CONFIDENCE CURVES

The uncertainty in time series represented by confidence curves can be read and extracted, but some metrics are still needed to measure the properties of a confidence curve and measure the similarity between confidence curves. When the parameter of interest is discrete, for instance the location of a CP, the property of a confidence curve can be measured by the actual versus nominal coverage probability, the actual found location of a CP when the null hypothesis holds, the actual found location of a CP when the null hypothesis fails, and the uncertainty measure (Un) of confidence curves.

When the parameter of interest is continuous, such as the dependence parameter in copulas, the property of a confidence curve can be measured by the actual versus nominal coverage probability, the estimated dependence parameter versus the true dependence parameter, and the width of confidence intervals with a given confidence level.

The similarity between confidence curves can be measured by the similarity index, which shows the overlap between two curves. The measure considered in this research could provide some guidance for studies which need to consider the properties of curves. The uncertainty measure (Un) can explicitly show the uncertainty of CP detection by a given method, and if the value of uncertainty is close to 1, it indicates there is no useful information about the location of CP.

7.2. LIMITATIONS AND RECOMMENDATION FOR FUTURE RESEARCH

7.2.1. LIMITATIONS

The confidence curve for a parameter of interest utilized, as proposed in this PhD dissertation, relies heavily on parameter estimation and sampling from relative short subseries. This will be problematic because parameter estimation and sampling are often affected by the sample length. For a time series with a short sample length, the confidence curves for the location of CP might have large uncertainty. For instance, when a time series is short, and if bootstrapping is considered to draw new samples from subseries, there will be many duplicated samples.

The CP is detected for the change in the sample mean, which is not often true in

reality. The change may occur in standard deviation, or other statistical characteristics.

7.2.2. RECOMMENDATION FOR FUTURE RESEARCH

CONFIDENCE CURVES FOR QUANTILES WITH A GIVEN RETURN PERIOD

Quantiles are the magnitude of a design flood or rainfall, and they are of great importance in hydological frequency analysis. A design flood/precipitation T−year return period is a continuous parameter, and there are many methods to estimate it, including parametric or non-parametric methods. If a log-likelihood function for the T−year quantile is derived, then a confidence curve for the quantile can be approximated by following the same steps as for constructing confidence curves for the dependence parameter in copulas. With confidence curves for quantiles, the uncertainty of frequency analysis can be graphically represented by confidence curves. This could provide some insights into the uncertainty of design quantiles. For instance confidence intervals for a 100-year annual maximum design quantile can be extracted from a confidence curve with specific confidence levels, therefore a confidence curve will be very informative about the uncertainty of the final design quantile.

CONFIDENCE CURVES FOR PARAMETERS IN PARAMETRIC DISTRIBUTIONS

Parameter estimation in parametric models plays an indispensable role in statistical models, and the estimates of parameters based on observations determine the output of the model. If a deviance function can be derived for parameters, with simulations or referring to Wilks' theorem, approximate confidence curves for parameters can be constructed.

Parametric distributions are the basis of hydrological frequency analysis, and parameters in distributions are blocks to estimate the quantile. Compared to quantiles, parameters in parametric distributions are often taken as nuisance parameters, but their estimates will greatly influence the estimation of quantiles directly. Therefore, conducting uncertainty analysis to nuisance parameters can provide a better understanding to the uncertainty in quantile estimation.

CONFIDENCE CURVES AND HYPOTHESIS TESTING STATISTICS

The confidence curves can also be constructed by using hypothesis testing statistics, for instance the 'Method A' in Cunen *et al.* [1]. If a statistics for the hypothesis testing problem and a confidence level are given, one can use sampling to rebuild confidence sets for the null hypothesis with the given confidence level. According to Schweder and Hjort [2], a confidence curve is a collection of confidence sets at all confidence levels.

CONFIDENCE CURVES IN OTHER FIELDS

As a statistical tool, a confidence curve is based on a frequentist framework and it can be used in many other fields to analyze the distribution of any parameter of interest, for instance, in financial, and medical fields.

CONFIDENCE CURVES FOR CHANGING WORLD

In a changing world, the hydrological processes are intensified by human activities and climate change. The uncertainty in the prediction for future events should be considered

seriously. Therefore, it is useful to analyze uncertainty before taking measures to adapt to the environment. It would be more informative to quantify the uncertainty in a graphical way, for instance in the form of confidence curves.

REFERENCES

[1] C. Cunen, G. Hermansen, and N. L. Hjort, *Confidence distributions for change-points and regime shifts,* Journal of Statistical Planning and Inference **195**, 14 (2018).

[2] T. Schweder and N. L. Hjort, *Confidence, likelihood, probability. Statistical inference with confidence distributions.* (Cambridge: Cambridge University Press, 2016) pp. xx + 500.

A

NOTATIONS

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

$$\Pr(E)$$

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.

The *probability density function* (pdf) of a member of the distribution family will be referred to as $f(\cdot;\cdot)$, and the *cumulative distribution function* (cdf) will be denoted by $F(\cdot;\cdot)$.

A.1. INDICATOR FUNCTION

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. For a set A it is traditionally written as $\mathbf{1}_A$ and defined by

$$\mathbf{1}_A(x) = \begin{cases} 0 & x \notin A \\ 1 & x \in A \end{cases}$$

There is a simpler and more general approach to translating expressions such as $x \in A$ that evaluate to true or false into numbers. It was proposed by Knuth [1], who in turn cited Iverson [2] 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 $[\![\cdot]\!]$ are used.

Examples are:

$$\begin{aligned}
[\![1 \le 4 \le 3]\!] &= 0 \\
[\![1 \le 2 \le 3]\!] &= 1 \\
[\![1 \le x \le 3]\!] &= \begin{cases} 0 & x < 1 \\ 1 & 1 \le x \le 3 \\ 0 & x > 3 \end{cases}
\end{aligned} \tag{A.1}$$

The indicator function of a set A applied to a variable x can now be written as $[\![x \in A]\!]$. The empirical cumulative distribution function (ecdf) is based on the set membership functions for sets of the form $\{r \in \mathbb{R} : r \le t\}$. Bracket notation simplifies the notation of such functions

$$\mathbf{1}_{\{r \in \mathbb{R}: r \le t\}}(x) = [\![x \in A]\!]$$

With this notation the ecdf F_n of a random sample of size n can be written as

$$F_n(t) = \frac{1}{n} \sum_{i=1}^{n} [\![X_i \le t]\!]$$

Please note that for each fixed value of t the expression $F_n(t)$ is itself a random variable.In this study the properties of statistical methods will be examined as follows. A given method for change point detection will be applied to a large number of time series generated pseudo-randomly. Certain aspects of these outcomes, for instance, a point estimate of the change point location or the width of a specific confidence interval will then be calculated for each time series. The results will be reported graphically by plots of the cumulative frequency of, for instance, the reported change point. These cumulative frequencies will be reported in percentages of the total number of observations.

A.2. SIGN FUNCTION

The sign function is defined as

$$\operatorname{sgn}(x) = \begin{cases} -1 & if \quad x < 0 \\ 0 & if \quad x = 0 \\ 1 & if \quad x > 0 \end{cases} \tag{A.2}$$

The sign function is used to define the Pettitt statistic [3]. Note that the sign function can be expressed in terms set membership:

$$\operatorname{sgn}(X_i - X_j) = [\![X_j \le X_i]\!] - [\![X_i \le X_j]\!] \tag{A.3}$$

B

TRADITIONAL CHANGE POINT STATISTICS AND SENSITIVITY TO SCALE CHANGES

B.1. CHANGE-POINT STATISTICS UNDER SCALING AND SHIFTING

For CvM-CP, the calculation of the change point statistic of a sample $(x_1, x_2, \ldots, x_n)$ depends only on the values of $[\![x_i \leq x_j]\!]$ for all pairs $i, j = 1, 2, \ldots, n$ with $i \neq j$. Shifting the entire sample does not change the value of these expressions, and neither does scaling the entire sample by a strictly positive value. As a result, the value of the statistic does not change if we shift and scale the entire sample. For Pet-CP we can use (A.3) to express the sign function, and then the same reasoning holds. For CUSUM-CP the calculation of the change point statistic of a sample depends only on $[\![c \leq x_j]\!]$ for all $j = 1, 2, \ldots, n$ and c the sample median. Again, shifting the entire sample does not change the value of this function, and neither does scaling the sample by a strictly positive value. As a result, the value of the statistic does not change if we shift and scale the entire sample.

Now, suppose that the random variables in the time series are from the same distribution family, and that this family is a location-scale family $F(\xi, \zeta)$, with location parameter ξ and scale parameter ζ. In that case $X_h = \zeta_h H_h + \xi_h$, with Y_h the independent identically distributed (*i.i.d.*) random variables for $h = 1, 2, \ldots, n$. We see that, for all three test statistics, the statistics for a series where X_i has parameters (ξ_L, ζ_L) for $i \leq \tau$ and (ξ_R, ζ_R) for $i > \tau$ is equivalent to a series with location zero and scale one up to τ, but location $(\xi_R - \xi_L)/\zeta_L$ and scale ζ_R/ζ_L beyond that point. This implies that, for a location scale family, the distribution of the test statistic, when a change point is present, depends only on the properties of Y_h and the quantities $(\xi_R - \xi_L)/\zeta_L$ and ζ_R/ζ_L For the normal distribution, the mean is the location parameter, and the standard deviation is the scale parameter.

For the GEV distributions and a change in the mean, the distribution of the test statistic when a change point is present will depend only on $(\mu_R - \mu_L)/\sigma_L$. If there is a change in the standard deviation while the mean value stays the same, then this corresponds to a change in both the scale and the location of the original distribution. After scaling, it turns out the change in the location is constant, and the change in distribution depends on this constant and σ_R/σ_L.

B.2. SENSITIVITY OF THE PETTITT TEST STATISTIC TO SCALE CHANGES

Suppose that the random variables in the time series are from a location-scale family that is symmetric with respect to the median, such as the normal distribution. In that case, it is possible to show that the probability distribution of the sign function for the difference of two of different random variables taken from the series does not depend on the scale. This can be done as follows.

Suppose $i \neq j$ and that at the change point only the scale changes. Shifting all random variables in the series to place the median of at zero does not change the distribution of any of the random variables. Now, for $i, j \leq \tau$ or $i, j > \tau$, we have $f_i = f_j$, so:

$$\begin{aligned}
&\Pr\{S_{ij} = 1\} = \Pr\{X_i \leq X_j\} = \int_{x_j=-\infty}^{\infty} \int_{x_i=-\infty}^{x_j} f_i(x_i) f_j(x_j) dx_i dx_j \\
&= \int_{x_j=-\infty}^{\infty} \int_{x_i=-\infty}^{x_j} f_i(x_j) f_i(x_j) dx_i dx_j \\
&= \int_{x_j=-\infty}^{\infty} f_i(x_j) \int_{y=0}^{F(x_j)} dy dx_j \\
&= \int_{x_j=-\infty}^{\infty} f_i(x_j) F_i(x_j) dx_j \\
&= \int_{z=0}^{1} z dz = \tfrac{1}{2}
\end{aligned}$$

For $i \leq \tau < j$ (similar reasoning holds for $j \leq \tau < i$) the following holds:

$$\Pr\{S_{ij} = 1\} = \Pr\{X_i \leq X_j\} = \int_{x_j=-\infty}^{\infty} \int_{x_i=-\infty}^{\infty} [\![X_i \leq X_j]\!] f_i(X_i) f_j(X_j) dX_i dX_j$$

We split the integration into the four quadrants to obtain:

$$\begin{aligned}
&\Pr\{S_{ij} = 1\} = \int_{x_j=0}^{\infty} \int_{x_i=0}^{x_j} [\![X_i \leq X_j]\!] f_i(x_i) f_j(x_j) dx_i dx_j \\
&+ \int_{x_j=-\infty}^{0} \int_{x_i=-\infty}^{0} [\![X_i \leq X_j]\!] f_i(x_j) f_i(x_j) dx_i dx_j \\
&+ \int_{x_j=-\infty}^{0} \int_{x_j=-\infty}^{x_i=0} [\![X_i \leq X_j]\!] f(x_i) f_j(x_j) dx_i dx_j \\
&+ \int_{x_j=0}^{\infty} \int_{x_j=-\infty}^{0} [\![X_i \leq X_j]\!] f(x_j) F_i(x_j) dx_j
\end{aligned}$$

For all x_i and x_j within the integration bounds of the fourth integral, the function $[\![X_i \leq X_j]\!]$ in the integrand equals equals one. In the third integral on the right hand side $[\![X_i \leq X_j]\!]$ equals zero. This allows us to write:

$$\begin{aligned}
&\Pr\{S_{ij} = 1\} = \int_{x_j=0}^{\infty} \int_{x_i=0}^{\infty} [\![X_i \leq X_j]\!] f_i(x_i) f_j(x_j) dx_i dx_j \\
&+ \int_{x_j=-\infty}^{0} \int_{x_i=-\infty}^{0} [\![X_i \leq X_j]\!] f_i(x_j) f_i(x_j) dx_i dx_j \\
&+ \int_{x_j=0}^{\infty} \int_{x_j=0}^{\infty} f_i(x_j) f_i(x_j) dx_i dx_j
\end{aligned}$$

Next, we introduce a new integration variable $y_i = -x_i$ whenever there is a negative integration boundary:

$$\begin{aligned}\Pr\{S_{ij}=1\} &= \int_{x_j=0}^{\infty}\int_{x_i=0}^{\infty}[\![X_i \le X_j]\!]\, f_i(x_i) f_j(x_j)\, dx_i\, dx_j \\ &+ \int_{y_j=0}^{\infty}\int_{y_i=0}^{\infty}[\![-y_i \le -y_j]\!]\, f_i(-y_i) f_j(-y_j)\, dy_i\, dy_j \\ &+ \int_{x_j=0}^{\infty}\int_{y_i=0}^{\infty} f_i(-y_i) f_j(x_j)\, dy_i\, dx_j\end{aligned}$$

We use symmetry around zero to replace $f_i(-y_i)$ by $f_i(y_i)$ in the second and third integrals and rewrite the inequality in the second integral to obtain:

$$\begin{aligned}\Pr\{S_{ij}=1\} &= \int_{x_j=0}^{\infty}\int_{x_i=0}^{x_j}[\![X_i \le X_j]\!]\, f(x_i) f_j(x_j)\, dx_i\, dx_j \\ &+ \int_{y_j=0}^{\infty}\int_{y_i=0}^{\infty}[\![y_i \le y_j]\!]\, f(y_i) f_j(y_j)\, dy_i\, dy_j \\ &+ \int_{x_j=0}^{\infty}\int_{y_i=0}^{\infty} f_i(y_i) f_j(x_j)\, dy_i\, dx_j\end{aligned}$$

We then rename the integration variables y_i and y_j to obtain:

$$\begin{aligned}\Pr\{S_{ij}=1\} &= \int_{x_j=0}^{\infty}\int_{x_i=0}^{\infty}[\![X_i \le X_j]\!]\, f(x_i) f_j(x_j)\, dx_i\, dx_j \\ &+ \int_{y_j=0}^{\infty}\int_{y_i=0}^{\infty}[\![x_i \le x_j]\!]\, f_i(x_i) f_j(x_j)\, dx_i\, dx_j \\ &+ \int_{x_j=0}^{\infty}\int_{y_i=0}^{\infty} f_i(y_i) f_j(x_j)\, dy_i\, dx_j\end{aligned}$$

By combining the first and second integral we obtain:

$$\begin{aligned}\Pr\{S_{ij}=1\} &= \int_{x_j=0}^{\infty}\int_{x_i=0}^{x_j}[\![X_i \le X_j]\!]\, f_i(x_i) f_j(x_j)\, dx_i\, dx_j \\ &+ \int_{x_j=0}^{\infty}\int_{y_i=0}^{\infty} f_i(y_i) f_j(x_j)\, dy_i\, dx_j\end{aligned}$$

By symmetry, both remaining integrals equal $\frac{1}{4}$, so $\Pr\{S_{ij}=1\} = \frac{1}{2}$ irrespective of the change in scale. While this does not prove that the distribution of the test statistic is independent of the scale change, it does indicate that any recoverable information on a change in scale can only be in the correlation structure between the S_{ij}.

B.3. THE EFFECT OF SCALING OF SHIFTING OR SCALING THE TIME SERIES ON THE CONFIDENCE CURVE

B.3.1. LOCATION-SCALE DISTRIBUTION FAMILIES

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

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

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 AED based on MoM and LMo methods, a similar equality holds for D_{pseu} under the condition that the estimates of the parameters satisfy

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

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

If $D(\tau, ay+b) = D(\tau, y)$, then tests on synthetic time series with $\theta_{\text{L}} = (0;1)$ while varying θ_{R} are representative for the performance of the method.

B.3.2. DISTRIBUTION FAMILIES WITH A SCALE PARAMETER

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

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

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 CLMo(CMoM) method a similar equality holds for D_{pseu} under the condition that the estimates of the parameters satisfy

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

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

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

C

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.* [4] 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.

C.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 [5]. The likelihood ratio for the AMOC problem is given by

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

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 [5] state that it is now natural to consider

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

and reject the null hypothesis of no change when this is large. From (3.10) and (C.1) it follows that

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

or, using (3.11),

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

The value of

$$\sup_{\theta,\zeta}\ell(\tau,\theta,\theta,\zeta;y)$$

is independent of τ, so

$$\begin{aligned}&-2\log\Lambda_{\hat\tau(y)}(y) + 2\log\Lambda_\tau(y)\\ &= 2\left(\ell_{\text{prof}}(\hat\tau(y);y) - \sup_{\theta,\zeta}\ell(\hat\tau(y),\theta,\theta,\zeta;y)\right)\\ &-2\left(\ell_{\text{prof}}(\tau;y) - \sup_{\theta,\zeta}\ell(\tau,\theta,\theta,\zeta;y)\right)\\ &= 2\left(\ell_{\text{prof}}(\hat\tau(y);y) - \ell_{\text{prof}}(\tau;y)\right) = D(\tau,y)\end{aligned} \tag{C.3}$$

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.

C.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 [6, 7]. For a change point in the mean, such an approach is presented, for instance, in Zou *et al.* [8] and Shen [9]. In Hall and La Scala [10] 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

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

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

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

The empirical likelihood ratio is derived by dividing L by

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

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

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

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

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

and

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

with likelihood ratio

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

For the change point problem with a change in the mean, Zou *et al.* [8] proposed the empirical likelihood ratio

$$\Lambda_{\text{emp}}(\tau; y) = \frac{\sup_{p \in \mathscr{S}_\tau} \left\{ \prod_{i=1}^{n} p_i : \sum_{i=1}^{\tau} p_i y_i = \sum_{i=\tau+1}^{n} p_i y_i \right\}}{\sup_{p \in \mathscr{S}_\tau} \prod_{i=1}^{n} p_i} \tag{C.4}$$

where

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

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 (C.4) 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

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

and their statistic is

$$Z_{\text{emp}} = \max_{1 \le \tau < n} \ell_{\text{emp}}(\tau; y)$$

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

$$D_{\text{emp}}(\tau, y) = 2\left(\ell_{\text{emp}}(\hat{\tau}_{\text{emp}}(y); y) - \ell_{\text{emp}}(\tau; y)\right)$$

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

$$K_{\text{emp},\tau}(r) = \Pr\left(D_{\text{emp}}(\tau, Y) < r\right)$$

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 [9] derived the following approximation formula for the logarithm of the empirical likelihood ratio for scalar y_i

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

for $n_{\min} < \tau < n - n_{\min}$ where $n_{\min}$ 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 $O_p\left(n_{\min}^{-1/2}\right)$ is present because the approximation does not hold for τ near the start or the end of the series. We will use $n_{\min}$ as given by Chapter 3. For the distributions used in the tests in this research the condition on the third moment is always satisfied for the log-normal and the gamma distribution; for Fréchet as parametrized in (F.40) 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 [5, page 85].

D

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 [11] who in turn refer to Jaccard [12]. For two sets, $A = \{a_1, a_2, ..., a_{n_A}\}$ and $B = \{b_1, b_2, ..., b_{n_B}\}$, the Jaccard index is given by

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

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

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

$$R_\gamma \cap R'_\gamma = \{\tau : \max(\mathrm{cc}(\tau, y_{\mathrm{obs}}), \mathrm{cc}'(\tau, y_{\mathrm{obs}})) \le \gamma\}$$

and

$$R_\gamma \cup R'_\gamma = \{\tau : \min(\mathrm{cc}(\tau, y_{\mathrm{obs}}), \mathrm{cc}'(\tau, y_{\mathrm{obs}})) \le \gamma\}$$

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

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

Table D.1: Quantiles of similarity index $\tilde{J}$ for random pairs of curves for different sample lengths.

γ \ 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

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

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

and

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

so

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

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

$$\tilde{J} = \frac{\sum_{\tau=1}^{n} \min\left(1 - \mathrm{cc}(\tau, y_{\mathrm{obs}}), 1 - \mathrm{cc}'(\tau, y_{\mathrm{obs}})\right)}{\sum_{\tau=1}^{n} \max\left(1 - \mathrm{cc}(\tau, y_{\mathrm{obs}}), 1 - \mathrm{cc}'(\tau, y_{\mathrm{obs}})\right)} \tag{D.2}$$

which will be used to compare the similarity of pairs of confidence curves. It is similar to the Ružička index [11]. 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 (D.2) relates to similarity, the following experiment was performed. For $n = 10, 20, \ldots, 100$ 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 $\tilde{J}$ for these pairs provides some indication of the range of values of $\tilde{J}$ that may occur for curves that were constructed to be unrelated to each other.

Figure D.1 shows the cumulative frequency distribution of $\tilde{J}$ for different sample lengths n. For a larger n, the distribution of $\tilde{J}$ approaches a step function. Figure D.2 shows the 90%, 95%, and 99% quantiles for $\tilde{J}$ as a function of sample size. To aid in the interpretation of Fig. D.2, Table D.1 is provided. For instance, if we have two random data sets with sample length $n = 100$, then the 95% quantile of the similarity index $\tilde{J}$ is 0.55. The actual similarity index $\tilde{J}_{\mathrm{actual}}$ between two confidence curves with a sample length $n = 100$ follows from (D.2). If $\tilde{J}_{\mathrm{actual}}$ is higher than 0.55, then we can be reasonably confident that the curves are similar.

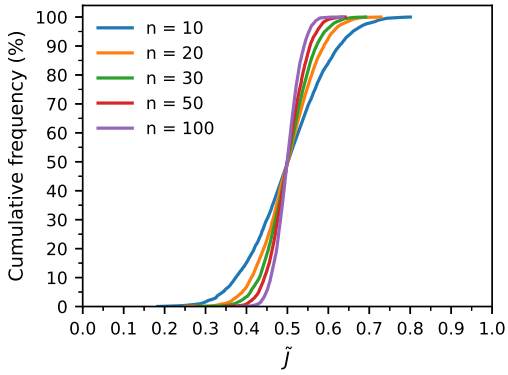


Figure D.1: Cumulative frequency for values of similarity index $\tilde{J}$ for pairs of randomly generated curves for different sample lengths.

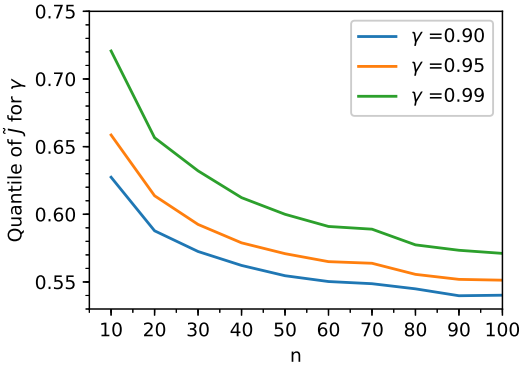


Figure D.2: Quantiles for values of similarity index $\tilde{J}$ for pairs of randomly generated curves for different sample lengths.

E

A MEASURE OF TOTAL UNCERTAINTY FOR A CONFIDENCE CURVE

To have a reference for the size of confidence sets for change points 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 $0 \leq \gamma \leq 1$, the following equations hold

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

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

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

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

$$S_{\text{CP}} = \{n_{\min}, n_{\min}+1, \ldots, n-n_{\min}\} \tag{E.3}$$

where $n_{\min}$ is the minimum sub-series length, it is easy to define random sets such that the probability that the true CP lies in the set is approximately, independently of the properties of the sample. Simply take

$$\Re_\gamma = \text{Choice}\left(\lceil \gamma(n-2n_{\min}+1) \rceil ; S_{\text{CP}}\right) \tag{E.4}$$

which has a coverage probability of

$$\Pr(\tau_{\text{true}} \in \Re_\gamma(X)) = \frac{\lceil \gamma(n-2n_{\min}+1) \rceil}{n-2n_{\min}+1} \geq \gamma \tag{E.5}$$

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

$$\text{Un}(R_\gamma) = \frac{\#R_\gamma - 1}{\gamma(n - 2n_{\min})} \tag{E.6}$$

Now for large n, the value of $\text{Un}(R_\gamma)$ is zero for a one point set, approximately one for a realization of R_γ 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 confidence curve a level needs to be selected. The highest confidence level for which a non-trivial R_γ can be constructed for which the equals sign holds is

$$\gamma_{\max} = \frac{n - 2n_{\min}}{n - 2n_{\min} + 1} \tag{E.7}$$

so that is the level that will be used. We take

$$\text{Un}(\text{cc}) = \frac{\sum_{k=n_{\min}}^{n-n_{\min}} [\![\text{cc}(k) \leq \gamma_{\max}]\!] - 1}{n - 2n_{\min}} \tag{E.8}$$

F

PARAMETRIC DISTRIBUTION FUNCTIONS AND ESTIMATORS

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

F.1. PARAMETER ESTIMATES: METHOD OF MOMENTS AND LINEAR MOMENTS METHOD

The definitions of the mean, the variance, and their unbiased estimators are well-known, which is often called Method of Moments (MoM). The same might not hold for the definitions of the Linear-moments (L-moments) and their estimators. Therefore, the definitions for the mean, the variance, the first two L-moments, and their estimators are included here.

$$\mu_X = \mathsf{E}[X] \tag{F.1}$$

and the unbiased estimator for a sample X_i, $i = 1, 2, \ldots, n$ is

$$\hat{\mu}(X) = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{F.2}$$

The variance is

$$var_X = \mathsf{E}\left[\left(X - \mu_X\right)^2\right] \tag{F.3}$$

with unbiased estimator for a sample X_i, $i = 1, 2, \ldots, n$

$$\widehat{var}(X) = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \hat{\mu}(X)\right)^2 \tag{F.4}$$

The estimates for a given sample $x_1, x_2, \ldots, x_n$ are

$$\hat{\mu}(x) = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{F.5}$$

$$\widehat{var}(x) = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \hat{\mu}(x)\right)^2 \tag{F.6}$$

We will also use the notation

$$\sigma_X = \sqrt{var_X} \tag{F.7}$$

$$\hat{\sigma}(x) = \sqrt{\widehat{var}(x)} \tag{F.8}$$

Note that for $y_i = ax_i + b$ the relations $\hat{\mu}(y) = a\hat{\mu}(x) + b$ and $\hat{\sigma}(y) = a\hat{\sigma}(x)$ hold. In connection with the behaviour of the Gumbel distribution, two more concepts are of interest, namely the skewness [13] and excess kurtosis [14]. For the Gumbel distribution, these two statistics are constant and independent of the parameters. Usually skewness refers to the standardized third moment,

$$\mathsf{E}\left[\left(\frac{X-\mu_X}{\sigma_X}\right)^3\right] \tag{F.9}$$

kurtosis refers to

$$\mathsf{E}\left[\left(\frac{X-\mu_X}{\sigma_X}\right)^4\right] \tag{F.10}$$

and excess kurtosis refers to

$$\mathsf{E}\left[\left(\frac{X-\mu_X}{\sigma_X}\right)^4\right] - 3 \tag{F.11}$$

F

Skewness is usually estimated by

$$\frac{n}{(n-1)(n-2)}\sum_{i=1}^{n}\left(\frac{x_i - \hat{\mu}(x)}{\hat{\sigma}(x)}\right)^3$$

while excess kurtosis is often estimated by

$$\frac{n(n+1)}{(n-1)(n-2)(n-3)}\sum_{i=1}^{n}\left(\frac{x_i - \hat{\mu}(x)}{\hat{\sigma}(x)}\right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)}$$

[15].

In Hosking [16], L-moments, a form of power weighted moments, are defined in terms of the order statistics $X_{1:n} \le X_{2:n} \le \cdots \le X_{n:n}$ of a random sample of size n drawn from X. The r-th L-moment is given by

$$\lambda_r = r^{-1}\sum_{k=0}^{r-1}(-1)^k\binom{r-1}{k}\mathsf{E}\left[X_{r-k:r}\right] \tag{F.12}$$

so

$$\lambda_1 = \mathsf{E}\left[X_{1:1}\right] = \mu \tag{F.13}$$

$$\lambda_2 = \frac{1}{2}\left(\mathsf{E}\left[X_{2:2}\right] - \mathsf{E}\left[X_{1:2}\right]\right) \tag{F.14}$$

The estimates for the first two L-moments for a given sorted sample $x_{1:n}, x_{2:n}, \ldots, x_{n:n}$ are

$$l_1(x) = \sum_{i=1}^{n} x_{i:n} \tag{F.15}$$

$$l_2(x) = \frac{1}{2}\frac{1}{\binom{n}{2}}\sum_{i=1}^{n}\sum_{j=1}^{i-1}\left(x_{i:n} - x_{j:n}\right) \tag{F.16}$$

Note that for $y_i = ax_i + b$ the relations $l_1(y) = al_1(x) + b$ and $l_2(y) = al_2(x)$ hold.

F.1.1. THE GUMBEL DISTRIBUTION

For the Gumbel distribution, the pdf is

$$f(x;\xi,\varsigma) = \frac{1}{\varsigma}\exp\left(-\frac{x-\xi}{\varsigma} - \exp\left(-\frac{x-\xi}{\varsigma}\right)\right) \tag{F.17}$$

where ς is the scale parameter, and ξ is the location parameter. The mean μ and variance σ^2 are

$$\mu = \xi + \varsigma\gamma_{\text{EM}};\ \sigma^2 = \frac{\pi^2}{6}\varsigma^2 \tag{F.18}$$

where γ_{EM} (≈ 0.577) is the Euler-Mascheroni constant. The parameters can be expressed in terms of the moments as follows

$$\xi = \mu - \gamma_{\text{EM}}\frac{\sqrt{6}}{\pi}\sigma;\ \varsigma = \frac{\sqrt{6}}{\pi}\sigma \tag{F.19}$$

If the estimates

$$\tilde{\xi}(x) = \hat{\mu}(x) - \gamma_{\text{EM}}\frac{\sqrt{6}}{\pi}\hat{\sigma}(x);\ \tilde{\varsigma}(x) = \frac{\sqrt{6}}{\pi}\hat{\sigma}(x) \tag{F.20}$$

are used, then $\tilde{\xi}(ax+b) = a\tilde{\xi}(x) + b$ and $\tilde{\varsigma}(ax+b) = a\tilde{\varsigma}(x)$. According to Hosking [16], the first two L-moments are

$$\lambda_1 = \xi + \varsigma\gamma_{\text{EM}};\ \lambda_2 = \varsigma\log 2 \tag{F.21}$$

so

$$\xi = \lambda_1 - \gamma_{\text{EM}}\frac{1}{\log 2}\lambda_2;\ \varsigma = \frac{1}{\log 2}\lambda_2 \tag{F.22}$$

If the estimates

$$\tilde{\xi}(x) = l_1(x) - \gamma_{\text{EM}}\frac{1}{\log 2}l_2(x);\ \tilde{\varsigma}(x) = \frac{1}{\log 2}l_2(x) \tag{F.23}$$

are used, then $\tilde{\xi}(ax+b) = a\tilde{\xi}(x) + b$ and $\tilde{\varsigma}(ax+b) = a\tilde{\varsigma}(x)$.

F.1.2. THE LOG-NORMAL DISTRIBUTION

For the log-normal distribution, the pdf is

$$f(x;\varsigma,\eta) = \begin{cases} 0 & x \le 0 \\ \frac{1}{x\eta\sqrt{2\pi}}\exp\left(-\frac{(\log(x/\varsigma))^2}{2\eta^2}\right) & x > 0 \end{cases} \tag{F.24}$$

The mean μ and variance σ^2 are

$$\mu = \varsigma\exp\left(\frac{\eta^2}{2}\right);\ \sigma^2 = \varsigma^2\left(\exp\left(\eta^2\right) - 1\right)\exp\left(\eta^2\right) \tag{F.25}$$

The parameters can be expressed in terms of the moments as follows

$$\varsigma = \frac{\mu}{\sqrt{\frac{\sigma^2}{\mu^2}+1}};\ \eta = \sqrt{\log\left(\frac{\sigma^2}{\mu^2}+1\right)} \tag{F.26}$$

If the estimates

$$\tilde{\varsigma}(x) = \frac{\hat{\mu}(x)}{\sqrt{\left(\frac{\hat{\sigma}(x)}{\hat{\mu}(x)}\right)^2+1}};\ \tilde{\eta}(x) = \sqrt{\log\left(\left(\frac{\hat{\sigma}(x)}{\hat{\mu}(x)}\right)^2+1\right)} \tag{F.27}$$

are used, then $\tilde{\varsigma}(ax) = a\tilde{\varsigma}(x)$ and $\tilde{\eta}(ax) = \tilde{\eta}(x)$. According to Hosking [16], the first two L-moments are

$$\lambda_1 = \varsigma\exp\left(\frac{\eta^2}{2}\right);\ \lambda_2 = \varsigma\exp\left(\frac{\eta^2}{2}\right)\mathrm{erf}\left(\frac{\eta}{2}\right) \tag{F.28}$$

where erf is given by

$$\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}}\int_{t=0}^{z}\exp(-t^2)\,dt \tag{F.29}$$

The parameters are related to the L-moments as follows

$$\varsigma = \frac{\lambda_1}{\exp(\frac{\eta^2}{2})};\ \eta = 2\mathrm{erf}^{-1}(\frac{\lambda_2}{\lambda_1}) \tag{F.30}$$

If the estimates

$$\tilde{\varsigma}(x) = \frac{l_1(x)}{\exp\left(\frac{[\tilde{\eta}(x)]^2}{2}\right)};\ \tilde{\eta}(x) = 2\mathrm{erf}^{-1}\left(\frac{l_2(x)}{l_1(x)}\right) \tag{F.31}$$

are used, then $\tilde{\varsigma}(ax) = a\tilde{\varsigma}(x)$ and $\tilde{\eta}(ax) = a\tilde{\eta}(x)$.

F.1.3. THE GAMMA DISTRIBUTION

For the gamma distribution, the pdf is

$$f(x;\varsigma,\eta) = \begin{cases} 0 & x \le 0 \\ \frac{1}{\Gamma(\eta)\varsigma^\eta}x^{\eta-1}\exp\left(\frac{x}{\varsigma}\right) & x > 0 \end{cases} \tag{F.32}$$

The mean μ and variance σ^2 are

$$\mu = \eta\varsigma;\ \sigma^2 = \eta\varsigma^2 \tag{F.33}$$

The parameters can be expressed in terms of the moments as follows

$$\eta = \frac{\mu^2}{\sigma^2};\ \varsigma = \frac{\sigma^2}{\mu} \tag{F.34}$$

If the estimates

$$\tilde{\eta}(x) = \left(\frac{\hat{\mu}(x)}{\hat{\sigma}(x)}\right)^2;\ \tilde{\varsigma}(x) = \frac{(\hat{\sigma}(x))^2}{\hat{\mu}(x)} \tag{F.35}$$

are used, then $\tilde{\eta}(ax) = \tilde{\eta}(x)$ and $\tilde{\varsigma}(ax) = a\tilde{\varsigma}(x)$. According to Hosking [16], the first two L-moments are

$$\lambda_1 = \eta\varsigma;\ \lambda_2 = \frac{1}{\sqrt{\pi}}\varsigma\frac{\Gamma\left(\eta+\frac{1}{2}\right)}{\Gamma(\eta)} \tag{F.36}$$

Now the function f_η defined by

$$f_\eta^{-1}(\eta) = \frac{1}{\sqrt{\pi}}\frac{\Gamma\left(\eta+\frac{1}{2}\right)}{\eta\Gamma(\eta)} \tag{F.37}$$

is needed to express the parameters in terms of the L-moments

$$\varsigma = \frac{\lambda_1}{\eta};\ \eta = f_\eta\left(\frac{\lambda_2}{\lambda_1}\right) \tag{F.38}$$

To obtain parameter estimates, f_η is approximated as in Hosking [16]. If the estimates

$$\tilde{\eta}(x) = f_\eta\left(\frac{l_2(x)}{l_1(x)}\right);\ \tilde{\varsigma}(x) = \frac{l_1(x)}{\tilde{\eta}(x)} \tag{F.39}$$

are used, then $\tilde{\eta}(ax) = \tilde{\eta}(x)$ and $\tilde{\varsigma}(ax) = a\tilde{\varsigma}(x)$.

F.1.4. FRÉCHET DISTRIBUTION AND THE GENERALIZED EXTREME VALUE DISTRIBUTION

For Generalized Extreme Value (GEV) distribution, the value of the shape parameter decides the type the distribution. According to Tyralis *et al.* [17], when GEV distribution is used to model annual streamflow maxima, shape parameter mostly depends on climatic indices.

$$f(x,m,s,\alpha) = \begin{cases} 0 & x \le m \\ \frac{1}{s}(1+\alpha(\frac{x-m}{s}))^{-\alpha-1}\exp\left(-(1+\alpha(\frac{x-m}{s}))^{-1/\alpha}\right) & x > m, \alpha \neq 0 \\ \frac{1}{s}\exp\left(-\exp(-\frac{x-m}{s}) - \frac{x-m}{s}\right) & x > m, \alpha = 0 \end{cases} \tag{F.40}$$

where α is the shape parameter, s is the scale parameter, m is the location parameter, the mean (μ) and standard deviation (σ). The immediate relations between sample moments and parameters in parametric distributions are

$$\begin{cases} m = \mu - \sigma\frac{\Gamma(1-\alpha)-1}{\sqrt{\Gamma(1-2\alpha)-\Gamma^2(1-\alpha)}}; s = \sigma\frac{\alpha}{\sqrt{\Gamma(1-2\alpha)-\Gamma^2(1-\alpha)}} & x > m, \alpha \neq 0 \\ m = \mu - \frac{\gamma\sqrt{6}}{\pi}\sigma; s = \frac{\sqrt{6}}{\pi}\sigma & x > m, \alpha = 0 \end{cases} \tag{F.41}$$

When the shape parameter $k = 0$, the GEV is a standard Gumbel distribution; when $k < 0$, the GEV is a reverse Weibull distribution; when $k > 0$, the GEV is a Fréchet distribution. Because when the shape parameter $k \neq 0$, Fréchet and reverse Weibull distributions have three parameters.

In Chapter 2, GEV distribution with shape parameter equals to –0.15 (the three-parameter reverse Weibull distribution with shape 20/3); GEV with shape 0 (the Gumbel distribution); and GEV with shape 0.15 (the three-parameter Fréchet distribution with shape

20/3) are used. The value 0.15 was chosen as representative for thick-tailed GEV distributions.

In Chapter 4, the Gumbel distribution is considered for study the performance of the two parametric methods (CML and CLMo) in constructing confidence curves for the location of a CP, where maximum likelihood estimate (mle) and pseudo maximum likelihood estimate (pmle) are used. In CLMo, the pseudo maximum likelihood is combined the LMo method for the estimation of nuisance parameters and maximum likelihood estimate for the parameter of interest.

In Chapter 5, Fréchet distribution with a constant shape parameter is used to generate synthetic data to explore properties of confidence curves based on approximate empirical likelihood ratio method. In Koutsoyiannis [18], the suggested value for the shape parameter in type II or GEV distribution "is expected to belong to a short range, approximately from 0 to 0.23 with confidence level 99%" for the global extreme precipitation observations. After that Ragulina and Reitan [19] extended that research and they suggested to use 0.139 as the shape parameter for Fréchet distribution. Therefore, for Fréchet distributions, the suggested shape parameter $k = 0.139$ is used in Chapter 5.

G

THE COMPUTATIONAL COST OF THE PARAMETRIC METHODS

For a time series of length n, with N MC runs for distribution approximation, all three methods (CML, CLMO, CMoM) need $(1+N)(n-2n_{\min}+1)$ deviance calculations. Each deviance calculation needs $2(n-2n_{\min}+1)$ parameter estimates and $(n-2n_{\min}+1)\times n$ calculations of the logarithm of the pdf. Each pair of mle 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 the mle 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

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

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 mle 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 the mle may well take anywhere from 10 to 200 times as long. For GU and GA, where the mle problems correspond to a one dimensional search for the point where a nonlinear function is zero, in practice the cost of the mle 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 analysed 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 $O\left(10^{12}\right)$ 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 [20] 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.

H

AN UNINFORMATIVE CONFIDENCE INTERVAL FOR KENDALL'S τ

It is known that $\tau = [-1,1]$. Now suppose that $\gamma \in (0,1)$ and there is no a-priori information on the location of τ. If a point y is selected at random in the interval [-1, 1-2γ], then

$$\begin{aligned}\Pr(\tau \in [y, y+2\gamma]) &= \int_{x=-1}^{1} \int_{y=-1}^{1-2\gamma} [\![y \le x \le y+2\gamma]\!] \tfrac{1}{2} dx \tfrac{1}{2-2\gamma} dy \\ &= \tfrac{1}{4(1-\gamma)} \int_{y=-1}^{1-2\gamma} \int_{x=-1}^{1} [\![y \le x \le y+2\gamma]\!] \, dx dy \\ &= \tfrac{1}{4(1-\gamma)} \int_{y=-1}^{1-2\gamma} \int_{x=y}^{\min(y+2\gamma,1)} [\![y \le x \le y+2\gamma]\!] \, dx dy\end{aligned} \tag{H.1}$$

so

$$\begin{aligned}\Pr(\tau \in [y, y+2\gamma]) &= \tfrac{1}{4(1-\gamma)} \int_{y=-1}^{1-2\gamma} \int_{x=y}^{y+2\gamma} dx dy \\ &= \tfrac{1}{4(1-\gamma)} \int_{y=-1}^{1-2\gamma} 2\gamma \, dy \\ &= \tfrac{1}{4(1-\gamma)} 2\gamma(2-2\gamma) = \gamma\end{aligned} \tag{H.2}$$

This implies that, as long as there is no a-priori reason to assume the parameter has a certain value, it is possible to obtain a confidence interval at level γ without using the sample, as long as an interval length of (at least) 2γ is allowed.

I

COPULAS

Before stating the conditions that a function must satisfy to be a copula and how such functions relate to multivariate distributions it is helpful to state the analogous conditions and facts for a univariate cumulative distribution function (cdf).

I.1. THE ONE DIMENSIONAL CUMULATIVE DISTRIBUTION FUNCTION

According to, for instance, Klenke [21], if $F(x)$ is the cdf of a real-valued random variable X, then $F(x) = \Pr(X \leq x)$ and F has the following properties

1. If $x_1 < x_2$, then $F(x_1) \leq F(x_2$ (F is non decreasing)
2. $\lim_{x \downarrow x_0} = F(x_0)$
3. $\lim_{x \to -\infty} = 0$
4. $\lim_{x \to \infty} = 1$

Moreover, for any function with these properties there is a random variable for which it is the cdf [21, 22]. Please note that, while $F(x) = \Pr(X < x)$ instead, in which case condition 2 should be changed to $\lim_{x \uparrow x_0} = F(x_0)$ (F is left continuous).

I.2. THE TWO DIMENSIONAL CUMULATIVE DISTRIBUTION FUNCTION

To keep the notation as simple as possible, only the two dimensional (2D) case is described. Vectors will be denoted by bold italic letters. The following shorthand will be used: $\boldsymbol{a} < \boldsymbol{b}$ means that $a_1 < b_1$ and $a_2 < b_2$. Now by definition

$$\Pr(\boldsymbol{a} < \boldsymbol{X} \leq \boldsymbol{b}) = \Pr(\boldsymbol{X} \leq \boldsymbol{b}) - \Pr(\boldsymbol{X} \leq (a_2, b_1)) - \Pr(\boldsymbol{X} \leq (a_1, b_2)) + \Pr(\boldsymbol{X} \leq \boldsymbol{a}) \tag{I.1}$$

This can be used to define a function that assigns a 'volume' to a bounded rectangle

$$B = (\boldsymbol{a}, \boldsymbol{b}] = \{(x, y) : a_1 < x \leq b_1, a_2 < y \leq b_2\} \tag{I.2}$$

for any $G : \mathbb{R} \to \mathbb{R}$ by setting

$$V_G(B) = G(\boldsymbol{b}) - G(a_2, b_1) - G(a_1, b_2) + G(\boldsymbol{a}) \tag{I.3}$$

This volume function is used to define 2-increasing functions as functions that satisfy $V_G(B) \geq 0$ or all rectangles. The definition of a volume function in n dimensions follows the same principle, but requires a more complex notation.

If F is the joint cdf of a 2D real-valued random vector $\boldsymbol{X} = (X_1, X_2)$, then $F(\boldsymbol{x}) = \Pr(X_1 \leq x_1, X_2 \leq x_2)$.

1. F is 2-increasing
2. $\lim_{x_1 \downarrow a_1} F(x_1, a_2) = F(\boldsymbol{a})$ and $\lim_{x_2 \downarrow a_2} F(a_1, x_2) = F(\boldsymbol{a})$ (F is right continuous for each vector component)
3. $\lim_{x_1 \to -\infty} F(x_1, a_2) = 0; \lim_{x_2 \to -\infty} F(a_1, x_2) = 0$
4. $\lim_{x \to (\infty, \infty)} F(\boldsymbol{x}) = 1$

According to Joe [22] these conditions are also sufficient for F to be a bivariate cdf.

I.2.1. A TWO DIMENSIONAL COPULA

An $2D$ copula is a function C from $[0,1]^2$ to $[0,1]$ that is continuous and non-decreasing such that

1. C is 2-increasing
2. C is right continuous
3. $C(x_1, 0) = 0; C(0, x_2) = 0$
4. $C(x_1, 1) = x_1; C(1, x_2) = x_2$

A shorter definition can be formulated by using the definition of a 2D cdf: a 2D copula is a 2D cdf on the unit square with uniform marginals. The central theorem about copulas (stated for the 2D case) is the following

Theorem 4. *If H is an 2D cdf with marginal distributions F_1, F_2 then there exists a copula C such that*

$$H(x_1, x_2) = C(F_1(x_1), F_2(x_2)) \tag{I.4}$$

and if F_1, F_2 are cdfs and C is a 2D copula, then $C(F_1(x_1), F_2(x_2))$ is a 2D cdf [22, 23].

I.3. SOME ARCHIMEDEAN COPULAS

I.3.1. FRANK

The cdf for the Frank copula is

$$C_{\mathrm{F}}(u,v;\theta) = -\frac{1}{\theta}\log\left[1+\frac{\left(\exp(-\theta u)-1\right)\left(\exp(-\theta v)-1\right)}{\exp(-\theta)-1}\right] \tag{I.5}$$

where $-\infty < \theta < 0$ or $0 < \theta < \infty$. The pdf for the Frank copula is

$$c_{\mathrm{F}}(u,v;\theta) = \frac{\theta(1-\exp(-\theta))\exp(-\theta[u+v])}{\left(\exp(-\theta)-\exp(-\theta u)-\exp(-\theta v)+\exp(-\theta[u+v])\right)^2} \tag{I.6}$$

I.3.2. CLAYTON

The cdf for the Clayton copula is

$$C_{\mathrm{C}}(u,v;\theta) = [\max(u^{-\theta}+v^{-\theta}-1,0)]^{-1/\theta} \tag{I.7}$$

where $-1 \le \theta < \infty$, The pdf for thr Clayton is

$$c_{\mathrm{C}(u,v;\theta)} = \begin{cases} 0 & u^{-\theta}+v^{-\theta}-1<0 \\ 0 & u^{-\theta}+v^{-\theta}-1=0, -1<\theta<0 \\ (1+\theta)u^{-1-\theta}v^{-1-\theta}(u^{-\theta}+v^{-\theta}-1)^{-2-1/\theta} & u^{-\theta}+v^{-\theta}-1>0 \end{cases} \tag{I.8}$$

I.3.3. GUMBEL

The cdf for the Gumbel copula is

$$C_{\mathrm{G}}(u,v;\theta) = \exp\left(-[-(-\log u)^{\theta}+(-\log v)^{\theta}]^{1/\theta}\right) \tag{I.9}$$

where $-1 \le \theta < \infty$. The pdf for the Gumbel copula is

$$c_{\mathrm{G}}(u,v;\theta) = \frac{C_{\mathrm{G}}(u,v;\theta)}{uv}(\log u \log v)^{\theta-1}\left((-\log u)^{\theta}+(-\log v)^{\theta}\right)^{1/\theta-2}\left(\left((-\log u)^{\theta}+(-\log v)^{\theta}\right)^{1/\theta}+\theta-1\right) \tag{I.10}$$

I.4. SOME EXTREME VALUE COPULAS

A bivariate extreme value copula is a copula that satisfies

$$C(u^t,v^t) = C^t(u,v) \tag{I.11}$$

The Gumbel copula described earlier is an extreme value copula.

I.4.1. GALAMBOS

The cdf for the Galambos copula is

$$C_{\mathrm{Ga}}(u,v;\theta) = uv\exp\left([(-\log u)^{-\theta}+(-\log v)^{-\theta}]-1/\theta\right) \tag{I.12}$$

where $0 \le \theta < \infty$.

I.4.2. HUESLER-REISS

The cdf for the Huesler-Reiss copula is

$$C_{\mathrm{HR}}(u,v;\theta) = \exp\left[(\log u)\Phi\left(\frac{1}{\theta}+\frac{\theta}{2}\log\left(\frac{\log u}{\log v}\right)\right)+(\log v)\Phi\left(\frac{1}{\theta}+\frac{\theta}{2}\log\left(\frac{\log v}{\log u}\right)\right)\right] \tag{I.13}$$

where Φ is the cdf of the univariate standard normal distribution.

REFERENCES

[1] D. E. Knuth, *Two notes on notation,* The American Mathematical Monthly **99**, 403 (1992).

[2] K. E. Iverson, *A programming language,* in *Proceedings of the May 1-3, 1962, spring joint computer conference* (1962) pp. 345–351.

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

[4] C. Cunen, G. Hermansen, and N. L. Hjort, *Confidence distributions for change-points and regime shifts,* Journal of Statistical Planning and Inference **195**, 14 (2018).

[5] M. Csörgö and L. Horváth, *Limit theorems in change-point analysis*, Vol. 18 (John Wiley & Sons Inc, 1997).

[6] A. B. Owen, *Empirical likelihood ratio confidence intervals for a single functional,* Biometrika **75**, 237 (1988).

[7] A. Owen *et al.*, *Empirical likelihood ratio confidence regions,* The Annals of Statistics **18**, 90 (1990).

[8] C. Zou, Y. Liu, P. Qin, and Z. Wang, *Empirical likelihood ratio test for the change-point problem,* Statistics & Probability Letters **77**, 374 (2007).

[9] G. Shen, *On empirical likelihood inference of a change-point,* Statistics & Probability Letters **83**, 1662 (2013).

[10] P. Hall and B. La Scala, *Methodology and algorithms of empirical likelihood,* International Statistical Review/Revue Internationale de Statistique , 109 (1990).

[11] A. Schubert and A. Telcs, *A note on the Jaccardized Czekanowski similarity index,* Scientometrics **98**, 1397 (2014).

[12] P. Jaccard, *Étude comparative de la distribution florale dans une portion des Alpes et des Jura,* Bull Soc Vaudoise Sci Nat **37**, 547 (1901).

[13] P. von Hippel, *Skewness,* International Encyclopedia of Statistical Science , 1340 (2011).

[14] E. Seier, *Kurtosis: An overview,* International Encyclopedia of Statistical Science , 1340 (2011).

[15] D. N. Joanes and C. A. Gill, *Comparing measures of sample skewness and kurtosis,* Journal of the Royal Statistical Society: Series D (The Statistician) **47**, 183 (1998).

[16] J. R. M. Hosking, *L-moments: analysis and estimation of distributions using linear combinations of order statistics,* J. Roy. Statist. Soc. Ser. B **52**, 105 (1990).

[17] H. Tyralis, G. Papacharalampous, and S. Tantanee, *How to explain and predict the shape parameter of the generalized extreme value distribution of streamflow extremes using a big dataset,* Journal of Hydrology **574**, 628 (2019).

[18] D. Koutsoyiannis, *Statistics of extremes and estimation of extreme rainfall: II. Empirical investigation of long rainfall records/Statistiques de valeurs extrêmes et estimation de précipitations extrêmes: II. Recherche empirique sur de longues séries de précipitations,* Hydrological Sciences Journal **49** (2004).

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

[20] Intel, *APP Metrics for Intel® Microprocessors: Intel® Xeon® Processor* (Technical Report Rev. 3. Intel Corporation, Santa Clara, CA, 2020).

[21] A. Klenke, *Probability theory,* Universitext. third ed., Springer International Publishing (2020), 10.1007/978-3-030-56402-5.

[22] H. Joe, *Dependence modeling with copulas* (CRC press, 2014).

[23] R. B. Nelsen, *An introduction to copulas* (Springer Science & Business Media, 2007).

ACKNOWLEDGEMENTS

Throughout the PhD research period I have received a great deal of support and assistance.

I would first like to thank my promotor, Professor Nick van de Giesen, for valuable guidance, critical comments, professional advices, and all of the opportunities I was given. Your insightful feedback pushed me to sharpen my thinking and brought my work to a higher level. Your wise advices for my future helped me have a better understanding about what I am capable of.

I would also like to thank my co-promotor, Dr. Ronald van Nooijen, whose expertise was invaluable in formulating the research questions and methodology. You statistical thinking, efficient programming, patient support, and all encouragements helped me learn, grow, and become more confident. Our discussions and ideas exchanges were sources always of my inspirations.

I would like to acknowledge Dr. Alla Kolechkina for her critical thinking, wise counsel, and sympathetic ear. You are always there for me. In addition, I would like to thank my parents and my family. They are always my rock, especially my little nephew. Thanks for your constant support and numerous happiness.

I would also like to thank my boyfriend, Waleed Hassan, for your constant support, love, and caring.

Finally, I could not have completed this dissertation without the support of my colleagues and friends. I will always remember our girls' night, stimulating discussions, as well as happy distractions to rest my mind outside of my research.

CURRICULUM VITÆ

PERSONAL INFORMATION

Name:	Changrang Zhou
Gender:	Female
Location:	Delft, Netherlands
Hometown:	China
Contact info:	(+31)0655549622
E-mail:	C.Zhou-1@tudelft.nl; zcrhhu@gmail.com
Page:	https://www.researchgate.net/profile/Changrang-Zhou-2

EDUCATION

09/2010–06/2014 **Bachelor of Engineering: Hyrology & Water Resources Engineering**

College of Earth Science and Engineering,
Shandong University of Science and Technology, Qingdao, China

Topic: The potential problems concerning recharging groundwater with waste water in mining areas

Supervisor: Dr. Shengtang Zhang

09/2014–06/2017 **Master of Engineering: Hydrology & Water Resources**

College of Hydrology and Water Resources,
Hohai University, Nanjing, China

Topic: Method of Higher Order Moments and its application in hydrological time series analysis

Promotor: Prof. Yuanfang Chen; Associate Prof. Shenghua Gu

09/2017–present **Ph.D. candidate: Water Management**

Department of Water Management, Technology University of Delft,
Delft, The Netherlands

Thesis: Use confidence curves to represent uncertainty in hydro-meteorological time series analysis

Promotor: Prof. dr. Nick van de Giesen

Copromotor: Dr. Ronald van Nooijen

PUBLICATIONS (PHD PERIOD)

5. van Nooijen, R., **Zhou, C.**, Kolechkina, A., *A method to quantify the uncertainty in copula parameters when studying dependence structures of time series*,in proceedings of the 39th IAHR World Congress, 19-24 June 2022, Granada, Spain. In press, 2022.

4. **Zhou, C.**, van Nooijen, R., Kolechkina, A., Gargouri-Ellouze, E., and van de Giesen, N., *Using confidence curves to capture the uncertainty in dependence structure in copula models*, Hydrological Sciences Journal, under review, 2021.

3. **Zhou, C.**, van Nooijen, R., and Kolechkina, A., *Capturing the uncertainty about a sudden change in the properties of time series with confidence curves*, Journal of Hydrology, under review, 2021.

2. **Zhou, C.**, van Nooijen, R., Kolechkina, A. and van de Giesen, N., *Confidence curves for change points in hydrometeorological time series*, Journal of Hydrology, Page: 1-19, **590**, 2020.

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

CONFERENCE ABSTRACTS & PRESENTATIONS

9. van Nooijen, R., Kolechkina, A., and **Zhou, C.**. *Building confidence curves with classical change point tests*, IAHS2022, **2022** (Abstract).

8. van Nooijen, R., Kolechkina, A., and **Zhou, C.**. *Confidence curves condensed to a number*, ICSH-STAHY 2021, **2021** (Abstract).

7. **Zhou, C.**, van Nooijen, R., Kolechkina, A. and van de Giesen, N.. *Distribution free methods to represent uncertainty in change point estimates*, STAHY 2021 Workshop, virtual, **2021** (Abstract, Oral Presentation online).

6. **Zhou, C.**, van Nooijen, R., Kolechkina, A. and van de Giesen, N.. *Assessing the uncertainty in change point detection with confidence curves*, 63rd ISI World Statistics Congress 2021, CPS 236, virtual, **2021** (Abstract, Oral Presentation online).

5. van Nooijen, R., Kolechkina, A., and **Zhou, C.**. *How hard is it to detect abrupt changes in the statistics of time series*, 62nd ISI World Statistics Congress, Page: 43, Kuala Lumpur, Malaysia, **2019** (Abstract).

4. **Zhou, C.**, van Nooijen, R., Kolechkina, A. and van de Giesen, N.. *Distribution free methods to represent uncertainty in change point estimates*, STAHY 2019 Workshop, Nanjing, China, **2019** (Abstract, Oral Presentation).

3. **Zhou, C.**, van Nooijen, R., van de Giesen, N. and Kolechkina, A.. *Confidence Distributions for change point estimates: parametric versus empirical likelihood*, European Geoscience Union General Assembly 2019-4320, Vienna, Austria, **2019** (Abstract, PICO presentation).

2. **Zhou, C.**, van Nooijen, R., and van de Giesen, N.. *The outcome of statistical change-point detection in the light of the historical*, European Geoscience Union General Assembly 2018-9951, Vienna, Austria, **2018** (Abstract, Oral Presentation).

1. **Zhou, C.**, van Nooijen, R., and Kolechkina, A.. *Effectiveness of some change-point detection methods when applied to synthetic hydrological time series*, European Geoscience Union General Assembly 2018-16992-1, Vienna, Austria, **2018** (Abstract, Poster Presentation).