Czech University of Life Sciences Prague
Department of Water Resources and Environmental modelling
Landscape Engineering — Environmental Modelling

Statistical models for drought indices

Doctoral Dissertation of:
Filip Strnad

Supervisor:
prof. Ing. Martin Hanel, Ph.D.

Co-Supervisor:
doc. Mgr. Ing. Ioannis Markonis, Ph.D.

I hereby declare that I have independently elaborated the thesis with the topic of "Statistical models for drought indices" and that I have cited all the information sources that I used in the thesis and that are also listed at the end of the thesis in the list of used information sources.

.....

Abstract

Modelling of hydrological extremes and drought modelling in particular has received much attention over recent decades. The thesis gives an overview of the drought definitions and drought indices and approaches to extreme value analysis including the regional frequency analysis with focus on drought. The main aim of the thesis is the application and comparison of regional frequency analysis models for drought characteristics based on various parameter estimation methods, that involves modifying the methods in order to work with intermittent variables. The attention is be paid to the reduction of uncertainty in the estimated return levels, in the periods of drought events and in the parameters of the extremal model. The goodness-of-fit of the models is be evaluated through discordance analysis, as well as the Anderson–Darling tests, with the critical values estimated by a bootstrap procedures. Performance of presented methods is evaluated by designed experiments focusing on behaviour of parameter estimation algorithms with increasing sample size in various situations. Experiment is also assessing the index flood method performance with increasing p_0 (probability of zero values, e.g. the fraction of years without drought). As part of the thesis a case study is presented that deals with development of an index flood model for deficit volumes for 133 catchments in the Czech Republic (1901–2015) that are simulated by hydrological model BILAN. The parameters of the regional distribution are estimated using L-moments. The goodness-of-fit of the statistical model is assessed by the *Anderson–Darling test* (A^2) test. For the estimation of critical values, sampling methods allowing for handling of years without drought were used.

Preface

It has been a little over six years since I have started working on the problems concerning drought modelling, little did I know about about the lion's den I am throwing my self into. Stochastic modelling of drought is an complex process with no unified set of methodologies. Therefore I tried approaches used for other extremes with little to no success. My state of mind halfway thru my studies could be neatly surmised by quote from Dante's Divine Comedy:

> *All hope abandon, ye who enter here.*
>
> *Dante Alighieri*

Now at the end of my journey, overlooking the path I walked (stumbling), I wish I could tell my former self that "there indeed is light at the end of the tunnel". This is the biggest motivation write the dissertations in this form - as a guide for people who would want to walk on the same road as I did telling them (borrowing again from Dante):

> *The devil is not as black as he is painted.*
>
> *Dante Alighieri*

Number of other publications preceded the dissertation serving as a stepping stone for this work (list of selected publications can be found in appendix A) offering an insight into the drought issue.

Although these publications have their own list of authors, there is still a number of people deserving acknowledgement and big thanks, notably my brother and my mother, my office buddy Vašek, my good friends Petr Bašta and Honza Čadek, my partner in crime Ivana, bartenders of the former restaurant Na Farmě (although "Farma" restaurant no longer exists it still holds a special place in my heart) Andrea and Bolek and last but not least the big bosses of the department - Petr Máca, Yannis Markonis and Martin Hanel.

Contents

List of Figures

List of Tables

List of Acronyms

A^2 Anderson–Darling test iv

CDF Cumulative Distribution Function 1

GEV Generalised Extreme Value distribution 15

GPD Generalised Pareto Distribution 15

IQR Inter Quantile Range 45

mMSE modified Mean Squared Error 20

PDSI Palmer Drought Severity Index 9

PDF Probability Density Function

RDI Reconnaissance Drought Index 9

RFA Regional Frequency Analysis 4

SOM Self-Organising Map 33

SWSI Soil Water Supply Index 10

CHAPTER 1

Introduction

When it comes to hydrological extremes, flooding events receive most attention, both in the news and in scientific literature, due to their fast, clearly visible, and dramatic consequences (Bloeschl et al., 2019). Drought events however - also called *the creeping disaster* - develop slowly and are often unnoticed and have diverse and indirect consequences.

Droughts can, however, cover extensive areas and can last for months to years, with devastating impacts on the Earth system linked to many economic sectors (Ciais et al., 2005). The probabilistic nature of drought phenomenon requires development of drought forecasting probabilistic frameworks extending current available knowledge. The crucial part of drought modelling is the description of probability distribution.

Estimating the type and parameters of the *Cumulative Distribution Function* (CDF) of the available data-sets is therefore a fundamental goal in estimating the risk of occurrence of a particular events. Probability

distribution models are useful tools for the statistical description of high quantiles.

Statistical modelling of extremes in general is subjected to large uncertainties due to the rarity of extreme events or problems with their measurement. This applies especially to droughts since drought does not occur every year and thus the length of series typically available for hydrological analysis provides only limited information. This can be, at least partly, overcome by *trading space for time*, i.e., combining data from several sites over homogeneous regions. The effect of adding sites/catchments is maximal when the data are independent. This is seldom true, however, thus the real reduction of uncertainty not only depends on the number of data but also on the spatial covariance structure of the analysed data.

1.1 The thesis aim

The main focus of this work is exploring various methods used for quantifying hydrological extremes, with a specific emphasis on drought events. The goal is to provide a comprehensive assessment of the methods and their performance in the context of drought quantification, with the ultimate aim of reducing the uncertainties in high quantiles estimations. The methods presented in this work are chosen with clear focus on their applicability in real-world situations. By providing a thorough evaluation of these methods and their strengths and weaknesses, this work seeks to contribute to the ongoing efforts to improve our understanding of hydrological extremes and to develop more effective strategies for managing and mitigating their impacts, which is crucial for effective water resource management and planning.

The thesis also aims to evaluate the performance of different methods and algorithms for high quantile estimation of intermittent variables i. e.

drought events. The focus of this work is on modifying existing methods to effectively deal with intermittent data, which can provide more accurate and reliable estimates of high quantiles. Through experiments, the strengths and weaknesses of various approaches will be identified, and guidance will be provided on selecting the most appropriate methods for different scenarios. The ultimate goal is to reduce uncertainties associated with drought quantification, thereby contributing to more accurate and reliable estimates of high quantiles, which are crucial for effective drought risk assessment and management.

To summarise the aims of this thesis, they can be listed as follows:

- To evaluate the performance of different methods and algorithms for drought high quantile estimation, with a focus on modifying the methods to work with intermittent variables.
- To identify the strengths and weaknesses of different approaches for high quantile estimation, and provide guidance on selecting the most appropriate methods for different scenarios.
- To reduce uncertainties associated with drought quantification, particularly in the estimation of high quantiles, which are crucial for drought risk assessment and management.
- To investigate the limitations of the extreme value paradigm, specifically in the context of hydrological modelling, and provide insights into ways to overcome these limitations.
- To develop and test new bootstrapping methods for extreme value analysis that account for spatial dependence and compound distributions.

1.2 The structure of the thesis

The work is divided into four main parts - Methods (Chapter 2); Simulations (Chapter 3); Case Study (Chapter 4) and Thesis summary (Chapter 5).

Chapter 2, is giving a summary of methods used to define and describe drought and methods used in extreme events estimation. There is a list of widely recognised drought definitions (section 2.1) as well as indices used to identify and quantify drought events. Section 2.2 is dealing with subjective choices one can come across when choosing the right distributions for drought indices and indicators and section 2.3 is listing methods for distribution parameters estimation.

Methods, more specifically distribution parameter fitting methods and index flood, are then tested in Chapter 3 by Monte-Carlo simulations in order to evaluate their behaviour in various situations.

In Chapter 4, there is a case study, where indicators of drought (maximum deficit volumes) were estimated and validated for the period 1900–2015 over the Czech Republic. Index flood method is then employed in order to reduce uncertainties in deficit volume return levels estimation.

Final part of the thesis - Chapter 5 - is an introspection of methods used in previous parts of the thesis, giving recommendations for drought high quantile estimation based on the results presented in Chapter 3 with comments on each step within the *Regional Frequency Analysis* (RFA) scheme.

CHAPTER 2

Methods used for high quantile estimation & drought quantification

Since no universal definition of drought exists (Lloyd-Hughes, 2014) (nor should exist, according to Wilhite and Glantz (1985)) and drought affects so many sectors in environment and society, there is a need for different definitions. The particular problem under study, the data availability and the climatic and regional characteristics are among the factors influencing the choice of drought event definition. Wilhite and Glantz (1985) found more than 150 published definitions of drought, which might be classified in a number of ways. Some of the most common drought definitions are summarised in Demuth and Bakenhus (1994); Dracup et al. (1980); Svoboda et al. (2016); Tate and Gustard (2000); Van Loon (2015).

Estimating the occurrence of drought events is a challenging task that is plagued by significant uncertainty. The problem with data availability inherent to extreme value analysis (the extremes are rare by definition) is even more serious pronounced for drought compared to extreme rainfall, since the drought does not occur every year. Thus the statistical analysis is often challenging due to short record lengths. In fact, many regions may experience drought only sporadically, with the most severe and long-lasting events occurring only once every few decades. This variability and unpredictability make it difficult to accurately estimate the likelihood and impact of drought. This uncertainty can however be reduced by employing methods like extreme value theory in conjunction with *Regional Frequency Analysis* (RFA).

Extreme value theory has emerged as one of the most important statistical disciplines for the applied sciences over the past 70 years. Extreme value techniques are also being used in many other disciplines. The distinguishing feature of an extreme value analysis is the objective to quantify the stochastic behaviour of a process at unusually large or small levels. In particular, extreme value analyses usually require estimation of the probability of events that are more extreme than any that have already been observed. Extreme value theory provides a framework that enables extrapolations of this type. This chapter describes methods widely used for estimation of high quantiles and methods employed in lowering their uncertainties.

2.1 Drought definition and quantification

Most common classification of drought, based on a disciplinary perspective can be found in Dracup et al. (1980), where droughts are related to precipitation (meteorological), stream-flow (hydrological), soil moisture (agricultural)

or any combination of the three. A similar classification can be found in Wilhite and Glantz (1985), where four categories are identified:

- Meteorological drought: departure of precipitation from normal over some period of time. Reflects one of the primary causes of a drought.
- Hydrological drought: deficiency of surface and subsurface water supplies. Reflects effects and impacts of droughts.
- Agricultural drought: low soil moisture, sometimes related to need of a particular crop at a particular time.
- Socio-economic drought: defined as a situation in which the water supply fails to satisfy water demand, thus resulting in negative consequences for society, the economy, and the environment.

To identify and quantify drought, several drought indicators can be used. Mawdsley et al. (1994) defined two classes or types of indicators:

- Environmental indicators are those hydro-meteorological and hydrological indicators, which measure the direct effect on the hydrological cycle. The nature of the water deficit might be related to precipitation, stream-flow or soil moisture. These indicators can help identifying the duration and/or severity of a drought and can be used to analyse the drought frequency. Environmental definitions usually determine the degree of departure from average conditions.
- Water resource indicators measure severity in terms of the impact of the drought on the use of water in its broadest sense, for example, impact on water supply for domestic or agricultural use, impact on groundwater recharge, abstractions and surface levels, impact on fisheries or impact

on recreation. This implies that an element of human interference as an increased water demand or mismanagement of water supply, as well as a lack of rainfall or runoff determines the drought. Hence, there is shortage of water to meet water supply needs.

Wilhite and Glantz (1985) categorise drought definitions into conceptual (definitions formulated in general terms) not applicable to current (i.e., real time) drought assessments, and operational. The latter category includes definitions attempting to identify the onset, severity and termination of drought episodes. In some publications (e.g. Tate and Gustard, 2000) the term operational drought is applied equivalent to water resource indicators, hence not consistent with the broad definition of Wilhite and Glantz (1985).

Throughout this work, the following definition of drought, proposed by Tallaksen and Van Lanen (2004) is considered:

> *Drought is a sustained period of below-normal water availability. It is a recurring and worldwide phenomenon, with spatial and temporal characteristics that vary significantly from one region to another*
>
> *Tallaksen and Van Lanen (2004)*

2.1.1 Drought indices

In Beran et al. (1985) a distinction is made between stream-flow droughts and low flows. The main feature of a drought is the deficit of water for some specific purpose. Low flows are normally experienced during a drought, but they feature only one element of the drought, i.e. the drought magnitude. Low flow studies are described as being analyses aimed at understanding

the physical development of flows at a point along a river at a short-term (e.g. daily time resolution). Hydrological drought analyses in terms of stream-flow deficits are studies over a season or longer time periods and in a regional context. However, also short-term (less than a season) stream-flow deficits might be defined as droughts and treated at a fixed point in space.

The most widely used drought indicators are the *Palmer Drought Severity Index* (PDSI) (Alley, 1984; Wayne, 1965), with temperature and precipitation as an input; the *Standardised Precipitation Index* (SPI) (McKee et al., 1993) using only precipitation; the *Standardised Precipitation Evapotranspiration Index* (SPEI) (Vicente-Serrano et al., 2010) facilitating both – the sensitivity of PDSI and the simplicity of the SPI calculation; and the *Reconnaissance Drought Index* (RDI) (Tsakiris and Vangelis, 2005) incorporating directly potential evapotranspiration.

The PDSI is a measure of long-term drought that takes into account both precipitation and temperature, and is based on the water balance equation. The PDSI is calculated for a given location using monthly temperature and precipitation data, and takes into account the soil moisture conditions at the start of the month.

In contrast, the SPI is a drought index that is based only on precipitation data. The SPI is calculated using monthly precipitation data and is designed to provide a measure of drought severity that is standardised across different time scales and locations.

Another modification of SPI was made by (Nalbantis and Tsakiris, 2009), introducing an analogous approach for stream-flow and thus capturing the hydrological droughts. (Mishra and Singh, 2010) offered a review of multiple climatological and hydrological parameters concerning drought and summarized drought modelling methods in (Mishra and Singh, 2011). (Myronidis et al., 2018b) performed stream-flow and hydrological drought

trend analysis using *Streamflow Drought Index* (SDI) (Nalbantis and Tsakiris, 2009).

The *Soil Water Supply Index* (SWSI) is designed to be an indicator of surface water conditions, Hayes (2006) described the index as "mountain water dependent," in which snowpack is a major component. The objective of the SWSI is to incorporate both hydrological and climatological features into a single index value resembling the Palmer Index for each major river basin in the state of Colorado. These values would be standardised to allow comparisons between basins. Four inputs are required for the SWSI calculation: snowpack, stream-flow, precipitation, and reservoir storage. The SWSI is computed from only the snowpack, precipitation, and reservoir storage in the winter. During the summer months, stream-flow replaces snowpack as a component within the SWSI equation. Each component has a monthly weight assigned to it depending on its typical contribution to the surface water within that basin, and these weighted components are summed to determine a SWSI value representing the entire basin.

As such, standardised indices derived by quantile mapping (figure 2.1) are better suited for cross temporal and spatial comparison rather than high quantile estimation. Depending on the viewpoint and more importantly the needs of a stakeholder it is sometimes more appropriate to employ indicators such as deficit volumes derived from the threshold method that are in absolute values to estimate return levels of drought events.

Myronidis et al. (2018a) compared ten most widely used meteorological drought indices and tracked the indicated effect of drought on stream-flow. The study found that the Standardized Precipitation Index (SPI) and the Standardized Precipitation Evapotranspiration Index (SPEI) performed best in capturing the impact of drought on streamflow, while other indices such as the Palmer Drought Severity Index (PDSI) and the Reconnaissance

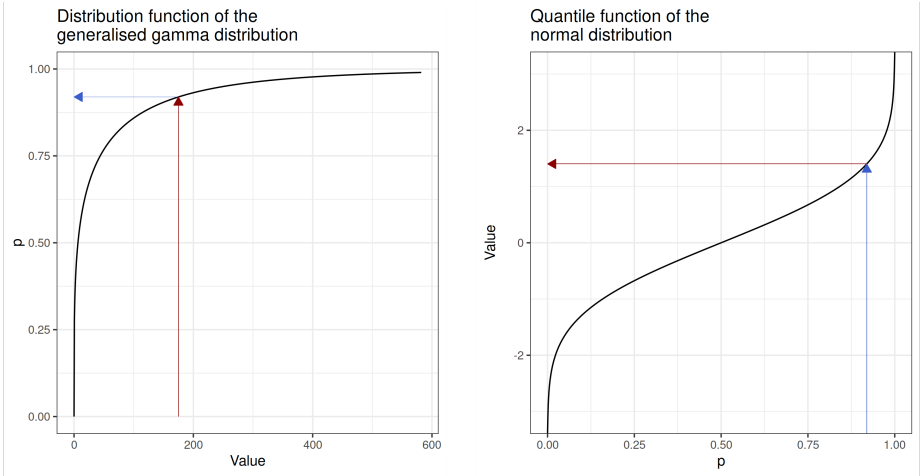

FIGURE 2.1: The process of quantile mapping - 1st probability is calculated by supplying corresponding quantile to the CDF of the precipitation (generalised gamma distribution in this specific case), 2nd calculated probability is supplied to the quantile function of the normal distribution - $\mathcal{N}(0, 1)$ - which will give corresponding quantile of the standardised index.

Drought Index (RDI) showed weaker performance. The study also found that the choice of index can have a significant impact on the severity of drought identified, with some indices underestimating the severity of drought compared to others. Finally, the authors recommend using multiple drought indices in combination to provide a more comprehensive understanding of drought conditions and their impact on streamflow.

2.1.2 Threshold method

The most frequently applied quantitative definition of a drought is based on defining a threshold, q_0, below which the river flow is considered as a drought.

The threshold level method generally study runs below or above a given threshold and was originally named method of crossing theory (Tallaksen, 2000). The method is relevant for storage/yield analysis and is associated with hydrological design and operation of reservoir storage systems. Important areas of application are hydropower and water management, water supply systems and irrigation schemes.

The method was first developed by Rice (1945) and later extended and summarised by Leadbetter (1967). Early application of crossing theory in hydrology includes Yevjevich et al. (1967), where the method is based on the statistical theory of runs for analysing a sequential time series. Statistical properties of the distribution of water deficits, run-length (drought duration), run-sum (deficit volume or severity) are recommended as parameters for at-site drought definition. Simultaneously it is possible to consider the minimum flow, and time of occurrence. The minimum flow can be regarded as a low flow measure, one of several characteristics of a stream-flow drought event. The time of drought occurrence has been given different definitions as for instance the starting date of the drought, the mean of the onset and the termination date or the date of the minimum flow. Often another drought index, the drought intensity, is defined as the ratio between drought deficit volume and drought duration.

The threshold might be chosen in a number of ways and the choice is amongst other a function of the type of water deficit to be studied (Dracup et al., 1980). In some applications the threshold is a well-defined flow quantity, e.g. a reservoir specific yield (Bonacci, 1993).

2.2 Choice of distribution

> *[...] it is important to be aware of the limitations implied by adoption of the extreme value paradigm. First, the models are developed using asymptotic arguments, and care is needed in treating them as exact results for finite samples. Second, the models themselves are derived under idealised circumstances, which may not be exact (or even reasonable) for a process under study. Third, the models may lead to a wastage of information when implemented in practice.*
>
> *Coles et al. (2001)*

The statement "it is important to be aware of the limitations implied by adoption of the extreme value paradigm" is in line with the arguments presented in Vít Klemeš' papers "Tall Tales about Tails of Hydrological Distributions" (Klemeš, 2000a,b). In this paper, Klemeš argues that there are many misconceptions and misunderstandings surrounding the use of extreme value theory in hydrology, and that caution must be exercised when applying this theory to real-world situations. Klemeš also emphasises the need to consider the limitations of extreme value models, such as the asymptotic nature of the models and the potential wastage of information when implementing them in practice. Therefore, both the statement and Klemeš' paper highlight the importance of understanding the limitations of extreme value models and exercising caution when applying them to real-world hydrological problems.

Both Boughton (1980) & Laursen (1983) saw need for only bounded distributions to be used for high quantile estimation since some numerical values are so unlikely as to be physically impossible, however Hosking and

Wallis (1997) maintain that this argument is misguided. They argue that the aim of an analysis is to estimate quantiles of return periods up to 100 years, that the estimated quantile at return period 100,000 years is "physically impossible" is of no relevance and should not be any cause for concern. Imposing the requirement that the distribution have a physically realistic upper bound may compromise the accuracy of quantile estimates at the return periods that are of real interest.

Two most common ways of assessing usability of a distribution visually are L-moment ratio diagrams and Gumbel plots. The L-moment ratio diagrams are constructed by plotting the estimated sample L-moment (see section 2.3.2 for L-moment definition) ratios versus the theoretical L-moment ratio curves for the candidate distributions. Gumbel plot is a quantile function with transformed Gumbel variate ($-log(-log(F))$) instead of probability (F) on the horizontal axis. This transformation is done in order to better visualise values with high return periods. Then, F, which is cumulative probability P of non exceedance of the mth value in n order ranked observations, can be calculated by the plotting position

$$F = \frac{m - 0.3}{n + 0.4}. \tag{2.1}$$

where m is the rank from the smallest ($m = 1$) to the largest ($m = n$) observation and n is the number of observations. This specific plotting position method (or empirical ranking method) was initially proposed by Beard (1943) and assumes that the data come from Gumble distribution. Chegodaev (1953) then rounded the numerical values of the parameters to $\alpha = 0.3$ and $\beta = 0.4$. However, different plotting positions are available and can be used in the case that the data deviate from Gunble distribution.

Another approach to see whether selected distribution is usable, is

to employ statistical tests (test description and critical values estimation methods are described in sections 2.5.2 and 2.5.2.1). According to extreme value theorem, properly scaled and independent seasonal and annual (i.e. block) maxima, converge towards *Generalised Extreme Value distribution* (GEV). For threshold methods *Generalised Pareto Distribution* (GPD) is much better suited due to inherent intermittent nature of threshold method results. (Tallaksen and Hisdal, 1997) used Generalized Extreme Value distribution, Generalized Pareto, three-parameter lognormal and Pearson type III distributions to describe drought durations and deficit volumes. This study did not add much in the context of different drought indices discussed here, but it is important for the discussion of which CDF to use to estimate large quantiles of said variables.

As discussed droughts are more rare than floods and an extreme event does not necessarily occur every year. It there are block maxima with no drought event, it is necessary to make a correction for years without droughts. This can be done by following Stedinger (1993) that considers the model with a probability mass concentrated in zero

$$F^*(x) = \begin{cases} p_0 & \text{if} \quad x = 0 \\ p_0 + (1 - p_0)F(x) & \text{if} \quad x > 0 \, . \end{cases} \tag{2.2}$$

where p_0 is the probability of year without drought. We estimate p_0 as the proportion of zero drought years (Engeland et al., 2004).

2.2.1 Generalised Extreme Value distribution

The GEV distribution is a family of continuous probability distributions developed within extreme value theory to combine the Gumbel, Fréchet and reverse Weibull distributions. By the extreme value theorem the GEV distri-

bution is the only possible limit distribution of properly scaled maxima of a sequence of independent and identically distributed random variables (Hosking and Wallis, 2005).

GEV is a three-parameter distribution given by

$$F(x) = exp \begin{cases} -\left[1 + \kappa(\frac{x-\xi}{\alpha})\right]^{-\frac{1}{\kappa}}, & \kappa \neq 0, \\ -exp\left[-(\frac{x-\xi}{\alpha})\right], & \kappa = 0. \end{cases} \tag{2.3}$$

where $\xi \in R$ is location, $\alpha > 0$ scale and $\kappa \in R$ shape parameter with range of x: $\xi \leq x \leq \xi + \alpha/\kappa$ if $\kappa > 0$ and $\xi \leq x \leq \infty$ if $\kappa \leq 0$.

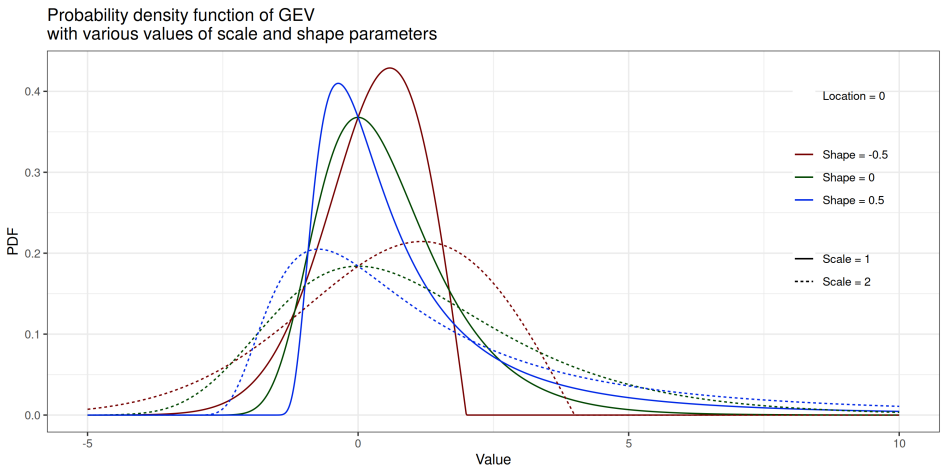

FIGURE 2.2: *Probability Density Function* (PDF) of the *Generalised Extreme Value distribution* (GEV) showing various values of parameters

2.2.2 Generalised Pareto Distribution

Generalised Pareto Distribution (GPD) is a continuous probability distri-

butions, which has been often used to model the tails of another distribution (e.g., Papalexiou et al., 2013). Although it is defined by three parameters: location, scale and shape parameters (Coles et al., 2001; Hosking and Wallis, 1997), it has been shown that can be defined by only scale and shape or just by its shape parameter (Hosking and Wallis, 1987). The three-parameter GPD is formulated as

$$F(x) = \alpha^{-1} e^{-(1-\kappa)y}, \quad y = \begin{cases} -\kappa^{-1} \log\left[1 - \frac{\kappa(x-\xi)}{\alpha}\right], & \kappa \neq 0, \\ \frac{(x-\xi)}{\alpha}, & \kappa = 0. \end{cases} \tag{2.4}$$

where $\xi \in R$ is location, $\alpha > 0$ scale and $\kappa \in R$ shape parameter with range of x: $\xi \leq x \leq \xi + \alpha/\kappa$ if $\kappa > 0$ and $\xi \leq x \leq \infty$ if $\kappa \leq 0$.

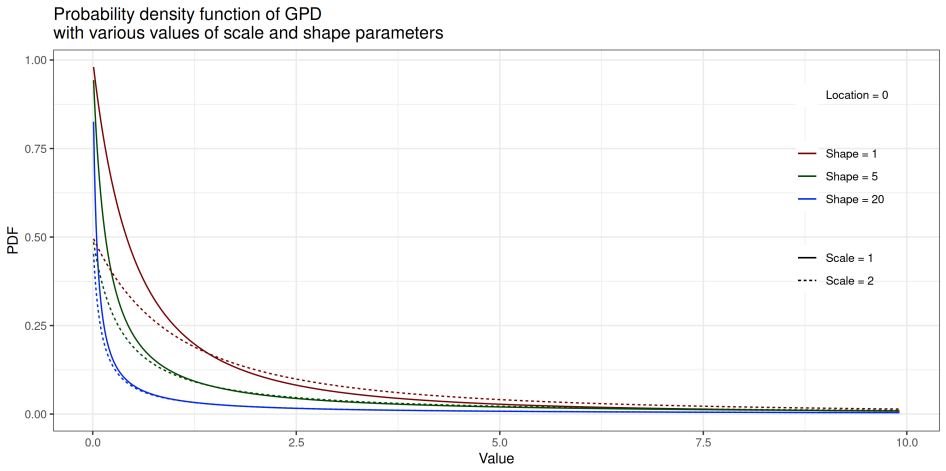

FIGURE 2.3: *Probability Density Function* (PDF) of the *Generalised Pareto Distribution* (GPD) showing various values of parameters

2.3 Parameter estimations methods

Several methods may be used for parameter estimation. Methods can be divided into the parametric and non-parametric ones. This work deals only with the parametric methods. These methods may produce quite different results with the small sample of extremes that is usually available, even though their results become asymptotically identical as the number of observed extremes becomes large (Von Storch and Zwiers, 2001).

2.3.1 Maximum likelihood

A general and flexible method of estimation of the unknown parameter θ_0 within a family F is maximum likelihood. Each value of $\theta \in \Theta$ defines a model in F that attaches (potentially) different probabilities (or probability densities) to the observed data. The probability of the observed data as a function of θ is the likelihood function. Values of θ that have high likelihood correspond to models which give high probability to the observed data. The principle of maximum likelihood estimation is to adopt the model with greatest likelihood, since this is the one that assigns highest probability to the observed data.

Suppose $x_1, ..., x_n$ are independent realisations of a random variable having *Probability Density Function* (PDF) $f(x; \theta_0)$, the likelihood function is

$$L(\theta) = \prod_{i=1}^{n} f(x; \theta_0). \tag{2.5}$$

It is often more convenient to take logarithms and work with the log-

likelihood function

$$l(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f(x; \theta_0). \tag{2.6}$$

Both equation 2.5 and 2.6 generalise the situation where the X_i are independent, but not necessarily with identical distributions. In this case, denoting the PDF of X_i by $f_i(x_i; \theta)$,

$$L(\theta) = \prod_{i=1}^{n} f_i(x_i; \theta_0). \tag{2.7}$$

and

$$l(\theta) = \sum_{i=1}^{n} \log f_i(x_i; \theta_0). \tag{2.8}$$

is obtained. Generally, if $F = \{f(x; \theta_0) : \theta \in \Theta\}$ denotes a family of joint PDFs for a set of (not necessarily independent) observations $x = \{x_1, ..., x_n)$, then the likelihood is $L(\theta) = f(x; \theta_0)$, regarded as a function of θ with x fixed at the observed value.

The maximum likelihood estimator $\hat{\theta_0}$ of θ_0 is defined as the value that maximises the appropriate likelihood function. Since the logarithm function is monotonic, the log-likelihood takes its maximum at the same point as the likelihood function, so that the maximum likelihood estimator also maximises the corresponding log-likelihood function. In practise (software implementation of the maximum likelihood) however values of log-likelihood are transformed to its negatives and are minimised.

2.3.2 L-Moments

L-moments are measures of the location, scale and shape of probability distributions. They are analogous to the conventional moments but can be estimated by linear combinations of order statistics. L-moments are related to expected values of order statistics.

Let X be a random variable and let $X_{j:n}$ denote an order statistic, a random variable distributed as the jth smallest element of a random sample of size n drawn from the distribution of X. Hosking (1990) defined the L-moments of X to be the quantities

$$\lambda_r = r^{-1} \sum_{j-0}^{r-1} (-1)^j \begin{pmatrix} r-1 \\ j \end{pmatrix} E(X_{r-j:r}). \tag{2.9}$$

where $X_{j:n}$ denotes the jth order statistic in an independent sample of size n from the distribution of X and E is the expected value.

2.3.3 Norms

Another convenient yet not widely used method is minimisation of the *modified Mean Squared Error* (mMSE) norms introduced in Papalexiou et al. (2013)

$$N1 = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\overline{F}(x_{(i)})}{\overline{F}_n(x_{(i)})} - 1 \right)^2, \tag{2.10}$$

$$N2 = \frac{1}{n} \sum_{i=1}^{n} \left(\overline{F}(x_{(i)}) - \overline{F}_n(x_{(i)}) \right)^2, \tag{2.11}$$

$$N3 = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{x_u}{x(i)} - 1 \right)^2, \tag{2.12}$$

$$N4 = \frac{1}{n}\sum_{i=1}^{n}\left(x_u - x(i)\right)^2. \tag{2.13}$$

where $\overline{F}(x_{(i)})$ are theoretical values of the distribution under study, $\overline{F}_n(x_{(i)})$ are empirical values calculated using Weibull plotting position and, $x_u = Q(u)$ is value predicted by the quantile function Q of the distribution for u equal to the empirical probability of $x_{(i)}$ according to the Weibull plotting position. These norms are implemented in Strnad et al. (2020) using Nelder-Mead Simplex algorithm for minimisation.

2.4 Regional frequency analysis

Since drought is a phenomenon that needs a long period of time to evolve and is an intermittent process, another limitation is usually short length of available observed data (Fekete et al., 2000). Due to the rarity of extreme events, modelling of drought extremes is related to large uncertainties. One possible way how to prolong the study period is the use reconstructed climate fields or climate models to obtain sufficient period length, this may however introduce new uncertainty. To cope with the uncertainty issue, it is possible to employ methods such as *Regional Frequency Analysis* (RFA) (Hosking and Wallis, 1997).

RFA uses data from a number of measuring sites. A "region" is a group of sites each of which is assumed to have data drawn from the same frequency distribution after scaling the at-site data with corresponding scaling factor (Hosking and Wallis, 1993). The RFA consists of two steps. In the first one the homogeneous regions are identified. The second establishes a regional frequency distribution curve for each region. A region is considered to be homogeneous when the sites belonging to it exhibit a similar behaviour when

the non-dimensional local frequency distribution curves have similar shapes within a sampling error.

A convenient way of pooling summary statistics for RFA from different data samples is the index-flood technique (Dalrymple, 1960). The term index-flood was coined because early applications of the pooling algorithm were to flood data in hydrology. The application of the method to low flows was termed *regional frequency analysis* (Tallaksen and Van Lanen, 2004) and *index low flow method* (Blöschl et al., 2013).

2.4.1 Index-flood

Suppose the data are available at N sites, with site i having sample size n_i and observed data Q_{ij},$j = 1, ..., n_i$. Let Qt $(F), 0 < F < 1$, be the quantile function of the frequency distribution at site i. The key assumption of an index-flood procedure is that the sites form a homogeneous region, that is, that the frequency distributions of the N sites are identical apart from a site-specific scaling factor, the index flood. This can be written as

$$Q_i(F) = \mu_i q(F), \qquad i = 1, ..., N \tag{2.14}$$

here the index flood μ_i is considered to be the mean of the at-site frequency distribution, however any location parameter of the distribution can be used instead. For example the 90% quantile $Q_i(0.9)$ was used by Smith (1989). The remaining factor in 2.14, $q(F)$, is the regional growth curve, a dimensionless quantile function common to every site. It is the quantile function of the regional frequency distribution, the common distribution of the Q_{ij}/μ_i (Hosking and Wallis, 2005).

The index flood is naturally estimated by the sample mean of the data at site i. Other location estimators such as the median or a trimmed mean

could be used instead. The dimensionless rescaled data $q_{ij} = Q_{ij}/\mu_i, j = 1, ..., n_i, i = 1, ..., N$, are the basis for estimating the regional growth curve $q(F), 0 < F < 1$. It is usually assumed that the form of $q(F)$ is known apart from p undetermined parameters $\theta_1, ..., \theta_p$ so it can be written $q(F)$ as $q(F; \theta_1, ..., \theta_p)$. These parameters may for example be the coefficient of variation and the skewness of the distribution, or the L-moment ratios $\tau, \tau_3, ...,$. The mean of the regional frequency distribution is not an unknown parameter, because by taking μ_i in 2.14 to be the mean of the frequency distribution at site i it is ensured that the regional frequency distribution has mean of 1. In this approach the parameters are estimated separately at each site, the site-i estimate of θ_k being denoted by $\hat{\theta}_k^{(i)}$ These at-site estimates are combined to give regional estimates

$$\hat{\theta}_k^R = \frac{\sum_{i=1}^N n_i \hat{\theta}_k^{(i)}}{\sum_{i=1}^N n_i} \tag{2.15}$$

This is a weighted average, with the site-i estimate given weight proportional to n_i because for regular statistical models the variance of $\hat{\theta}_k^{(i)}$ is inversely proportional to n_i. Substituting these estimates into $q(F)$ gives the estimated regional growth curve $\hat{q}(F) = q(F; \hat{\theta}_1^R, ..., \hat{\theta}_p^R)$. This method of obtaining regional estimates is essentially that of Wallis (1980), except that the weighting proportional to n_i is a later addition, suggested by Wallis (1982). Somewhat different methods were used by (Dalrymple, 1960).

The quantile estimates at site i are obtained by combining the estimates μ_i and $q(F)$

$$\hat{Q}_i(F) = \hat{\mu}_i \hat{q}(F) \tag{2.16}$$

This index-flood procedure makes the following assumptions.

- Observations at any given site are identically distributed,
- observations at any given site are serially independent,
- observations at different sites are independent,
- frequency distributions at different sites are identical apart from a scale factor and
- the mathematical form of the regional growth curve is correctly specified.

The first two assumptions are plausible for many kinds of data, particularly for annual totals or extremes, which are free from seasonal variations. It is a basic assumption of most methods of frequency analysis that the events observed in the past are likely to be typical of what may be expected in the future. This assumption may be undermined when obvious sources of time trends are present; frequency distributions for stream-flow data, for example, are affected by changes in land use and by artificial regulation of the flow. When sites affected by such obvious sources of non-stationarity are removed from the data-set, the assumption of identical distributions for a site's observations is often reasonable (Hosking and Wallis, 2005).

The effect of serial dependence on at-site frequency analysis has been investigated by Landwehr et al. (1979) and MacMahon and Srikanthan (1982). They considered frequency distributions of extreme-value type I and log-Pearson type III, respectively, and found that serial dependence caused a small amount of bias and a small increase in the standard error of quantile estimates. It can be concluded that a small amount of serial dependence in annual data series has little effect on the quality of quantile estimates (Hosking and Wallis, 2005).

The last three assumptions are unlikely to be satisfied by environmental data. Correlation between nearby sites may be expected for many kinds of data. Meteorological events such as storms and droughts typically affect an area large enough to contain more than one measuring site, and the event magnitudes at neighbouring sites are therefore likely to be positively correlated. The last two assumptions will never be exactly valid in practice. At best they may be approximately attained, by careful selection of the sites that are to be regarded as forming a region and by careful choice of a frequency distribution that is consistent with the data. Therefore an index-flood procedure can be appropriate only if it is robust to physically plausible departures from these three assumptions. Research conducted by Hosking and Wallis (1988); Hosking et al. (1985); Lettenmaier and Potter (1985); Lettenmaier et al. (1987); Wallis and Wood (1985) has shown that it is possible to construct index-flood procedures that yield suitably robust and accurate quantile estimates.

An alternative approach to regional estimation is to model $\log Q$ rather than Q, basing the analysis on log-transformed data. Taking logarithms in 2.14 gives

$$\log Q_i(F) = \log \mu_i + \log q(F) \tag{2.17}$$

The index flood enters as an additive term, which makes some aspects of the analysis easier. For example, if unbiased estimators of $\log \mu_i$ and $\log q(F)$ can be found, their sum will be an unbiased estimator of $\log Q_i(F)$. The disadvantage of using log-transformed data is that low data values may become low outliers after logarithmic transformation and have an undue influence on the estimates. In applications in which estimation of quantiles in the upper tail of the distribution is of principal importance, it is particularly

unfortunate for low data values to have a strong effect on the upper tail of the estimated frequency distribution. For this reason it is generally preferable to work with the original untransformed data (Hosking and Wallis, 2005).

2.4.2 Selection of homogeneous regions

As descried in the previous section, the main assumption of the index flood method is that the probability distributions of extremes at different sites in the region are identical, except for a scaling factor. Therefore the selection of homogeneous regions is an important step in RFA.

Nathan and McMahon (1990) claims that spatial pooling methods traditionally encompassed geographically contiguous areas. However, if the subregions are defined on the basis of hydrological similarity or basin characteristics it is not necessary for region within a given grouping to be geographically contiguous.

There is a number of problems associated with spatial pooling in general. The first issue common to all pooling techniques is related to the selection of variables used to assess the degree of similarity between different regions. Furthermore, literally any group of variables is capable of generating clusters, and it is necessary to select variables according to their relevance to the problem.

The selection of the most suitable algorithm for a given data-set and optimisation criterion is critical, and requires careful consideration by the user. Nevertheless, these issues can be mitigated by incorporating prior knowledge of the data into the spatial pooling process. Furthermore, determining the optimal number of regions for partitioning the data is a common challenge faced by most algorithms. However, this can be addressed by utilising objective approaches like gap statistics. It is worth noting that some methods,

such as region of influence, may not require explicit determination of the number of regions, and thus can alleviate this issue.

The methodology used to identify a set of homogeneous regions used to estimate regional quantiles is summarised by the steps below

- Selection of predictor variables
- Selection of the most appropriate cluster analysis
- Identification of homogeneous regions
- Evaluation of homogeneous groups

2.4.2.1 Gap statistics

Gap statistic is an objective method for estimating the optimal number of clusters in a clustering algorithm proposed by Tibshirani et al. (2001). The goal of gap statistic is to identify the number of clusters that best captures the underlying structure of the data, without overfitting or underfitting. The method works by comparing the *Within-cluster Sum of Squares* (WSS) of the data to its expected value under a null reference distribution. WSS is calculated as the sum of the squared Euclidean distances between each data point and the centroid of the cluster it belongs to. The null reference distribution is created by generating a large number of random data sets that are similar to the original data, but have no underlying structure or clustering. The expected WSS value of the null reference data is then used as a benchmark to compare the WSS of the original data. The optimal number of clusters is identified as the point where the gap between the observed WSS and the expected WSS is maximized as seen on figure 2.4.

To compute the gap statistic, the algorithm first determines the optimal number of clusters for a range of values. For each number of clusters, it computes the WSS of the original data, as well as the expected WSS of the null reference data. The gap statistic is then calculated as the difference between the observed WSS and the expected WSS, minus a correction term that accounts for the variability in the null reference data. The optimal number of clusters is identified as the value where the gap statistic is maximised. The gap statistic method has been shown to be effective in a variety of clustering applications and is widely used in practice.

The algorithm can be surmised as follows

1. Cluster the observed data, varying the number of clusters from $k = 1, ..., k_{max}$, and compute the corresponding total WSS variation W_k, which is a measure of the clustering quality. This step helps to identify the range of k values that should be considered for further analysis.

2. Generate B reference data-sets with a random uniform distribution. Cluster each of these reference data-sets with varying number of clusters $k = 1, ..., k_{max}$, and compute the corresponding total intra-cluster variation for each value of k (W_{kb}).

3. The estimated gap statistic is computed as the difference between the observed WSS for the observed data and the expected WSS for the reference data-sets under the null hypothesis: $Gap(k) = \frac{1}{B}\sum_{b=1}^{B} log(W_{kb}^{*}) - log(W_k)$. The expected WSS is the average of the WSS values obtained from the reference data-sets. The gap statistic is computed for each value of k, and the standard deviation s of the statistics is also computed.

4. Choose the number of clusters as the smallest value of k such that the gap statistic is within one standard deviation s of the gap at $k+1$: $Gap(k) \geq Gap(k+1) - s_{k+1}$. This step helps to identify the number of clusters that provides the best balance between clustering quality and model complexity.

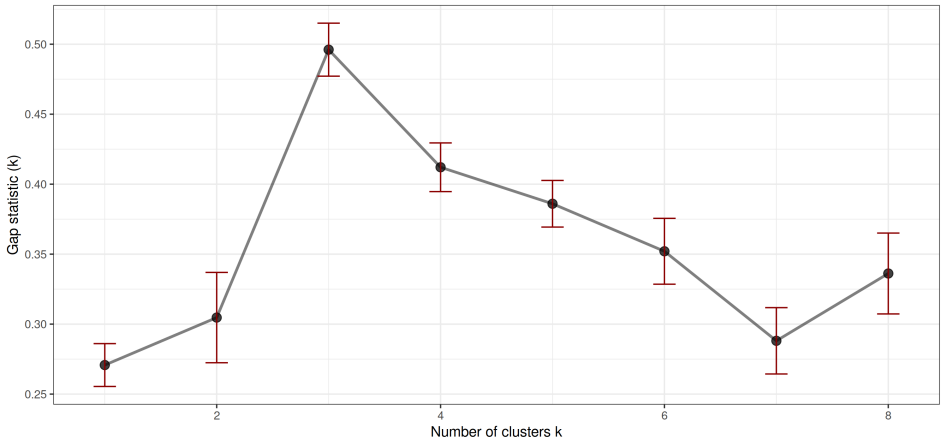

FIGURE 2.4: Gap statistic estimated for the randomly generated data-set, determining the ideal number of clusters for K-means method. (Clustered data-set can be seen on figure 2.5)

2.4.2.2 K-means

K-means clustering is a popular unsupervised learning algorithm, developed between years 1975 and 1977 and surmised in Hartigan and Wong (1979) and used to partition a given data-set into K clusters or groups. The goal of this

algorithm is to minimise the sum of squared distances between data points and their corresponding cluster centres. The algorithm starts by randomly selecting K initial cluster centres, which are usually chosen from the data points themselves. It then iteratively performs two steps until convergence is reached. In the first step, each data point is assigned to the cluster with the nearest centroid based on the Euclidean distance between the data point and the centroid. In the second step, the centroid of each cluster is updated as the mean of all the data points assigned to that cluster. The algorithm terminates when the centroids no longer change or a predefined maximum number of iterations is reached.

The choice of K is an important decision when applying k-means clustering. If K is too small, the clusters may be too broad and not capture the underlying structure of the data. If K is too large, the algorithm may create smaller, less meaningful clusters. In addition to the above described gap statistic, various other methods can be used to estimate the optimal number of clusters such as the elbow method or silhouette method. In addition, k-means clustering can be sensitive to the initial random selection of cluster centres and may converge to sub-optimal solutions. To mitigate this issue, the algorithm is often run multiple times with different initialisation, and the best result is chosen based on some predefined criteria (Arthur and Vassilvitskii, 2007; Jain, 2010).

Step by step the algorithm can be described as

1. Choose the number of clusters - K.
2. Select random K points - the centroids
3. Assign each data point to closest centroid that forms K clusters.
4. Compute and place the new centroid of each region.

5. Reassign each data point to new cluster.

6. Repeat steps 3 and 5 until convergence, which occurs when the assignments of data points to centroids no longer change.

FIGURE 2.5: Left panel show randomly generated data-set drawn from multivariate normal distribution consisting of three groups (denoted by shapes); Right panel shows the same data-set clustered using K-means algorithm (clusters are highlighted using various colours)

2.4.2.3 Hierarchical clustering

Hierarchical clustering is as simple as K-means, but instead of there being a fixed number of clusters, the number changes in every iteration. It creates a tree-like structure of nested clusters, also called a dendrogram, where the top-level node represents all the data, and each subsequent level represents a subset of the data that is increasingly similar. There are two main types of

hierarchical clustering: agglomerative and divisive. Agglomerative clustering starts with each data point as a separate cluster and iteratively merges the closest pairs of clusters until only one cluster remains. Divisive clustering starts with all data in one cluster and iteratively splits it into smaller clusters until each data point is in its own cluster.

In agglomerative clustering, each data point starts as a separate cluster. Then, the algorithm computes a distance matrix between all pairs of clusters and merges the two closest clusters. The distance between clusters can be computed in various ways, such as the minimum distance between any two points in the two clusters, the maximum distance, or the average distance. The algorithm continues merging the closest clusters until all the data points are in one cluster, or until a stopping criterion is met, such as a desired number of clusters or a threshold distance. The result is a dendrogram that shows the hierarchical structure of the clusters. The dendrogram can be cut at any level to obtain a specific number of clusters. Hierarchical clustering has the advantage of being able to capture the nested structure of clusters and can be useful in exploratory data analysis, but it can be computationally demanding for large data-sets (Szekely et al., 2005).

Put in steps, the algorithm works as follows:

1. Begin by considering each data point as a separate cluster.
2. Calculate the distance between all pairs of data points.
3. Merge the two closest clusters into a single cluster.
4. Recalculate the distance between the new cluster and all other clusters.
5. Repeat steps 3 and 4 until all data points are in a single cluster.
6. Plot a dendrogram to visualise the clustering process.

7. Choose a cutoff point on the dendrogram to determine the number of clusters.

2.4.2.4 Self-Organizing Maps

Self-Organising Map (SOM) algorithm is an iterative process, that transforms the original data-set to a smaller representative set of nodes. The subset is usually presented through a two-dimensional output layer, where each node corresponds to a group of members of the original data-set that share some features (unified-distance matrix or U-Matrix; Ultsch and Siemon (1990)). After a large number of iterations, each data point would be assigned to a specific node, with adjacent or neighbouring nodes of the U-Matrix representing points being more similar than the distant ones. This allows for easier visualisations of the data space, not only presenting clusters with similar properties, but also their non-linear relationships. Another interesting concept in SOM is that not only the number of classes is presented, but also the features that define each class.

1. Initialise the SOM network:
    - Set the number of neurons
    - Set neuron initial weights and topology
    - Set the learning parameters
2. Select a data point randomly from the data-set.
3. Calculate the distance between the input vector and each neuron's weight vector.
4. Find the neuron with the smallest distance to the input vector, known as the *Best Matching Unit* (BMU).

5. Update the BMU and its neighbouring neurons' weights to move closer to the input vector. The amount of weight adjustment decreases with distance from the BMU.

6. Repeat steps 2-5 for each input vector in the data-set, adjusting the learning rate and neighbourhood size at each iteration.

7. Visualise the resulting SOM to identify clusters and patterns in the data.

An advantage of the method is that neither the number of classes nor their range is determined a priory but results from the process itself. The number of nodes of the SOM is predefined though, with no single method for its determination. The most common practice is based on the comparison of differently sized SOMs and the selection of the one that minimises homogeneity measure, while at the same time preserving noticeable levels of clustering and offers a substantial comprehensibility (Chang et al., 2010; Ley et al., 2011; Rousi et al., 2017).

2.5 Model assessment

Various tools can be used to check the resulting RFA model. In particular, the homogeneity of the regions and the goodness-of-fit for the distribution used in the RFA should be assessed.

2.5.1 Discordancy

One way to test the homogeneity assumption of pooled regions is through the local discordancy measure, here denoted as D_i, as proposed by (Hosking and Wallis, 1993). This measure assesses the similarity between the L-moment

ratios of a site i and those of the entire pooling group, and identifies sites with L-moment ratios that are significantly different from the group average.

A formal definition of discordance (Hosking and Wallis, 1997) for N sites is

$$\bar{\mathbf{u}} = N^{-1} \sum_{i=1}^{N} \mathbf{u}_i. \tag{2.18}$$

which is the group average with $\mathbf{u}_i$ being a transpose vector containing values of L-moment ratios τ, τ_3 and τ_4 for site i,

$$\mathbf{A} = \sum_{i=1}^{N} (\mathbf{u}_i - \bar{\mathbf{u}})(\mathbf{u}_i - \bar{\mathbf{u}})^T. \tag{2.19}$$

is matrix of sums of squares and cross-products and

$$D_i = \frac{1}{3} N (\mathbf{u}_i - \bar{\mathbf{u}})^T \mathbf{A}^{-1} (\mathbf{u}_i - \bar{\mathbf{u}}). \tag{2.20}$$

is discordance measure D for site i. The criterion for finding site as discordant is an increasing function of the number of sites in the region. This is because large regions are more likely to contain sites with large values of D. However it is recommended to regard any site with $D_i > 3$ as discordant, since such sites have the L-moment ratios that are markedly different from the average for the other sites in the region (Hosking and Wallis, 2005).

2.5.2 Anderson-Darling test

In contrast to the discordancy measure which serves as homogeneity check, *Anderson–Darling test* (A^2) is a test that is primarily used to asses the consistency between the data and fitted distribution (it can however be used in context of homogeneity test as well within the region of influence scope).

A^2 is a modification of the Cramér–von Mises test (Cramér, 1928; Smirnov, 1936; Von Mises, 1931). It differs from the Cramér–von Mises test in such a way that it gives more weight to the tails of the distribution (Farrell and Rogers-Stewart, 2006). A^2 is the most powerful empirical CDF test (Masqat, 2003). The A^2 statistic belongs to the quadratic class of the empirical CDF statistic in which it is based on the squared difference $[F_n(x) - F(x)]^2$.

Anderson and Darling (1954) defined the statistic test as

$$A^2 = n \int_{-\infty}^{\infty} \frac{[F_n(x) - F(x)]^2}{F(x)[1 - F(x)]} \mathrm{d}F(x). \tag{2.21}$$

where F is theoretical CDF under the null hypothesis and F_n is empirical cumulative distribution function.

2.5.2.1 Estimation of critical values

Critical values for the *Anderson–Darling test* (A^2) depend on the values of the parameters of the distribution and therefore are often derived using bootstrap resampling as suggested by Davison and Hinkley (1997) and used by Hanel et al. (2009). In addition, in the RFA context the at-site tests do not provide information whether the model is adequate since even in the case of a correct model it is expected that some sites may not pass the test. Therefore more complex resampling is needed to provide "global" critical value.

Let $t(s)$ be the value of A^2 calculated for site $s (s = 1, ..., S)$ and let $t_b^*(s)$ be the value of A^2 from bootstrap sample b $(b = 1, ..., B)$ for this site. For a chosen significance level α_{LOC}, the local critical values $c^{\alpha_{LOC}}(s)$ are obtained for each site as the kth smallest value $t_{(k)}^*(s)$ of the $t_b^*(s)$, where $k = (1 - \alpha_{LOC})(B + 1)$.

The determination of the global critical values of the A^2 requires simulation from the model under the null hypothesis. In particular, the preservation of spatial dependence is important. This is done by bootstraping the data for a certain year simultaneously, rather than bootstrapping the data of the sites individually (Faulkner and Jones, 1999; Kharin et al., 2007). Let $c_{-b}^{\alpha_{LOC}}(s)$ be the local critical values that we get if bootstrap sample b is excluded. The bootstrap estimate of the global error rate α_{GLOB} is obtained as

$$\alpha_{GLOB} = \frac{\#\{b : \left[t_b^*(s) \geq c_{-b}^{\alpha_{LOC}}(s), \text{ for any } s\right]\}}{b} \tag{2.22}$$

where $\#\{b : A_b\}$ is the number of b for which A_b is true. This error rate can be calculated using the fact that bootstrap sample b fulfils the condition $\left[t_b^*(s) \geq c_{-b}^{\alpha_{LOC}}(s), \text{ for any } s\right]$ if and only if rank $\left[t_b^*(s)\right] \geq k = (1 - \alpha_{LOC})(B + 1)$ for at least one s. Thus, if the values of $t_b^*(s)$ are stored in a matrix with stations in columns and bootstrap samples in rows, then we first calculate the columnwise ranks and subsequently the proportion of rows in which the maximum rank is greater than or equal to k. The value of k is chosen such that α_{GLOB} is as close as possible to the desired global significance level.

CHAPTER 3

Simulation Experiments

In order to assess methods described in sections 2.2–2.4.1 simulation (bootstrapping) experiments were designed. The experiments aim to evaluate the performance of parameter estimation algorithms in different scenarios. This can help identify the strengths and weaknesses of these algorithms and provide insights into their accuracy and reliability. The experiments examine the effects of sample size on parameter estimation. Larger sample sizes are often associated with more accurate parameter estimates, however the scarcity of data usually leads to unreliable results and higher uncertainty thus requiring careful consideration. Understanding the relationship between sample size and estimation accuracy can help improve the accuracy predictions and its application in real world scenarios.

The bootstrap is a resampling method that can be used to estimate the distribution of a test statistic or estimator. Essentially, it involves treating the available data as if they represent the entire population, and

then generating multiple "bootstrapped" samples by drawing observations from this data with replacement. Under certain conditions, the bootstrap can yield a reliable approximation to the distribution of interest, and this approximation is often more accurate than those obtained from traditional asymptotic theory. As such, the bootstrap is a useful tool for situations where calculating the asymptotic distribution of an estimator or test statistic is difficult or impractical, and provides a way to substitute computational methods for more traditional mathematical analysis (Efron and Tibshirani, 1994).

3.1 Fitting methods

The first part of the experiment aims to investigate the behaviour of parameter estimation algorithms. Specifically, the study focuses on evaluating the performance of these algorithms as sample size increases. Additionally, the experiment assesses the algorithm's performance in cases where the "wrong" target distribution is chosen. This can occur due to the subjective nature of selecting the appropriate distribution for a given data-set, and it is important to evaluate the robustness of the algorithms in such scenarios. The results of this part of the experiment will provide insights into the behaviour of parameter estimation algorithms and their limitations in different scenarios, contributing to the development of more accurate and reliable methods for drought high quantile estimation.

Methods described in previous chapter come with their own strengths and weaknesses. Some methods may be more appropriate for certain types of data or modelling scenarios, while others may be more robust or efficient. Understanding the strengths and weaknesses of fitting methods can help make informed choices when selecting a method for their particular application.

Maximum Likelihood:

Pros:

- Often results in the most efficient estimators (in terms of being unbiased and having the smallest variance).
- Well-established method with a large body of literature.
- Can be used with a wide range of distributional assumptions.
- Can be easily modified to account for non-stationarity.

Cons:

- Assumes a specific parametric form for the distribution, which may not always be appropriate or accurate.
- Can be sensitive to outliers.
- Can be computationally demanding, especially for complex distributions.

L-moments:

Pros:

- Provides a method to estimate distributional parameters without specifying a particular distribution.
- Generally less sensitive to outliers than maximum likelihood.
- Can be used with a wide range of distributions.
- Works well with smaller sample sizes

Cons:

- Requires more computation than maximum likelihood, particularly for distributions with more than three parameters.
- Requires estimation of higher-order moments, which can be difficult with limited data.

Norms

Pros:

- Simple to compute and implement.
- Can be used with a wide range of distributions.

Cons:

- May not be as efficient as other methods.
- Can be sensitive to the choice of reference distribution and specific norm.

The performance of the three fitting methods described in Section 2.3 was evaluated in various conditions using a designed experiment, which involved the following steps:

1. A data sample of size n was generated with specified set of distribution parameters.
2. The sample was fitted using the three methods described in Section 2.3.
3. For each fit (fitting method), the Anderson-Darling statistic (A^2) was estimated according to Sections 2.5.2 and 2.5.2.1, as follows:
   a) A^2 was calculated for the fitted set of parameters.
   b) New samples were generated using the fitted set of parameters.
   c) The new samples were fitted using the corresponding method.
   d) A^2 was calculated for the new sample parameters.

e) The p-values were estimated using Equation 2.22.

4. Steps 1-4 were repeated 1000 times for various sample sizes n.

5. Calculate The 5th, 25th, 50th, 75th, and 95th quantiles of the fitted parameters were calculated.

3.2 Index flood and uncertainty reduction

The choice of spatial pooling method prior to regional frequency analysis can significantly impact the results of the analysis. Spatial pooling involves aggregating data from multiple sites within a region in order to estimate a single set of frequency parameters. The choice of spatial pooling method should be made carefully based on the available data and the characteristics of the study area in order to ensure the accuracy and precision of the estimated frequency parameters. Additionally, it is important to assess the advantages and limitations of each method.

K-means clustering:

Pros:

- Computationally efficient for large data-sets.
- Easy to implement and widely used.
- Results can be easily interpreted and visualised.
- Can handle non-hierarchical structures.

Cons:

- Can be sensitive to initial cluster centres and may converge to local optima.
- Assumes clusters are spherical and equally sized.
- May not work well with noisy or overlapping data.
- Requires the user to specify the number of clusters beforehand.

Hierarchical clustering:

Pros:

- Can handle any shape and size of clusters.
- No need to specify the number of clusters beforehand.
- Results can be easily interpreted and visualized.
- Provides a hierarchy of clusters that can be useful in some applications.

Cons:

- Can be computationally expensive for large datasets.
- May not work well with noisy or overlapping data.
- Results can be affected by the choice of linkage method and distance metric.
- Once a merge is made, it cannot be undone.

***Self-Organising Map* (SOM)**

Pros:

- Can handle high-dimensional data and non-linear structures.
- Provides a topological representation of the data.
- Can handle missing data and noisy data.
- Can be used for data visualisation and dimensionality reduction.

Cons:

- Can be computationally expensive for large data-sets.
- May not work well with small data-sets.
- Can be sensitive to the choice of the SOM size and topology.
- Results can be difficult to interpret compared to other clustering methods.

To assess the impact of pooling methods on uncertainty reduction, we designed the following experiment:

1. Generate multivariate sample:
   a) Draw data from multivariate normal distribution using a set correlation structure with set seed to ensure the "base" sample is constant over the course of the experiment
   b) Use probability integral transformation to convert the sample to GEV
2. Employ a pre-defined pooling method with a specific cluster number and split the data into clusters

3. For each cluster than

    a) Fit the GEV parameters to the sample using L-moments

    b) Generate a bootstrap sample using one the methods described bellow

    c) Fit the GEV parameters to the bootstrap sample using the L-moments

    d) Estimate quantiles using the at-site quantile functions and the regional quantile function for probabilities 0.80, 0.90, 0.95, 0.98, 0.99 (i. e. return periods of 5, 10, 20, 50 and 100 years)

    e) Calculate *Inter Quantile Range* (IQR) of return levels and evaluate the relative difference between the at-site and the regional quantile function IQRs in order to quantify the reduction of uncertainties in return level (high quantile) estimation

4. Repeat steps 2 and 3 for all pooling methods

5. Repeat all of the previous steps until desired number of samples is reached (in this case 3000 times)

The bootstrapping methods employed in this study were developed to account for the spatial dependence of block maxima, a common approach in extreme value analysis (e. g. Hanel et al., 2009). However, previous work has only evaluated these methods in the absence of compound distributions (probability mass concentrated in zero with probability p_0).

To assess the performance of the bootstrapping methods for index flood estimation, the most stable process identified in the parameter fitting test described in Section 3.1 was selected. Specifically, the L-moments method for

parameter estimation and the generalized extreme value (GEV) distribution was used. The experiment consisted of the following steps:

1. Generate multivariate sample:
   a) Draw data from multivariate normal distribution using a set correlation structure with set seed to ensure the "base" sample is constant over the course of the experiment
   b) Introduce p_0 to the sample
   c) Use probability integral transformation to convert the sample to GEV
2. Fit the GEV parameters to the sample using L-moments
3. Generate a bootstrap sample using one of the methods described bellow
4. Fit the GEV parameters to the bootstrap sample using the L-moments
5. Estimate quantiles using the at-site quantile functions and the regional quantile function for probabilities 0.80, 0.90, 0.95, 0.98, 0.99 (i. e. return periods of 5, 10, 20, 50 and 100 years)
6. Calculate *Inter Quantile Range* (IQR) of return levels and evaluate the relative difference between the at-site and the regional quantile function IQRs in order to quantify the reduction of uncertainties in return level (high quantile) estimation
7. Repeat all of the previous steps until desired number of samples is reached (in this case 3000 times)

Results of the parameter estimation for different methods can be seen in figure 3.7 and the reduction of uncertainties in figure 3.8.

3.2.1 Bootstrapping methods

The three bootstrapping methods used here are defined as follows:

Method 1 (nonparametric):

1. Shuffle the block maxima of the original sample in a way that preserves spatial correlation - keep the corresponding block maxima for all sites intact when shuffling

Method 2 (parametric - p_0 structure from original sample):

1. Fit the statistical model to the original sample
2. Calculate standard normal residuals with the parameter estimates from step 1 using quantile mapping
3. Calculate the average correlation $\hat{\rho}$ of the standard normal residuals
4. Generate a sample of equicorrelated standard normal variables with correlation $\hat{\rho}$
5. Transform the sample from step 4 to the GEV scale using the parameter estimates from step 1
6. Transfer the p_0 structure from the original sample to the bootstrap sample

Method 3 (parametric - p_0 structure created using the CDF of the normal distribution):

1. Fit the statistical model to the original sample
2. Calculate standard normal residuals with the parameter estimates from step 1 using quantile mapping

3. Calculate the average correlation $\hat{\rho}$ of the standard normal residuals

4. Generate a sample of equicorrelated standard normal variables with correlation $\hat{\rho}$

5. Introduce p_0 to the bootstrap sample using $\frac{p-p_0}{1-p_0}$ where p is the probability calculated using CDF of the normal distribution and p_0 is the desired probability of missing value, after the calculation negative probabilities are replaced by zeros

6. Transform the sample from step 5 to the GEV scale using the parameter estimates from step 1

3.3 Experiments summary

Two main experiments were performed - experiment testing the performance of fitting methods and an experiment to assess the index flood method used for intermittent processes.

For the fitting methods experiments the base sample parameter values were set as follows: parameter values for GPD being location = 1, scale = 2 and shape = .1; parameters for GEV location = 10, scale = 1.5 and shape = .1; reasoning for the selection of parameter values in the fitting methods experiments for the GPD and GEV was provided by following work of Beirlant et al. (2006); Coles et al. (2001); El Adlouni and Ouarda (2008); Hosking and Wallis (1987); Hosking et al. (1985); Salvadori and Michele (2006).

The index flood ran with following setting: number of sites = 17, sample size for each site = 50, scaling factor values ranging from 12.21 to 17.12 and regional parameters location = 0.78, scale = 0.33 and shape = -0.06,

these values were taken directly from data-set used in Hanel et al. (2009). Correlation structure of the base sample can be seen in figure 3.1.

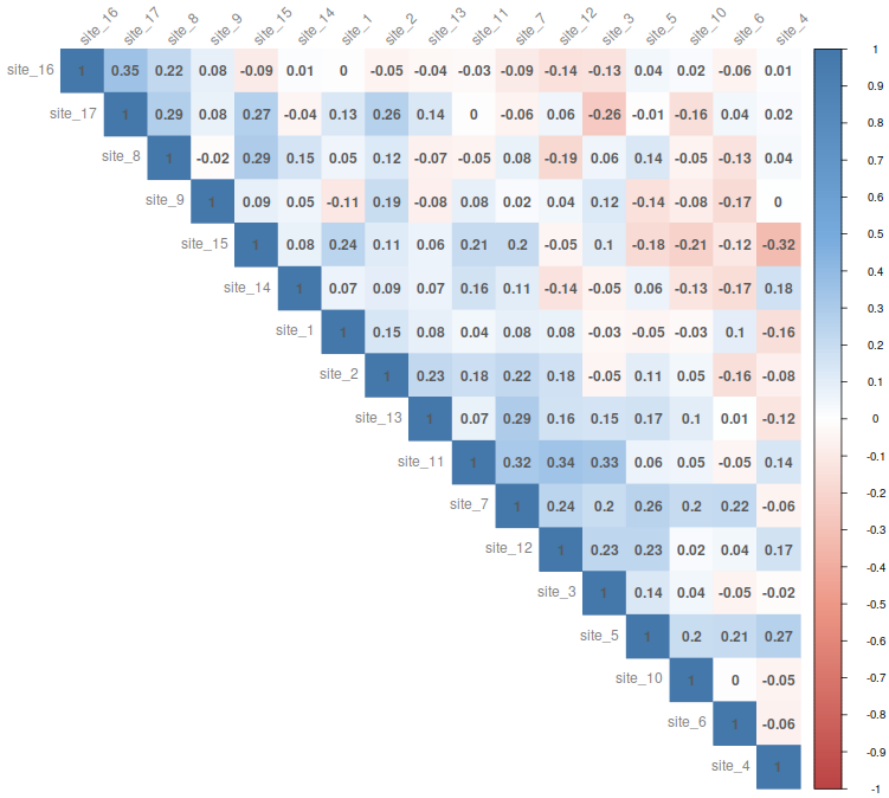

FIGURE 3.1: Base sample correlation structure used throughout the index flood experiment

Fitting methods of L-moments and the fitting norms performed as anticipated for all scenarios (converging to the same value of the parameters with with sample size increasing), maximum likelihood however exhibited unstable behaviour whenever PDF of the GPD was used. This could be caused by several factors, first being the infinite nature of the distribution tail, second is the implementation of the maximum likelihood estimator algorithm itself, another possibility is the implementation of the distribution PDF and the fact that maximum likelihood does not always have a solution for GPD (Grimshaw, 1993; Hüsler et al., 2011).

In Figure 3.2, the upper panel illustrates the expected behaviour of the fitting methods. As the sample size increases, the uncertainty decreases, and there is consistency between the fitting methods. On the other hand, the lower panel shows the p-values of the Anderson-Darling goodness-of-fit test, which is used to assess the performance and robustness of parameter estimation for different fitting methods. The maximum likelihood method shows good performance and consistency, as reflected by the low p-values obtained for various sample sizes.

Very similar image can be seen on figure 3.3 with two differences being the nonstandard behaviour of the maximum likelihood method and noticeable underperformance of fitting norms in estimation of the shape parameter for lower sample sizes (50 and less) that can be caused by the infinite variance of the GPD with the shape parameter being higher that zero or flexibility of the fitting norms that are employing minimisation of the parametric functions (CDF and/or quantile function) in order to estimate the parameters, this behaviour of the fitting norms can be seen throughout the entire experiment.

Figure 3.4 shows the erratic behaviour of the maximum likelihood when used with the GPD PDF (fitted to the GEV sample), the other two methods exhibit a bias in parameters trying to compensate in order to fit the "wrong"

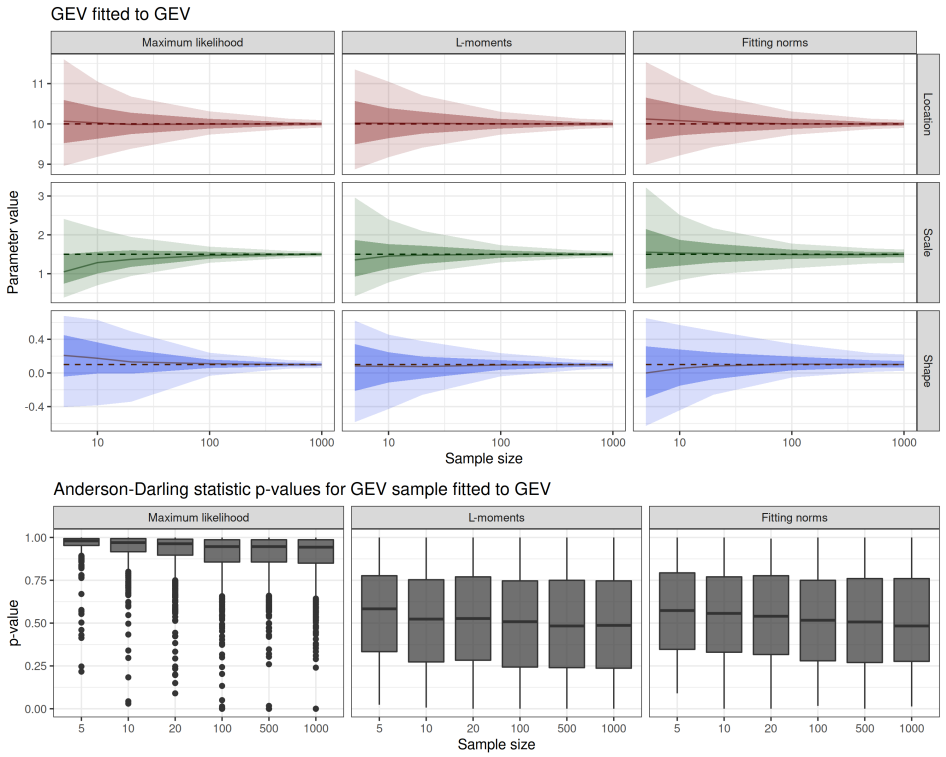

FIGURE 3.2: GEV fitted to GEV; brighter colour on the upper panel represents 90% confidence interval, darker colour 50% confidence interval, coloured line represents the median and the dashed line is the original parameter value.

distribution. When looking at the A^2, for smaller sample sizes the null hypotheses (i.e. that the data being drawn from GEV) cannot be rejected. This corresponds to the envelope of the 50th percent confidence interval overlaying the original parameter value on the panel showing the fitted parameters.

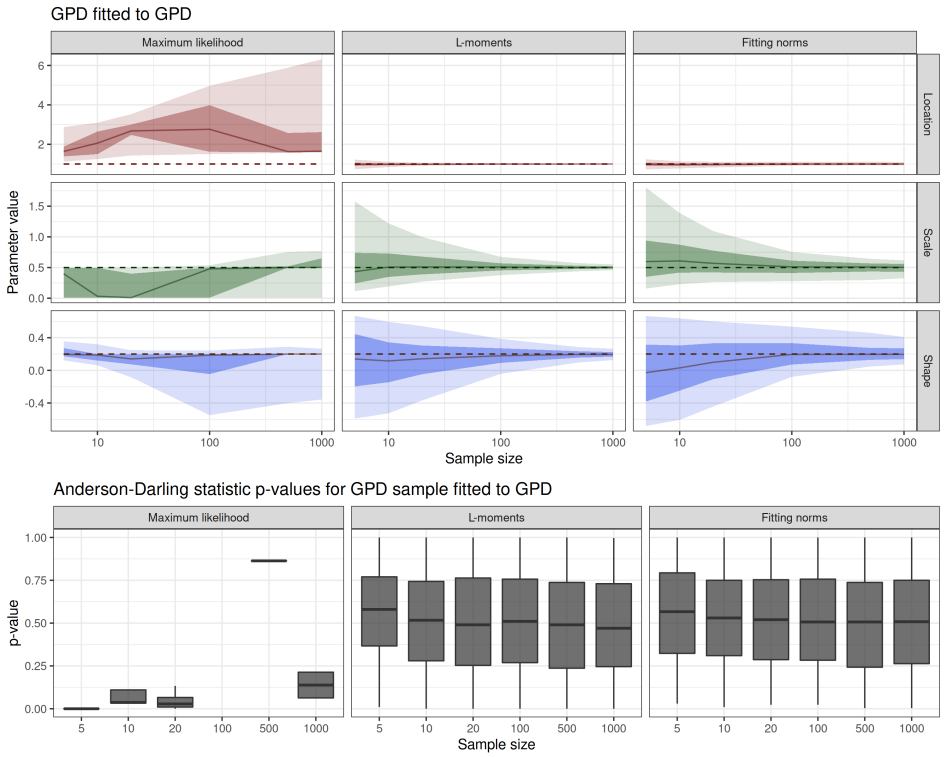

FIGURE 3.3: GPD fitted to GPD; brighter colour on the upper panel represents 90% confidence interval, darker colour 50% confidence interval, coloured line represents the median and the dashed line is the original parameter value.

Comparable behaviour can be seen on figure 3.5 with the difference of stable maximum likelihood method and increased chance of rejecting the null hypothesis of the test caused by the arguably higher flexibility of the GEV.

As mentioned previously the most stable process from the fitting method

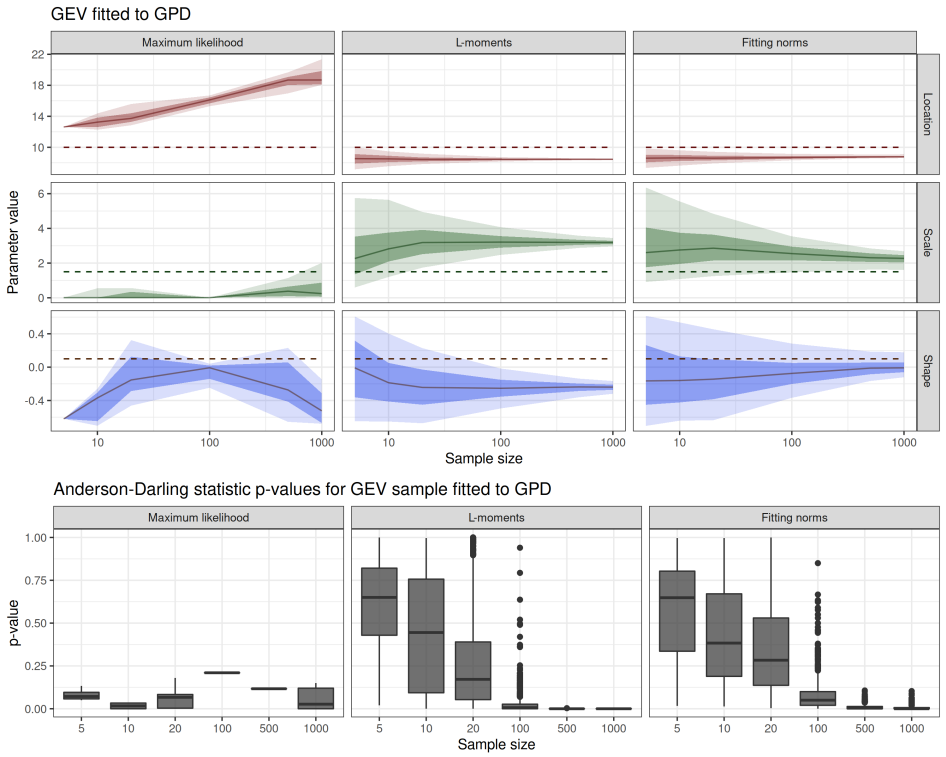

FIGURE 3.4: GEV fitted to GPD; brighter colour on the upper panel represents 90% confidence interval, darker colour 50% confidence interval, coloured line represents the median and the dashed line is the original parameter value.

tests section was chosen for the index flood method assessment - L-moments fitting of the GEV. The base sample for bootstrapping consisted of 17 sites and had a 50 year record length of simulated block maxima.

Parameter fits can be seen on figure 3.7, the figure shows performance of index flood method for regional parameters estimates and as expected

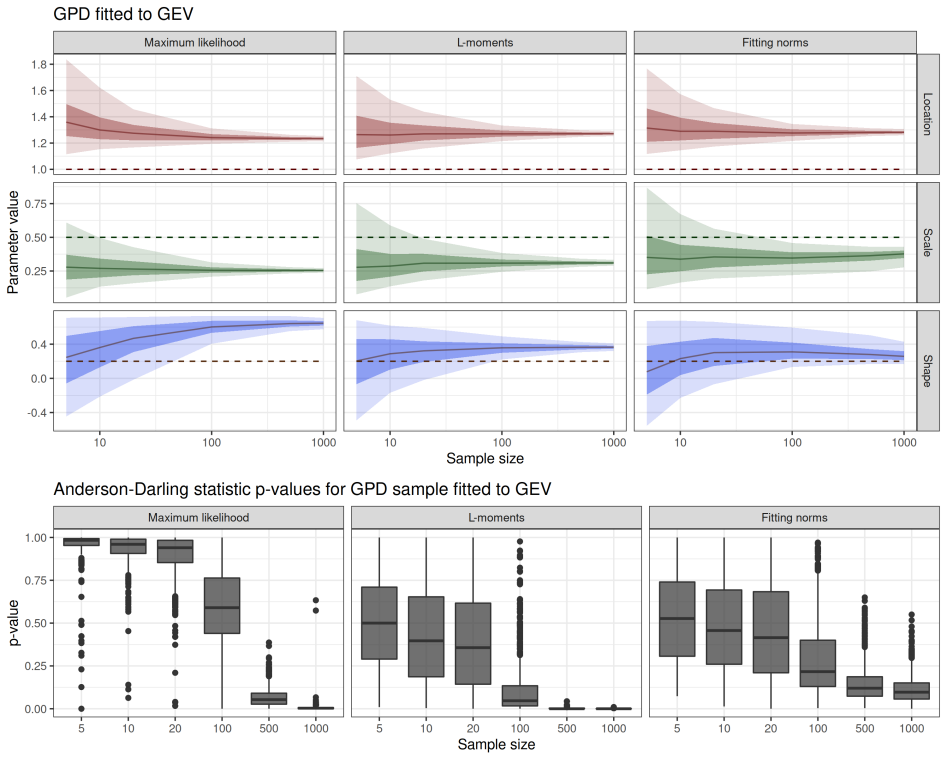

FIGURE 3.5: GPD fitted to GEV; brighter colour on the upper panel represents 90% confidence interval, darker colour 50% confidence interval, coloured line represents the median and the dashed line is the original parameter value.

with increasing p_0 (that shortens the sample size) the range of the estimated parameters is increasing, biases in the parameters are comparable to those on figure 3.2 for lower sample sizes.

Figure 3.6 provides a clear visual representation of the uncertainty reduction achieved by the different pooling methods and can be used as a

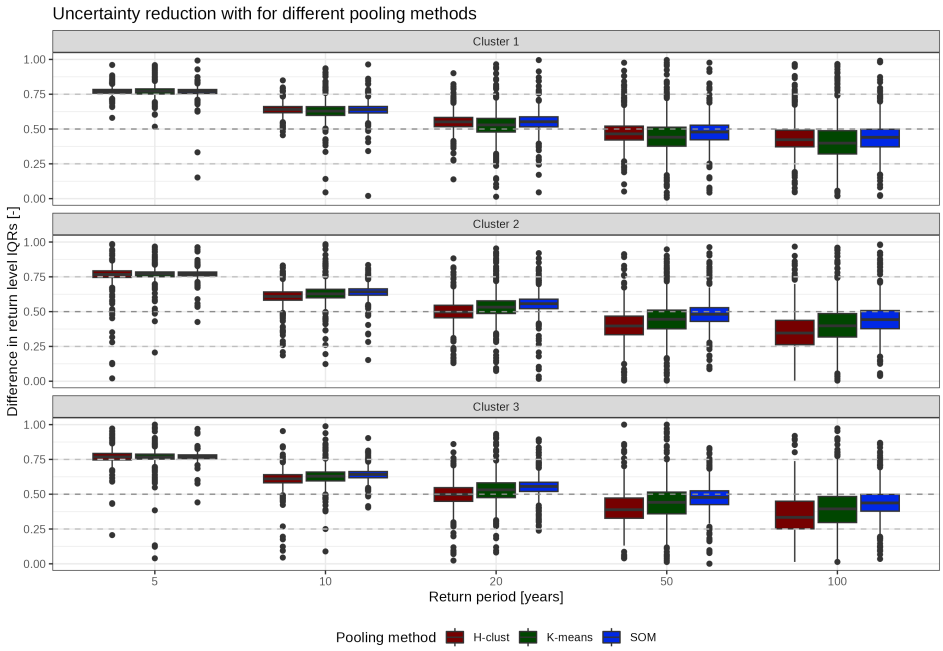

FIGURE 3.6: Uncertainty reduction in return levels for return periods of 5, 10, 20, 50 and 100 years, calculated from the relative difference of the IQRs for the at-site and the regional quantile functions estimates with use of different pooling methods.

reference for future studies in the field of extreme value analysis. While the results of the pooling experiment show that the different pooling methods perform similarly across the board, with a slightly higher uncertainty reduction for SOM, it is important to keep in mind that the experiment did not follow the recommendation of Hosking and Wallis (2005) to use at-site characteristics as inputs for the pooling, but rather at-site statistics (block maxima) since the test was designed for set conditions. Therefore, when

choosing a pooling method, one should rely on the recommendations made earlier in the literature.

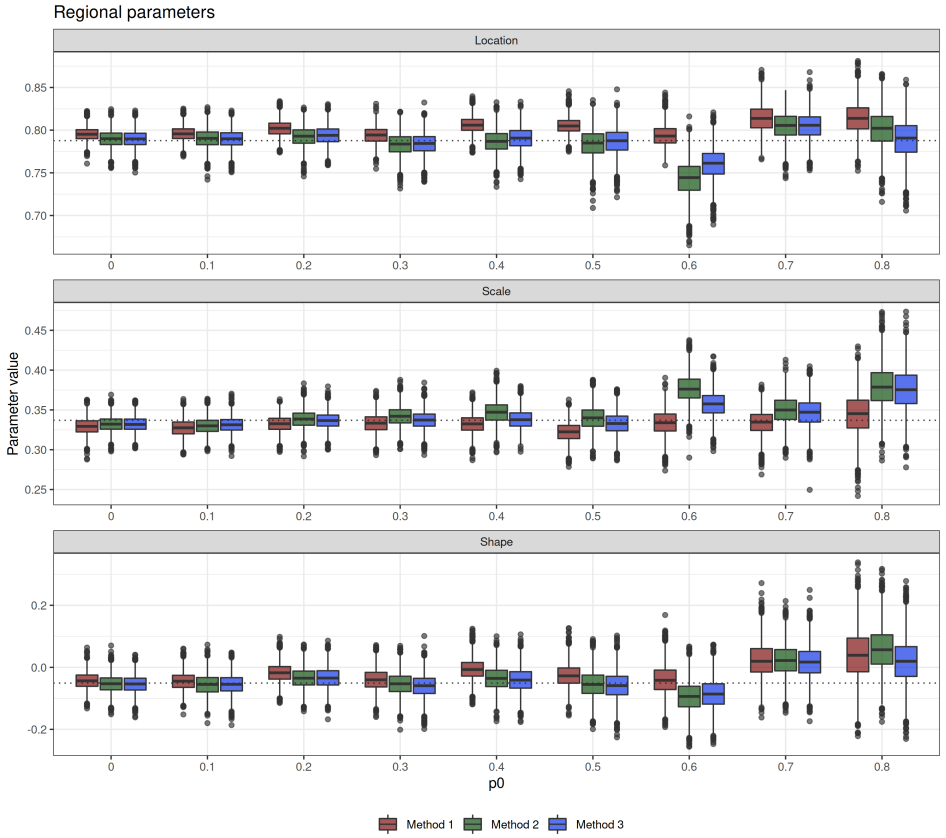

FIGURE 3.7: Estimates of regional parameters for the three bootstrapping methods and varying values of p_0.

When comparing the lowering accuracy of the parameter estimation (figure 3.7) to the uncertainty reduction (figure 3.8) with increasing p_0, it

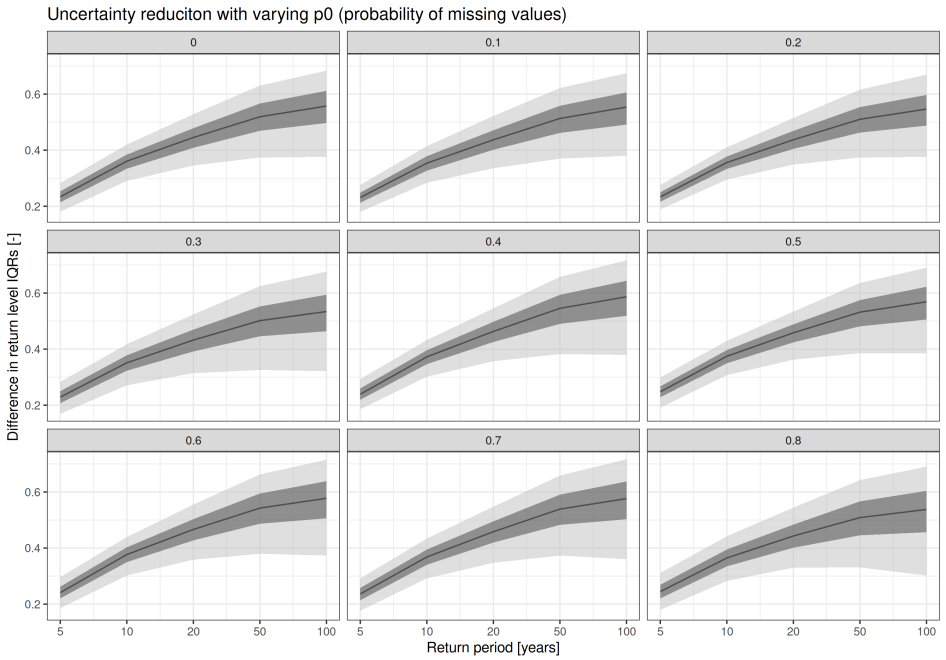

FIGURE 3.8: Percentual reduction of uncertainty in return levels for return periods of 5, 10, 20, 50 and 100 years, calculated from the relative difference of the IQRs for the at-site and the regional quantile functions estimates for the bootstrap samples, brighter colour represents 90% confidence interval, darker colour 50% confidence interval and the coloured line represents the median.

can be seen that the percentual reduction of uncertainty is not affected the same way the parameters are (although there is an apparent increase of the 90% confidence interval with increasing p_0).

It can be concluded that the uncertainty reduction effect of the index flood method remains relatively constant with increasing p_0 however the

absolute values of the return level (high quantile) estimates can be skewed due to the bias in parameter estimation.

CHAPTER 4

Case study

Central Europe recently experienced a number severe drought events (e.g., 2000, 2003, 2015, 2018; e.g. Fink et al. (2004); Ionita et al. (2017); Laaha et al. (2016)). These events attracted public, media and scientific attention as well as stimulated drought research, development of drought legislation and adaptation strategies. Many of these activities require assessment of drought characteristics (like severity, intensity, duration and frequency). While these characteristics are routinely estimated for heavy precipitation events and floods (e.g., Blazkov and Beven (1997); Burn (1990); Dalrymple (1960); Iacobellis et al. (2010)), the applications in the drought context are less common.

This could be at least partly attributed to the vague definition of drought (described in section 2.1) potentially leading to contradicting assessments. A common definition of such drought is the deficit of water with respect to variable of interest or specific water use.

In contrast to extreme precipitation or runoff, the definition of drought is not straightforward and various definitions do exist. In the present study, deficit volume is considered, due to its clear physical interpretation. On the other hand, one may also consider drought indices, based on cumulative deviation from the mean, e.g., Drought Severity Index (Phillips and McGregor, 1998) or indices inspired by the *Standardised Precipitation Index* (SPI). The use of the latter within regional frequency analysis, however, is complex since often the temporal dimension of drought is characterised by different time-scales for which the SPI is calculated.

Moreover, even the definition of deficit volume allows for several subjective choices like threshold level, form of the threshold (variable or fixed within a year), number of days/months needed for the discharge to be above threshold to end the drought event etc. This increases the uncertainty in the estimation of drought characteristics.

In this chapter the application of an RFA model based on L-moments for estimation of drought characteristics is presented, more specifically the distribution of maximum deficit volumes for the period 1900–2015 over the Czech Republic is assessed. The model aims at reduction of uncertainty in the estimated return levels, in the periods of drought events and in the parameters of the extremal model. The goodness-of-fit of the model is evaluated through discordance analysis, as well as the A^2, with the critical values estimated by a bootstrap procedure.

As described in section 2.4, RFA uses spatial pooling of data from a homogeneous region to reduce the standard error of the estimates, i.e., it trades time for space. The vast majority of its applications is for runoff (e.g., Clausen and Pearson (1995); Noto and La Loggia (2009)), precipitation (e.g., Fowler and Kilsby (2003); Modarres (2010); Santos et al. (2011)) or temperature maxima (e.g., Brown and Katz (1995)), while the applications

in the drought context are rare. Some noteworthy exceptions are Madsen and Rosbjerg (1998) who carried out RFA of deficit runoff volumes, or Chen et al. (2006) who presented regional analysis of low flows over South China. The RFA method based on L-moments was carried out by Núnez et al. (2011) and Abdi et al. (2017).

Identification of the homogeneous regions for RFA requires the greatest amount of subjective judgement of all stages of regional frequency analysis. When using K-mean clustering, methods uncertainties stem from the choice of the number of clusters, which can actually be mitigated by using methods like gap statistics (Tibshirani et al., 2001; Yan and Ye, 2007). However in this study we had a predefined number of clusters from the very beginning since the initial idea was to classify the catchments into three groups based on the level of threat by drought. Another ways to proceed with spatial pooling would be by using *Self-Organising Map* (SOM) (Kohonen, 1998; Lin and Chen, 2006), dimensionality reduction technique (Kraemer et al., 2018), or pooling methods first suggested by (Acreman and Wiltshire, 1987) and (Acreman, 1987) with subsequent implementation of the method referred to as the region of influence approach by (Burn, 1990).

4.1 Study Area — Czech Republic

Although Czech Republic is a small country in central Europe, weather conditions differs markedly among its various regions. The variability of the weather is strongly driven by the unstable location and magnitude of two main pressure centres. In particular during the warm period of the year, the expansion of the high pressure projection into Czech Republic causes warmer temperatures and dry weather, whereas the Icelandic Low manifests itself with a greater number of atmospheric fronts bringing more clouds

and precipitation.

The average air temperature is strongly dependent on the altitude and ranges from 0.4 °C on the highest elevation point (mountain Sněžka; 1603 m) to almost 10 °C in the lowlands of southeast Moravia. The annual rainfall is also strongly dependent on the altitude and orography. The wettest areas are the mountain ranges with steep slopes facing northwest in Jizerské hory (Jizera Mountains) with average total rainfall exceeding 1700 millimetres. On the other hand, the driest regions are the lowlands in southeast Moravia and northwest Bohemia receiving approximately 400 mm on average (the latter is influenced by rain shadow east of the Krušné hory (Ore Mountains) Figure 4.1 left.

For the purpose of the study all of the 133 catchments defined by (Zítek, 1965) are considered covering the entirety of the Czech Republic (Figure 4.1 right) with respective areas ranging from 154 to 1928 km^2. The catchments are based on hydrological division of the Czech Republic as provided by the Czech Hydrometeorological Institute, which is also considered in the application of water management policies.

4.2 Data & methods

4.2.1 Data

Since 80 out of the 133 catchments over the study area are ungauged, the BILAN hydrological model (Horáček et al., 2009; Kašpárek et al., 2016; Tallaksen and Van Lanen, 2004; Vizina et al., 2015) was used to estimate runoff from each catchment. BILAN has been frequently applied in various hydrological studies, as well as for the assessment of possible climate change impacts on water resources in the Czech Republic (Beran and Hanel, 2015;

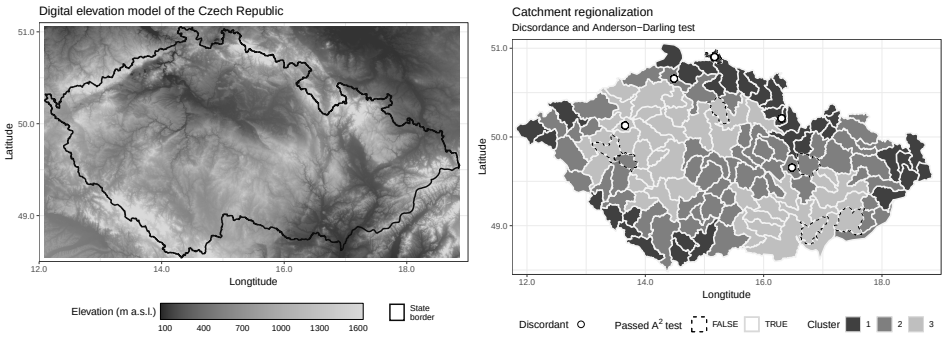

FIGURE 4.1: Left panel: Digital Elevation model of the Czech Republic; Right panel: Resulting clusters, discordance measure and results of at-site A^2 test.

Beran et al., 2016; Hanel et al., 2013; Horáček et al., 2008). It is a lumped hydrological model for assessment of water balance components in monthly or daily step. The catchment is schematised as a system of reservoirs and flows, with catchment precipitation, air temperature and relative air humidity as inputs and total stream-flow as output.

Precipitation and air temperature from the HadCRU-TS3.21 (Harris et al., 2014) data-set was used for the period 1900–1960 and the gridded data-set of precipitation and temperature provided by (Štěpánek et al., 2011) for the period 1961–2015. The latter is derived from a larger number of stations and therefore the HadCRU-TS3.21 data-set was adjusted to have same monthly mean over the period 1961–2015. The gridded data were transferred to the river catchment areas using a weighted average, with weights proportional to the area of the intersection between the catchment and grid boxes.

Since the spatial resolution of the gridded data set used for derivation of catchment precipitation and temperature might be too coarse for smaller

catchments with large altitude differences, the mean monthly catchment precipitation and temperature were finally corrected for error in long-term mean (1980–2010) by comparing the long-term average of the derived catchment data with long-term average precipitation and temperature calculated from fine-scale (1 km) gridded product provided by Czech Hydrometeorological Institute (Tolasz et al., 2007).

The BILAN model simulates water balance at three vertical levels: on the land surface, in the soil layer and in groundwater aquifer. The three water balance algorithms that are applied were developed for winter conditions, snow melting and summer conditions. Surface water balance depends on evapotranspiration, which is derived using temperature based empirical formula derived by (Oudin et al., 2005). Excess water (precipitation minus evapotranspiration) forms direct runoff or infiltrates to a deeper zone, where it is divided into inter-flow and groundwater recharge (Horáček et al., 2008).

To estimate water balance at ungauged catchments, a database of calibrated BILAN models available for more than 300 catchments in the Czech Republic was used. For each catchment of interest models from catchments intersecting the catchment area were transferred . The simulated runoff for the catchment of interest was calculated as a weighted average of runoff from transferred models. The weights were proportional to the area of intersection between the catchment of interest and the transferred model. Thus, for each catchment a time series for the period 1900–2015 was obtained.

4.2.2 Drought Definition

To analyse drought characteristics, a cumulative deficit volume below a preselected threshold is considered (Luo et al., 2017; Tallaksen, 2000; Van Loon, 2015). The volume was first developed by (Rice, 1945) and later extended

and summarised by (Leadbetter, 1967). An early application in hydrology includes (Yevjevich et al., 1967), where the method is based on the statistical theory of runs for analysing a sequential time series.

The threshold level is either representing a certain water demand, e.g., power plants or water supply, or the boundary between normal and unusually low stream flow conditions. The threshold level can be fixed or varying over the year to reflect seasonal variability of hydrological regime or water demands, and can be chosen in a number of ways.

In the present study a varying monthly 80% quantile of the flow exceedance curve was chosen, similarly to (Hisdal et al., 2000) or (Fleig et al., 2006). Basic characteristics describing the deficit event include:

- event severity (deficit volume), D [mm or m^3];
- event length, L [months];
- event intensity, $I = D/L$ [mm/month or m^3/month];
- relative severity (i.e., deficit volume to monthly runoff ratio), rD [-];
- relative event intensity, $rI = rD/L$ [t^{-1}].

While all of the above mentioned characteristics were used for the evaluation of the simulation of hydrological model, for the extreme value analysis annual maximum event severity (deficit volume) were considered.

4.2.3 Statistical Model

The RFA approach based on L-moments (Hosking and Wallis, 2005) was applied in order to estimate quantiles of the distribution of maximum deficit volumes.

In this study, the Hartigan–Wong k-means algorithm Hartigan and Wong (1979) described in section 2.4.2.2 was used to identify the homogeneous regions. K-means algorithm identifies k number of centroids, and then allocates every data point to the nearest centroid, while keeping the clusters as small as possible. The input to the algorithm is a set of points defined by the coordinates in the n-dimensional space, and the number k, defining the number of clusters. The cluster analysis was carried out with scaled data of runoff minus potential evapotranspiration.

The parameters of the regional distribution are estimated using the L-moments method (Papalexiou and Koutsoyiannis, 2016). The at-site L-moments for annual maximum deficit volumes were estimated using the algorithm developed by Hosking (2017). (Section 2.3.2)

Following Stedinger (1993) we consider the model with a probability mass concentrated in zero

$$F^*(x) = \begin{cases} p_0 & \text{if} \quad x = 0 \\ p_0 + (1 - p_0)F(x) & \text{if} \quad x > 0 \ . \end{cases} \tag{4.1}$$

where p_0 is the probability of year without drought. We estimate p_0 as the proportion of zero drought years (Engeland et al., 2004). Further we describe the model for the distribution of non-zero deficit volumes, $F(x)$.

4.3 Model Assessment & Results

Model was checked visually by the Ratio diagrams and Gumbel plots. The ratio diagrams ere constructed by plotting the estimated sample L-moment ratios versus the theoretical L-moment ratio curves for the candidate distributions. Gumbel plot is a quantile function with transformed Gumbel variate ($-log(-log(F))$) instead of probability (F) on the horizontal axis.

A discordance analysis was performed in order to assess whether the distributions of at-site deficit volumes within each cluster were acceptably similar. *Anderson–Darling test* (A^2) test was chosen over goodness-of-fit framework within Hosking and Wallis (1997) based on the findings presented by Viglione et al. (2007) which specifically compares the A^2 with methods used in Hosking and Wallis (1997) in order to make recommendations for test selection based on the assumed skewness of the data. Methods are described in section 2.2

The characteristics of simulated deficit events in four successive 30-year (climatic) periods starting in 1901 are given in Table 4.1. The average values of event severity (D), intensity (I), length (L), relative severity (rD) and relative intensity (rI) are varying over the periods with largest values of event severity in the periods 1931–1960 and 1961–1990. These periods are in good agreement with the extreme droughts that manifested in 1947, 1953–1954, 1959, 1963–1964, 1973–1974, 1983 (Blinka, 2005; Brázdil et al., 2016; Hanel et al., 2018; Spinoni et al., 2015; Treml, 2011).

Table 4.1: Average values of severity (D), intensity (I), length (L), relative severity (rD) and relative intensity (rI) of deficit events derived from simulated data.

Period	$\boldsymbol{D}$	$\boldsymbol{I}$	$\boldsymbol{L}$	$\boldsymbol{rD}$	$\boldsymbol{rI}$
1901–1930	4.46	1.70	2.34	0.24	0.09
1931–1960	6.01	1.97	2.76	0.36	0.11
1961–1990	6.68	2.19	2.95	0.44	0.12
1991–2015	4.74	1.79	2.38	0.29	0.10

The relatively lower values of all variables in the last period might be linked with the rather wet conditions that prevailed in Central Europe

(Markonis et al., 2018). In addition, the current dry period over the Czech Republic spans the years 2014–2018, so considerable part is not considered here. A steady decrease in soil moisture has been reported for the same period (Trnka et al., 2015), due to the increasing temperature and consequently to the rising evapotranspiration. The latter can be also seen in the drought representation by the SPEI index (Potopova et al., 2014).

In the 53 gauged catchments, the properties of simulated deficit volumes for the period 1980–2010 were compared to the observational records. The validation showed that the characteristics of simulated deficit volumes correspond well to those based on observed data, as shown in Figure 4.2 and Table 4.2. In Figure 4.2, the simulated event severities and lengths are well represented through the median, as well as through the confidence interval in all ranges. The simulated low event intensities correspond quite well to the observed ones, despite the overestimation of the high intensities by the model. The simulated relative severity and intensity are slightly overestimated in the whole range, due to the cumulative effect in their computation. This overestimation pattern is well shown in Table 4.2 through the average of the individual variables.

Table 4.2: Validation of simulated deficit volumes. D: severity, I: intensity, L: length, rD: relative severity and rI: relative intensity.

	$\boldsymbol{D}$	$\boldsymbol{I}$	$\boldsymbol{L}$	$\boldsymbol{rD}$	$\boldsymbol{rI}$
Observed runoff	5.25	1.94	2.29	0.24	0.09
Simulated runoff	6.15	2.35	2.36	0.28	0.10

In the spatial pooling step of RFA, input to K-means algorithm was mean runoff and mean potential evapotranspiration for each catchment which resulted in three clusters of catchments. The algorithm ran ten times, each

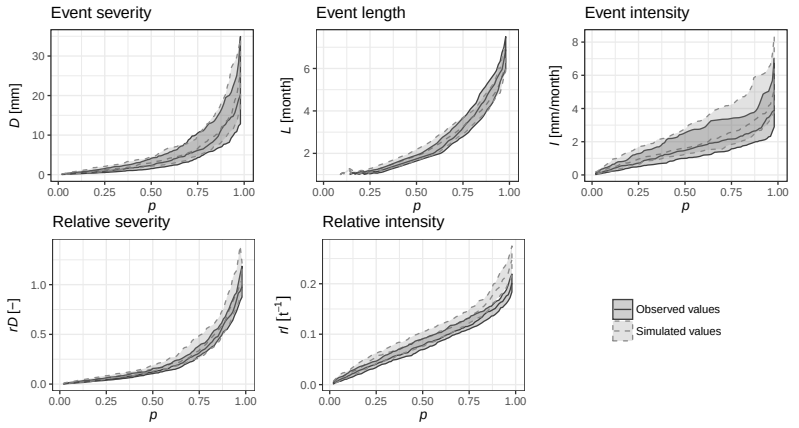

FIGURE 4.2: Comparison of drought characteristics for observed and simulated runoff. The empirical quantiles of the individual characteristics are indicated on the horizontal axis, the vertical axis shows values of drought characteristics, the polygons correspond to interquartile range.

time starting with cluster centres in a different random position. Within fifty iterations, each run converged to a locally-optimal solution. Cluster 1 represents the catchments at high elevations with a lot of precipitation (see Table 4.3 for average precipitation for individual clusters), low land dry catchments with limited precipitation form cluster 3, while cluster 2 is a transition between the low drought risk cluster 1 and severe drought event risk cluster 3. Table 4.3 reports also the probability of year without drought. It may be surprising that the low-risk cluster 1 has the lowest probability of year without drought (0.3), while this probability is 0.49 for severe drought event risk cluster 3. However, it has to be noted (and is demonstrated further) that the tail of the distribution of deficit volume is much heavier in cluster 3 than in cluster 1 (see, e.g., κ parameter in Table 4.4 or the quantile

functions in Figure 4.4).

Although we used unsupervised clustering algorithm it is worth noting that the resulting regions used for regional frequency analysis shown in Figure 4.1, correspond well with the distribution of hydro-climatic variables relevant to drought such as aridity index (Tolasz et al., 2007), which supports the relevance of the clustering algorithm.

Table 4.3: Mean values of annual precipitation sum (P[mm]) for each cluster, average deficit volumes DV [mm] for each cluster and probabilities p_0 of year without drought event per cluster.

	$\boldsymbol{P}$ **[mm]**	$\boldsymbol{DV}$ **[mm]**	$\boldsymbol{p_0}$
Cluster 1	993.87	21.65	0.30
Cluster 2	699.80	10.73	0.36
Cluster 3	574.50	6.43	0.49

At-site distributions were chosen on the basis of L-moment ratio diagrams and at-site A^2. The diagrams were constructed by plotting the estimated sample L-moment ratios versus the theoretical L-moment ratio curves for the candidate distributions (Figure 4.3). From the considered distributions, the estimated L-moment ratios for deficit volumes correspond best to those of the *Generalised Pareto Distribution* (GPD). In addition, the A^2 at the significance level $\alpha_{LOC} = 0.05$ rejected the GPD only at six out of 133 catchments, which is very close to the nominal level of the test.

For each cluster a stationary index flood model for scaled deficit volumes was developed. The scaling was performed by the at-site first L-moment, with the scaling factors varying between 1.94 and 23.5 mm. The fitted regional parameters of the model are presented in Table 4.4. It is evident that the cluster 3 (dry catchments) exhibits quite different behaviour than

the other two clusters. In particular, the low value of the shape parameter indicates heavy tail. In addition, the smaller scale parameter also points towards dry regime prone to heavy extremes.

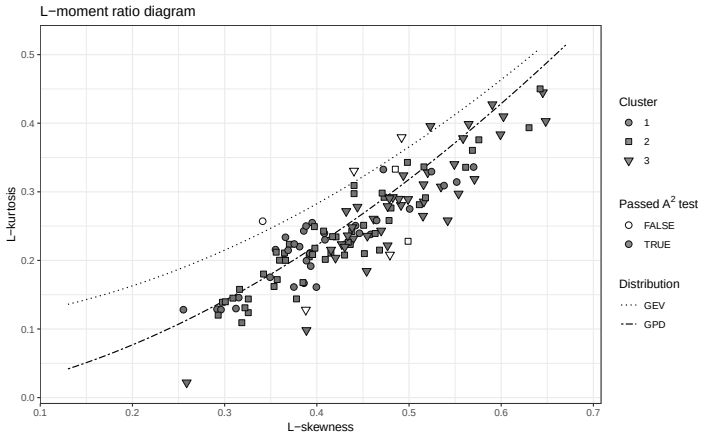

FIGURE 4.3: L-moment ratio diagram with highlighted clusters and results of at-site A^2. The dashed lines show the theoretical L-moment ratio for *Generalised Pareto Distribution* (GPD) and *Generalised Extreme Value distribution* (GEV) and the points the L-moment ratios for each catchment.

Table 4.4: Fitted regional parameters with estimated regional A^2 critical values.

	$\boldsymbol{\xi}$	$\boldsymbol{\alpha}$	$\boldsymbol{\kappa}$	$\mathbf{A^2}$ **Critical Value**
Cluster 1	−0.01	0.86	−0.15	2.42
Cluster 2	−0.02	0.83	−0.19	2.64
Cluster 3	−0.04	0.71	−0.32	2.79

The goodness-of-fit was assessed using Gumbel plots, discordance measure

and regional A^2 (Figure 4.4). It is clear that the regional model fits the deficit volumes scaled by the first L-moment well. The same figure, highlights 1–2 catchments in every cluster demonstrating different behaviour than the rest of the cluster (discordant catchments). Regions were checked for within-cluster discordance based on a critical value set at 3 with a 10% significance level as defined by Hosking and Wallis (2005), and five catchments in total were found discordant. In the A^2, the regional critical values were estimated using the methods described above with 3000 bootstrap samples for each region (Table 4.4). All clusters passed the regional A^2 for significance level $\alpha_{GLOB} = 0.10$.

The bootstrapping method used here is described by (Hanel et al., 2009) (and in section 2.5.2.1) in steps as:

1. Fit the statistical model to the original sample.
2. Calculate standard normal residuals with the parameter estimates from step using quantile mapping.
3. Calculate the average correlation $\hat{\rho}$ of the standard normal residuals.
4. Generate a sample of S equicorrelated standard normal variables with correlation $\hat{\rho}$.
5. Transform the sample from step 4 back to the original scale using the parameter estimates from step 1.
6. Fit the statistical model again.
7. Calculate the A^2 statistics.
8. Repeat steps 4–7 until the desired number of bootstrap samples is obtained.

The annual maximum deficit volumes analysed in the present study cannot be regarded as standard block maxima, since there is often only one drought event (and only seldom more than two) for individual year and catchment. Therefore the annual maximum deficit volume are not theoretically expected to follow *Generalised Extreme Value distribution* (GEV). Indeed, the results suggest that for most stations the GPD is appropriate for the description of the distribution of the annual maximum deficit volume (though generalised normal and generalised extreme value distributions could be also good candidates for stations that did not pass the A^2 or are being an outliers in Figure 4.3).

Similar results can be seen in (Tallaksen and Hisdal, 1997), where annual deficit volumes were fitted to various distributions and GPD presented the best results. However, no spatial pooling was employed in that study. In another study that employed RFA for deficit volumes (Madsen and Rosbjerg, 1998) the Generalised Exponential Distribution was used, which is a reparameterisation of the Generalised Pareto Distribution.

In addition, the analyses conducted within searching for the optimal at-site distribution revealed that Generalised Extreme Value distribution cannot be used to characterise the distribution of deficit volumes, although it is very often found appropriate for maximum discharges or heavy precipitation indices. This result can be, at least partly, region-specific, therefore the at-site distribution should be always checked prior the regional frequency analysis.

To assess the increase in precision of the parameter estimates owing to spatial pooling, GPD parameters were fitted for each individual catchment and the 25th and 75th percentiles of the parameter estimates were calculated using 500 bootstrap samples. Then, for each region and each parameter the average interquartile range was obtained as the difference between the average 75th and 25th percentile of the estimates. These average interquartile

ranges were compared with those of the regional model. Results are shown in Table 4.5.

Increase in precision for the return levels was calculated by substituting the estimated parameters of the bootstrap sample to GPD quantile function with corresponding probability p, $p = 1 - 1/T$, where T is the return period in years. The estimated return levels for each cluster together with calculated confidence intervals can be seen in Figure 4.5.

Table 4.5: Percentage decrease in uncertainties in parameter and return level estimation.

	$\boldsymbol{\alpha}$	$\boldsymbol{\kappa}$	$\mathbf{2}_{\boldsymbol{yr}}$	$\mathbf{50}_{\boldsymbol{yr}}$
Cluster 1	99.86	69.97	67.99	66.44
Cluster 2	99.84	75.03	74.95	72.82
Cluster 3	99.40	55.94	56.28	52.04

Another option how to increase the sample size is to consider reconstructed climate data (e.g., (Dobrovolný et al., 2015)) in combination with a hydrological model. This introduces additional sources of uncertainty, though, through the reconstructed climate fields and the parameterisation of the hydrological model. On the other hand, the spatial and temporal scales relevant for drought may allow to obtain reliable information even based on data with limited spatial coverage.

4.4 Case study summary

Statistical model using index-flood method based on L-moments was used on simulated runoff series for the period 1900–2015 for 133 catchments in the Czech Republic. Goodness-of-fit of the model was assessed using

Gumbel plots and A^2. Critical values of the test were estimated by bootstrap resampling, which also provided the estimate of the confidence intervals allowing for calculation of the reduction in uncertainties of the regional parameter and return level estimation.

The main conclusions that can be drawn are:

- Regional frequency analysis reduces uncertainty of estimated drought characteristics and parameters of its distribution.
- Use of *Generalised Pareto Distribution* (GPD) is appropriate to describe the deficit volumes on majority of catchments, which is not the case for *Generalised Extreme Value distribution* (GEV). However, it is not clear to what extent this result depends on characteristics of the area under study and other parameters of the analysis like the threshold defining drought.
- The most subjective part of the regional frequency analysis is the definition of homogeneous regions—methods such as region of influence or *Self-Organising Map* (SOM) could be considered to minimise the subjective decisions within the regional frequency analysis.

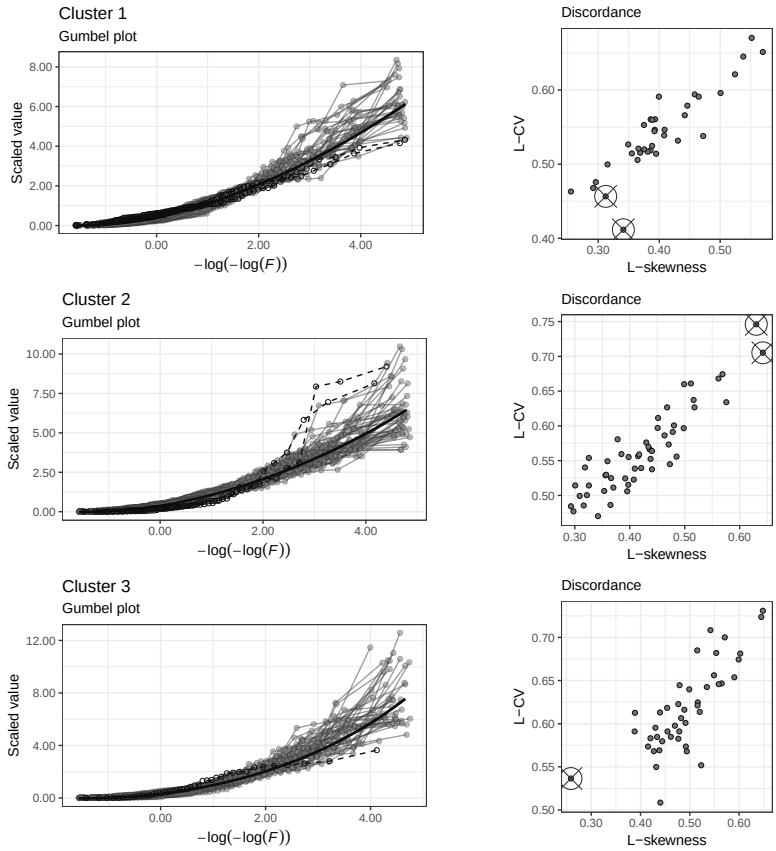

FIGURE 4.4: Left panels: Gumbel plot—Continuous black lines represent fitted regional quantile functions, grey points with lines are scaled deficit volumes with probabilities calculated using plotting position, dashed lines highlight discordant catchments; Right panels: Discordance measure showing ratio between coefficient of L-variation and L-skewness, discordant ratios lie outside the notional ellipsis (critical value) that would be drawn around the concordant values.

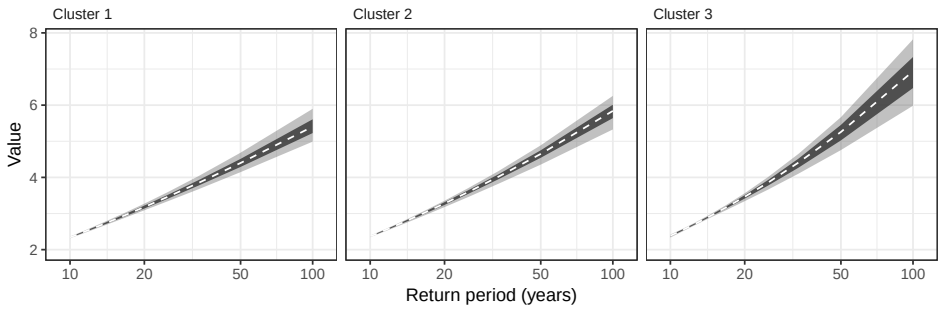

FIGURE 4.5: Estimated return periods of deficit volumes. Dashed line shows regional quantile function, darker area the 25th and 75th, light area 5th and 95th percent quantile calculated from the bootstrap samples.

CHAPTER 5

Discussion & Conclusions

Estimation of extreme events is subject to considerable uncertainty. In the context of drought, this uncertainty is strengthen by the fact that drought does not occur every year and more significant droughts occur in only a small fraction of years. Since there is no unified drought definition, additional uncertainty is introduced by the choice of the indicator used for drought quantification, which has to be always based on the objective of the study or the needs of the stakeholders.

When selecting the most suitable probability distribution for drought indicators/indices, two distributions are commonly considered: the *Generalised Extreme Value distribution* (GEV) distribution and the *Generalised Pareto Distribution* (GPD). The choice between these two distributions is relatively straightforward as it is based on the nature of the data. The extreme value theorem dictates that the GEV distribution should be used to describe properly scaled block maxima, while threshold-based indices cannot

be treated as standard block maxima. As there is often only one drought event (and only seldom more than two) for individual year and catchment in threshold-based indices, the GPD distribution is the appropriate choice.

It should be noted that while GEV and GPD are commonly used distributions for drought indices, they are not the only two distributions that can be used. Other distributions such as Weibull or Lognormal distributions can also be used to model drought events. However, the choice of distribution should be based on the characteristics of the drought data and the purpose of the analysis. The selection of the appropriate distribution should always be verified prior to proceeding with *Regional Frequency Analysis* (RFA). This can be done through a variety of methods such as Gumbel plots and other visual aids, as well as statistical tests such as the Anderson-Darling test (A^2). This verification step is essential to ensure accurate and reliable results in RFA.

Each of the three presented methods (Maximum likelihood, L-moments and fitting norms) has its strengths and weaknesses based on which an appropriate method can be chosen for specific situations.

When searching for the optimal distribution of the drought indicator/indices, it is important to consider the strengths and weaknesses of each fitting method, as presented in previous sections.

Maximum likelihood method is appropriate for higher sample sizes and can also be easily modified to account for non-stationarity employing methods like local likelihood smoothing (Davison and Ramesh, 2000) or parametric trend estimates. The implementation of the algorithm can, however, be unstable in some instances. The likelihood functions have to be specifically derived for a given distribution and estimation problem. The mathematics is often non-trivial, particularly if confidence intervals for the parameters are required. It is also important to note that maximum likelihood does not

always have a solution for GPD (Grimshaw, 1993; Hüsler et al., 2011).

L-moments perform better than the maximum likelihood for the smaller sample sizes, the difference between the two methods however should asymptotically approach zero with increasing sample size. L-moments relationship to the distribution parameters has to be defined a priory which can be a major drawback in certain situations. Definitions for selected two and three parameter distributions (including GEV and GPD) can be found in appendix of Hosking and Wallis (2005). The approach of estimating the parameters for four parameter distributions (or distributions with two shape parameters) can be found in Papalexiou and Koutsoyiannis (2012). An advantage of the L-moment method worth noting is the computational time of the estimation that can be, depending on the complexity of the PDF or CDF used by other methods, orders of magnitude faster since the method is estimating linear combination of moments using simple order statistics.

Unlike the other two methods discussed previously, norms method can be used with any distribution that has an analytical solution for its distribution and/or quantile function (Papalexiou et al., 2013). This means that the user has more freedom to choose the distribution that best fits the data, without being restricted to a specific set of distributions as in the case of maximum likelihood or L-moments. However, the success of the fitting norms method is dependent on the specific implementation of the minimising algorithm, and the user must carefully select appropriate criteria for evaluating the goodness of fit. In terms of computational efficiency, the fitting norms method may be the slowest of the three methods, especially if the quantile or distribution function used for fitting is complex.

When employing the index flood method in order to reduce the uncertainty, one is basically pooling available data from different sites/catchments (that are considered to be homogeneous), in order to increase the sample

size. The effect of these additions is maximal when the data are independent. This is seldom true, however, therefore the real reduction of uncertainty not only depends on the number of data but also on the dependence structure of the analysed data, therefore the selection of homogeneous regions (spatial pooling) should receive extra attentions.

Spatial pooling is a subjective process as there is no single "correct" method for pooling data, and the choice of algorithm depends on various factors such as the spatial structure of the data, the sampling design, the type of response variable, and the research question being investigated.

If there are data of the site characteristics available one should employ either K-means, Hierarchical clustering or Self organising maps. If only the data of the extreme events are available, region of influence method (Burn, 1990) can be used. If the number of clusters is known in advance, K-means method is recommended (Govender and Sivakumar, 2020). K-means method is also the fastest algorithm. If the similarities between the sites are based on an a priory knowledge, hierarchical clustering should be used to pool the data (Govender and Sivakumar, 2020). Hierarchical clustering produces a dendrogram (i.e., a graph which shows the order with which segments are grouped together) that helps to determine the number of clusters. Self-organising maps can be used when there is a large number of input (site characteristics) data as it creates clusters that are ordered on a two dimensional lattice (map). However, a drawback is that complete data for all characteristics at all sites is required to generate the map (Chon, 2011).

Main aim of this work was presentation of approaches used for quantification of extremes and pragmatic assesment of the applicability of presented methods.

Performance of selected methods was evaluated in chapter 3 where

experiment focusing on behaviour of parameter estimation algorithms with increasing sample size in various situations was designed. Other experiment was designed to see how the index flood method performs with increasing p_0 (probability of zero values, e.g. the fraction of years without drought).

In chapter 4 a case study was presented that dealt with development of an index flood model for deficit volumes for 133 catchments in the Czech Republic (1901–2015) that were simulated by hydrological model BILAN. The parameters of the regional distribution were estimated using L-moments. The goodness-of-fit of the statistical model was assessed by the A^2 test. For the estimation of critical values, sampling methods allowing for handling of years without drought were used.

There are still some areas that might deserve further study. One being the impact of using various spatial pooling methods in the index flood performance since there is still a lot of subjective decisions being made in selection of homogeneous regions. Other is the performance of different bootstrapping methods in context of the regional *Anderson–Darling test* (A^2). Both of these areas would be a good stepping stone for a case study where non-stationarity of drought indices is examined in context of the climate change in order to tackle the most urgent needs of drought estimation (Brunner et al., 2021).

In addition, it is important to keep in mind the concerns raised by Klemeš (2000a,b) regarding the high quantile modelling in hydrology. While there are many areas that deserve further study, it is crucial that researchers do not simply apply statistical methods without a solid understanding of the underlying physical processes. This is particularly relevant for the analysis of hydrological extremes, where the complexities of the processes involved can be easily oversimplified. Therefore, it is important to approach the study of hydrological extremes with caution and to strive for a balance between the

use of statistical methods and the physical understanding of the underlying processes.

The focus on flooding events in the field of hydrology is understandable, given their immediate and striking impacts. However, it is important to recognise that drought can be just as, if not more, damaging in the long term. Unlike floods, droughts can have insidious effects on natural and human systems, often resulting in complex, indirect, and long-lasting consequences. For example, prolonged droughts can lead to reduced crop yields, water scarcity, increased risk of wildfire, and adverse impacts on ecosystems and biodiversity. Therefore, it is crucial to devote more attention to the study of drought and its impacts, including the development of more effective monitoring, prediction, and management strategies. By doing so, we can better prepare for and mitigate the effects of droughts, and ensure the sustainable use and management of water resources.

APPENDIX A

List of selected publications

Regionalization of deficit runoff volumes in the Czech Republic

Available at https://doi.org/10.46555/VTEI.2017.05.002

Abstract

The aim of the study is the regionalization of the Czech Republic with respect to drought characteristics in individual catchments. The regionalization was done for a set of 133 catchments and for the period 1901–2015. The basic index describing the hydrological drought is the deficit discharge, i.e. the cumulative volume bellow 20th percentile of the monthly runoff distribution. The regionalization of drought characteristics was based on the mean precipitation, evaporation, total and base flow and hydrogeological districts. This regionalization was revised by experts. The deficit volumes in the simulation of the Bilan model were estimated and the statistical model for estimation of N-year deficit volumes was developed and validated

An Index-Flood Statistical Model for Hydrological Drought Assessment

Available at `https://doi.org/10.3390/w12041213`

Abstract

Modelling of hydrological extremes and drought modelling in particular has received much attention over recent decades. The main aim of this study is to apply a statistical model for drought estimation (in this case deficit volume) using extreme value theory and the index-flood method and to reduce the uncertainties in estimation of drought event return levels. Deficit volumes for 133 catchments in the Czech Republic (1901–2015) were simulated by hydrological model BILAN. The validation of severity, intensity and length of simulated drought events revealed good match with the available observed data. To estimate return levels of the deficit volumes, it is assumed (in accord with the index-flood method), that the deficit volumes within a homogeneous region are identically distributed after scaling with a site-specific factor. The parameters of the scaled regional distribution are estimated using L-moments. The goodness-of-fit of the statistical model is assessed by Anderson–Darling test. For the estimation of critical values, sampling methods allowing for handling of years without drought were used. It is shown, that the index-flood model with a Generalized Pareto distribution performs well and substantially reduces the uncertainty related to the estimation of the shape parameter and of the large deficit volume quantiles.

Representation of European hydroclimatic patterns with self-organizing maps

Available at `https://doi.org/10.1177/0959683620913924`

Abstract

Self-organizing maps provide a powerful, non-linear technique of dimensionality reduction that can be used to identify clusters with similar attributes. Here, they were constructed from a 1000-year-long gridded palaeoclimatic dataset, namely the Old World Drought Atlas, to detect regions of homogeneous hydroclimatic variability across the European continent. A classification scheme of 10 regions was found to describe most efficiently the spatial properties of Europe's hydroclimate. These regions were mainly divided into a northern and a southern subset, linked together with a northwest-to-southeast orientation. Further analysis of the classification scheme with complex networks confirmed the divergence between the northern and southern components of European hydroclimate, also revealing that is not strongly correlated to the Iberian Peninsula. On the contrary, the region covering the British Isles, France and Germany appeared to be linked to both branches, implying links of hydroclimate with atmospheric/oceanic circulation.

CoSMoS v2.0: Making Time Series Generation Simple

Available at https://doi.org/10.5194/egusphere-egu2020-22357 & https://cran.r-project.org/web/packages/CoSMoS/index.html

Abstract

Many physically based models aiming to quantify the vulnerability and risk of hydrologic and geomorphic hazards need as input or forcing time series of processes such as precipitation, temperature, humidity, etc. The reliability of their output depends on how realistic the inputs are. CoSMoS is a multi-platform software that generates reliable time series from hydroclimatic variables (precipitation, temperature, wind, relative humidity, streamflow, etc.). It is developed in R (version 2.0) as well as in other platforms (Matlab, Mathematica, Excel). It can be used to generate univariate and multivariate time series at any time scale by reproducing the marginal distributions and the linear correlation structures (including intermittency) of the process under investigation. CoSMoS implements a unified stochastic modelling scheme that expands and enhances a generic modelling approach based on the transformation of "parent" Gaussian time series. By design it aims to offer a simple and easy-to-apply solution to the user requesting minimal information, such as the target marginal distribution and the correlation structure. The software is accompanied by a complete users' manual.

APPENDIX B

Experiments & case study implementation

The codes used to calculate all the results for the dissertation are available on the GitHub repository at `https://github.com/strnda/dissertation-codes`. The repository includes all the codes used in the data analysis, visualization, and modeling. The codes are written in R (R Core Team, 2023). The GitHub repository is publicly accessible, and anyone can download, use, and modify the codes under the GPL3 license. The GPL3 license ensures that the codes are open-source and free to use, but any modified versions of the code must also be made available under the same license. Therefore, the codes can be used not only to reproduce the results presented in the dissertation but also to build upon them and develop new research directions.

Bibliography

Amin Abdi, Yousef Hassanzadeh, Siamak Talatahari, Ahmad Fakheri-Fard, and Rasoul Mirabbasi. Regional drought frequency analysis using l-moments and adjusted charged system search. *Journal of Hydroinformatics*, 19(3):426–442, 2017.

MC Acreman. Regional flood frequency analysis in the UK: Recent research-new ideas. *Institute of Hydrology, Wallingford, UK*, 1987.

MC Acreman and SE Wiltshire. Identification of regions for regional flood frequency analysis. *Eos*, 68(44):1262, 1987.

William M Alley. The Palmer drought severity index: limitations and assumptions. *Journal of climate and applied meteorology*, 23(7):1100–1109, 1984.

Theodore W Anderson and Donald A Darling. A test of goodness of fit. *Journal of the American statistical association*, 49(268):765–769, 1954.

David Arthur and Sergei Vassilvitskii. k-means++: the advantages of careful seeding, p 1027–1035. *Society for Industrial and Applied Mathematics*, 2007.

Leo R Beard. Statistical analysis in hydrology. *Transactions of the American Society of Civil Engineers*, 108(1):1110–1121, 1943.

Jan Beirlant, Yuri Goegebeur, Johan Segers, and Jozef L Teugels. *Statistics of extremes: theory and applications*. John Wiley & Sons, 2006.

Adam Beran and Martin Hanel. Identification of regions vulnerable to deficits in water resources in the Czech Republic. *Water Management Technical and Economical Information Journal*, 57(4–5):23–26, 2015.

Adam Beran, Martin Hanel, and Magdalena Nesládková. Changes in the hydrological balance caused by climate change impacts in the Karlovy Vary district. *Water Management Technical and Economical Information Journal*, 58(5):20–25, 2016.

Max A Beran, Jean A Rodier, et al. *Hydrological aspects of drought: a contribution to the International Hydrological Programme*, volume 39. Unesco, 1985.

Sarka Blazkov and Keith Beven. Flood frequency prediction for data limited catchments in the czech republic using a stochastic rainfall model and topmodel. *Journal of Hydrology*, 195(1-4):256–278, 1997.

P Blinka. Climatological evaluation of drought and dry periods on the territory of Czech Republic in the years 1876-2002. *The Meteorological Bulletin*, 58:10–18, 2005.

Guenter Bloeschl, Julia Hall, Alberto Viglione, Rui A. P. Perdigao, Juraj Parajka, Bruno Merz, David Lun, Berit Arheimer, Giuseppe T. Aronica, Ardian Bilibashi, Milon Bohac, Ognjen Bonacci, Marco Borga, Ivan Canjevac, Attilio Castellarin, Giovanni B. Chirico, Pierluigi Claps, Natalia

Frolova, Daniele Ganora, Liudmyla Gorbachova, Ali Gul, Jamie Hannaford, Shaun Harrigan, Maria Kireeva, Andrea Kiss, Thomas R. Kjeldsen, Silvia Kohnova, Jarkko J. Koskela, Ondrej Ledvinka, Neil Macdonald, Maria Mavrova-Guirguinova, Luis Mediero, Ralf Merz, Peter Molnar, Alberto Montanari, Conor Murphy, Marzena Osuch, Valeryia Ovcharuk, Ivan Radevski, Jose L. Salinas, Eric Sauquet, Mojca Sraj, Jan Szolgay, Elena Volpi, Donna Wilson, Klodian Zaimi, and Nenad Zivkovic. Changing climate both increases and decreases European river floods. *Nature*, 573 (7772), SEP 5 2019.

Günter Blöschl, Murugesu Sivapalan, Thorsten Wagener, Hubert Savenije, and Alberto Viglione. *Runoff prediction in ungauged basins: synthesis across processes, places and scales.* Cambridge University Press, 2013.

Ognjen Bonacci. Hydrological identification of drought. *Hydrological Processes*, 7(3):249–262, 1993.

Walter C Boughton. A frequency distribution for annual floods. *Water Resources Research*, 16(2):347–354, 1980.

Rudolf Brázdil, Miroslav Trnka, Pavel Zahradníček, Petr Dobrovolný, Ladislava Řezníčková, Pavel Treml, Zdeněk Stachoň, et al. The Central European drought of 1947: causes and consequences, with particular reference to the Czech Lands. *Climate Research*, 70(2-3):161–178, 2016.

Barbara G Brown and Richard W Katz. Regional analysis of temperature extremes: Spatial analog for climate change? *Journal of Climate*, 8(1): 108–119, 1995.

Manuela I Brunner, Louise Slater, Lena M Tallaksen, and Martyn Clark. Challenges in modeling and predicting floods and droughts: A review. *Wiley Interdisciplinary Reviews: Water*, 8(3):e1520, 2021.

Donald H Burn. Evaluation of regional flood frequency analysis with a region of influence approach. *Water Resources Research*, 26(10):2257–2265, 1990.

Fi-John Chang, Li-Chiu Chang, Huey-Shan Kao, and Gwo-Ru Wu. Assessing the effort of meteorological variables for evaporation estimation by self-organizing map neural network. *Journal of Hydrology*, 384(1):118–129, 2010.

NN Chegodaev. Computation of surface runoff on small catchments. *All-Union Scientific Research Institute for Railway Construction and Planning Rep*, 37, 1953.

Yongqin David Chen, Guoru Huang, Quanxi Shao, and Chong-Yu Xu. Regional analysis of low flow using L-moments for Dongjiang basin, South China. *Hydrological Sciences Journal*, 51(6):1051–1064, 2006.

Tae-Soo Chon. Self-organizing maps applied to ecological sciences. *Ecological Informatics*, 6(1):50–61, 2011.

P Ciais, M Reichstein, N Viovy, A Granier, J Ogee, V Allard, M Aubinet, N Buchmann, C Bernhofer, A Carrara, F Chevallier, N De Noblet, AD Friend, P Friedlingstein, T Grunwald, B Heinesch, P Keronen, A Knohl, G Krinner, D Loustau, G Manca, G Matteucci, F Miglietta, JM Ourcival, D Papale, K Pilegaard, S Rambal, G Seufert, JF Soussana, MJ Sanz, ED Schulze, T Vesala, and R Valentini. Europe-wide reduction in primary productivity caused by the heat and drought in 2003. *Nature*, 437(7058), SEP 22 2005.

B Clausen and CP Pearson. Regional frequency analysis of annual maximum streamflow drought. *Journal of Hydrology*, 173(1-4):111–130, 1995.

Stuart Coles, Joanna Bawa, Lesley Trenner, and Pat Dorazio. *An introduction to statistical modeling of extreme values*, volume 208. Springer, 2001.

Harald Cramér. On the composition of elementary errors: First paper: Mathematical deductions. *Scandinavian Actuarial Journal*, 1928(1):13–74, 1928.

Tate Dalrymple. Flood-frequency analyses, manual of hydrology: Part 3. Technical report, USGPO,, 1960.

Dante Alighieri. *The divine comedy*. Blackwell for the Shakespeare Head Press, 1972.

Anthony C Davison and NI Ramesh. Local likelihood smoothing of sample extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(1):191–208, 2000.

Anthony Christopher Davison and David Victor Hinkley. *Bootstrap methods and their application*, volume 1. Cambridge university press, 1997.

S Demuth and A Bakenhus. Hydrological drought-a literature review. *University of Freiburg: Freiburg, Germany*, 1994.

Petr Dobrovolný, Rudolf Brázdil, Mirek Trnka, Oldřich Kotyza, and Hubert Valášek. Precipitation reconstruction for the Czech Lands, AD 1501-2010. *International Journal of Climatology*, 35(1):1–14, 2015.

John A Dracup, Kil Seong Lee, and Edwin G Paulson. On the definition of droughts. *Water resources research*, 16(2):297–302, 1980.

Bradley Efron and Robert J Tibshirani. *An introduction to the bootstrap*. CRC press, 1994.

S El Adlouni and TBMJ Ouarda. Comparison of methods for estimating the parameters of the non-stationary gev model. *Revue des Sciences de l'Eau*, 21(1):35–50, 2008.

Kolbjørn Engeland, Hege Hisdal, and Arnoldo Frigessi. Practical extreme value modelling of hydrological floods and droughts: a case study. *Extremes*, 7(1):5–30, 2004.

Patrick J Farrell and Katrina Rogers-Stewart. Comprehensive study of tests for normality and symmetry: extending the Spiegelhalter test. *Journal of Statistical Computation and Simulation*, 76(9):803–816, 2006.

DS Faulkner and DA Jones. The FORGEX method of rainfall growth estimation III: Examples and confidence intervals. *Hydrology and Earth System Sciences Discussions*, 3(2):205–212, 1999.

BM Fekete, CJ Vörösmarty, and W Grabs. Global composite runoff data set (v1. 0). *Complex Systems Research Center, University of New Hampshire, Durham, New Hampshire*, 2000.

Andreas H Fink, Tim Brücher, Andreas Krüger, Gregor C Leckebusch, Joaquim G Pinto, and Uwe Ulbrich. The 2003 European summer heatwaves and drought–synoptic diagnosis and impacts. *Weather*, 59(8):209–216, 2004.

Anne K Fleig, Lena M Tallaksen, Hege Hisdal, and Siegfried Demuth. A global evaluation of streamflow drought characteristics. *Hydrology and Earth System Sciences Discussions*, 10(4):535–552, 2006.

HJ Fowler and CG Kilsby. A regional frequency analysis of United Kingdom extreme rainfall from 1961 to 2000. *International Journal of Climatology*, 23(11):1313–1334, 2003.

Paulene Govender and Venkataraman Sivakumar. Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019). *Atmospheric Pollution Research*, 11(1):40–56, 2020.

Scott D Grimshaw. Computing maximum likelihood estimates for the generalized pareto distribution. *Technometrics*, 35(2):185–191, 1993.

Martin Hanel, T Adri Buishand, and Christopher AT Ferro. A nonstationary index flood model for precipitation extremes in transient regional climate model simulations. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 114(D15), 2009.

Martin Hanel, Magdalena Mrkvičková, Petr Máca, Adam Vizina, and Pavel Pech. Evaluation of simple statistical downscaling methods for monthly regional climate model simulations with respect to the estimated changes in runoff in the czech republic. *Water resources management*, 27(15): 5261–5279, 2013.

Martin Hanel, Oldřich Rakovec, Yannis Markonis, Petr Máca, Luis Samaniego, Jan Kyselý, and Rohini Kumar. Revisiting the recent European droughts from a long-term perspective. *Scientific Reports*, 8(1):9499, 2018.

IPDJ Harris, PD Jones, TJ Osborn, and DH Lister. Updated high-resolution grids of monthly climatic observations–the CRU TS3. 10 Dataset. *International Journal of Climatology*, 34(3):623–642, 2014.

John A Hartigan and Manchek A Wong. Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1):100–108, 1979.

Michael J Hayes. *Drought indices*. Wiley Online Library, 2006.

H Hisdal, LM Tallaksen, E Peters, K Stahl, and M Zaidman. Drought event definition. *ARIDE Technical Rep*, 6:15, 2000.

Stanislav Horáček, Ladislav Kašpárek, and Oldřich Novický. Estimation of climate change impact on water resources by using Bilan water balance model. In *IOP conference series: earth and environmental science*, volume 4, page 012023. IOP Publishing, 2008.

Stanislav Horáček, Oldřich Rakovec, Ladislav Kašpárek, and Adam Vizina. Development of the hydrological balance model BILAN. *Water Management Technical and Economical Information Journal*, 51(1):2–5, 2009.

Jonathan RM Hosking. L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society: Series B (Methodological)*, 52(1):105–124, 1990.

JRM Hosking. *L-Moments*, 2017. URL `https://CRAN.R-project.org/package=lmom`. R package, version 2.6.

JRM Hosking and JR Wallis. Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics*, 29(3):339–349, 1987.

JRM Hosking and JR Wallis. The effect of intersite dependence on regional flood frequency analysis. *Water Resources Research*, 24(4):588–600, 1988.

JRM Hosking and JR Wallis. Some statistics useful in regional frequency analysis. *Water resources research*, 29(2):271–281, 1993.

JRM Hosking and JR Wallis. *Regional frequency analysis: An Approach Based on L-Moments*. Cambridge University Press, 1997.

JRM Hosking and JR Wallis. *Regional frequency analysis: An Approach Based on L-Moments*. Cambridge University Press, 2005.

JRM Hosking, JR Wallis, and Eric F Wood. An appraisal of the regional flood frequency procedure in the uk flood studies report. *Hydrological Sciences Journal*, 30(1):85–109, 1985.

Jürg Hüsler, Deyuan Li, and Mathias Raschke. Estimation for the generalized pareto distribution using maximum likelihood and goodness of fit. *Communications in statistics-theory and methods*, 40(14):2500–2510, 2011.

Vito Iacobellis, Mauro Fiorentino, Andrea Gioia, and Salvatore Manfreda. Best fit and selection of theoretical flood frequency distributions based on different runoff generation mechanisms. *Water*, 2(2):239–256, 2010.

Monica Ionita, Lena Tallaksen, Daniel Kingston, James Stagge, Gregor Laaha, Henny Van Lanen, Patrick Scholz, Silvia Chelcea, and Klaus Haslinger. The European 2015 drought from a climatological perspective. *Hydrology and Earth System Sciences*, 21:1397–1419, 2017.

Anil K Jain. Data clustering: 50 years beyond k-means. *Pattern recognition letters*, 31(8):651–666, 2010.

Ladislav Kašpárek, Martin Hanel, Stanislav Horáček, Petr Máca, and Adam Vizina. *Bilan water balance model*, 2016. R package version 2016-10-20.

Viatcheslav V Kharin, Francis W Zwiers, Xuebin Zhang, and Gabriele C Hegerl. Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations. *Journal of Climate*, 20(8): 1419–1444, 2007.

Vít Klemeš. Tall tales about tails of hydrological distributions. i. *Journal of Hydrologic Engineering*, 5(3):227–231, 2000a.

Vít Klemeš. Tall tales about tails of hydrological distributions. ii. *Journal of Hydrologic Engineering*, 5(3):232–239, 2000b.

Teuvo Kohonen. The self-organizing map. *Neurocomputing*, 21(1-3):1–6, 1998.

Guido Kraemer, Markus Reichstein, and Miguel D Mahecha. dimRed and coRanking—unifying dimensionality reduction in R. *R Journal*, 10(1): 342–358, 2018.

Gregor Laaha, Tobias Gauster, L Tallaksen, Jean-Philippe Vidal, Kerstin Stahl, Christel Prudhomme, Benedikt Heudorfer, Radek Vlnas, Monica Ionita, Henny AJ Van Lanen, et al. The European 2015 drought from a hydrological perspective. *Hydrology and Earth System Sciences*, 21(6): 3001–3024, 2016.

J Maciunas Landwehr, NC Matalas, and JR Wallis. Probability weighted moments compared with some traditional techniques in estimating gumbel parameters and quantiles. *Water Resources Research*, 15(5):1055–1064, 1979.

Emmett M Laursen. Comment on "paleohydrology of southwestern texas" by r. craig kochel, victor r. baker, and peter c. patton. *Water Resources Research*, 19(5):1339–1339, 1983.

H Cramer-MR Leadbetter. Stationary and related stochastic processes. *New York*, 1967.

Dennis P Lettenmaier and Kenneth W Potter. Testing flood frequency estimation methods using a regional flood generation model. *Water Resources Research*, 21(12):1903–1914, 1985.

Dennis P Lettenmaier, JR Wallis, and Eric F Wood. Effect of regional heterogeneity on flood frequency estimation. *Water Resources Research*, 23(2):313–323, 1987.

Rita Ley, MC Casper, Hugo Hellebrand, and Ralf Merz. Catchment classification by runoff behaviour with self-organizing maps (som). *Hydrology and Earth System Sciences*, 15(9):2947, 2011.

Gwo-Fong Lin and Lu-Hsien Chen. Identification of homogeneous regions for regional frequency analysis using the self-organizing map. *Journal of Hydrology*, 324(1-4):1–9, 2006.

Benjamin Lloyd-Hughes. The impracticality of a universal drought definition. *Theoretical and Applied Climatology*, 117(3):607–611, 2014.

Lifeng Luo, Deanna Apps, Samuel Arcand, Huating Xu, Ming Pan, and Martin Hoerling. Contribution of temperature and precipitation anomalies to the California drought during 2012-2015. *Geophysical Research Letters*, 2017. ISSN 1944-8007. doi: 10.1002/2016GL072027.

TA MacMahon and R Srikanthan. Log pearson type 3 distribution effect of dependence, distribution parameters and sample size on peak annual flood estimates. *J. Hydrol*, 52:815–826, 1982.

Henrik Madsen and Dan Rosbjerg. A regional Bayesian method for estimation of extreme streamflow droughts. In *Statistical and Bayesian methods in hydrological sciences*, pages 327–340. UNESCO, 1998.

Y Markonis, M Hanel, P Máca, J Kyselý, and ER Cook. Persistent multi-scale fluctuations shift European hydroclimate to its millennial boundaries. *Nature communications*, 9(1):1767, 2018.

Oman Masqat. Anderson Darling and modified Anderson Darling tests for generalized Pareto distribution. *Pakistan Journal of Applied Sciences*, 3 (2):85–88, 2003.

John Mawdsley, Geoffrey E Petts, and Susan Walker. *Assessment of drought severity*. Institute of Hydrology, 1994.

Thomas B McKee, Nolan J Doesken, John Kleist, et al. The relationship of drought frequency and duration to time scales. In *Proceedings of the 8th Conference on Applied Climatology*, volume 17, pages 179–183. Boston, 1993.

Ashok K Mishra and Vijay P Singh. A review of drought concepts. *Journal of hydrology*, 391(1-2):202–216, 2010.

Ashok K Mishra and Vijay P Singh. Drought modeling - A review. *Journal of Hydrology*, 403(1-2):157–175, 2011.

Reza Modarres. Regional dry spells frequency analysis by L-moment and multivariate analysis. *Water resources management*, 24(10):2365–2380, 2010.

Dimitrios Myronidis, Dimitrios Fotakis, Konstantinos Ioannou, and Konstantina Sgouropoulou. Comparison of ten notable meteorological drought

indices on tracking the effect of drought on streamflow. *Hydrological sciences journal*, 63(15-16):2005–2019, 2018a.

Dimitrios Myronidis, Konstantinos Ioannou, Dimitrios Fotakis, and Gerald Dörflinger. Streamflow and hydrological drought trend analysis and forecasting in cyprus. *Water resources management*, 32(5):1759–1776, 2018b.

I Nalbantis and G Tsakiris. Assessment of hydrological drought revisited. *Water Resources Management*, 23(5):881–897, 2009.

RJ Nathan and TA McMahon. Identification of homogeneous regions for the purposes of regionalisation. *Journal of Hydrology*, 121(1-4):217–238, 1990.

Leonardo V Noto and Goffredo La Loggia. Use of L-moments approach for regional flood frequency analysis in Sicily, Italy. *Water resources management*, 23(11):2207–2229, 2009.

Jorge H Núnez, Koen Verbist, Jim R Wallis, Mel G Schaefer, Luis Morales, and WM Cornelis. Regional frequency analysis for mapping drought events in north-central chile. *Journal of hydrology*, 405(3-4):352–366, 2011.

Ludovic Oudin, Frédéric Hervieu, Claude Michel, Charles Perrin, Vazken Andréassian, François Anctil, and Cécile Loumagne. Which potential evapotranspiration input for a lumped rainfall–runoff model?: Part 2—towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling. *Journal of hydrology*, 303(1-4):290–306, 2005.

Simon Michael Papalexiou and Demetris Koutsoyiannis. Entropy based derivation of probability distributions: A case study to daily rainfall. *Advances in Water Resources*, 45:51–57, 2012.

Simon Michael Papalexiou and Demetris Koutsoyiannis. A global survey on the seasonal variation of the marginal distribution of daily precipitation. *Advances in water resources*, 94:131–145, 2016.

SM Papalexiou, D Koutsoyiannis, and C Makropoulos. How extreme is extreme? An assessment of daily rainfall distribution tails. *Hydrology and Earth System Sciences*, 17(2):851–862, 2013.

Ian D Phillips and Glenn R McGregor. The utility of a drought index for assessing the drought hazard in Devon and Cornwall, South West England. *Meteorological Applications*, 5(4):359–372, 1998.

Vera Potopova, Constanţa Boroneanţ, Martin Možný, Petr Štěpánek, and Petr Skalák. Observed spatiotemporal characteristics of drought on various time scales over the Czech Republic. *Theoretical and applied climatology*, 115(3-4):563–581, 2014.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2023. URL `https://www.R-project.org/`.

Stephen O Rice. Mathematical analysis of random noise. *The Bell System Technical Journal*, 24(1):46–156, 1945.

E. Rousi, U. Ulbrich, H. W. Rust, and C. Anagnostolpoulou. An NAO Climatology in Reanalysis Data with the Use of Self-organizing Maps. In *Perspectives on Atmospheric Sciences*, pages 719–724. Springer, Cham, 2017. DOI: 10.1007/978-3-319-35095-0_103.

Gianfausto Salvadori and Carlo De Michele. Statistical characterization of temporal structure of storms. *Advances in Water Resources*, 29:827–842, 2006.

João Filipe Santos, Maria Manuela Portela, and Inmaculada Pulido-Calvo. Regional frequency analysis of droughts in Portugal. *Water Resources Management*, 25(14):3537, 2011.

Nikolai Vasil'evich Smirnov. Sur la distribution de w2. *Comp. Rend. Acad. Sci*, 202:449–452, 1936.

James A Smith. Regional flood frequency analysis using extreme order statistics of the annual peak record. *Water Resources Research*, 25(2): 311–317, 1989.

Jonathan Spinoni, Gustavo Naumann, Jürgen V Vogt, and Paulo Barbosa. The biggest drought events in Europe from 1950 to 2012. *Journal of Hydrology: Regional Studies*, 3:509–524, 2015.

J R Stedinger. Frequency analysis of extreme events. *in Handbook of Hydrology.*, 1993.

Petr Štěpánek, Pavel Zahradníček, and Radan Huth. Interpolation techniques used for data quality control and calculation of technical series: an example of a Central European daily time series. *Idojaras*, 115(1-2):87–98, 2011.

Filip Strnad, Simon Papalexiou, Yanis Markonis, Francesco Serinaldi, and Kevin Shook. *CoSMoS: Complete Stochastic Modelling Solution*, 2020. URL `https://CRAN.R-project.org/package=CoSMoS`. R package version 2.0.0.

Mark Svoboda, Brian Fuchs, et al. Handbook of drought indicators and indices. *Drought and water crises: Integrating science, management, and policy*, pages 155–208, 2016.

Gabor J Szekely, Maria L Rizzo, et al. Hierarchical clustering via joint between-within distances: Extending ward's minimum variance method. *Journal of classification*, 22(2):151–184, 2005.

Lena M Tallaksen. Streamflow drought frequency analysis. In *Drought and drought mitigation in Europe*, pages 103–117. Springer, 2000.

Lena M Tallaksen and HEGE Hisdal. Regional analysis of extreme streamflow drought duration and deficit volume. *IAHS Publication*, 246:141–150, 1997.

Lena M Tallaksen and Henny AJ Van Lanen. *Hydrological drought: processes and estimation methods for streamflow and groundwater*, volume 48. Elsevier, 2004.

EL Tate and Alan Gustard. Drought definition: a hydrological perspective. In *Drought and drought mitigation in Europe*, pages 23–48. Springer, 2000.

Robert Tibshirani, Guenther Walther, and Trevor Hastie. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):411–423, 2001.

Radim Tolasz, R Brázdil, O Bulíř, P Dobrovolnỳ, M Dubrovskỳ, L Hájková, O Halásková, J Hostỳnek, M Janouch, M Kohut, et al. Atlas podnebí česka. 1. vydání. *Praha, Olomouc, Českỳ hydrometeorologickỳ ústav, Universita Palackého*, 2007.

P Treml. The largest droughts in the Czech Republic in the period 1875–2010. *The Meteorological Bulletin*, 64:168–176, 2011.

M Trnka, R Brázdil, M Možný, P Štěpánek, P Dobrovolný, P Zahradníček, J Balek, D Semerádová, M Dubrovský, P Hlavinka, et al. Soil moisture

trends in the Czech Republic between 1961 and 2012. *International Journal of Climatology*, 35(13):3733–3747, 2015.

G Tsakiris and H Vangelis. Establishing a drought index incorporating evapotranspiration. *European Water*, 9(10):3–11, 2005.

Alfred Ultsch and H. Peter Siemon. Kohonen's self organizing feature maps for exploratory data analysis. In Bernard Widrow and Bernard Angeniol, editors, *Proceedings of the International Neural Network Conference (INNC-90), Paris, France, July 9–13, 1990 1. Dordrecht, Netherlands*, volume 1, pages 305–308, Dordrecht, Netherlands, 1990. Kluwer Academic Press. URL `http://www.uni-marburg.de/fb12/datenbionik/pdf/pubs/1990/UltschSiemon90`.

Anne F Van Loon. Hydrological drought explained. *Wiley Interdisciplinary Reviews: Water*, 2(4):359–392, 2015.

Sergio M Vicente-Serrano, Santiago Beguería, and Juan I López-Moreno. A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. *Journal of climate*, 23(7):1696–1718, 2010.

A Viglione, F Laio, and P Claps. A comparison of homogeneity tests for regional frequency analysis. *Water Resources Research*, 43(3), 2007.

Adam Vizina, Stanislav Horáček, Martin Hanel, et al. Recent developments of the BILAN model. *Water Management Technical and Economical Information Journal*, 57(4–5):7–10, 2015.

Richard Von Mises. *Vorlesungen aus dem Gebiete der angewandten Mathematik: Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik*. F. Deuticke, 1931.

Hans Von Storch and Francis W Zwiers. *Statistical analysis in climate research*. Cambridge university press, 2001.

JR Wallis. *Risk and uncertainties in the evaluation of flood events for the design of hydrologic structures*. IBM Thomas J. Watson Research Division, 1980.

JR Wallis. Hydrologic problems associated with oilshale development. *Environmental Systems and Management, edited by S. Rinaldi*, pages 85–102, 1982.

JR Wallis and Eric F Wood. Relative accuracy of log pearson iii procedures. *Journal of Hydraulic Engineering*, 111(7):1043–1056, 1985.

C Palmer Wayne. Meteorological drought. *US Weather Bureau research paper*, 30(45), 1965.

Donald A Wilhite and Michael H Glantz. Understanding the Drought Phenomenon: The Role of Definitions. *Water international*, 10(3):111–120, 1985.

Mingjin Yan and Keying Ye. Determining the number of clusters using the weighted gap statistic. *Biometrics*, 63(4):1031–1037, 2007.

Vujica M Yevjevich et al. An objective approach to definitions and investigations of continental hydrologic droughts. *Hydrology papers (Colorado State University); no. 23.*, 1967.

Josef Zítek. *Hydrologické poměry ČSSR*. Hydrometeorologický ústav, 1965.