

**Estimation, Testing and Pooling
in Block Maxima Models
for Climate Extremes**

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Leandra Zanger
aus Düsseldorf

Düsseldorf, März 2023

aus dem Mathematischen Institut
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Berichterstatter:

1. Prof. Dr. Axel Bücher
2. Jun.-Prof. Dr. Marco Oesting

Tag der mündlichen Prüfung: 22. Juni 2023

Abstract

Modelling climate extremes can be challenging due to short observation periods, which leads to unfavourably large estimation uncertainties. In the widely used block maxima method, it has been found in several scenarios that estimation variance may be reduced if sliding blocks are used instead of disjoint blocks. This line of research is extended by examining the probability-weighted moment estimator for the parameter vector of the generalised extreme value distribution, based on both disjoint and sliding block maxima of univariate observations. In contrast to other results on the probability-weighted moment estimator from the literature, which usually consider independent and identically distributed observations, the assumptions on the underlying random variables are adapted to the setting of environmental applications, by assuming either stationarity or some kind of piecewise stationarity. For the latter setting, a proof of concept is provided that encourages the use of sliding block maxima despite their non-stationarity. The estimators are analysed both theoretically in an asymptotic framework and sub-asymptotically in a simulation study, showing increased efficiency of the sliding version in both analyses.

For spatial data, the pooling approach is common for reducing estimation variance. It consists of combining spatial observations that are assumed to have some sort of homogeneous probabilistic behaviour. To avoid biased estimators, it is important to validate this homogeneity assumption in advance. New statistical significance tests for testing corresponding hypotheses are provided, which are based on multivariate generalised extreme value models. The underlying random variables are assumed to be serially, but not necessarily spatially, independent. Unlike many competing tests in the literature, the proposed tests take into account possible cross-correlations of the data. Tests are provided for the case of stationary models as well as models that exhibit a certain type of trend in their extremes. They are based on limiting distributions that are derived for estimators of the parameter vectors, and reliable p -values are obtained by means of parametric bootstrap procedures. Finite-sample properties are investigated in a simulation study. Further, a method for selecting a region that can be assumed homogeneous is provided, which is based on multiple testing.

Acknowledgements

First and foremost, I would like to express my gratitude to my advisor Prof. Dr. Axel Bücher for his invaluable support and his guidance throughout the supervision of this thesis. Thank you for many fruitful discussions and your infectious enthusiasm.

I would also like to thank Jun.-Prof. Dr. Marco Oesting for reviewing this thesis and for all the valuable suggestions and input from many online meetings we had during ClimXtreme.

Further, I wish to thank my current and former colleagues from the Chair of Mathematical Statistics and Probability Theory for the pleasant working environment.

Moreover, I am grateful for the financial support of the Bundesministerium für Bildung und Forschung.

Finally, I thank my friends, especially Laura, Alex and K., who always believed in me.

Contents

1	Introduction	1
2	Included articles	11
2.1	On the Disjoint and Sliding Block Maxima method for piecewise stationary time series	13
2.2	A note on statistical tests for homogeneities in multivariate extreme value models for block maxima	83
2.3	Regional Pooling in Extreme Event Attribution Studies: an Approach Based on Multiple Statistical Testing	117
3	Outlook	167
	Author Contribution Statement	179

1 Introduction

Modelling climate extremes constitutes one of the main applications of extreme value theory, which deals with the description of the probabilistic behaviour of extreme events. In meteorology, hydrology and climatology, the extremes of a wide variety of variables are studied, for example temperature, precipitation, river run-offs and wind speeds, see Coles (2001) for numerous examples. Since extreme values of any of these variables can have tremendous impacts on health, infrastructure, agriculture and hence on human life, it is of interest to assess the probability of their occurrence and to predict their magnitude. Based on such predictions, precautions can be taken in order to anticipate or mitigate these impacts. For risk assessment in this context, the concept of *return levels* (RLs) and *return periods* (RPs) is regularly applied. In a stationary climate, the T -year return level corresponds to the level that can be expected to be exceeded once in T years, and the m -return period corresponds to the length of the time interval within which one exceedance of m can be expected, and these quantities are often the estimation target within an extreme value analysis. For example, one might be interested in estimating the magnitude of a 1000-year flood event at a particular location in order to take appropriate flood protection measures. Now assume that the water discharge at the site in question is stationary and has been observed over a period of 100 years. The challenge is to make a statement about extreme values that might occur within the next 1000 years, based on the much shorter observation period. Therefore, extrapolation into the tail of the distribution is necessary, for which extreme value theory provides the appropriate framework. A gentle introduction to the theory can be found in Coles (2001) or Beirlant et al. (2004).

In extreme value theory, there are two main approaches to modelling extremes. The first one, dating back to Gumbel (1958) and known as the *block maxima* method, consists of fitting the Generalised Extreme Value (GEV) distribution to a sample of block maxima (e.g., seasonal or annual maxima). For observations X_1, \dots, X_n , $n \in \mathbb{N}$, and a block size $r \in \mathbb{N}$ with $r < n$, the block maxima sample traditionally consists of disjoint block maxima

$$M_{r,j}^{(\text{db})} = \max \{ X_{(j-1)r+1}, \dots, X_{jr} \}, \quad j = 1, \dots, \left\lfloor \frac{n}{r} \right\rfloor. \quad (1)$$

The GEV distribution is a continuous three-parameter distribution, with a location parameter $\mu \in \mathbb{R}$, a scale parameter $\sigma > 0$ and a shape parameter $\gamma \in \mathbb{R}$, with cumulative distribution

function (c.d.f.) $G_{(\mu,\sigma,\gamma)}$ given by

$$G_{(\mu,\sigma,\gamma)}(x) = \exp \left\{ - \left(1 + \gamma \frac{x - \mu}{\sigma} \right)^{-\frac{1}{\gamma}} \right\}, \quad 1 + \gamma \frac{x - \mu}{\sigma} > 0. \quad (2)$$

The induced family of GEV distributions combines the Weibull, the Gumbel and the Fréchet distribution families, where the distribution type is determined by the shape parameter: a negative shape parameter refers to Weibull distributions and thus implies a finite right end point of the distribution, while a positive value refers to Fréchet distributions and thus implies a heavy tail. A shape parameter that equals zero implies the light-tailed Gumbel case. The choice of the GEV distribution is motivated by the Fisher-Tippett-Gnedenko Theorem, which states that the only possible non-degenerate limit distribution of the properly rescaled maximum of a sequence of independent and identically distributed (i.i.d.) random variables is a GEV distribution, see e.g. Theorem 3.1 in Coles (2001). The assumption of independence can even be relaxed, as shown by Leadbetter (1983). It may thus be argued that if the block size r is large, the distribution of a block maximum can be approximated by a GEV distribution. Corresponding parameters can be estimated based on the block maxima sample; classical estimation methods include the maximum likelihood (ML) method and the probability weighted moment (PWM) method. Respective theory, based on the assumption of independent maxima that are exactly GEV-distributed, dates back to Smith (1985) in case of ML estimation, with some major corrections in Bücher and Segers (2017), and to Hosking et al. (1985) for PWM estimation. More recently, theoretical results have been obtained for the case of limiting rather than exact GEV distributions: Dombry (2015) treated the existence and consistency of ML estimators involving block maxima of underlying i.i.d. observations, followed by Dombry and Ferreira (2019) showing asymptotic normality in this context. Theory for the PWM estimator based on block maxima can be found in Ferreira and de Haan (2015).

The second main approach to modelling extremes is the so-called *peaks-over-threshold* (POT) approach, where excesses of random variables above a previously chosen high threshold are modelled by the Generalised Pareto distribution. It goes back to Balkema and de Haan (1974) and Pickands (1975), and theory on classical estimators within this framework can for example be found in de Haan and Ferreira (2006) and Embrechts et al. (1997).

It is often argued in the literature that the latter approach allows for a more efficient use of the available data (see e.g. Naveau et al. (2005)), since all ‘large’ observations contribute, whereas with the BM method, some of the rather large observations might be disregarded. For example, in the case where the largest and the second largest observation fall into the same block, only the largest observation will appear in the BM sample, while both observations will appear in the POT sample (provided the POT sample consists of at least two

observations). However, Bücher and Zhou (2021) recently carried out a thorough comparison of the two approaches and found that both methods may outperform the other in certain scenarios, depending on the process that generated the underlying observations.

As indicated before, a difficulty that arises from the rarity of extreme events and from the classical block maxima method itself is that the data situation is often poor. Suppose a hydrologist wants to model the distribution of annual maximum discharge. When daily observations of maximum discharges are available for n years, applying the BM method results in a sample size of n . Often, one is confronted with record lengths of $n < 100$, and even $n < 50$ is not unusual. This inevitably leads to large uncertainties in parameter, return level and return period estimates, which is undesirable. For example, if the estimated 95%-confidence interval of a 100-year flood event spans from $70 \text{ m}^3\text{s}^{-1}$ to $500 \text{ m}^3\text{s}^{-1}$ with a point estimate of $285 \text{ m}^3\text{s}^{-1}$, it is ambiguous against which level to take flood protection measures.

This thesis presents three articles that provide theoretical results and methods on two approaches that address this shortcoming and hence allow for more accurate estimation. Thereby, we will concentrate on the BM method, which is regularly preferred by applied scientists in environmental studies (e.g. Section 4.2.2 in Philip et al. (2020)). This preference has several reasons: first of all, for many climate-related questions, the choice of the block size parameter used within the BM method seems more natural than the choice of a threshold parameter needed within the POT method, since the seasonality of the data naturally suggests the use of annual or seasonal maxima. Secondly, if threshold exceedances are modelled and they tend to occur in clusters, a further step is required to estimate this very tendency. It can be captured by the extremal index; an introduction to this parameter and corresponding estimators is e.g. given in Beirlant et al. (2004). A discussion about its relevance for the estimation of return levels can be found in Bücher and Zhou (2021). Last but not least, data is not always available at timescales that would be required for the application of the POT method: e.g., climate models often output several sample characteristics such as monthly means, minima and maxima rather than daily observations.

The first approach to reducing estimation variance is to make better use of the available data by modifying the traditional block maxima method. Instead of dividing the underlying observations into equally sized disjoint blocks and extracting the maximum within each block as done in Equation (1), a block is slid through the observations in steps of one, and the maximum is calculated for each of these *sliding blocks*. The idea of using sliding blocks can be attributed to Robert et al. (2009), who first applied it for the estimation of the extremal index. To be precise, recall the sample X_1, \dots, X_n , $n \in \mathbb{N}$, of random variables and the block size parameter $r \in \mathbb{N}$ with $r < n$. Then, the sample of sliding block maxima is denoted by

$\mathcal{M}^{(\text{sb})} = (M_{r,j}^{(\text{sb})})_j$ and defined through

$$M_{r,j}^{(\text{sb})} = \max\{X_j, \dots, X_{j+r-1}\} \quad (3)$$

for $j = 1, \dots, n - r + 1$. Analogously, the sample of disjoint block maxima is denoted by $\mathcal{M}^{(\text{db})}$. In case the sequence X_1, \dots, X_n is a segment of a stationary time series $(X_t)_t$, this stationarity transfers to the sample of block maxima for both sampling methods. Therefore, the marginal distributions of the block maxima can be approximated by one common GEV distribution.

The heuristic rationale for using the sliding block rather than the disjoint block sample is as follows. First of all, each disjoint block maximum does also appear in the sample of sliding block maxima, so information provided by disjoint block maxima is guaranteed to be present. Further, those observations which are large but do not represent the maximum of any disjoint block now have a chance to appear in the sample as well. Many of the sliding block maxima will appear repeatedly in the sliding blocks sample because they are maxima of several of the sliding blocks. This repetition of observations within the sample may act like a weighting procedure, so that overall, the sliding blocks sample contains more diverse and (naturally) weighted observations and hence more information about the distribution of block maxima. An illustration of the two block maxima samples can be found in Figure 1, where daily maximum temperatures (in $^{\circ}\text{C}$) observed at station Essen-Bredene¹ during summer months (June, July, August) of 1988 – 2002 are presented. For a block size of $r = 92$, corresponding to the season's length, the sample of disjoint BM, represented by blue points, is shown in the top row, and an illustration of the sample of sliding BM is given in the bottom row. For the latter, each point corresponds to a sliding block maximum, with the point's size chosen proportional to its frequency of occurrence within the sample.

For any estimation method that is performed based on a sample \mathcal{M} , the idea is thus to employ $\mathcal{M} = \mathcal{M}^{(\text{sb})}$ instead of $\mathcal{M} = \mathcal{M}^{(\text{db})}$. When applying this principle, increased efficiency of estimators based on sliding blocks has been shown for several scenarios, e.g. for the ML estimator of Fréchet parameters in the case of heavy tails in Bücher and Segers (2018a), for estimators of the extremal index in Berghaus and Bücher (2018) and Bücher and Jennessen (2020) and for estimation of extreme value copulas of multivariate data in Zou et al. (2021).

The article presented first in Chapter 2 extends the theory by studying the PWM estimator for GEV parameters based on disjoint as well as sliding blocks. Probability-weighted moments were first introduced by Greenwood et al. (1979), and for a random variable X with c.d.f. F they are defined as

$$\beta_{p,q,s} = \mathbb{E}[X^p F^q(X)(1 - F(X))^s], \quad p, q, s \in \mathbb{R}, \quad (4)$$

¹Data basis: Deutscher Wetterdienst, Climate Data Center. https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/daily/kl/historical/

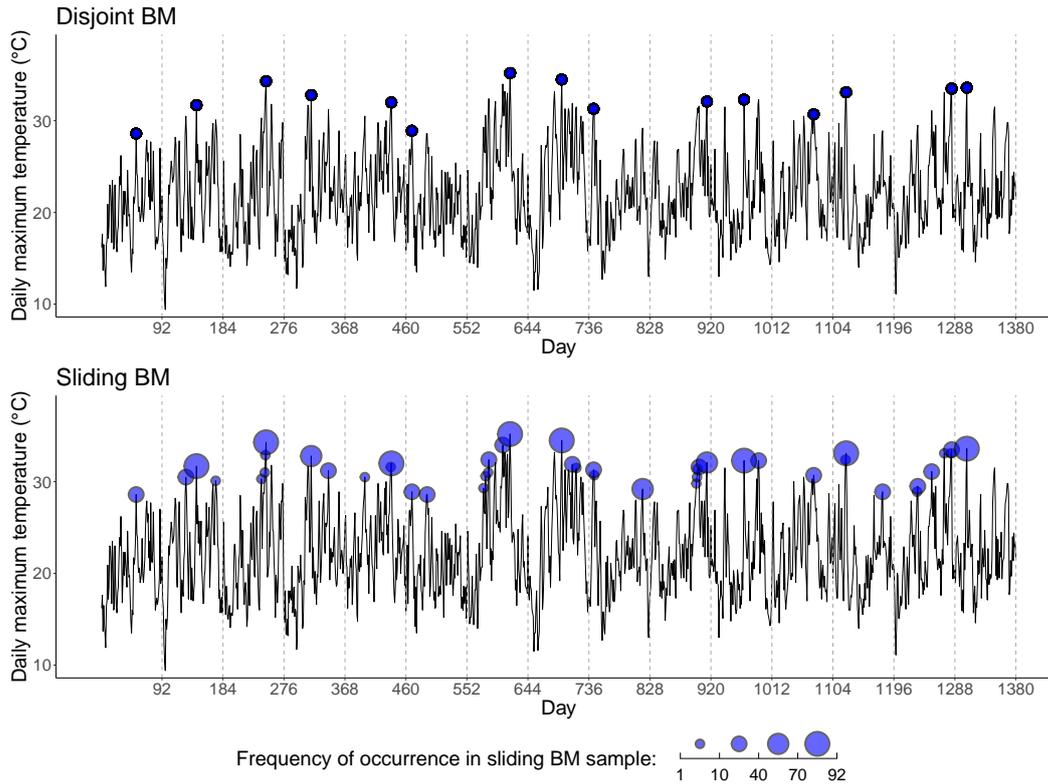


Figure 1: Daily maximum temperatures (in $^{\circ}\text{C}$) observed at station Essen-Bredene during summer months (June, July, August) of 1988 – 2002, along with the sample of disjoint BM (top) and an illustration of the sample of sliding BM (bottom).

but often, only $p, q, s \in \mathbb{N}_0$ are considered. For a GEV-distributed random variable, Hosking et al. (1985) show the unique relation between the three PWMs $\beta_{1,0,0}$, $\beta_{1,1,0}$ and $\beta_{1,2,0}$ and the parameters of the respective GEV distribution. The PWM estimator puts this relation to use and determines parameter estimates based on empirical versions of the PWMs. The asymptotic normality results provided in the article again yield an efficiency gain of the estimator based on sliding blocks, in the sense that the sliding blocks version has a smaller asymptotic covariance matrix than the disjoint blocks version with respect to the Loewner-ordering, while the bias does not change. In addition to elaborating the asymptotics for stationary sequences that satisfy some specific mixing conditions, some emphasis is placed on working with assumptions on the underlying sample X_1, \dots, X_n that correspond to practically relevant situations encountered when analysing climate extremes, where typically attention is restricted to a certain season's maximum, e.g. the maximum temperature observed during winter months. In such applications, the underlying sample X_1, \dots, X_n consists of concatenated observations from the season of interest, observed over several years. In the article, a suitable framework for this kind of sampling scheme is formulated, which assumes that the observations from different seasons are i.i.d. copies of a segment of a time series that may exhibit some sort of weak serial dependence. As a result, neither the underlying sample nor

the resulting sample of sliding block maxima can be assumed to be stationary. However, it is shown that the marginal distributions of sliding block maxima can be approximated by one common GEV distribution, which justifies the application of sliding blocks even in this scenario. To complement the asymptotic results, a large-scale Monte Carlo simulation study is conducted to assess the finite-sample properties of the proposed estimators, with the result that the theoretically derived asymptotic gain of efficiency can already be found sub-asymptotically in many situations.

A second approach that addresses the inconvenience of large estimation uncertainty is the *pooling* approach, which involves pooling samples from multiple sources to one large sample that can then be used in subsequent analyses. The idea behind this is that combining data from different sources leads to a higher information content in the pooled sample. For example, when one has access to several samples from one common population, the distribution of that population can be estimated more accurately by using the pooled sample instead of a single sample. In environmental applications, such a pooling approach is most often applied in a spatial context, i.e., samples from several measuring stations that are located close to each other or in a certain region are pooled, as for example suggested in Philip et al. (2020) and carried out in van Oldenborgh et al. (2017).

The pooling approach is, however, not restricted to the case of one common parent population, but also allows for less stringent assumptions. A famous example from hydrology, first proposed by Dalrymple (1960) and used for flood frequency analysis, is the so-called *index flood* model. The model is based on the assumption that within a certain region, the quantiles of annual maximum discharge at several sites within that region can be decomposed into a regional quantile, which coincides for all sites within the region, and a site-specific factor. Mathematically, for a set of $D \in \mathbb{N}$, $D \geq 2$, sites, random variables M_1, \dots, M_D , where M_d has c.d.f. F_d and describes the annual maximum discharge at site d for $d = 1, \dots, D$, this assumption can be expressed as the hypothesis

$$\mathcal{H}_{0,IF} : \begin{cases} \exists \text{ c.d.f. } G \text{ and constants } s_d = s(F_d) > 0 \text{ such that} \\ F_d^{-1} = s_d \cdot G^{-1} \text{ for all } d = 1, \dots, D. \end{cases} \quad (5)$$

The site-specific factors s_d are often referred to as index floods, and a set of sites for which Equation (5) holds is called *homogeneous*. When such a model holds, the regional quantile can be estimated based on all D sites, and thus it is plausible that any estimation procedure based on this pooled sample will be more accurate and encounter less variance. On the other hand, in case the model assumptions do not hold, imposing the model wrongly may lead to a large estimation bias, as pointed out in Lettenmaier et al. (1987).

It is thus important to test whether Equation (5) is satisfied for a set of locations prior to fitting the model. Procedures that aim at testing whether a region is homogeneous are referred

to as *homogeneity tests*. An overview of the most common tests applied in this context is given in Viglione et al. (2007). The most widely used approach is the heterogeneity measure of Hosking and Wallis (Hosking and Wallis (1993)). It is based on a comparison of observed and expected variability of L-moment (LM) ratios. LMs, introduced in Hosking (1990), are certain linear combinations of expectations of order statistics, and they can also be expressed as linear combinations of PWMs as defined in Equation (4) with $p = 1$ and $s = 0$. They can be used to summarise characteristics of probability distributions which are comparable to e.g. location or skewness, and for a range of distributions, many of which are regularly considered in hydrological applications, there is an explicit relation between the distribution parameters and L-moments (a list can be found in Hosking (1990)). However, as pointed out in Viglione et al. (2007), the Hosking-Wallis (HW) heterogeneity measure, as well as the other standard approaches, have the major drawback that inter-site dependence is not accounted for, even though spatial proximity suggests that observations are not independent in space. Mentionable exceptions are Castellarin et al. (2008), who propose to add a correction term to the test statistic of the HW heterogeneity measure, or Lilienthal et al. (2018), who elaborated a generalisation of the HW heterogeneity measure that accounts for inter-site dependence through copula modelling.

Common homogeneity tests such as the HW measure do not assume that margins are GEV-distributed, but are rather rank-based or assume that the margins follow kappa distributions, which are four-parametric distributions that include the GEV distributions as special cases. However, when working with block maxima, approximating the marginal distributions with GEV distributions is often reasonable. When directly assuming GEV margins, i.e. $F_d = G_{\boldsymbol{\vartheta}_d}$, $\boldsymbol{\vartheta}_d \in \Theta = \mathbb{R} \times (0, \infty) \times \mathbb{R}$ for each $d = 1, \dots, D$, the homogeneity hypothesis of Equation (5) can be expressed as

$$H_0 : h(\boldsymbol{\vartheta}) = 0 \quad \text{vs.} \quad H_1 : h(\boldsymbol{\vartheta}) \neq 0, \quad (6)$$

where $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}'_1, \dots, \boldsymbol{\vartheta}'_D)'$ and where $h : \mathbb{R}^{3D} \rightarrow \mathbb{R}^{2(D-1)}$ is a known continuously differentiable function.

The second article presented in Chapter 2 exploits this representation to construct new homogeneity tests, or, more general, arbitrary tests that can be expressed through Equation (6) for a continuously differentiable function $h : \mathbb{R}^{3D} \rightarrow \mathbb{R}^q$, $q \in \mathbb{N}$, based on limiting distributions of estimators for $\boldsymbol{\vartheta}$. Thereby, three different estimators are considered: the LM estimator, the *trimmed* L-moment (TLM) estimator and a (pseudo) maximum likelihood estimator. TLMs, which were introduced by Elamir and Seheult (2003) and which can also exist for heavy-tailed distributions with non-existing mean, provided that trimming parameters are chosen appropriately, are generalisations of LMs, making LMs a special case of TLMs. As with the PWM method, the TLM method establishes a relationship between the GEV parameters and the TLMs, and empirical versions of the TLMs can be used to obtain cor-

responding parameter estimates. For univariate i.i.d. observations, asymptotic normality for empirical TLMS has been worked out in Elamir and Seheult (2003) and Hosking (2007), from which asymptotic normality of GEV parameter estimation can be obtained with the delta method. For ML estimation, asymptotic normality has been elaborated in Bücher and Segers (2017). These results are now expanded to the multivariate case, taking possible inter-site dependencies into account. Based on these limiting distributions, asymptotic level- α -tests for hypotheses as in Equation (6) are developed. Since the index flood hypothesis $\mathcal{H}_{0,IF}$ can be expressed in this way, a simulation study is conducted for the derived tests as well as the heterogeneity measure of Hosking and Wallis. Their finite-sample performances in terms of empirical type I error under the null hypothesis and empirical power under the alternative are compared. Due to poor performance for short to medium sample sizes, a bootstrap scheme based on max-stable process models is proposed. Max-stable processes constitute the class of possible limit processes that can arise as the pointwise limit of normalised maxima of independent and identically distributed random fields $\{Y_i(x) : x \in \mathbb{R}^p\}$ (de Haan (1984)), and are thus the natural choice when modelling spatial extremes (see e.g. Smith (1990) and Davison et al. (2012)). In a further simulation study, this bootstrap procedure is found to be more powerful than the asymptotic tests.

So far, all models and methods presented have assumed some kind of temporal stationarity for the observations. However, there is consensus in the climate research community that our climate is changing, see for example Alexander et al. (2006) or Seneviratne et al. (2012), and the latest report of the intergovernmental panel on climate change (IPCC) even states that human life has unequivocally influenced global warming (IPCC (2021)). A new field of study, the *extreme event attribution* (EEA) science, has recently emerged, dedicated to the task of attributing observed weather extremes to climate change. As climate change is suspected to alter the frequency and severity of extreme weather events in particular (Stocker et al. (2013)), there has been some interest in and activity on modelling non-stationary extremes lately, e.g. with the block maxima method and GEV models in El Adlouni et al. (2007) (for ML estimation), Ribereau et al. (2008) (for PWM estimation), Cheng et al. (2014) (using Bayesian inference), or with the POT method in Eastoe and Tawn (2009). Also, spatio-temporal models have been considered, e.g. in Westra and Sisson (2011) for detecting trends. Several suggestions about possible interpretations of return levels in a non-stationary climate have been made, e.g. in Cooley (2012) and Rootzén and Katz (2013).

One building block within EEA studies is trend detection, which is usually carried out on observational data (rather than simulated data from climate models, which constitutes a second building block of EEA). To account for non-stationarities in extremes, the approach which has become standard is to model block maxima with a GEV distribution whose parameters $\mu(t)$, $\sigma(t)$ and $\gamma(t)$ are allowed to depend on a temporal covariate in a predefined manner (Section 4 in Philip et al. (2020)). This covariate, being an indicator for climate

change, is often chosen as smoothed global mean surface temperature (GMST) anomaly or atmospheric CO₂ concentration. Depending on the variable of interest, the GEV distribution is assumed to either shift or scale with respect to the covariate. When denoting the covariate's value for time index t with $c^{(t)}$, the shift model postulates that the block maximum $M^{(t)}$ of block t (e.g., year or season t) follows a GEV distribution with a linear trend in the location parameter, i.e.

$$\mu(t) = \mu_0 + \alpha c^{(t)}, \quad \sigma(t) \equiv \sigma_0, \quad \gamma(t) \equiv \gamma_0, \quad (7)$$

for some parameters $\mu_0, \gamma_0, \alpha \in \mathbb{R}$ and $\sigma_0 > 0$. The shift model is most commonly used for modelling temperature extremes, as done e.g. in van Oldenborgh et al. (2018). For precipitation and wind extremes, a scale model, as e.g. used in van der Wiel et al. (2017) and van Oldenborgh et al. (2017), is the standard choice. It assumes that, for μ_0 and $\sigma_0 > 0$, γ and $\alpha \in \mathbb{R}$, the t -th block maximum $M^{(t)}$ can be modelled with a GEV distribution with parameters

$$\mu(t) = \mu_0 \exp\left(\frac{\alpha}{\mu_0} c^{(t)}\right), \quad \sigma(t) = \sigma_0 \exp\left(\frac{\alpha}{\mu_0} c^{(t)}\right), \quad \gamma(t) \equiv \gamma_0, \quad (8)$$

which is inspired by the Clausius-Clapeyron relation from thermodynamics (see van der Wiel et al. (2017) for more details on this). Once a suitable model has been fitted, one attempts to answer the attribution question, for example based on whether the trend parameter α is found to be significantly different from zero. Further, one can compare the return periods for a specific event m for several reference climates t_i , e.g., pre-industrial and current, by estimating values of

$$\text{RP}_{t_i}(m) = \{1 - G_{(\mu(t_i), \sigma(t_i), \gamma(t_i))}(m)\}^{-1} \quad (9)$$

to illustrate how risk has changed. Similar calculations can be made for return levels.

As in the stationary case, estimation uncertainty poses a problem, and it might even be exacerbated due to the additional parameter that is estimated. Therefore, in case there are several measuring stations near the location of interest, the pooling approach has become popular to improve the data situation, as e.g. applied in Vautard et al. (2015) and Eden et al. (2016). Further, pooling has been used on gridded data sets, for example in Kreienkamp et al. (2021). The pooled data is usually treated as one sample from a GEV distribution with unknown parameters. Therefore, the inherent assumption of coinciding parameters should be checked carefully before fitting and relying on such a model. The state of the art method for validating the model assumptions consists of several makeshift diagnostic methods, such as a simple comparison of location-wise estimated GEV parameters (Kreienkamp et al. (2021)).

The article presented in Chapter 2.3 offers a statistically more sound approach to this problem, again based on homogeneity tests. In this context, a set of locations $\{1, \dots, D\}$ is called homogeneous when their location-wise GEV parameters coincide, i.e., when the hypothesis of ‘equal distribution’

$$\mathcal{H}_{0,ED} : \exists (\mu, \sigma, \gamma, \alpha)' \in \Theta \forall d \in \{1, \dots, D\} : (\mu_d, \sigma_d, \gamma_d, \alpha_d)' = (\mu, \sigma, \gamma, \alpha)' \quad (10)$$

with Θ the respective parameter space, is satisfied.

Besides only testing whether a given set of locations is homogeneous, a new multiple testing approach can be applied for detecting a region that can be considered homogeneous. The proposed procedures are based on a pseudo ML estimator for the $4D$ -dimensional parameter vector $(\mu_1, \sigma_1, \gamma_1, \alpha_1, \dots, \mu_D, \sigma_D, \gamma_D, \alpha_D)'$, for which approximate normality is motivated, so that a Wald-type test statistic can be derived. In order to obtain reliable p -values, a bootstrap method is proposed that involves either max-stable process models or, in case $D = 2$, bivariate extreme value distributions, and can thus capture the (spatial) dependence of the observations. Modelling the dependence structure of the observations has yet another advantage: based on the model fit, one can estimate the *regional m -return period* for a fixed reference climate, which corresponds to the period within which one exceedance of m for at least one of the locations that make up the region can be expected, assuming a stationary climate with conditions of the reference climate. The proposed methods are tested in a Monte Carlo simulation study and then applied in a case study about heavy precipitation.

The remainder of this thesis is structured as follows. The previously mentioned articles to which the author of this thesis contributed are listed in section 2: Section 2.1 starts with the research paper on the sliding block maxima method for piecewise stationary time series with application to the PWM estimator; Sections 2.2 and 2.3 provide the articles on homogeneity tests. Here, the article on procedures for stationary model assumptions is presented first, followed by the article for the non-stationary framework.

Section 3 gives a brief outlook on possible extensions of the presented research. The author’s contribution statement is given in the appendix.

2 Included articles

The following list contains the articles that are included in this thesis.

- 2.1 Bücher, A. and Zanger, L. (2023). On the Disjoint and Sliding Block Maxima method for piecewise stationary time series.
Accepted for publication in *The Annals of Statistics*. Previous version publicly available at arXiv:2110.15576v1.
- 2.2 Lilienthal, J., Zanger, L., Bücher, A. and Fried, R. (2022). A note on statistical tests for homogeneities in multivariate extreme value models for block maxima.
Environmetrics, 33(7):e2746 (DOI: 10.1002/env.2746).
- 2.3 Zanger, L., Bücher, A., Kreienkamp, F., Lorenz, P. and Tradowsky, J. (2023). Regional Pooling in Extreme Event Attribution Studies: an Approach Based on Multiple Statistical Testing.
Submitted to *Extremes* and publicly available at arXiv:2301.06295v1.

ON THE DISJOINT AND SLIDING BLOCK MAXIMA METHOD FOR PIECEWISE STATIONARY TIME SERIES

BY AXEL BÜCHER¹ AND LEANDRA ZANGER²

¹*Heinrich-Heine-University, Düsseldorf, axel.buecher@hhu.de*

²*Heinrich-Heine-University, Düsseldorf, Leandra.Zanger@hhu.de*

Modeling univariate block maxima by the generalized extreme value distribution constitutes one of the most widely applied approaches in extreme value statistics. It has recently been found that, for an underlying stationary time series, respective estimators may be improved by calculating block maxima in an overlapping way. A proof of concept is provided that the latter finding also holds in situations that involve certain piecewise stationarities. A weak convergence result for an empirical process of central interest is provided, and further details are exemplarily worked out for the probability weighted moment estimator. Irrespective of the serial dependence, the asymptotic estimation variance is shown to be smaller for the new estimator. In extensive simulation experiments, the finite-sample variance was typically found to be smaller as well, while the bias stays approximately the same. The results are illustrated by Monte Carlo simulation experiments and are applied to a common situation involving temperature extremes in a changing climate.

1. Introduction. The annual or seasonal maximum of a certain variable of interest is a common target distribution, in particular in environmental statistics [25, 1]. For instance, hydrologists are interested in maximal river discharges to facilitate flood protection, while meteorologists and climatologists study maximal temperatures, precipitation or wind speeds, collected over certain spatial or temporal regions. The latter comprises the emerging field of extreme event attribution studies [35], which aim at exploring how the probability of certain extreme events evolve in the context of a changing climate due to anthropogenic activities.

The underlying statistical principle is known as the block maxima method and dates back to [22], see also the monographs [11, 1]. In its simplest form, it is postulated that a sample of successive (annual) block maxima constitutes an independent and identically distributed (i.i.d.) sample from the generalized extreme value (GEV) distribution, as suggested by the asymptotics formulated in the Fisher-Tippet-Gnedenko Theorem [21]. The model may then be fitted by any method of choice, the most popular approaches being maximum likelihood [32, 6] and the probability weighted moment (PWM) method [24].

Considering the validation of statistical methodology (like proving consistency and asymptotic normality of estimators), it has long been assumed that the block maxima sample is a genuine independent sample from the GEV-distribution. From a mathematical viewpoint, this assumption seems overly simplified: neither does it allow to quantify a possible bias due to the fact that block maxima are only asymptotically GEV-distributed, nor does it quantify to what extent possible temporal dependencies in the underlying sample are negligible. Notable exceptions are [16, 20, 17], who investigate respective methods under the assumption that block maxima of size $r = r_n \rightarrow \infty, r_n = o(n)$ are calculated based on an underlying i.i.d. series of length n (corresponding to, say, daily observations). The latter is however still not really fitting to typical applications of the block maxima method, where serial independence

AMS 2000 subject classifications: Primary 62G32, 62F12; secondary 62P12, 60G70.

Keywords and phrases: Asymptotic normality, extreme value statistics, Marshall–Olkin distribution, return level, temperature extremes.

of a daily time series is rarely the case (another nuisance are potential seasonalities, which will be discussed below). Extensions to the case of a strictly stationary time series have been worked out in [5, 7], for the estimation of extreme value copulas in a multivariate context and estimation of Fréchet parameters in a univariate heavy tailed situation, respectively. The new viewpoint has also led to methodological improvements, as it allows to study estimators which are based on block maxima calculated from sliding (overlapping) blocks of observations. Perhaps surprisingly, respective estimators have been shown to be more efficient than their disjoint blocks counterparts in certain general situations [8, 36]. In the i.i.d case, [30] recently provided a further methodological improvement based on what has been called the all-block maxima method; a method that is, however, not easily transferable to the time series case except the extremal index [29] is one. Furthermore, [18, 10] study the use of sliding blocks with POT-type estimators.

This paper’s main contribution is a surprising proof of concept that the sliding block maxima method may even yield more efficient estimators when applied to datasets that result in a non-stationary behavior of the sample of sliding block maxima. More precisely, suitable asymptotic theory is developed for a sampling scheme that involves an underlying triangular array consisting of independent and identically distributed stretches of observations extracted from a stationary time series. The framework is designed to asymptotically mimic the practically relevant situation encountered in environmental statistics where, due to seasonalities, stationarity can only be (approximately) guaranteed for, say, daily observations collected throughout the summer months.

Under the prescribed sampling scheme, as well as under a classical sampling scheme involving a plain stationary time series, asymptotic theory is developed for (1) an empirical process of pure theoretical interest as well as for (2) the PWM estimator of practical interest. It is worthwhile to mention that the restriction to PWM estimators is partly arbitrary, and that similar findings can be expected to hold for other estimators of practical interest. One of the reasons we opted for PWM is that we extend, as a by-product, results from [20] on the disjoint blocks maxima PWM estimator in an i.i.d. context.

The asymptotic results are similar but not the same as in [8, 36]: it is found that, despite non-stationarity, the sliding blocks method *works* and yields smaller asymptotic variances than the disjoint blocks method. However, the asymptotic bias is only guaranteed to be the same for stationary data. Within extensive simulation experiments on the PWM estimators, it is found that the overall improvement of the sliding blocks version over its disjoint blocks counterpart is remarkably large for negative shape parameters, while only small improvements are visible for positive shapes. For situations involving strong serial dependencies and comparably small block sizes, one may even have a worse performance; however, such unfavorable combinations appear to be untypical for practical applications. The improvement for negative shape parameters is illustrated in Figure 1, where we depict the mean squared estimation error for the estimation of the shape parameter γ for a fixed block size $r = 90$ (roughly corresponding to the length of a season) and increasing number of seasons.

Negative shape parameters are often found when analyzing temperature extremes, for which shapes are typically within the range -0.4 to -0.2 . A respective case study is worked out, where we also deal with non-stationarities in the location parameter of the GEV-model. The considered model is commonly employed in extreme event attribution studies, see [31]. A parametric bootstrap device is proposed to assess estimation uncertainties.

The (sliding) blocks method exhibits an important methodological advantage over the competing peaks-over-threshold (POT) approach [14] when the ultimate goal consists of assessing return levels or return periods. Indeed, for the latter purpose, methods based on the POT approach typically require an application of a declustering approach (or an estimator of the extremal index) to take care of the time series’ serial dependence. This is not necessary

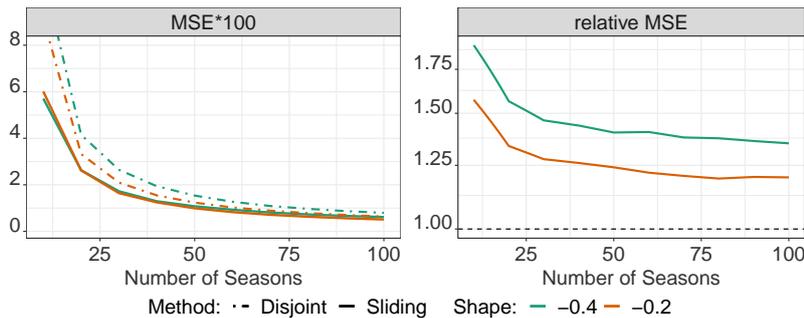


FIG 1. Mean squared error for the estimation of γ for the disjoint and sliding blocks estimator (left), and the ratio $\text{MSE}(\text{disjoint})/\text{MSE}(\text{sliding})$ (right). The data generating process is an $\text{AR}(0.5)\text{-GPD}(\gamma)$ -model, sampling scheme (S2) with fixed block size $r = 90$, as described in Section 4.

for the (sliding) blocks method, where the serial dependence only shows up in the scaling sequences associated with the max-domain of attraction condition (see Condition 2.1 below for details), both of which are automatically estimated by the method. Respective details are worked out in Section 3.2.

This paper is organized as follows. Section 2 contains details on the basic model assumptions and a weak convergence result on an empirical process of central interest. Details on the PWM estimator are worked out in Section 3. A large scale Monte Carlo simulation study is presented in Section 4. The case study on temperature extremes can be found in Section 5, followed by a conclusion in Section 6. The most important proofs are worked out in Sections 7, with some lengthy parts and some less central parts postponed to a supplementary material [9]. Theoretical results from the supplement are numbered by capital letters; e.g., Lemma B.3. All convergences are for $n \rightarrow \infty$, if not mentioned otherwise. The generalized (left-continuous) inverse of a cumulative distribution function (c.d.f.) F is denoted by F^{\leftarrow} .

2. A new sampling scheme and some general theoretical results. Recall the Generalized Extreme Value (GEV) distribution with parameters μ (location), σ (scale) and γ (shape), defined by its cumulative distribution function

$$G_{(\mu, \sigma, \gamma)}(x) := \exp \left[- \left\{ 1 + \gamma \left(\frac{x - \mu}{\sigma} \right) \right\}^{-\frac{1}{\gamma}} \right], \quad 1 + \gamma \frac{x - \mu}{\sigma} > 0.$$

If $\theta = (\mu, \sigma, \gamma)' = (0, 1, \gamma)'$, we will use the abbreviation $G_{(0,1,\gamma)} = G_\gamma$. The support of G_γ is denoted by $S_\gamma = \{x \in \mathbb{R} : 1 + \gamma x > 0\}$.

An extension of the classical extremal types theorem to strictly stationary time series [29] implies that, under suitable conditions, affinely standardized maxima extracted from a stationary time series converge to the GEV-distribution. We make this an assumption, and additionally require the scaling sequences to exhibit some common regularity inspired by the max-domain of attraction condition in the i.i.d. case [14].

CONDITION 2.1 (Max-domain of attraction). Let $(X_t)_{t \in \mathbb{Z}}$ denote a stationary time series. There exist sequences $(a_r)_r \subset \mathbb{R}_+$, $(b_r)_r \subset \mathbb{R}$ and $\gamma \in \mathbb{R}$, such that, for any $s > 0$,

$$(1) \quad \lim_{r \rightarrow \infty} \frac{a_{\lfloor rs \rfloor}}{a_r} = s^\gamma, \quad \lim_{r \rightarrow \infty} \frac{b_{\lfloor rs \rfloor} - b_r}{a_r} = \frac{s^\gamma - 1}{\gamma},$$

where the second limit is interpreted as $\log(s)$ if $\gamma = 0$. Moreover, for $r \rightarrow \infty$,

$$(2) \quad Z_r = \frac{\max(X_1, \dots, X_r) - b_r}{a_r} \xrightarrow{d} Z \sim G_\gamma.$$

Note that (1) and (2) may for instance be deduced from Leadbetter's $D(u_n)$ condition, a domain-of-attraction condition on the associated i.i.d. sequence and a weak requirement on the convergence of the c.d.f. of Z_r , see Theorem 10.4 in [1].

The max-domain of attraction condition allows to formulate two sampling mechanisms used throughout this paper.

CONDITION 2.2 (Observation scheme). For sample size $n \in \mathbb{N}$, we have observations $X_{1,n}, \dots, X_{n,n}$ that do not contain ties with probability one, such that either (S1) or (S2) holds, where:

(S1) $(X_{1,n}, \dots, X_{n,n}) = (X_1, \dots, X_n)$ is an excerpt from a strictly stationary time series satisfying Condition 2.1 with continuous marginal c.d.f F .

(S2) For some block length sequence $(r_n)_n \subset \mathbb{N}$ diverging to infinity such that $r_n = o(n)$, we have

$$(X_{1,n}, \dots, X_{n,n}) = (Y_1^{(1)}, \dots, Y_{r_n}^{(1)}, Y_1^{(2)}, \dots, Y_{r_n}^{(2)}, \dots, Y_1^{(m)}, \dots, Y_{r_n}^{(m)}, Y_1^{(m+1)}, \dots, Y_{n-mr_n}^{(m+1)}),$$

where $m = m_n = \lfloor n/r_n \rfloor$ and where $(Y_t^{(1)})_t, (Y_t^{(2)})_t, \dots$ denote i.i.d. copies from a stationary time series satisfying Condition 2.1 with continuous marginal c.d.f F . Note that $Y_t^{(j)}$ should be regarded as the t th observation in the j th season.

Sampling scheme (S2) shall represent typical environmental applications which are subject to seasonalities. The parameter r_n may correspond to, say, the number of daily observations within the summer months. For such a situation, it appears reasonable to assume strict stationarity within a particular summer, and stochastic independence and distributional equality between multiple summers. In order to obtain meaningful asymptotic results, which in particular cover a sliding blocks version, r_n must be assumed to go to infinity.

REMARK 2.3 (Possible relaxations of Condition 2.2). It is worthwhile to mention that sampling scheme (S2) has been chosen as a starting point for this paper because it is, on the one hand, reasonably general to capture typical real data situations and, on the other hand, simple enough to allow for accessible proofs. It may be extended in various ways: first of all, different 'seasons' may be assumed to be serially dependent and to satisfy certain mixing conditions; the necessary changes in the proofs would mostly require bringing together arguments from the (S1) and (S2) case. Next, sampling schemes (S1) and (S2) may be subsumed under a more general condition: denoting by $S(s_n)$ the sampling scheme that consists of concatenating independent 'seasonal blocks' of size s_n , we observe that (S1) is the same as $S(n)$, while (S2) is the same as $S(r_n)$. At the cost of more sophisticated conditions and proofs (in particular, a more complex version of Condition 3.2 below would be needed), one may extend the results in this paper to the case where $r_n/s_n \rightarrow c \in (0, 1]$ (which, in practice, may represent monthly maxima, $r_n = 30$, after concatenating seasons, $s_n = 90$). The higher level of complexity needed for handling this situation results from the fact that even the disjoint blocks maxima sample may not be stationary anymore (for instance, if $s_n = 1.5r_n$). Finally, the no-tie assumption in Condition 2.2 is merely made for convenience. At the cost of more sophisticated proofs and conditions on the serial dependence, it can possibly be dispensed with. Similar arguments apply to the continuity assumption on F .

2.1. *Two approximate block maxima samples from the GEV distribution.* Subsequently, we write $X_i = X_{i,n}$ for simplicity. The block maxima method with block size parameter $r = r_n \in \{1, \dots, n\}$ is traditionally based on the sample of disjoint block maxima $\mathcal{M}_r^{(\text{db})} = \mathcal{M}_{r,n}^{(\text{db})} = \{M_{r,1}^{(\text{db})}, \dots, M_{r,m}^{(\text{db})}\}$, where the block maxima are defined by

$$M_{r,j}^{(\text{db})} := \max(X_{(j-1)r+1}, \dots, X_{jr}), \quad j \in \{1, \dots, m\},$$

and where $m = m_n = \lfloor n/r_n \rfloor$ denotes the number of disjoint blocks that fit into $\{1, \dots, n\}$. For data arising from one of the sampling schemes in Condition 2.2, it follows that the sample $\mathcal{M}_{r,n}^{(\text{db})}$ is stationary with marginal c.d.f. denoted by

$$(3) \quad F_r(x) = \mathbb{P}(\max(X_1, \dots, X_r) \leq x), \quad x \in \mathbb{R}.$$

As a consequence of Condition 2.1, we have $F_r(x) \approx G_{(b_r, a_r, \gamma)}$, whence the parameters (b_r, a_r, γ) may be estimated by any method of choice for fitting the GEV-distribution.

As mentioned in the introduction, the sample of sliding block maxima $\mathcal{M}_r^{(\text{sb})} = \mathcal{M}_{r,n}^{(\text{sb})} = \{M_{r,1}^{(\text{sb})}, \dots, M_{r,n-r+1}^{(\text{sb})}\}$ defined by

$$M_{r,j}^{(\text{sb})} := \max(X_j, \dots, X_{j+r-1}), \quad j \in \{1, \dots, n-r+1\}$$

provides an attractive alternative to the sample $\mathcal{M}_{r,n}^{(\text{db})}$. In fact, under sampling scheme (S1), we have $M_{r,j}^{(\text{sb})} \sim F_r$ for all $j \in \{1, \dots, n-r+1\}$ as well, whence respective estimators can be expected to work, in particular when based on the method of moments. Note that the asymptotic analysis becomes substantially more difficult due to the strong serial dependence between the sliding block maxima.

In this paper, we also advocate the use of sliding block maxima under the possibly more realistic sampling scheme (S2). Compared to (S1), an obstacle occurs: the c.d.f. of $M_{r,j}^{(\text{sb})}$,

$$(4) \quad F_{r,j}(x) = \mathbb{P}(\max(X_j, \dots, X_{j+r-1}) \leq x), \quad x \in \mathbb{R},$$

is in general no longer independent of j . Perhaps surprisingly, it can be shown that $F_{r,j} \approx G_{(b_r, a_r, \gamma)}$ for all j and sufficiently large r :

LEMMA 2.4 (Asymptotic stationarity of sliding block maxima). *Suppose one of the sampling schemes from Condition 2.2 is met. Then, for every $\xi \in [0, 1]$ and $z \in \mathbb{R}$,*

$$\lim_{r \rightarrow \infty} F_{r,1+\lfloor r\xi \rfloor}(a_r z + b_r) = G_\gamma(z).$$

As a consequence of this lemma, estimators based on $\mathcal{M}_{r,n}^{(\text{sb})}$ can still be expected to work under (S2), provided the block size r is reasonably large. Further extensions to joint convergence of two block maxima are provided in Lemmas B.3 and B.4. A sub-asymptotic discussion for a model involving ARMAX-dynamics is provided in Section D.

2.2. *An empirical process associated with rescaled block maxima.* A central theoretical ingredient for all subsequent results (and, presumably, for possible future results on other estimators involving the sliding block maxima method) is weak convergence of the centered empirical process associated with the empirical distribution function of the (unobservable) rescaled block maxima samples $Z_{r,1}^{(\text{db})}, \dots, Z_{r,m}^{(\text{db})}$ and $Z_{r,1}^{(\text{sb})}, \dots, Z_{r,n-r+1}^{(\text{sb})}$, where

$$(5) \quad Z_{r,j}^{(\text{db})} = \frac{M_{r,j}^{(\text{db})} - b_r}{a_r}, \quad Z_{r,j}^{(\text{sb})} = \frac{M_{r,j}^{(\text{sb})} - b_r}{a_r}.$$

Throughout its proof, we are going to apply common blocking techniques, whence the block length r must be well-adapted to the serial dependence of the time series. Suitable control may be provided by mixing conditions that were also imposed in related situations in [5, 7, 8].

CONDITION 2.5. For the block size sequence $(r_n)_n$ it holds that

- (i) $r_n \rightarrow \infty$ and $r_n = o(n)$.
- (ii) $\left(\frac{n}{r_n}\right)^{1/2} \beta(r_n) = o(1)$ and $\left(\frac{n}{r_n}\right)^{1+\omega} \alpha(r_n) = o(1)$ for some $\omega > 0$.
- (iii) There exists a sequence $(\ell_n)_n \subset \mathbb{N}$ such that $\ell_n \rightarrow \infty$, $\ell_n = o(r_n)$, $\frac{n}{r_n} \alpha(\ell_n) = o(1)$ and $\frac{r_n}{\ell_n} \alpha(\ell_n) = o(1)$.

Here, α and β denote the α - and β -mixing coefficients of the time series $(X_t)_t$ that was introduced in Condition 2.1 (see [3] for a precise definition). Note that Conditions (ii) and (iii) imply that the block length sequence r_n must not be too small.

Subsequently, for $z \in \mathbb{R}$, let

$$\hat{H}_r^{(\text{db})}(z) = \frac{1}{m} \sum_{j=1}^m \mathbf{1}(Z_{r,j}^{(\text{db})} \leq z), \quad \hat{H}_r^{(\text{sb})}(z) = \frac{1}{n-r+1} \sum_{j=1}^{n-r+1} \mathbf{1}(Z_{r,j}^{(\text{sb})} \leq z)$$

and

$$(6) \quad \bar{H}_r(z) = \frac{1}{r} \sum_{j=1}^r H_{r,j}(z), \quad H_{r,j}(z) = \mathbb{P}(Z_{r,j}^{(\text{sb})} \leq z).$$

Note that $\mathbb{E}[\hat{H}_r^{(\text{mb})}(z)] = \bar{H}_r(z)$, unless $\text{mb} = \text{sb}$ and sampling scheme (S2) is met, in which case we have $\mathbb{E}[\hat{H}_r^{(\text{sb})}(z)] = \bar{H}_r(z) + O(m^{-1})$. The following central result is similar to Theorem 2.10 in [36], despite under different assumptions (in particular sampling scheme (S2)).

THEOREM 2.6. Consider one of the sampling schemes from Condition 2.2. Under Condition 2.5, we have for $\text{mb} \in \{\text{db}, \text{sb}\}$

$$\mathbb{H}_r^{(\text{mb})} = \sqrt{\frac{n}{r}} \left(\hat{H}_r^{(\text{mb})} - \bar{H}_r \right) \xrightarrow{d} \mathbb{H}^{(\text{mb})} = \mathbb{C}^{(\text{mb})} \circ G_\gamma$$

in $\ell^\infty(\mathbb{R})$ equipped with the supremum metric, where $\mathbb{C}^{(\text{db})}$ is a standard Brownian bridge on $[0, 1]$ and where $\mathbb{C}^{(\text{sb})}$ is a centered Gaussian process with covariance function

$$(7) \quad \text{Cov}(\mathbb{C}^{(\text{sb})}(u), \mathbb{C}^{(\text{sb})}(v)) = 2 \left(\frac{uv - u \wedge v}{\ln(u \vee v)} - uv \right), \quad u, v \in (0, 1).$$

Moreover, the limit processes $\mathbb{H}^{(\text{mb})}$ are almost surely contained in $C_b(\mathbb{R})$ (the space of continuous and bounded real-valued functions on \mathbb{R}) and satisfy

$$(8) \quad \text{Cov}(\mathbb{C}^{(\text{sb})}(u_1), \dots, \mathbb{C}^{(\text{sb})}(u_d)) \leq_L \text{Cov}(\mathbb{C}^{(\text{db})}(u_1), \dots, \mathbb{C}^{(\text{db})}(u_d))$$

for all $u_1, \dots, u_d \in (0, 1)$ and $d \in \mathbb{N}$, where \leq_L denotes the Loewner-ordering between symmetric matrices.

Convergence of the finite-dimensional distributions in Theorem 2.6 is a consequence of a more general multivariate central limit theorem of independent interest, see Theorem B.1 in the supplement [9].

3. PWM estimators based on the block maxima method. Throughout this section, let M denote a GEV-distributed random variable with parameter $\theta = (\mu, \sigma, \gamma)'$. For $\gamma < 1$, the first three probability weighted moments of M are given by

$$(9) \quad \beta_{\theta,k} := \mathbb{E}[MG_{(\mu,\sigma,\gamma)}^k(M)] = \frac{1}{k+1} \left[\mu - \frac{\sigma}{\gamma} \{1 - (k+1)^\gamma \Gamma(1-\gamma)\} \right],$$

where $k \in \{0, 1, 2\}$. As shown by [24], we obtain the following equation system between θ and $(\beta_{\theta,0}, \beta_{\theta,1}, \beta_{\theta,2})$:

$$(10) \quad \begin{cases} \gamma = g_1^{-1} \left(\frac{3\beta_{\theta,2} - \beta_{\theta,0}}{2\beta_{\theta,1} - \beta_{\theta,0}} \right) \\ \sigma = g_2(\gamma) (2\beta_{\theta,1} - \beta_{\theta,0}) \\ \mu = \beta_{\theta,0} + \sigma g_3(\gamma) \end{cases}$$

where

$$g_1(\gamma) = \frac{3^\gamma - 1}{2^\gamma - 1}, \quad g_2(\gamma) = \frac{\gamma}{\Gamma(1-\gamma)(2^\gamma - 1)}, \quad g_3(\gamma) = \frac{1 - \Gamma(1-\gamma)}{\gamma}$$

with $g_1(0) = \log 3 / \log 2$, $g_2(0) = 1 / \log 2$ and $g_3(0) = -\gamma_{\text{EM}}$ defined by continuity. Here, Γ denotes the Gamma function and γ_{EM} is the Euler-Mascheroni constant. The PWM estimator is then defined by replacing the respective moments on the right-hand side of (10) by empirical versions and successively solving for γ , σ and μ . Several (asymptotic equivalent) empirical versions suggest itself, and throughout this paper we opt for the version proposed in [28], that is,

$$\hat{\beta}_0(\mathcal{M}) = \frac{1}{n} \sum_{i=1}^n M_i, \quad \hat{\beta}_1(\mathcal{M}) = \frac{1}{n} \sum_{i=1}^n \frac{i-1}{n-1} M_{(i)}, \quad \hat{\beta}_2(\mathcal{M}) = \frac{1}{n} \sum_{i=1}^n \frac{(i-1)(i-2)}{(n-1)(n-2)} M_{(i)},$$

where $M_{(1)} \leq \dots \leq M_{(n)}$ is the order statistic of a sample $\mathcal{M} = \{M_1, \dots, M_n\}$ which is to be fitted to the GEV-distribution. It is worthwhile to mention that these estimators are unbiased in case \mathcal{M} is an i.i.d. sample. Indeed, we may rewrite $\hat{\beta}_1(\mathcal{M}) = \{n(n-1)\}^{-1} \sum_{i \neq j} M_i \mathbf{1}(M_j \leq M_i)$, whence $\mathbb{E}[\hat{\beta}_1(\mathcal{M})] = \mathbb{E}[M_i \mathbf{1}(M_j \leq M_i)] = \beta_{\theta,1}$, and a similar calculation can be made for $\hat{\beta}_2(\mathcal{M})$. The resulting estimator for θ based on solving (10) will be denoted by $\hat{\theta}(\mathcal{M})$. The estimators of ultimate interest in this paper are

$$(11) \quad \hat{\theta}_r^{(\text{db})} = \hat{\theta}(\mathcal{M}_{r,n}^{(\text{db})}), \quad \hat{\theta}_r^{(\text{sb})} = \hat{\theta}(\mathcal{M}_{r,n}^{(\text{sb})}),$$

which are derived from the empirical weighted moments $\hat{\beta}_{r,k}^{(\text{mb})} = \hat{\beta}_k(\mathcal{M}_{r,n}^{(\text{mb})})$ for $\text{mb} \in \{\text{db}, \text{sb}\}$ and are to be considered as estimators for $\theta_r = (b_r, a_r, \gamma)'$.

REMARK 3.1 (Bias-reduced sliding blocks estimator). Well-known heuristics suggest that block maxima are asymptotically independent when calculated based on non-overlapping time periods, and that they are asymptotically dependent otherwise (see Lemma B.3 for a rigorous result). As a consequence, the sliding blocks empirical PWMs from (11) may exhibit a certain ‘dependency’ bias. To remove this bias, one may alternatively consider the estimators $\tilde{\beta}_{r,0}^{(\text{sb})} = \tilde{\beta}_{r,0}^{(\text{db})}$,

$$\tilde{\beta}_{r,1}^{(\text{sb})} = \frac{1}{|D_n(2)|} \sum_{(i,j) \in D_n(2)} M_{r,i}^{(\text{sb})} \mathbf{1}(M_{r,j}^{(\text{sb})} \leq M_{r,i}^{(\text{sb})})$$

$$\tilde{\beta}_{r,2}^{(\text{db})} = \frac{1}{|D_n(3)|} \sum_{(i,j,j') \in D_n(3)} M_{r,i}^{(\text{sb})} \mathbf{1}(M_{r,j}^{(\text{sb})} \leq M_{r,i}^{(\text{sb})}) \mathbf{1}(M_{r,j'}^{(\text{sb})} \leq M_{r,i}^{(\text{sb})})$$

where $D_n(2)$ denotes the set of all pairs $(i, j) \in \{1, \dots, n-r+1\}^2$ such that $I_i \cap I_j = \emptyset$ and where $D_n(3)$ is the set of all triples $(i, j, j') \in \{1, \dots, n-r+1\}^3$ such that $I_i \cap I_j = I_i \cap I_{j'} = I_j \cap I_{j'} = \emptyset$, with $I_i = \{i, \dots, i+r-1\}$. Obviously, the larger the block size r , the more $\tilde{\beta}_{r,k}^{(\text{sb})}$ deviates from $\hat{\beta}_{r,k}^{(\text{sb})}$. The difference between the two estimators is asymptotically negligible though, while the computational cost is substantially higher for the tilde version.

3.1. *Asymptotic normality of PWM estimators.* Before formulating explicit results, it is worthwhile to mention that asymptotic theory involving PWM estimators has hitherto been mostly worked out under the simplifying assumption that the (disjoint) block maxima provide a genuine i.i.d. sample from the GEV distribution (as noted in the introduction, [20] is a notable exception). The alternative viewpoint based on Condition 2.2 has at least three important advantages: it allows to explicitly describe potential bias terms, it does not neglect serial dependence between successive blocks (sampling scheme (S1)), and, perhaps most importantly, it makes possible the treatment of the more efficient sliding blocks version.

A number of regularity conditions is needed to derive consistency and asymptotic normality of the estimators in (11).

CONDITION 3.2 (Bias). For the normalizing sequences $(a_r)_{r \in \mathbb{N}}$ and $(b_r)_{r \in \mathbb{N}}$ from Condition 2.1 and for $k \in \{0, 1, 2\}$, $\text{mb} \in \{\text{sb}, \text{db}\}$ and $S \in \{S1, S2\}$, the limit

$$B_k^{(\text{mb}, S)} = \lim_{n \rightarrow \infty} B_{n,k}^{(\text{mb}, S)},$$

exists, where

$$B_{n,k}^{(\text{mb}, S)} = \begin{cases} \sqrt{\frac{n}{r}} \left\{ \mathbb{E}[Z_r H_r^k(Z_r)] - \mathbb{E}[Z G_\gamma^k(Z)] \right\}, & (\text{mb}, S) \neq (\text{sb}, S2), \\ \sqrt{\frac{n}{r}} \frac{1}{r} \sum_{j=1}^r \left\{ \mathbb{E} \left[Z_{r,j}^{(\text{sb})} \bar{H}_r^k(Z_{r,j}^{(\text{sb})}) \right] - \mathbb{E}[Z G_\gamma^k(Z)] \right\}, & (\text{mb}, S) = (\text{sb}, S2), \end{cases}$$

where Z_r from (2) has c.d.f. H_r , where $Z \sim G_\gamma$ and where $Z_{r,j}^{(\text{sb})}$ and \bar{H}_r are defined in (5) and (6), respectively.

It is worthwhile to mention that $B_{n,k}^{(\text{sb}, S1)} = B_{n,k}^{(\text{sb}, S2)}$ provided the underlying time series is serially independent; in fact, the entire sampling schemes coincide in this case. In typical cases of serial dependence, the simulation experiments in Section 4 and the supplement suggest that the difference between the two limits is small.

CONDITION 3.3 (Uniform integrability). There exists $\nu > \frac{2}{\omega}$ with ω from Condition 2.5(ii) such that, for the normalizing sequences $(a_r)_{r \in \mathbb{N}}$ and $(b_r)_{r \in \mathbb{N}}$ from Condition 2.1,

$$\limsup_{r \rightarrow \infty} \mathbb{E} \left[|Z_r|^{2+\nu} \right] < \infty.$$

The condition is used to deduce convergence of moments from convergence in distribution, which, for certain moments, is needed in view of the fact that the PWM estimators are based on the method of moments.

REMARK 3.4. In case $\gamma > 0$, Condition 3.3 together with Condition 2.2 and 2.5 implies additional constraints on γ and ω . Indeed, observing that $\mathbb{E}[|Z|^{2+\nu}] < \infty$ iff $\nu < 1/\gamma - 2$, Condition 3.3 can only be satisfied if $\gamma < 1/2$. Further, since $\nu > 2/\omega$, we must have $2/\omega < 1/\gamma - 2$, which is equivalent to $\omega > (2\gamma)/(1 - 2\gamma)$.

Our main result will be a corollary of the following theorem on the joint asymptotic properties of the empirical probability weighted moments. The following notations are needed for its formulation: let $f_0(x) = x$ and

$$(12) \quad f_1(x) = x G_\gamma(x) + \mathbb{E}[Z \mathbf{1}(Z > x)], \quad f_2(x) = x G_\gamma^2(x) + 2\mathbb{E}[Z G_\gamma(Z) \mathbf{1}(Z > x)],$$

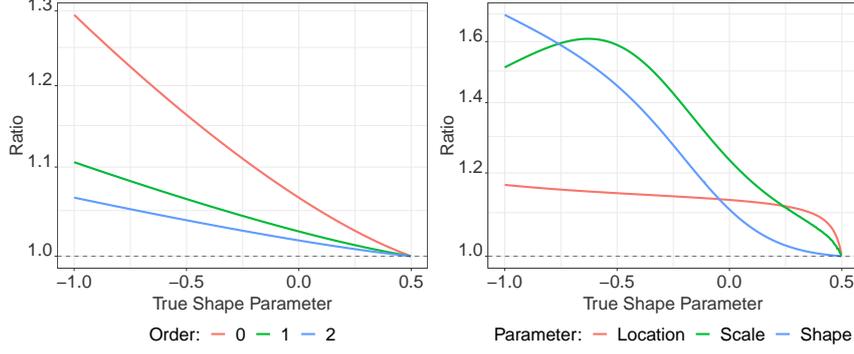


FIG 2. Left: Graph of $\gamma \mapsto \Omega_{k,k}^{(\text{db})} / \Omega_{k,k}^{(\text{sb})}$ for $k \in \{0, 1, 2\}$ and with $\Omega^{(\text{mb})}$ as in Theorem 3.5. Right: Graph of $\gamma \mapsto \Sigma_{\ell,\ell}^{(\text{db})} / \Sigma_{\ell,\ell}^{(\text{sb})}$ for $\ell = 1$ (shape), $\ell = 2$ (scale) and $\ell = 3$ (location) with $\Sigma^{(\text{mb})} = (\Sigma_{\ell,\ell'}^{(\text{mb})})_{\ell,\ell'=1,2,3}$ from Corollary 3.6.

where $Z \sim G_\gamma$ (note that the dependence on γ is suppressed in the notation f_k). Moreover, let $G_{\gamma,\xi}(x, y) = G_\gamma(x)G_\gamma(y)$ for $\xi > 1$ and

$$(13) \quad G_{\gamma,\xi}(x, y) = \exp \left[- \left\{ \xi(1 + \gamma x)^{-\frac{1}{\gamma}} + \xi(1 + \gamma y)^{-\frac{1}{\gamma}} + (1 - \xi)(1 + \gamma(x \wedge y))^{-\frac{1}{\gamma}} \right\} \right],$$

for $\xi \in [0, 1]$, where (x, y) is such that $1 + \gamma x > 0$ and $1 + \gamma y > 0$. Note that $G_{\gamma,\xi}$ defines a bivariate extreme value distribution with marginal c.d.f.s G_γ , irrespective of ξ , and with Pickands dependence function $A_\xi(w) = (1 \wedge \xi) + \{1 - (1 \wedge \xi)\}\{w \vee (1 - w)\}$.

THEOREM 3.5. *Suppose one of the sampling schemes from Condition 2.2 is met with $\gamma < 1/2$. Further, assume that Conditions 2.5, 3.2 and 3.3 are met, and write $\theta_r = (b_r, a_r, \gamma)'$ with respective PWMs $\beta_{\theta_r, k}$. Then, for $\text{mb} \in \{\text{db}, \text{sb}\}$ and $S \in (S1, S2)$,*

$$(14) \quad \left(\sqrt{\frac{n}{r}} \left(\frac{\hat{\beta}_{r,k}^{(\text{mb})} - \beta_{\theta_r, k}}{a_r} \right) \right)_{k=0,1,2} \xrightarrow{d} \mathcal{N}_3(\mathbf{B}^{(\text{mb}, S)}, \mathbf{\Omega}^{(\text{mb})}),$$

where $\mathbf{B}^{(\text{mb}, S)} = (B_k^{(\text{mb}, S)})_{k=0,1,2}$, $\mathbf{\Omega}^{(\text{mb})} = (\Omega_{k,k'}^{(\text{mb})})_{k,k'=0,1,2}$ and where, with $Z \sim G_\gamma$ and $(Z_{1\xi}, Z_{2\xi}) \sim G_{\gamma,\xi}$,

$$\Omega_{k,k'}^{(\text{db})} = \text{Cov}(f_k(Z), f_{k'}(Z)), \quad \Omega_{k,k'}^{(\text{sb})} = 2 \int_0^1 \text{Cov}(f_k(Z_{1\xi}), f_{k'}(Z_{2\xi})) d\xi$$

Moreover, with \leq_L denoting the Loewner-ordering between symmetric matrices, we have

$$(15) \quad \mathbf{\Omega}^{(\text{sb})} \leq_L \mathbf{\Omega}^{(\text{db})}.$$

Recall that the asymptotic bias is always the same, except under sampling scheme (S2) and for sliding blocks. It is worthwhile to mention that the theorem may be extended to arbitrary $k \geq 3$; in that case, f_k is given by $f_k = f_{k,1} + f_{k,2}$ with $f_{k,1}, f_{k,2}$ from (B.10) in the supplement [9]. Further, note that the integral in $\Omega_{k,k'}^{(\text{sb})}$ corresponds to the contribution introduced by the strong serial dependence between the sliding block maxima. More explicit expressions for the asymptotic covariances can be found in Appendix C in the supplementary material, see Lemma C.1. The graphs of the ratio of the variance curves $\gamma \mapsto \Omega_{k,k}^{(\text{db})} / \Omega_{k,k}^{(\text{sb})}$ are depicted in Figure 2, for $\gamma \in (-1, 1/2)$. As can be seen, the sliding blocks variances are universally smaller than the disjoint blocks counterparts, with a substantial improvement for $k = 0$ and negative γ .

Asymptotic normality of the PWM estimator for (b_r, a_r, γ) essentially follows from the above theorem and the delta method. Let

$$(16) \quad \phi : \mathcal{D}_\phi \rightarrow \mathbb{R}^3, \quad \boldsymbol{\beta} := (\beta_0, \beta_1, \beta_2)' \mapsto \begin{pmatrix} \phi_1(\boldsymbol{\beta}) \\ \phi_2(\boldsymbol{\beta}) \\ \phi_3(\boldsymbol{\beta}) \end{pmatrix} = \begin{pmatrix} g_1^{-1} \left(\frac{3\beta_2 - \beta_0}{2\beta_1 - \beta_0} \right) \\ g_2(\phi_1(\boldsymbol{\beta}))(2\beta_1 - \beta_0) \\ \beta_0 + \phi_2(\boldsymbol{\beta})g_3(\phi_1(\boldsymbol{\beta})) \end{pmatrix},$$

where $\mathcal{D}_\phi = \{\boldsymbol{\beta} \in \mathbb{R}^3 : 2\beta_1 - \beta_0 > 0, 3\beta_2 - 2\beta_1 > 0, -\beta_0 + 4\beta_1 - 3\beta_2 > 0\}$. Recall that $\theta_r = (b_r, a_r, \gamma)' = \phi(\beta_{\theta_r,0}, \beta_{\theta_r,1}, \beta_{\theta_r,2})$ for $\gamma < 1$ by (10), and that $\hat{\theta}_r^{(\text{mb})} = \phi(\hat{\boldsymbol{\beta}}_r^{(\text{mb})})$, where $\hat{\boldsymbol{\beta}}_r^{(\text{mb})} = (\hat{\beta}_{r,0}^{(\text{mb})}, \hat{\beta}_{r,1}^{(\text{mb})}, \hat{\beta}_{r,2}^{(\text{mb})})'$ and $\text{mb} \in \{\text{db}, \text{sb}\}$. Further, as shown in Proposition 2.1 in [26], we necessarily have $(\beta_{\theta_r,0}, \beta_{\theta_r,1}, \beta_{\theta_r,2})' \in \mathcal{D}_\phi$ for $\gamma < 1$. Theorem 3.5 then implies $\lim_{n \rightarrow \infty} \mathbb{P}(\hat{\boldsymbol{\beta}}_r^{(\text{mb})} \in \mathcal{D}_\phi) = 1$ after a simple calculation.

COROLLARY 3.6. Write $\hat{\theta}_r^{(\text{mb})} = (\hat{b}_r^{(\text{mb})}, \hat{a}_r^{(\text{mb})}, \hat{\gamma}_r^{(\text{mb})})'$. Under the conditions of Theorem 3.5, we have

$$(17) \quad \sqrt{\frac{n}{r}} \begin{pmatrix} \hat{\gamma}_r^{(\text{mb})} - \gamma \\ (\hat{a}_r^{(\text{mb})} - a_r)/a_r \\ (\hat{b}_r^{(\text{mb})} - b_r)/a_r \end{pmatrix} \xrightarrow{d} \mathcal{N}_3(\mathbf{CB}^{(\text{mb},S)}, \boldsymbol{\Sigma}^{(\text{mb})}),$$

where $\boldsymbol{\Sigma}^{(\text{mb})} = \mathbf{C}\boldsymbol{\Omega}^{(\text{mb})}\mathbf{C}'$ with $\mathbf{C} = (D\phi)(\boldsymbol{\beta}_\gamma)$ the Jacobian of ϕ evaluated at the true PWMs $\boldsymbol{\beta}_\gamma = (\beta_{\gamma,0}, \beta_{\gamma,1}, \beta_{\gamma,2})'$ of G_γ . Moreover, we have

$$(18) \quad \boldsymbol{\Sigma}^{(\text{sb})} \leq_L \boldsymbol{\Sigma}^{(\text{db})}.$$

Precise formulas for the matrix \mathbf{C} can be found in Lemma C.2. Together with respective formulas for $\boldsymbol{\Omega}^{(\text{mb})}$ (Lemma C.1) one may confirm that the disjoint blocks variance $\boldsymbol{\Sigma}^{(\text{db})}$ coincides with the one obtained in [20] for the i.i.d. case.

3.2. Application: return level estimation. A typical quantity of interest in environmental statistics is the return level (RL) of an extreme event. Formally, for a block size r (often a year or a season) and a target number of (disjoint) blocks T , the (T, r) -return level of the distribution F_r defined in (3) is defined as

$$\text{RL}(T, r) = F_r^{\leftarrow}(1 - 1/T) = \inf\{x \in \mathbb{R} : F_r(x) \geq 1 - 1/T\}.$$

Note that it will take on average T independent disjoint blocks of size r until the first such block whose maximum exceeds $\text{RL}(T, r)$. Now, by Condition 2.1, we have $F_r \approx G_{(b_r, a_r, \gamma)}$, whence $\text{RL}(T, r) \approx \text{RL}^\circ(T, r)$, where

$$\text{RL}^\circ(T, r) = G_{(b_r, a_r, \gamma)}^{\leftarrow}(1 - 1/T) = a_r \frac{c_T^{-\gamma} - 1}{\gamma} + b_r,$$

and where $c_T = -\log(1 - 1/T)$. We therefore obtain the estimators

$$\widehat{\text{RL}}^{(\text{mb})}(T, r) = \hat{a}_r^{(\text{mb})} \frac{c_T^{-\hat{\gamma}_r^{(\text{mb})}} - 1}{\hat{\gamma}_r^{(\text{mb})}} + \hat{b}_r^{(\text{mb})}, \quad \text{mb} \in \{\text{db}, \text{sb}\}.$$

COROLLARY 3.7. Under the conditions of Theorem 3.5, we have

$$\sqrt{n/r} \left(\frac{\widehat{\text{RL}}^{(\text{mb})}(T, r) - \text{RL}^\circ(T, r)}{a_r} \right) \xrightarrow{d} \mathcal{N} \left(q_T' \mathbf{CB}^{(\text{mb},S)}, q_T' \boldsymbol{\Sigma}^{(\text{mb})} q_T \right),$$

where $q_T = q_T(\gamma)$ is defined as $q_T(0) = (\log^2(c_T)/2, -\log(c_T), 1)'$ and

$$q_T(\gamma) = \left(\frac{1 - c_T^{-\gamma}(\gamma \ln(c_T) + 1)}{\gamma^2}, \frac{c_T^{-\gamma} - 1}{\gamma}, 1 \right)', \quad \gamma \neq 0.$$

The asymptotic variance in Corollary 3.7 being an explicit function of γ , it may easily be estimated by the plug-in principle; we denote the respective estimator by $\hat{\sigma}_T^{2,(\text{mb})}$. Corollary 3.7 then allows to construct asymptotic confidence intervals for $\text{RL}(T, r)$. Indeed, assuming that the block size r is chosen sufficiently large to guarantee that $\mathbf{B}^{(\text{mb}, \text{S})} = 0$ and that $\text{RL}^\circ(T, r) = \text{RL}(T, r) + o(\sqrt{r/na_r})$, we obtain that

$$\text{RL}(T, r) \in \left[\widehat{\text{RL}}^{(\text{mb})}(T, r) \mp \hat{a}_r^{(\text{mb})} \sqrt{\frac{r}{n} \hat{\sigma}_T^{2,(\text{mb})}} u_{1-\alpha/2} \right]$$

with asymptotic probability α , where $u_{1-\alpha/2}$ is the $(1 - \alpha/2)$ -quantile of the standard normal distribution. It follows from the bounds on the asymptotic variances in Corollary 3.6 that the confidence intervals are asymptotically more narrow for the sliding blocks method; an observation that will be confirmed by the case study in Section 5.

4. Simulation study. The finite-sample properties of the proposed estimators have been evaluated in a large scale Monte Carlo simulation study. Three target variables have been selected: the shape parameter γ , and two return levels, $\text{RL}(50, r)$ and $\text{RL}(100, r)$. The following central aspects have been investigated:

- (i) Performance of the PWM estimator when sub-asymptotic versions of sampling schemes (S1) and (S2) from Condition 2.2 with varying degree of extremal temporal dependence are met for fixed block size r (Section 4.1) and for fixed sample size n (Section E.2 of the supplementary material [9], summarized in Section 4.2).
- (ii) Performance of the PWM estimator when the seasonal stationarity from Condition 2.2 is violated (Section 4.3).
- (iii) Comparison of the PWM estimator to Maximum Likelihood estimators based on sliding blocks (Section E.6 of the supplementary material [9], summarized in Section 4.4).
- (iv) Performance under strong temporal dependence, in particular for sampling scheme (S2) (Section 4.5).

The data-generating processes that were used for (i), (iii) and (iv) are as follows:

(a) **Stationary distribution of X_t .** We opted for a model that allows for both positive and negative shape parameters in a continuous way, and hence chose five distributions from the generalized Pareto family, namely $\text{GPD}(0, 1, \gamma)$ with shape parameter γ in $\{-0.4, -0.2, 0, 0.2, 0.4\}$ with corresponding c.d.f.

$$(19) \quad F_\gamma(x) = \begin{cases} \left(1 - (1 + \gamma x)^{-\frac{1}{\gamma}}\right) \mathbf{1}(x \geq 0), & \gamma > 0, \\ \left(1 - (1 + \gamma x)^{-\frac{1}{\gamma}}\right) \mathbf{1}(0 \leq x \leq |\gamma|^{-1}), & \gamma < 0, \\ (1 - \exp(-x)) \mathbf{1}(x \geq 0), & \gamma = 0. \end{cases}$$

Note that an i.i.d. series from F_γ satisfies Condition 2.1 with shape parameter γ and scaling sequences $a_r = r^\gamma$ and $b_r = (r^\gamma - 1)/\gamma$, to be interpreted as $\log r$ for $\gamma = 0$.

Experiments involving a different family of distributions (where weak convergence of block maxima to the GEV is slower than for the GPD) have also been performed; the qualitatively similar results can be found in Section E.5 in the supplementary material [9].

(b) **Time series model.** Next to the i.i.d. case, we considered quantile transformed versions of the Gaussian AR(1) model (with extremal index 1), of an AR(1) process with heavy

tailed Cauchy(1) innovations and of the Fréchet ARMAX(1) model (the latter two having extremal index smaller than 1). Recall that the extremal index is a measure for the tendency of extreme observations to occur in clusters (the smaller θ , the larger that tendency), see Section 10.2.3 in [1] for a gentle introduction.

The transformed Gaussian AR-model is defined as follows: for given AR-parameter $|\phi| < 1$ (we chose $\phi \in \{0, 0.25, 0.5, 0.75, 0.9\}$; note that $\phi = 0$ corresponds to the i.i.d. case), consider the stationary solution $(Y_t)_t$ of the classical AR(1) recursion

$$(20) \quad Y_t = \phi Y_{t-1} + \epsilon_t, \quad t \in \mathbb{Z}, \quad (\epsilon_t)_t \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1).$$

The marginal distribution, say F_Y , is known to be centred normal with variance $1/(1 - \phi^2)$ [4] and the extremal index of $(Y_t)_t$ is known to be 1 [19]. As a consequence, $X_t = F_\gamma^{\leftarrow}(U_t)$ with $U_t = F_Y(Y_t)$ satisfies Condition 2.1 with shape parameter γ and extremal index 1.

For the Cauchy AR (CAR) model, the Gaussian innovations in (20) are replaced by i.i.d. Cauchy(1)-innovations. Proposition 13.3.2 in [4] yields the representation $Y_t = \sum_{j=1}^{\infty} \phi^j \epsilon_{t-j}$. For $\phi \in (0, 1)$ (we chose $\phi \in \{0.25, 0.5, 0.75, 0.9\}$), Example 8.1.1(d) in [19] then implies that the extremal index exists and is given by $\theta = 1 - \phi$. Moreover, a simple calculation based on characteristic functions shows that the marginal distribution F_Y of Y_t is Cauchy as well, with scale parameter $1/(1 - \phi)$. We may thus transform to uniform margins by letting $U_t = F_Y(Y_t)$ and may generate $X_t = F_\gamma^{\leftarrow}(U_t)$, which satisfies Condition 2.1 with shape parameter γ and extremal index $\theta = 1 - \phi$.

The transformed ARMAX-model is defined as follows: for given $b \in [0, 1)$ (we chose $b \in \{0.25, 0.5, 0.75, 0.9\}$), consider the stationary solution $(Y_t)_t$ of the ARMAX(1) recursion

$$Y_t := \max(bY_{t-1}, (1 - b)\epsilon_t), \quad t \in \mathbb{Z}, \quad (\epsilon_t)_t \stackrel{\text{i.i.d.}}{\sim} \text{Fréchet}(1).$$

The marginal distribution F_Y is known to be Fréchet(1) as well, and the extremal index is equal to $\theta = 1 - b$ (Section 10 in [1]). As a consequence, $X_t = F_\gamma^{\leftarrow}(U_t)$ with $U_t = F_Y(Y_t)$ satisfies Condition 2.1 with shape parameter γ and extremal index $\theta = 1 - b$.

4.1. Fixed block length r . In a first experiment, we considered each combination of the described time series model and the marginal distribution function in a situation where the block size is fixed and the overall sample size is increasing. We fixed $r = 90$, which could be interpreted as the number of daily observations within a three-month season; a common situation encountered in environmental applications. The number of seasons was chosen to vary between 10 and 100, yielding overall sample sizes of the underlying time series between 900 and 9000. We computed the PWM estimators based on disjoint and sliding block maxima, and the respective estimators for $\text{RL}(50, r)$ and $\text{RL}(100, r)$ from Section 3.2. The estimators have been evaluated in terms of their relative MSE based on $N = 5000$ simulation repetitions, i.e., we divided the MSE of the disjoint blocks estimator by the MSE of the sliding blocks counterparts. For the sake of brevity, we only report the results for a selection of time series models with at most moderate temporal dependence (which is typical for, e.g., daily temperature or precipitation data): the i.i.d. case as well as AR, CAR and ARMAX dependence structures with a time series parameter of 0.5. All other results can be found in the supplementary material [9], see in particular Section E.1.

Results for the estimation of γ are presented in Figure 3 (see also Figure 1 from the introduction), with remarkably similar results for the two sampling schemes (S1) and (S2). Note that sampling schemes (S1) and (S2) coincide for i.i.d. observations. The results reveal that the sliding blocks method is universally better than the disjoint blocks method for non-positive shape parameters, with large improvements for small sample sizes (note that situations of less than 50 seasons are not uncommon in environmental applications; in particular when restricting attention to stationary time periods). On the other hand, for positive

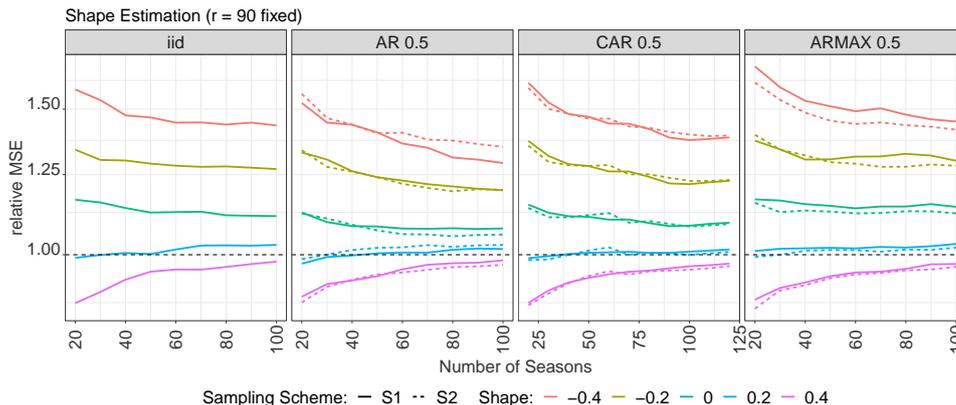


FIG 3. *Relative MSE (MSE of disjoint blocks estimator divided by MSE of sliding blocks estimator) for the estimation of γ in a selection of time series models with GPD-margins for sampling schemes (S1) (solid line) and (S2) (dashed line) and fixed block size $r = 90$.*

shape parameters, the disjoint blocks method may outperform the sliding blocks method for small sample sizes. This effect can mostly be resolved by considering the bias-reduced sliding blocks estimator from Remark 3.1, which, however, is computationally costly for situations involving overall sample sizes of up to $n = 9000$. A discussion of the latter estimator is postponed to Section E.3 of the supplementary material [9]. Finally, it is worthwhile to mention that the time series model does not have a huge impact on the qualitative results.

We next consider the estimation of return levels. For the evaluation of the respective estimators, ('true') population values for the return levels are needed. Since these are not known explicitly, they have been obtained by a preliminary simulation: after simulating 10^6 independent blocks of length r , we calculated the empirical $(1 - 1/T)$ -quantile of the obtained sample to obtain an accurate approximation for $RL(T, r)$. The respective values for block size $r = 90$ can be found in Table E.1 in Section E of the supplementary material. The results from the simulation experiment are presented in a similar way as for the shape estimation and can be found in Figure 4. For the sake of a clear presentation, we only consider sampling scheme (S2); the results for sampling scheme (S1) are similar and can be found in the supplementary material. Overall, the findings are quite similar to those for the estimation of γ . Compared to the latter target variable, slight advantages for the sliding blocks method are also visible for $\gamma = 0.2$, while we still observe a disadvantage for $\gamma = 0.4$. Finally, it is worthwhile to mention that the relative MSE is increasing in T for all considered situations.

4.2. *Fixed sample size n .* In a second experiment, we considered varying values of the block length parameter r for a fixed overall sample size of $n = 1000$. Under sampling scheme (S1), the block length can be considered a hyperparameter to be chosen by the statistician, which induces the common bias-variance tradeoff in extremes (note that r is given when considering sampling scheme (S2)). For the sake of brevity, detailed results and discussions are postponed to Sections E.2 (comparison of disjoint and plain sliding blocks estimation), E.3 (comparison of plain and bias-reduced sliding blocks estimation) and E.4 (comparison of estimation within (S1) and (S2)) in the supplement [9]. They can briefly be summarised as follows: Most MSE curves exhibit the common u-shape, with universally smaller values for the sliding blocks method compared to the disjoint blocks method for negative shape parameters, and hardly any difference for positive shape parameters.

4.3. *Deviation of the piecewise stationary setting.* In a third experiment, we investigate the performance within a situation that deviates from the piecewise stationary setting postulated in Condition 2.2. Since the previous simulation results suggest that the efficiency gain of

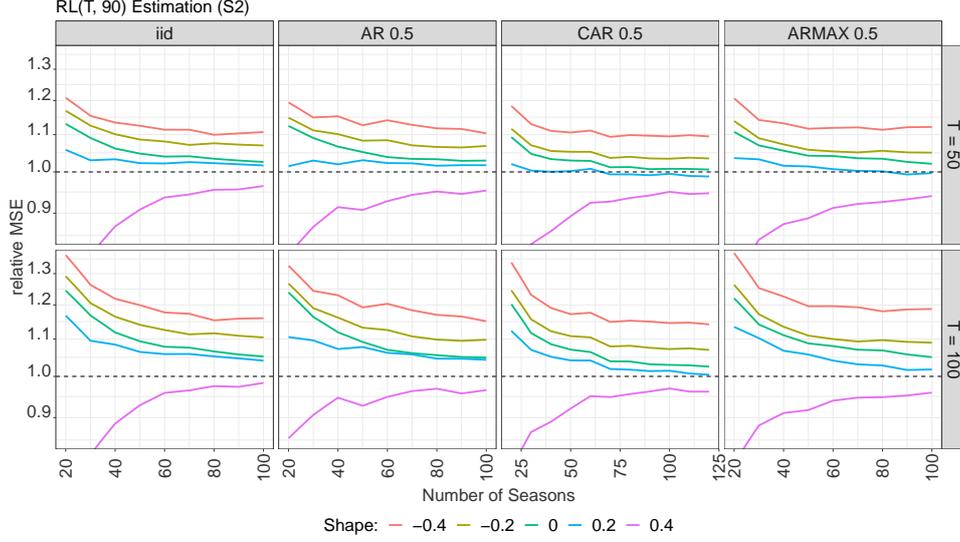


FIG 4. *Relative MSE (MSE of disjoint blocks estimator divided by MSE of sliding blocks estimator) for the estimation of $RL(T, 90)$ with $T = 50$ (top row) and $T = 100$ (bottom row) in a selection of time series models with GPD-margins under sampling scheme (S2) for fixed block size $r = 90$.*

using sliding blocks is largest for non-positive shape parameters, we aim for a model describing temperature extremes, since shape parameters of seasonal maxima are well-known to be negative for this kind of data. We may then rely on [34], where the asymptotic distribution of block maxima was investigated in a framework where the finite upper bound of the ‘daily observations’ was allowed to depend smoothly on (rescaled) inner-seasonal time. In the case of serially independent observations, the limiting distribution was found to be GEV again, despite with an unexpected shape parameter; see Theorem 1. Extensions to serial dependence were not worked out explicitly, but it was conjectured that similar phenomena arise.

We employ the marginal model described in the third paragraph on page 5 in [34]: for the i th day of the year (restricting attention to the first 90 days of the summer season corresponds to $i \in \{152, \dots, 241\}$), we denote by F_i the c.d.f. of the

$$\text{GPD}(u_i - (7 \cdot 10^7)^{\frac{1}{5}}, ((7 \cdot 10^7)^{\frac{1}{5}})/5, -0.2)$$

distribution, where $u_i = 111 - (i - 200)^2/400$.

We then apply the quantile transformation technique again: starting from one of the serial dependence structures of interest, we transform the marginals to the time dependent GPD F_i . We restrict attention to sampling scheme (S2), since this seems to be the natural choice here. Last but not least, note that the above model is in Fahrenheit, so we transform the simulated data to $^{\circ}C$ by multiplying by $5/9$ after subtracting 32.

We restrict attention to return level estimation (note that the true limiting shape parameter is only known for the i.i.d. case: it is $-2/11$ by Theorem 1 in [34]). Since ‘true’ return levels are not known explicitly either, they are approximated based on a preliminary Monte Carlo simulation involving $N = 10^6$ block maxima of size $r = 90$, from which the empirical 99%-quantile (i.e., the 100-season return level) is determined.

The results are compared to a situation without innerseasonal non-stationarities. To obtain observations of the same magnitude, we generate data with margins corresponding to $\text{GPD}(72.21, ((7 \cdot 10^7)^{\frac{1}{5}})/5, -0.2)$, since $\frac{1}{90} \sum_{i=1}^{90} (u_i - (7 \cdot 10^7)^{1/5}) = 72.21$ is the mean of the location parameters of the non-stationary counterparts. MSEs and relative MSEs observed in a selection of models with different dependence structures are shown in Figure 5. We observe that the innerseasonal non-stationarity does not have a significant influence on

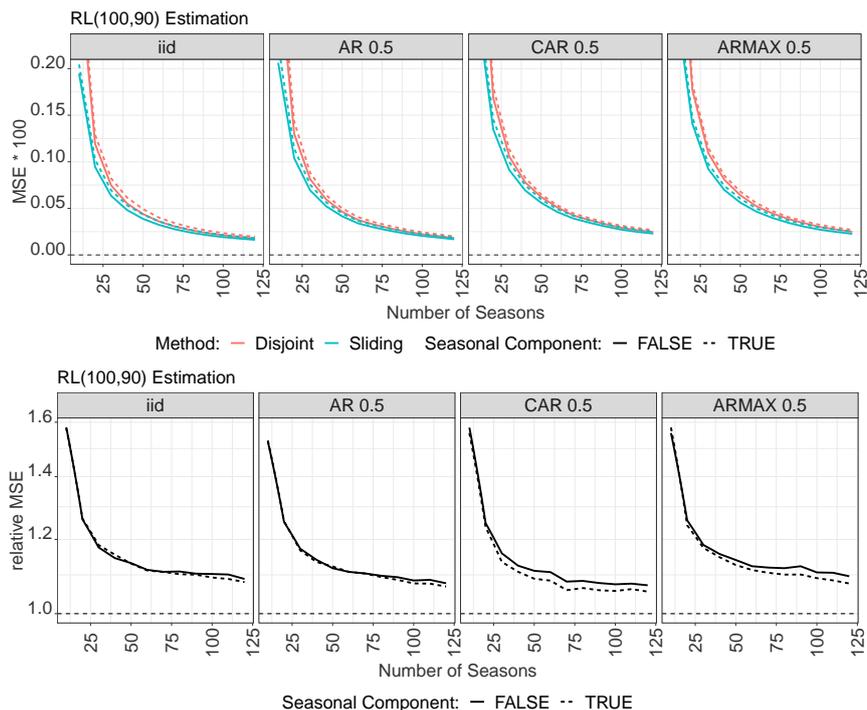


FIG 5. *MSE (top) and relative MSE (bottom; MSE of disjoint block estimation divided by MSE of sliding block estimation) of RL(100,90) estimation without (solid line) and with (dashed line) innerseasonal non-stationarity as a function of the observed number of seasons for fixed block size $r = 90$.*

the estimation performance, and that the advantage of sliding blocks over disjoint blocks remains. The sliding blocks method may hence be regarded as robust to certain deviations from the piecewise stationary setup.

4.4. Comparison with Maximum Likelihood Estimation. The sliding blocks PWM estimator has been compared to its counterpart based on (pseudo) maximum likelihood estimation, which is defined by maximizing the GEV likelihood function that arises from treating all sliding blocks as independent. For the sake of brevity, the results are illustrated in the supplementary material [9] only. They can be summarized as follows: the PWM estimator has a tendency to be superior for small sample sizes while the maximum likelihood estimator is superior for large sample sizes; to the best of our knowledge this is a usual view of the two estimators among applied statisticians. For shape estimation, smaller shapes yield better results for the PWM estimator, while for return level estimation, the picture is almost reversed.

4.5. Performance under strong temporal dependence. In practical applications, the extremal temporal dependence is typically at most moderate; for instance, the extremal index is around 45-50% for daily temperatures and above 80-90% for cumulative daily precipitation. In such situations, the sliding blocks method was found to outperform the disjoint blocks method in the previous sections, both under (S1) and (S2). In this section, we will analyse situations where the extremal temporal dependence is large in comparison to the block size. In such scenarios, one may heuristically expect a deterioration of the sliding blocks methods under sampling scheme (S2), perhaps due to a potentially substantial bias resulting from the non-stationarity of the sliding block maxima sample.

The deterioration of the sliding blocks method under (S2) can indeed be confirmed by simulation experiments: in the upper half of Figure 6, which is akin to Figure 4 but for the

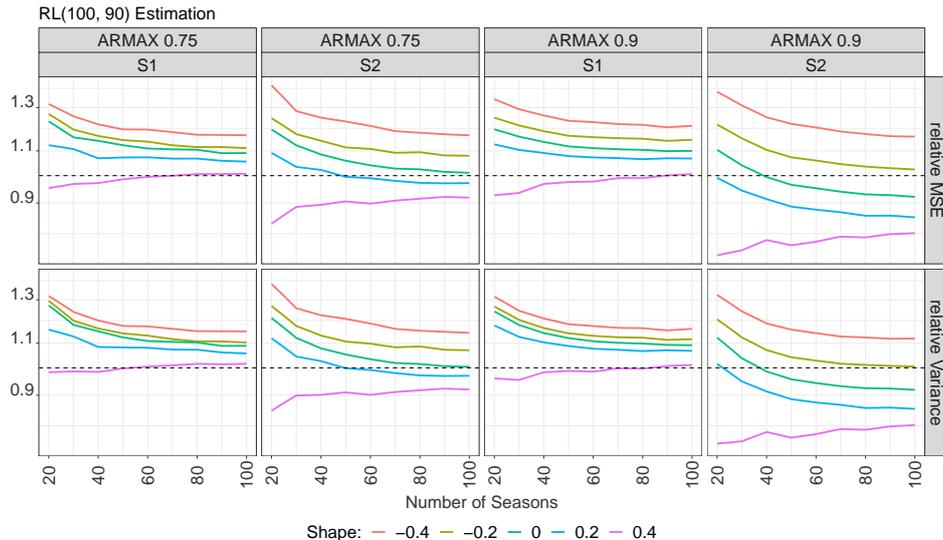


FIG 6. *Relative MSE (top row) and relative Variance (bottom row) (MSE (Variance) of disjoint blocks estimator divided by MSE (Variance) of sliding blocks estimator) for the estimation of RL(100,90) in a selection of time series models with GPD-margins for fixed block size $r = 90$.*

ARMAX model only and with larger values of $b = 1 - \theta$, it can be seen that the sliding blocks method does not provide an improvement over the disjoint blocks method for a broader range of shapes when increasing the temporal dependence to $\theta = 0.1$ under (S2). However, as illustrated in the lower half of Figure 6 and contrary to the above heuristics, this behavior is not due to a larger bias of the sliding blocks method, but rather due to the fact that the variance of the sliding blocks method is larger than that of its disjoint blocks counterpart.

The negligibility of potential bias terms is discussed in detail in the supplementary material, Section D. The observed variance behavior may be explained as follows: the disjoint block maxima are independent under (S2), but they are (subasymptotically) positively correlated under (S1), the more so the smaller the block size in comparison to the serial dependence of the underlying data. The positive correlation will eventually show up in the variance of respective estimators, making the finite-sample (S2) variance smaller than the (S1) variance when keeping all other parameters fixed. On the other hand, for the sliding blocks method, the majority of the pairs of sliding blocks that are based on non-overlapping, but neighboring blocks will exhibit some left-over dependence that will only vanish asymptotically ($r \rightarrow \infty$) even under (S2) and of course under (S1). Hence, the finite-sample (S2) variance will be close to the finite-sample (S1) variance when keeping all other parameters fixed. Overall, under (S2), we may have unfavorable combinations of the block size and the serial dependence for which the improvement “disjoint (S1) \rightarrow disjoint (S2)” will be larger than the improvement “disjoint (S1) \rightarrow sliding (S1) \approx sliding (S2)”, which will ultimately make the disjoint method more efficient than the sliding blocks method under (S2). Such unfavorable situations will always be reached when fixing the block size and increasing the serial dependence, which is exactly the phenomenon illustrated in Figure 6. In additional simulation experiments it was found that the potential superiority of disjoint over sliding blocks may be eliminated by increasing the block size. It was further found that the effect depends on the estimator and target quantity, with PWM return level estimation providing the most unfavorable case for the sliding blocks method; see also the results in the supplementary material.

5. Case study. Estimating return levels of the distribution of annual or seasonal maxima (of some meteorologic variable of interest) based on GEV-models constitutes one of

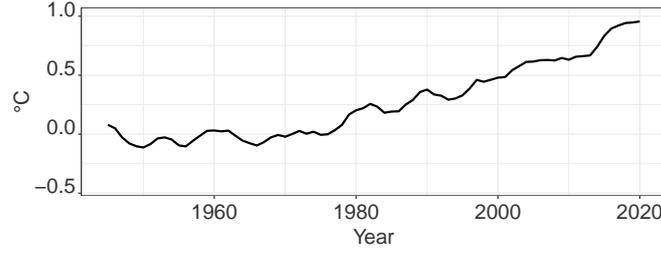


FIG 7. 4-year smoothed global mean surface temperature (sGMST) anomaly, with reference value being the average of GMST from 1951-1980.

the cornerstones of extreme weather event attribution studies [35]. Since the sliding blocks PWM estimator has been seen to provide the largest improvement over its disjoint counterpart for negative shape parameters, our case study concentrates on maximal air temperature data, for which shapes are usually within the range -0.4 to -0.2 . The data set to be analyzed consists of daily observations throughout the summer months (June, July, August) at four selected weather stations in Germany (Aachen, Essen-Bredeney, Frankfurt/Main, Kahler Asten), provided by the DWD (Deutsche Wetterdienst), and covers the sampling periods 1945–2010, 1948–2019, 1949–2019 and 1955–2019, respectively, resulting in sample lengths of 66, 72, 71, 65, with a block size equal to $r = 92$.

Maximal temperature data are non-stationary due to climate change (the average global surface temperature has roughly increased by about 1 degree celsius compared to pre-industrial times), whence a realistic model for maximal temperature must involve non-stationarities as well. Subsequently, let $T_1, \dots, T_{92}, T_{93}, \dots, T_{184}, \dots$ denote the concatenated sequence of daily temperatures throughout the summer months at a specific station, where the first observation corresponds to June 1 in a certain year. A standard GEV-model that is commonly applied within the context of extreme event attribution studies for maximal temperature data $M_t = \max(T_{92(t-1)+1}, \dots, T_{92t})$ in season t consists of imposing a simple linear model for the location parameter in terms of the 4-year smoothed global mean surface temperature (sGMST) anomaly, see, e.g., [31] and the references therein. More precisely,

$$(21) \quad M_t = cx_t + Z_t \sim \text{GEV}(b + cx_t, a, \gamma)$$

where $(x_t)_t$ denotes the yearly sequence of sGMST (see Figure 7), where $b, c, \gamma \in \mathbb{R}$ and $a > 0$ are the free parameters of the model and where $Z_t \sim \text{GEV}(b, a, \gamma)$ is stationary. After subtracting the global trend, it appears heuristically reasonable to assume that the (unobservable) detrended time series defined by concatenating the blocks

$$(22) \quad (Y_1^{(t)}, \dots, Y_{92}^{(t)}) = (T_{92(t-1)+1} - cx_t, \dots, T_{92t} - cx_t)$$

consists of independent and identically distributed blocks that are close to being stationary, with possibly some small deterministic seasonal component as investigated in Section 4.3. Recall that the latter inner-seasonal non-stationarity was found to have no big impact on estimation performance. An observable counterpart of (22) may be obtained by estimating the slope parameter c in model (21), for which we employ the widely used and robust method from Sen [33]. The respective parameter estimates $\hat{c} = \hat{c}(M_1, \dots, M_m)$ for the four stations of interest are stated in the third column of Table 2. The resulting sample

$$(23) \quad (\hat{T}_1, \dots, \hat{T}_{92}, \hat{T}_{93}, \dots) = (T_1 - \hat{c}x_1, \dots, T_{92} - \hat{c}x_1, T_{93} - \hat{c}x_2, \dots)$$

will be referred to as the sample of detrended daily observations.

For illustrative purposes, we proceed the analysis by ignoring any potential estimation error within \hat{c} (a more rigorous treatment can be found below), and implicitly assume that

TABLE 1
Theoretical 95% confidence intervals for the 100-year RL of the series of detrended summer maxima. Bootstrapped confidence interval bounds (10 000 repetitions) are shown in round brackets.

Station	Method	RL	CI lower Bound	CI upper Bound	CI Width
Aachen	Disjoint	36.66	35.10 (35.08)	38.22 (38.24)	3.12 (3.16)
	Sliding	36.67	35.22 (35.24)	38.11 (38.09)	2.88 (2.84)
Essen	Disjoint	35.43	34.02 (34.00)	36.84 (36.85)	2.82 (2.85)
	Sliding	35.24	34.05 (34.07)	36.43 (36.41)	2.38 (2.34)
Frankfurt	Disjoint	38.18	36.51 (36.47)	39.85 (39.88)	3.34 (3.41)
	Sliding	37.80	36.52 (36.54)	39.09 (39.07)	2.56 (2.52)
Kahler Asten	Disjoint	30.84	29.43 (29.42)	32.24 (32.25)	2.81 (2.83)
	Sliding	30.75	29.46 (29.48)	32.04 (32.03)	2.58 (2.55)

the sample defined in (23) meets the assumption of sampling scheme (S2). Note that the respective block maxima

$$(24) \quad Z_t^{(\text{db})} = \max(\hat{T}_{92(t-1)+1}, \dots, \hat{T}_{92t}), \quad Z_{92(t-1)+j}^{(\text{sb})} = \max(\hat{T}_{92(t-1)+j}, \dots, \hat{T}_{92t+j})$$

satisfy $Z_{92(t-1)+1}^{(\text{sb})} = Z_t^{(\text{db})} = M_t^{(\text{db})} - \hat{c}x_t$. In view of the fact that the (detrended daily) observations from the first and last disjoint block have a reduced chance of appearing multiple times within the sliding blocks sample (for instance, if the sample maximum is the very last observation, it only appears once in the sliding blocks sample, while it would appear r times if it was observed in the second to last season), we chose to tweak the underlying daily sample by attaching the first block to the last one (which is akin to the circular block bootstrap in time series analysis). The resulting sliding blocks sample has then a sample size of exactly 92 times the number of seasons. The disjoint and sliding block maxima can then be fitted to the GEV distribution based on the PWM methods. Estimated parameters are collected in Table 2, and a graphical check of the fit of the resulting distributions can be found in the supplementary material [9], Section F.

Respective estimates for the 100-season return level can be obtained as described in Section 3.2, including (asymptotic) confidence bounds. The results are summarized in Table 1 (the results in brackets will be explained below and can be skipped for the moment). As was to be expected from both the theoretical results and the simulation study, the confidence intervals based on sliding blocks method are always smaller than their disjoint counterparts, with a substantial margin between 0.23 and 0.78.

Note that point estimates for the return level in the climate of season t , say $\text{RL}_t(100, 92)$, may be obtained by simply adding $\hat{c}x_t$ to the values in the third column of Table 1 (here, the T -season return level of season t is the level which is expected to be exceeded only once in T years if the climate hypothetically remained constant/stationary equal to that of season t ; see also the discussion in [12]). However, simply adding $\hat{c}x_t$ to the confidence bounds in Table 1 does not provide valid confidence sets for $\text{RL}_t(100, 92)$, as the estimation error of \hat{c} has not been captured. The latter may be captured by suitable bootstrap devices, for which we propose the following parametric bootstrap scheme.

Given estimates $(\hat{c}, \hat{b}, \hat{a}, \hat{\gamma})$ (where the last three components may either be based on disjoint or sliding block maxima samples), we may generate, for each season t , an i.i.d. sample $T_{92(t-1)+1}^*, \dots, T_{92t}^*$ of size $r = 92$ from the GEV distribution with parameter

$$\tilde{b}_t = \hat{b} + \hat{c}x_t - \frac{\hat{a}(92^{\hat{\gamma}} - 1)}{92^{\hat{\gamma}}\hat{\gamma}}, \quad \tilde{a} = \frac{\hat{a}}{92^{\hat{\gamma}}}, \quad \tilde{\gamma} = \hat{\gamma}.$$

A simple calculation shows that the t th disjoint block maximum from the bootstrap sample is GEV-distributed with parameter $(\hat{b} + \hat{c}x_t, \hat{a}, \hat{\gamma})$. The fact that $T_{92(t-1)+1}^*, \dots, T_{92t}^*$ may be

simulated serially independent can be explained by the fact that the asymptotic distribution of the PWM estimator does not depend on the serial dependence of the underlying time series (except through the parameter sequences b_r and a_r , see Corollary 3.6; and under the assumption that the block length is sufficiently large to guarantee that the bias is negligible). Now we apply the same procedure as for the original observations T_1, T_2, \dots : first, we build disjoint block maxima and estimate the trend \hat{c}^* , then we use this estimate to detrend T_1^*, T_2^*, \dots as in (23), and we finally calculate the respective disjoint and sliding block maxima as in (24), based on which we ultimately obtain bootstrap estimates $(\hat{b}^*, \hat{a}^*, \hat{\gamma}^*)$ and $\widehat{\text{RL}}_t^*(100, 92)$. Repeating the bootstrap procedure $B = 10000$ times, we may obtain estimates of the standard error of $(\hat{c}, \hat{b}, \hat{a}, \hat{\gamma})$ by calculating the empirical standard deviation of the sample of bootstrap estimates. Likewise, we may obtain bootstrap confidence intervals for any parameter of interest based on the percentile method [13]. To obtain symmetric 95% confidence intervals with respect to the estimated return level, we rather solve $\hat{F}^*(\widehat{\text{RL}} + \epsilon) - \hat{F}^*(\widehat{\text{RL}} - \epsilon) = 0.95$ for ϵ , where \hat{F}^* is the empirical distribution function of the bootstrap estimates of return levels, and use $(\widehat{\text{RL}} - \epsilon, \widehat{\text{RL}} + \epsilon)$ as a confidence interval.

The bootstrap scheme has been applied to each station, both for the disjoint and the sliding blocks method. The results are summarized in Table 2 (standard deviation of the estimation of c, b, a and γ) and in Figure 8 (pointwise confidence intervals for the estimation of $\text{RL}_t(100, 92)$). Remarkably, at each station, the sliding blocks estimator yields slightly smaller estimates for the shape parameter and slightly larger estimates for the scale parameter. The resulting estimates for the 100-year return level are slightly smaller as well, except for station ‘Aachen’. By definition, the slope estimates at each station are the same, which explains why the difference between the sliding and disjoint blocks curves in Figure 8 is constant. In all cases, the confidence bands are smaller for the sliding blocks version, as expected from large-sample theory and the simulation experiments. Coincidentally, the lower bounds of the confidence intervals are almost the same for the stations for which the sliding block point estimates are smaller.

As promised above, we finally explain the values in brackets in Table 1, and return to the data sets from (24) which were considered to be arising from an underlying stationary time series. The values in brackets correspond to bootstrap confidence bounds based on a parametric bootstrap adapted to this simple stationary situation. More precisely, the bootstrap scheme is carried out as before, but with setting $\hat{c} = 0$ when generating the bootstrap samples, and therefore also omitting the detrending step. As can be seen from the results in Table 1, the bootstrap confidence bounds are very similar to the bounds obtained by the normal approximation and estimation of the theoretical asymptotic variance (see Section 3.2). These findings indicate that the bootstrap scheme is working well, and support its application to the non-stationary situation described above, where the normal approximation cannot be applied without major additional calculations regarding the propagation of uncertainty due to the initial estimation of the slope parameter. We would like to stress, however, that one should take this assessment with some care, as we did not conduct an extensive study on the coverage probabilities of either of the two methods to construct confidence intervals.

6. Conclusion and discussion. Large sample theory for univariate extreme value statistics based on the block maxima method has long been developed under the assumption that the block maxima constitute a genuine i.i.d. sample from the GEV distribution. Two more realistic sampling schemes were considered in this paper: either an underlying stationary time series, or a triangular array consisting of independent blocks extracted from a stationary time series model. The latter shall represent a typical situation encountered in environmental statistics, where stationarity can only be (approximately) guaranteed within seasons rather than years. Under certain additional regularity conditions, it was found that more efficient

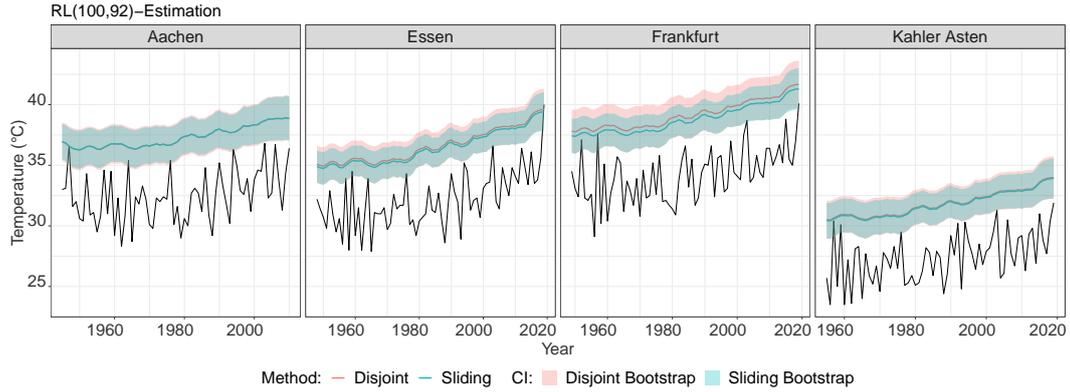


FIG 8. Estimated 100-year RL of summer months' maximal air temperature along with 95% confidence regions obtained from parametric bootstrap (10 000 bootstrap repetitions) at four stations in the western part of Germany.

TABLE 2
Estimated Parameters along with the standard deviation based on the bootstrap.

Station	Method	Slope	Location	Scale	Shape
Aachen	Disjoint	3.48 (1.07)	31.04 (0.32)	1.85 (0.18)	-0.19 (0.092)
	Sliding	3.48 (1.12)	31.05 (0.32)	1.96 (0.16)	-0.22 (0.081)
Essen	Disjoint	4.40 (0.71)	30.26 (0.29)	1.65 (0.16)	-0.18 (0.088)
	Sliding	4.40 (0.77)	30.19 (0.31)	1.84 (0.14)	-0.25 (0.076)
Frankfurt	Disjoint	3.69 (0.76)	32.42 (0.31)	1.71 (0.17)	-0.14 (0.089)
	Sliding	3.69 (0.82)	32.43 (0.33)	1.95 (0.15)	-0.25 (0.078)
Kahler Asten	Disjoint	3.33 (0.79)	25.68 (0.34)	1.75 (0.17)	-0.21 (0.093)
	Sliding	3.33 (0.83)	25.59 (0.34)	1.87 (0.15)	-0.25 (0.082)

estimators can be obtained by taking into account all successive, overlapping block maxima. The results are remarkable in view of the fact that the time series of sliding block maxima is non-stationary under the second sampling scheme. When restricted to the PWM estimator, the improvement was found to be substantial for negative shape parameters, both in large-sample theory and in finite-sample simulations. On the other hand, a deterioration may occur in finite-sample situations where the block size is small in comparison to the serial dependence (e.g., extremal index $\theta \leq 0.25$ with $r = 90$ and $\gamma \geq 0$), which, however, is untypical at least for certain environmental applications. As a consequence, subject to computational feasibility, the sliding blocks estimator should usually be preferred over its disjoint blocks version. A possible approach to deal with non-stationarities between seasons was worked out in a case study involving temperature extremes, including a bootstrap approach to assess estimation uncertainty. The paper suggests several important topics for future research:

- The new sampling scheme may be a worthwhile starting point for developing large-sample theory for other estimators commonly applied in extreme value statistics. Furthermore, in view of the simulation results in Section 4.3 and the theoretical results in [34], the sampling scheme may be generalized to certain forms of inner-seasonal non-stationarity.
- The developed theory shows that the sliding blocks method can be applied in situations where the respective sample is non-stationary (with constant GEV parameters). This suggests that the general method may also be applicable in situations involving non-stationary models for the GEV parameters, possibly to be estimated by maximum likelihood then.
- The proof of Lemma 2.4 suggests that the sliding blocks method may be generalized to some method involving an even larger subset of the set of all block maxima à la [30].
- The parametric bootstrap approach has not been studied theoretically. Likewise, possible alternative (nonparametric, block) bootstrap schemes could be investigated.

- (e) The asymptotic results may be used to derive more powerful formal tests for homogeneities within multivariate models, for instance involving a scaling model as imposed with the index flood assumption [23].
- (f) It was found that the sliding blocks method may be worse than its disjoint blocks counterpart under sampling scheme (S2) for certain unfavorable combinations of the block size and the serial dependence. A data-adaptive rule that provides a warning in such situations would be a useful statistical device.

7. Proofs. Within this section, we provide proofs for Lemma 2.4 on weak convergence of sliding block maxima and for Theorem 2.6 on the empirical process of rescaled block maxima. All further proofs and some intermediate results are postponed to the supplement [9].

PROOF OF LEMMA 2.4. Recalling the definition of $Z_{r,j}^{(\text{sb})}$ in (5), the assertion to be shown is equivalent to convergence in distribution of $Z_{r,1+\lfloor r\xi \rfloor}^{(\text{sb})}$ to $Z \sim G_\gamma$. We omit the upper index sb. Under sampling scheme (S1) the assertion holds by stationarity of the sliding block maxima. Consider sampling scheme (S2). For $\xi \in \{0, 1\}$, the assertion holds by assumption. Let $\xi \in (0, 1)$. By independence between and stationarity within blocks we get, for any $x \in S_\gamma$,

$$\begin{aligned} & \mathbb{P}(Z_{r,1+\lfloor r\xi \rfloor} \leq x) \\ &= \mathbb{P}(\max(X_{1+\lfloor r\xi \rfloor}, \dots, X_{r+\lfloor r\xi \rfloor}) \leq a_r x + b_r) \\ &= \mathbb{P}(\max(X_{1+\lfloor r\xi \rfloor}, \dots, X_r) \leq a_r x + b_r) \mathbb{P}(\max(X_{r+1}, \dots, X_{r+\lfloor r\xi \rfloor}) \leq a_r x + b_r) \\ &= \mathbb{P}\left(Z_{r-\lfloor r\xi \rfloor,1} \leq \frac{a_r}{a_{r-\lfloor r\xi \rfloor}} x + \frac{b_r - b_{r-\lfloor r\xi \rfloor}}{a_{r-\lfloor r\xi \rfloor}}\right) \mathbb{P}\left(Z_{\lfloor r\xi \rfloor,1} \leq \frac{a_r}{a_{\lfloor r\xi \rfloor}} x + \frac{b_r - b_{\lfloor r\xi \rfloor}}{a_{\lfloor r\xi \rfloor}}\right). \end{aligned}$$

Condition 2.1 implies that the expression in the previous display converges to

$$(25) \quad G_\gamma\left(\frac{x}{(1-\xi)^\gamma} + \frac{(1-\xi)^{-\gamma} - 1}{\gamma}\right) G_\gamma\left(\frac{x}{\xi^\gamma} + \frac{\xi^{-\gamma} - 1}{\gamma}\right) = G_\gamma(x),$$

where the last equation follows from a straightforward calculation. \square

PROOF OF THEOREM 2.6. First, consider the case where $\text{mb} = \text{db}$. Then we can write $\mathbb{H}_r^{(\text{db})} = \mathbb{C}_{n,r} \circ H_r$ almost surely, where

$$\mathbb{C}_{n,r}(u) = \sqrt{\frac{n}{r}} \frac{1}{m} \sum_{j=1}^m \left\{ \mathbf{1}(H_r(Z_{r,j}^{(\text{db})}) \leq u) - u \right\}, \quad u \in [0, 1].$$

Under sampling scheme (S1), we have $\mathbb{C}_{n,r} \xrightarrow{d} \mathbb{C}^{(\text{db})}$ in $\ell^\infty([0, 1])$ by Theorem 3.1 in [5], and similar (but simpler) arguments as in that proof show that the same convergence is met under sampling scheme (S2). Hence, by asymptotic equicontinuity, we obtain, $\mathbb{C}_{n,r} \circ H_r = \mathbb{C}_{n,r} \circ G_\gamma + o_{\mathbb{P}}(1) \xrightarrow{d} \mathbb{C}^{(\text{db})} \circ G_\gamma$ as asserted. Since the Brownian bridge $\mathbb{C}^{(\text{db})}$ has continuous trajectories almost surely, so does the limit process $\mathbb{H}^{(\text{db})}$.

Now, let $\text{mb} = \text{sb}$, omit the upper index sb, and note that we may redefine

$$\mathbb{H}_r^{(\text{sb})}(x) = \frac{1}{\sqrt{nr}} \sum_{j=1}^{n-r} \left\{ \mathbf{1}(Z_{r,j} \leq x) - \bar{H}_r(x) \right\}.$$

First, we are going to show asymptotic tightness. For simplicity, assume $r/n \in 3\mathbb{N}$. We may then write $\mathbb{H}_r(x) = \mathbb{H}_{r,1}(x) + \mathbb{H}_{r,2}(x) + \mathbb{H}_{r,3}(x)$, where

$$\mathbb{H}_{r,\ell}(x) = \frac{1}{\sqrt{nr}} \sum_{j \in J_r(\ell)} \sum_{i \in I_j} \left\{ \mathbf{1}(Z_{r,i} \leq x) - \bar{H}_r(x) \right\}$$

with $J_r(\ell) = \{j \in \{1, \dots, m-1\} : j \in 3\mathbb{N}_0 + \ell\}$ for $\ell = 1, 2, 3$ and $I_j = \{(j-1)r + 1, \dots, jr\}$ denoting the indices making up the j -th disjoint block of observations. It is sufficient to show asymptotic tightness of each $\mathbb{H}_{r,\ell}$, and since they all have the same distribution we only consider the case $\ell = 1$.

After successively applying Berbee's coupling lemma ([2], see also Lemma 4.1 in [15]), we can construct a triangular array $\{\tilde{Z}_{r,i}\}_{i \in I_1 \cup I_4 \cup \dots}$ for which the following hold:

- (i) For any $j \in J_r(1)$, we have $(\tilde{Z}_{r,i})_{i \in I_j} \stackrel{D}{=} (Z_{r,i})_{i \in I_j}$.
- (ii) For any $j \in J_r(1)$, we have $\mathbb{P}((\tilde{Z}_{r,i})_{i \in I_j} \neq (Z_{r,i})_{i \in I_j}) \leq \beta(r)$.
- (iii) $(\tilde{Z}_{r,i})_{i \in I_1}, (\tilde{Z}_{r,i})_{i \in I_4}, (\tilde{Z}_{r,i})_{i \in I_7} \dots$ is independent and identically distributed.

Let $\tilde{\mathbb{H}}_{r1}$ be defined in the same way as \mathbb{H}_{r1} , but in terms of $\{\tilde{Z}_{r,i}\}_{i \in I_1 \cup I_4 \cup \dots}$ instead of $\{Z_{r,i}\}_{i \in I_1 \cup I_4 \cup \dots}$. Asymptotic tightness of \mathbb{H}_{r1} follows once we show that

$$(26) \quad \|\mathbb{H}_{r1} - \tilde{\mathbb{H}}_{r1}\|_\infty = o_{\mathbb{P}}(1)$$

(where $\|H\|_\infty = \sup_{x \in \mathbb{R}} |H(x)|$) and that $\tilde{\mathbb{H}}_{r1}$ is asymptotically tight.

Regarding the latter assertion, note that

$$\tilde{\mathbb{H}}_{r1}(x) = \sum_{j \in J_r(1)} \{f_{r,j}(x) - \mathbb{E}[f_{r,j}(x)]\},$$

where

$$f_{r,j}(x) = \frac{1}{\sqrt{nr}} \sum_{l=1}^r \mathbf{1}(\tilde{Z}_{r,(j-1)r+l} \leq x)$$

Since the summands $f_{r,j}(x)$ making up $\tilde{\mathbb{H}}_{r1}$ are independent, we may apply classical results from empirical process theory for independent sequences. More precisely, asymptotic tightness follows from Theorem 11.16 in [27], once we show that $\{f_{r,j} : j \in J_r(1)\}$ is almost measurable Suslin (AMS) and that Conditions (A)-(E) from that Theorem are met. The AMS property follows from Lemma 11.15 in [27]; use $T_n = \mathbb{Q}$ as the a countable subset to deduce separability. The remaining items can be seen as follows:

- (a) Since $x \mapsto f_{r,j}(x)$ is monotone increasing, the discussion on p. 213 of [27] yields the manageability. The envelope functions can be chosen as

$$E_{r,j}(x) := \sqrt{r/n}, \quad j \in J_r(1),$$

which are trivially independent.

- (b) The limit $\lim_{n \rightarrow \infty} \mathbb{E}[\mathbb{H}_r(x)\tilde{\mathbb{H}}_r(y)]$ exists for all $x, y \in \mathbb{R}$. Indeed, since $f_{r,j}$ is independent of $f_{r,j'}$ when $j \neq j'$, we have

$$\begin{aligned} \mathbb{E}[\tilde{\mathbb{H}}_{r1}(x)\tilde{\mathbb{H}}_{r1}(y)] &= \frac{n}{3r} \text{Cov}(f_{r,1}(x), f_{r,1}(y)) \\ &= \frac{1}{3r} \sum_{l=1}^r \frac{1}{r} \sum_{m=1}^r \text{Cov}(\mathbf{1}(\tilde{Z}_{r,l} \leq x), \mathbf{1}(\tilde{Z}_{r,m} \leq y)) \\ &= \frac{1}{3} \int_0^1 \int_0^1 g_{r,x,y}(\xi, \xi') \, d\xi \, d\xi', \end{aligned}$$

where $g_{r,x,y}(\xi, \xi') = \text{Cov}(g_x(\tilde{Z}_{r,1+\lfloor r\xi \rfloor}), g_y(\tilde{Z}_{r,1+\lfloor r\xi' \rfloor}))$ for $g_x(z) = \mathbf{1}(z \leq x)$. Lemma B.3 (sampling scheme (S1)) and Lemma B.4 (sampling scheme (S2)) from the supplement yields $\lim_{r \rightarrow \infty} g_{r,x,y}(\xi, \xi') = G_{\gamma,|\xi-\xi'|}(x, y) - G_\gamma(x)G_\gamma(y)$ and by dominated convergence, we get

$$\lim_{n \rightarrow \infty} \mathbb{E}[\tilde{\mathbb{H}}_{r1}(x)\tilde{\mathbb{H}}_{r2}(y)] = \frac{1}{3} \int_0^1 \int_0^1 G_{\gamma,|\xi-\xi'|}(x, y) \, d\xi \, d\xi' - G_\gamma(x)G_\gamma(y).$$

- (c) Since $\sum_{j \in J_r(1)} \mathbb{E}[E_{r,j}^2] = \frac{1}{3}$, the sum of second moments of the envelopes is finite.
- (d) We have $\limsup_{n \rightarrow \infty} \sum_{j \in J_r(1)} \mathbb{E}[E_{r,j}^2 \mathbf{1}(E_{r,j} > \varepsilon)] = \limsup_{n \rightarrow \infty} \frac{1}{3} \mathbf{1}(\sqrt{r/n} > \varepsilon) = 0$ for every $\varepsilon > 0$.
- (e) For $x, y \in \mathbb{R}$, let

$$\rho_n(x, y) = \left\{ \sum_{j \in J_r(1)} \mathbb{E} \left[|f_{r,j}(x) - f_{r,j}(y)|^2 \right] \right\}^{1/2}.$$

We have to show that the pointwise limit of $\rho_n(x, y)$, say $\rho(x, y)$, exists and that, if $\lim_{n \rightarrow \infty} \rho(x_n, y_n) = 0$, then $\lim_{n \rightarrow \infty} \rho_n(x_n, y_n) = 0$. Without loss of generality assume $x \leq y$. Then

$$\begin{aligned} \rho_n(x, y)^2 &= \frac{1}{3r^2} \mathbb{E} \left[\left(\sum_{l=1}^r \mathbf{1}(x < \tilde{Z}_{r,l} \leq y) \right)^2 \right] \\ &= \frac{1}{2r^2} \sum_{l=1}^r \mathbb{P}(x < \tilde{Z}_{r,l} \leq y) + \frac{1}{r^2} \sum_{l=1}^r \sum_{h=l+1}^r \mathbb{P}(x < \tilde{Z}_{r,l} \leq y, x < \tilde{Z}_{r,h} \leq y). \end{aligned}$$

The first term is of order $1/r$ and thus converges to 0. The second one equals $\int_0^1 \int_\xi^1 \mathbb{P}(x < \tilde{Z}_{1+\lfloor r\xi \rfloor} \leq y, x < \tilde{Z}_{1+\lfloor r\xi' \rfloor} \leq y) d\xi' d\xi$. Due to Lemma B.4 and dominated convergence, this converges to

$$\rho(x, y)^2 = \int_0^1 \int_\xi^1 G_{\gamma, |\xi - \xi'|}(x, x) + G_{\gamma, |\xi - \xi'|}(y, y) - 2G_{\gamma, |\xi - \xi'|}(x, y) d\xi' d\xi.$$

This double integral can be calculated explicitly, where some care has to be taken on whether both, one, or none of the arguments x, y fall into the support of $G_{\gamma, |\xi - \xi'|}$. Since the first case is the most involved, we restrict to that case. For $x \leq y$ in such a way that $1 + \gamma x > 0$ and $1 + \gamma y > 0$, a straightforward calculation implies that

$$\rho(x, y)^2 = \frac{e^{-\tilde{x}}}{\tilde{x}} \left\{ 1 + \frac{e^{-\tilde{x}} - 1}{\tilde{x}} \right\} + \frac{e^{-\tilde{y}}}{\tilde{y}} \left\{ 1 + \frac{e^{-\tilde{y}} - 1}{\tilde{y}} \right\} - 2 \frac{e^{-\tilde{x}}}{\tilde{y}} \left\{ 1 + \frac{e^{-\tilde{y}} - 1}{\tilde{y}} \right\},$$

where $\tilde{x} := (1 + \gamma x)^{-1/\gamma} \geq \tilde{y} := (1 + \gamma y)^{-1/\gamma}$. Obviously, $\rho(x, y) = 0$ for $x = y$. Write $g(s) := \{1 + (e^{-s} - 1)/s\}/s$. Observing that $g(\tilde{x}) \leq g(\tilde{y})$, a careful calculation of derivatives shows that the function $[\tilde{y}, \infty) \rightarrow \mathbb{R}, \tilde{x} \mapsto g(\tilde{x})e^{-\tilde{x}} - 2g(\tilde{y})e^{-\tilde{x}}$ is strictly increasing. As a consequence, $\rho(x, y)$ is strictly decreasing in x (for $x \leq y$) and can therefore have only one root which must be at $x = y$. Altogether, $\rho(x, y) = 0$ iff $x = y$. But then, $\lim_{n \rightarrow \infty} \rho(x_n, y_n) = 0$ iff either $x = \lim_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} y_n$, the limit possibly being $\pm\infty$, or if $x_n = y_n$, eventually. In the latter case, $\rho_n(x_n, y_n) = 0$ eventually, while in the former case, we have

$$\begin{aligned} \rho_n(x_n, y_n) &\leq \frac{1}{r} \sum_{l=1}^r \left\| \mathbf{1}(x_n < \tilde{Z}_{r,l} \leq y_n) \mathbf{1}(x_n < y_n) + \mathbf{1}(y_n < \tilde{Z}_{r,l} \leq x_n) \mathbf{1}(y_n < x_n) \right\|_2 \\ &= \frac{1}{r} \sum_{l=1}^r |H_{r,l}(x_n) - H_{r,l}(y_n)| \\ &\leq |G_\gamma(x_n) - G_\gamma(y_n)| + \frac{2}{r} \sum_{l=1}^r \|H_{r,l} - G_\gamma\|_\infty, \end{aligned}$$

which converges to 0 by continuity of G_γ and Lemma B.5 (sampling scheme (S2)) or Condition 2.1 (sampling scheme (S1), which implies $H_{r,\ell} = H_r$).

Finally, (26) follows from

$$\begin{aligned} |\mathbb{H}_{r1}(x) - \tilde{\mathbb{H}}_{r1}(x)| &\leq \frac{1}{\sqrt{nr}} \sum_{j \in J_r(1)} \sum_{i \in I_j} |\mathbf{1}(Z_{r,i} \leq x) - \mathbf{1}(\tilde{Z}_{r,i} \leq x)| \\ &\leq \frac{1}{\sqrt{nr}} \sum_{j \in J_r(1)} \sum_{i \in I_j} \{\mathbf{1}(Z_{r,i} \neq \tilde{Z}_{r,i})\} \end{aligned}$$

for any $x \in \mathbb{R}$, which implies $\mathbb{P}(\|\mathbb{H}_{r1} - \tilde{\mathbb{H}}_{r1}\|_\infty > \varepsilon) \lesssim \sqrt{\frac{n}{r}}\beta(r) = o(1)$ for any $\varepsilon > 0$ by Markov's inequality and Condition 2.5(ii). The proof of asymptotic tightness of \mathbb{H}_{r1} and hence of \mathbb{H}_r is finished.

For the convergence of the finite dimensional distributions, note that indicator functions $g_x(z) = \mathbf{1}(z \leq x)$ are elements of the set \mathcal{G}' defined in Theorem B.1 and that we may write $\mathbb{H}_r^{(\text{sb})}(x) = \mathbb{G}_n^{(\text{sb})} g_x$. Therefore, Theorem B.1 yields convergence to a centered normal distribution. Further, a simple calculation shows that

$$\lim_{n \rightarrow \infty} \text{Cov}(\mathbb{G}_n^{(\text{sb})} g_x, \mathbb{G}_n^{(\text{sb})} g_y) = \text{Cov}(\mathbb{C}^{(\text{sb})}(G_\gamma(x)), \mathbb{C}^{(\text{sb})}(G_\gamma(y)))$$

with covariance function of $\mathbb{C}^{(\text{sb})}$ as defined in (7).

The inequality in (8) is a special case of the result in Theorem 2.12 in [36], and may also be deduced from Lemma B.10 in the supplement. \square

Acknowledgements. This work has been supported by the integrated project ‘‘Climate Change and Extreme Events - ClimXtreme Module B - Statistics (subproject B3.3)’’ funded by the German Federal Ministry of Education and Research (BMBF), which is gratefully acknowledged. Computational infrastructure and support were provided by the Centre for Information and Media Technology at Heinrich Heine University D usseldorf.

The authors are grateful to three unknown referees and an associate editor for their constructive comments that helped to improve the presentation substantially. The authors also appreciate valuable comments by Petra Friederichs, Marco Oesting and multiple participants of the Extreme Value Analysis (EVA) Conference 2021.

SUPPLEMENTARY MATERIAL

Supplement to ‘‘On the Disjoint and Sliding Block Maxima method for piecewise stationary time series’’

(doi: [COMPLETED BY THE TYPESETTER](#)). The supplement contains the remaining proofs for the results from the paper, as well as additional simulation results.

REFERENCES

- [1] BEIRLANT, J., GOEGEBEUR, Y., SEGERS, J. and TEUGELS, J. (2004). *Statistics of extremes: Theory and Applications*. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester.
- [2] BERBEE, H. C. P. (1979). *Random walks with stationary increments and renewal theory*. *Mathematical Centre Tracts* **112**. Mathematisch Centrum, Amsterdam. [MR547109 \(81e:60093\)](#)
- [3] BRADLEY, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.* **2** 107–144. Update of, and a supplement to, the 1986 original.
- [4] BROCKWELL, P. J. and DAVIS, R. A. (1987). *Time series: theory and methods*. *Springer Series in Statistics*. Springer-Verlag, New York. [MR868859](#)
- [5] B UCHER, A. and SEGERS, J. (2014). Extreme value copula estimation based on block maxima of a multivariate stationary time series. *Extremes* **17** 495–528.
- [6] B UCHER, A. and SEGERS, J. (2017). On the maximum likelihood estimator for the generalized extreme-value distribution. *Extremes* **20** 839–872. [MR3737387](#)
- [7] B UCHER, A. and SEGERS, J. (2018). Maximum likelihood estimation for the Fr chet distribution based on block maxima extracted from a time series. *Bernoulli* **24** 1427–1462.

- [8] BÜCHER, A. and SEGERS, J. (2018). Inference for heavy tailed stationary time series based on sliding blocks. *Electron. J. Statist.* **12** 1098–1125.
- [9] BÜCHER, A. and ZANGER, L. (2022). Supplement to “On the Disjoint and Sliding Block Maxima method for piecewise stationary time series”.
- [10] CISSOKHO, Y. and KULIK, R. (2021). Estimation of cluster functionals for regularly varying time series: sliding blocks estimators. *Electron. J. Stat.* **15** 2777–2831. [MR4280158](#)
- [11] COLES, S. (2001). *An introduction to statistical modeling of extreme values*. Springer Series in Statistics. Springer-Verlag London, Ltd., London. [MR1932132](#)
- [12] COOLEY, D. (2013). *Return Periods and Return Levels Under Climate Change* In *Extremes in a Changing Climate: Detection, Analysis and Uncertainty* 97–114. Springer Netherlands, Dordrecht.
- [13] DAVISON, A. C. and HINKLEY, D. V. (1997). *Bootstrap Methods and their Application*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.
- [14] DE HAAN, L. and FERREIRA, A. (2006). *Extreme value theory: an introduction*. Springer.
- [15] DEHLING, H. and PHILIPP, W. (2002). Empirical process techniques for dependent data. In *Empirical process techniques for dependent data* 3–113. Birkhäuser Boston, Boston, MA.
- [16] DOMBRY, C. (2015). Existence and consistency of the maximum likelihood estimators for the extreme value index within the block maxima framework. *Bernoulli* **21** 420–436.
- [17] DOMBRY, C. and FERREIRA, A. (2019). Maximum likelihood estimators based on the block maxima method. *Bernoulli* **25** 1690–1723.
- [18] DREES, H. and NEBLUNG, S. (2021). Asymptotics for sliding blocks estimators of rare events. *Bernoulli* **27** 1239–1269. [MR4255233](#)
- [19] EMBRECHTS, P., KLÜPPELBERG, C. and MIKOSCH, T. (1997). *Modelling extremal events: for insurance and finance*. Springer.
- [20] FERREIRA, A. and DE HAAN, L. (2015). On the block maxima method in extreme value theory: PWM estimators. *Ann. Statist.* **43** 276–298.
- [21] FISHER, R. A. and TIPPETT, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society* **24** 180–190.
- [22] GUMBEL, E. J. (1958). *Statistics of extremes*. Columbia University Press, New York.
- [23] HOSKING, J. R. M. and WALLIS, J. R. (1997). *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge University Press, Cambridge.
- [24] HOSKING, J. R. M., WALLIS, J. R. and WOOD, E. F. (1985). Estimation of the Generalized Extreme-Value Distribution by the Method of Probability-Weighted Moments. *Technometrics* **27** 251–261.
- [25] KATZ, R. W., PARLANGE, M. B. and NAVEAU, P. (2002). Statistics of extremes in hydrology. *Advances in Water Resources* **25** 1287 - 1304.
- [26] KOJADINOVIC, I. and NAVEAU, P. (2017). Detecting distributional changes in samples of independent block maxima using probability weighted moments. *Extremes* **20** 417–450.
- [27] KOSOROK, M. R. (2008). *Introduction to empirical processes and semiparametric inference*. Springer Series in Statistics. Springer, New York. [MR2724368](#) (2012b:62005)
- [28] LANDWEHR, J. M., MATALAS, N. C. and WALLIS, J. R. (1979). Probability weighted moments compared with some traditional techniques in estimating Gumbel Parameters and quantiles. *Water Resources Research* **15** 1055-1064.
- [29] LEADBETTER, M. R. (1983). Extremes and local dependence in stationary sequences. *Z. Wahrsch. Verw. Gebiete* **65** 291–306. [MR722133](#) (85b:60033)
- [30] OORSCHOT, J. and ZHOU, C. (2020). All Block Maxima method for estimating the extreme value index. *ArXiv preprints: 2010.15950*.
- [31] PHILIP, S., KEW, S., VAN OLDENBORGH, G. J., OTTO, F., VAUTARD, R., VAN DER WIEL, K., KING, A., LOTT, F., ARRIGHI, J., SINGH, R. et al. (2020). A protocol for probabilistic extreme event attribution analyses. *Advances in Statistical Climatology, Meteorology and Oceanography* **6** 177–203.
- [32] PRESCOTT, P. and WALDEN, A. T. (1980). Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. *Biometrika* **67** 723–724. [MR601119](#) (81m:62046)
- [33] SEN, P. K. (1968). Estimates of the regression coefficient based on Kendall’s tau. *Journal of the American statistical association* **63** 1379–1389.
- [34] STEIN, M. (2017). Should annual maximum temperatures follow a generalized extreme value distribution? *Biometrika* **104** 1–16.
- [35] STOTT, P. A., CHRISTIDIS, N., OTTO, F. E. L., SUN, Y., VANDERLINDEN, J.-P., VAN OLDENBORGH, G. J., VAUTARD, R., VON STORCH, H., WALTON, P., YIOU, P. and ZWIERS, F. W. (2016). Attribution of extreme weather and climate-related events. *WIREs Climate Change* **7** 23–41.
- [36] ZOU, N., VOLGUSHEV, S. and BÜCHER, A. (2021). Multiple block sizes and overlapping blocks for multivariate time series extremes. *Ann. Statist.* **49** 295–320.

**SUPPLEMENT TO THE PAPER:
“ON THE DISJOINT AND SLIDING BLOCK MAXIMA METHOD FOR
PIECEWISE STATIONARY TIME SERIES”**

BY AXEL BÜCHER¹ AND LEANDRA ZANGER²

¹*Heinrich-Heine-University, Düsseldorf, axel.buecher@hhu.de*

²*Heinrich-Heine-University, Düsseldorf, Leandra.Zanger@hhu.de*

Missing proofs for the results of the main paper are presented in Appendix A, with a couple of further theoretical results postponed to Appendix B. Appendix C contains explicit formulas for asymptotic covariance matrices appearing in the main paper. A finite sample comparison of sliding vs. disjoint block maxima samples is carried out in Appendix D. Appendix E contains additional simulation results. Finally, Appendix F contains a figure supporting the case study. References like Lemma 1.9, Figure 0, or Equation (4) always refer to the main paper.

The theoretical results are organized as follows: in Appendix A, we provide the missing proofs of Theorem 3.5 and Corollaries 3.6 and 3.7 from the main paper. Appendix B contains further theoretical results used throughout the proofs, and is decomposed into four sections:

- Section B.1 provides some central arguments used throughout the proofs of Theorem 2.6 and Theorem 3.5: Theorem B.1 is a general asymptotic normality result for a quite general class of functions (which might therefore be of interest on its own), while Proposition B.2 proves a key decomposition used within the proof of Theorem 3.5.
- Section B.2 is about weak convergence and moment convergence of sliding block maxima. Joint weak convergence of sliding block maxima is considered in Lemma B.3 (sampling scheme (S1)) and Lemma B.4 (sampling scheme (S2)); the results may be considered as bivariate extensions of Lemma 2.4 from the main paper and are later used for calculating asymptotic covariances. Lemma B.5 is about (uniform) convergence of the average cdf \bar{H}_r from Equation (6) under (S2); it is needed in the proofs of Theorem 2.6 and Proposition B.2. Moment convergence of block maxima is the content of Lemma B.6, which is deduced from weak convergence and uniform integrability, the latter being part of Lemma B.7.
- Section B.3 is about asymptotic covariances for empirical moments of block maxima, as required in the proof of the general asymptotic normality result in Theorem B.1. Sampling scheme (S1) is treated in Lemma B.8, while sampling scheme (S2) is treated in Lemma B.9. Finally, Lemma B.10 states that the sliding blocks limiting covariance in Theorem B.1 is never larger than its disjoint blocks counterpart.
- Section B.4 contains further auxiliary results. First, Lemma B.11 provides consistency of some abstract functionals which were employed in the proof of Proposition B.2. Next, Lemma B.12 provides Wasserstein consistency of $\hat{H}_r^{(mb)}$ for G_γ , a technical result needed in the proof of Proposition B.2 that eventually allows to dispense with arguments involving weighted weak convergence as used for deriving PWM asymptotics in [8]. Its proof may partly be generalized to a more abstract setting, which has been formulated in a separate Lemma B.13. Finally, Lemma B.14 and Lemma B.15 are simple adaptations of Lemma A.7 and A.8 in [3] which are needed for the blocking technique.

Section C contains two lemmas: Lemma C.1 provides formulas for the asymptotic covariance in Theorem 3.5, while Lemma C.2 provides formulas for the Jacobi matrix in Corollary 3.6.

Last but not least, Section E contains additional simulation results, collected in a sequence of subsections, i.e., additional simulation results for fixed block size (Section E.1), simulation results for fixed sample size (Section E.2), results for comparing the plain and the bias-reduced sliding blocks estimator (Section E.3), results for comparing sampling schemes (S1) and (S2) (Section E.4), additional simulation results for a different marginal distributions (Section E.5) and results for comparing ML and PWM estimation (Section E.6).

APPENDIX A: MISSING PROOFS FOR RESULTS FROM THE MAIN PAPER

PROOF OF THEOREM 3.5. Recall the definition of $Z_{r,j}^{(\text{db})}$ and $Z_{r,j}^{(\text{sb})}$ from (5) and note that, under sampling scheme (S1), both are approximately G_γ -distributed with PWMs $\beta_{\gamma,k} = \beta_{(0,1,\gamma),k}$ for large r (in particular, unlike $M_{r,j}^{(\cdot)}$, the variables $Z_{r,j}^{(\cdot)}$ are stochastically bounded for $r \rightarrow \infty$). The same in fact holds under sampling scheme (S2), see Lemma 2.4. For $\text{mb} \in \{\text{db}, \text{sb}\}$, $k \in \{0, 1, 2\}$, let

$$(A.1) \quad \tilde{\beta}_{r,k}^{(\text{db})} = \hat{\beta}_k(Z_{r,1}^{(\text{db})}, \dots, Z_{r,m}^{(\text{db})}), \quad \tilde{\beta}_{r,k}^{(\text{sb})} = \hat{\beta}_k(Z_{r,1}^{(\text{sb})}, \dots, Z_{r,n-r+1}^{(\text{sb})}).$$

Further, for $f : \mathbb{R} \rightarrow \mathbb{R}$ integrable with respect to G_γ , let

$$(A.2) \quad \mathbb{G}_n^{(\text{db})} f := \sqrt{\frac{n}{r}} \left(\frac{1}{m} \sum_{j=1}^m f(Z_{r,j}^{(\text{db})}) - \mathbb{E} \left[f(Z_{r,j}^{(\text{db})}) \right] \right),$$

$$(A.3) \quad \mathbb{G}_n^{(\text{sb})} f := \sqrt{\frac{n}{r}} \left(\frac{1}{n-r+1} \sum_{j=1}^{n-r+1} f(Z_{r,j}^{(\text{sb})}) - \mathbb{E} \left[f(Z_{r,j}^{(\text{sb})}) \right] \right),$$

where $Z \sim G_\gamma$. In view of Condition 3.2, the proof of (14) is finished once we show that

$$(A.4) \quad \sqrt{\frac{n}{r}} \left(\frac{\hat{\beta}_{r,k}^{(\text{mb})} - \beta_{\theta_r,k}}{a_r} \right) = \sqrt{\frac{n}{r}} \left(\tilde{\beta}_{r,k}^{(\text{mb})} - \beta_{\gamma,k} \right),$$

$$(A.5) \quad \sqrt{\frac{n}{r}} \left(\tilde{\beta}_{r,k}^{(\text{mb})} - \beta_{\gamma,k} \right) = \mathbb{G}_n^{(\text{mb})} f_k + B_{n,k}^{(\text{mb},\text{S})} + o_{\mathbb{P}}(1),$$

$$(A.6) \quad \left(\mathbb{G}_n^{(\text{mb})} f_k \right)_{k=0,1,2} \xrightarrow{d} \mathcal{N}_3(0, \mathbf{\Omega}^{(\text{mb})}).$$

Subsequently, we write $M_{r,j}$ and $Z_{r,j}$ instead of $M_{r,j}^{(\text{mb})}$ and $Z_{r,j}^{(\text{mb})}$, respectively, whenever an equation is correct both for the disjoint and the sliding blocks version.

We begin by proving (A.4), which holds irrespective of the sampling scheme. Note that $M_{r,(j)} = a_r Z_{r,(j)} + b_r$. Hence, since $\sum_{i=1}^n (i-1) = n(n-1)/2$ and $\sum_{i=1}^n (i-1)(i-2) = n(n-1)(n-2)/3$,

$$(A.7) \quad \hat{\beta}_{r,k}^{(\text{mb})} = a_r \tilde{\beta}_{r,k}^{(\text{mb})} + \frac{b_r}{k+1}, \quad k \in \{0, 1, 2\}.$$

Likewise, recalling the notation $\beta_{\gamma,k} = \beta_{(0,1,\gamma),k}$, a simple calculation shows that

$$(A.8) \quad \beta_{\theta_r,k} = a_r \beta_{\gamma,k} + \frac{b_r}{k+1},$$

This implies (A.4). The assertion in (A.5) is a consequence of Proposition B.2, and the weak convergence result in (A.6) follows from Theorem B.1. Finally, the assertion in (15) is a consequence of Lemma B.10. \square

PROOF OF COROLLARY 3.6. For the ease of reading, we omit the upper index mb. Recall ϕ defined in (16). Clearly, for $\beta = (\beta_0, \beta_1, \beta_2)' \in \mathcal{D}_\phi$,

$$\begin{pmatrix} \phi_1(\beta) \\ \frac{1}{a_r} \phi_2(\beta) \\ \frac{1}{a_r} \phi_3(\beta) - \frac{b_r}{a_r} \end{pmatrix} = \phi \begin{pmatrix} \frac{\beta_0 - b_r}{a_r} \\ \frac{\beta_1 - b_r/2}{a_r} \\ \frac{\beta_2 - b_r/3}{a_r} \end{pmatrix}.$$

As a consequence, by (A.7) and (A.8),

$$\begin{aligned} \sqrt{\frac{n}{r}} \begin{pmatrix} \hat{\gamma}_r - \gamma \\ (\hat{a}_r - a_r)/a_r \\ (\hat{b}_r - b_r)/a_r \end{pmatrix} &= \sqrt{\frac{n}{r}} \left\{ \begin{pmatrix} \phi_1(\hat{\beta}_r) \\ \frac{1}{a_r} \phi_2(\hat{\beta}_r) \\ \frac{1}{a_r} \phi_3(\hat{\beta}_r) - \frac{b_r}{a_r} \end{pmatrix} - \begin{pmatrix} \phi_1(\beta_{\theta_r}) \\ \frac{1}{a_r} \phi_2(\beta_{\theta_r}) \\ \frac{1}{a_r} \phi_3(\beta_{\theta_r}) - \frac{b_r}{a_r} \end{pmatrix} \right\} \\ &= \sqrt{\frac{n}{r}} \left\{ \phi(\hat{\beta}_r) - \phi(\beta_\gamma) \right\}, \end{aligned}$$

where $\beta_\gamma = (\beta_{\gamma,0}, \beta_{\gamma,1}, \beta_{\gamma,2})'$. The assertion in (17) is now a consequence of (A.5), (A.6), Condition 3.2 and the delta method. Finally, the assertion in (18) is an immediate consequence of (15). \square

PROOF OF COROLLARY 3.7. We omit the upper index mb. For $a > 0$ and $\gamma \in \mathbb{R}$ let $f(\gamma, a) = a \frac{c_T^{-\gamma} - 1}{\gamma}$. Note that $a_r^{-1} f(\gamma, a) = f(\gamma, a/a_r)$ and that $\nabla f(\gamma, 1)$ is equal to the first two coordinates of q_T . As a consequence, by the delta method,

$$\begin{aligned} \sqrt{n/r} \left(\frac{\widehat{\text{RL}}(T, r) - \text{RL}(T, r)}{a_r} \right) &= \sqrt{n/r} \left(f(\hat{\gamma}_r, \frac{\hat{a}_r}{a}) - f(\gamma, 1) + \frac{\hat{b}_r - b_r}{a_r} \right), \\ &= q_T' \sqrt{n/r} \begin{pmatrix} \hat{\gamma}_r - \gamma \\ (\hat{a}_r - a_r)/a_r \\ (\hat{b}_r - b_r)/a_r \end{pmatrix} + o_{\mathbb{P}}(1). \end{aligned}$$

The assertion then follows from Corollary 3.6. \square

APPENDIX B: ADDITIONAL THEORETICAL RESULTS USED THROUGHOUT THE PROOFS

B.1. Central arguments used for proving asymptotic normality of empirical PWMs.

The following result is a central ingredient in the proofs of Theorem 2.6 and Theorem 3.5, and may be of independent interest. Its proof is similar to the proof of Theorem 3.6 in [3] (disjoint blocks) and Theorem 2.6 in [4] (sliding blocks).

THEOREM B.1. *Assume that one of the sampling schemes from Condition 2.2 holds with $\gamma < 1/2$. Let*

$$(B.1) \quad \mathcal{G} = \{g : \mathbb{R} \rightarrow \mathbb{R} \text{ continuous} \mid \exists c, d \text{ such that } |g(x)| \leq c|x| + d \text{ for all } x \in \mathbb{R}\}.$$

If Conditions 2.5 and 3.3 hold, then, for arbitrary $g_1, \dots, g_p \in \mathcal{G}$, $p \in \mathbb{N}$, we have

$$\left(\mathbb{G}_n^{(\text{mb})} g_k \right)_{k=1, \dots, p} \xrightarrow{d} \mathcal{N}_p \left(\mathbf{0}, \left(\Lambda_{k, k'}^{(\text{mb})} \right)_{k, k'=1, \dots, p} \right),$$

where $\mathbb{G}_n^{(\text{mb})}$ is defined in (A.2) and (A.3) and where, with $Z \sim G_\gamma$ and $(Z_{1\xi}, Z_{2\xi}) \sim G_{\gamma, \xi}$,

$$\Lambda_{k, k'}^{(\text{db})} = \text{Cov}(g_k(Z), g_{k'}(Z)), \quad \Lambda_{k, k'}^{(\text{sb})} = 2 \int_0^1 \text{Cov}(g_k(Z_{1\xi}), g_{k'}(Z_{2\xi})) d\xi.$$

The same result holds with \mathcal{G} replaced by $\mathcal{G}' = \{\mathbf{1}_{(-\infty, t]} : t \in \mathbb{R}\}$; in that case, one may dispense with Condition 3.3.

PROOF OF THEOREM B.1. We start by considering the function class \mathcal{G} . The disjoint blocks case is a straightforward adaptation of the proof of Theorem 3.6 in [3] and is therefore omitted. For the sliding blocks case, we may follow the proof of Theorem 2.6 in [4], with substantial modifications for sampling scheme (S2). The basic idea consists of successively merging blocks of size r into a ‘big block of blocks’ followed by a ‘small block of blocks’ followed by a ‘big block of blocks’ and so on in such a way that the ‘small blocks of blocks’ are small enough to be asymptotically negligible for the asymptotics and at the same are large enough to make the ‘big blocks of blocks’ asymptotically independent, whence standard central limit theorems become available. We omit the upper index sb. Since \mathcal{G} is a vector space and by the Cramér-Wold-device, it suffices to show that, for any fixed $g \in \mathcal{G}$,

$$(B.2) \quad \mathbb{G}_n g \xrightarrow{d} \mathcal{N}(0, \sigma^2), \quad \sigma^2 = \int_0^1 \text{Cov}(g(Z_{1\xi}), g(Z_{2\xi})) d\xi.$$

For that purpose, let $I_j := \{(j-1)r+1, \dots, jr\}$, $j \in \{1, \dots, m-1\}$, denote the set of indices making up the j -th disjoint block of observations. Let $m^* = m_n^*$ be an integer sequence with $3 \leq m^* \leq m-1$ that converges to infinity and satisfies $m^* = o(m^{\delta/\{2(1+\delta)\}})$ for some $\delta \in (\frac{2}{\omega}, 2+\nu)$. For simplicity, assume that $q = (m-1)/m^* \in \mathbb{N}$. For $j \in \{1, \dots, q\}$, let

$$J_j^+ := I_{(j-1)m^*+1} \cup \dots \cup I_{jm^*-2}, \quad J_j^- := I_{jm^*-1} \cup I_{jm^*},$$

such that $|J_j^+| = (m^* - 2)r$ and $|J_j^-| = 2r$. Then, by (A.3),

$$\begin{aligned} \mathbb{G}_n g &= \sqrt{\frac{n}{r}} \left(\frac{1}{n-r+1} \sum_{j=1}^{n-r+1} (g(Z_{r,j}) - \mathbb{E}[g(Z_{r,j})]) \right) \\ &= (1 + o(1)) \frac{1}{\sqrt{nr}} \sum_{j=1}^q \left\{ \sum_{s \in J_j^+} (g(Z_{r,s}) - \mathbb{E}[g(Z_{r,s})]) + \sum_{s \in J_j^-} (g(Z_{r,s}) - \mathbb{E}[g(Z_{r,s})]) \right\} \\ &\quad + (1 + o(1)) \frac{1}{\sqrt{nr}} (g(Z_{r,n-r+1}) - \mathbb{E}[g(Z_{r,n-r+1})]) \end{aligned}$$

(B.3)

$$= (1 + o(1)) \left\{ \frac{1}{\sqrt{q}} \sum_{j=1}^q S_{n,j}^+ + \frac{1}{\sqrt{q}} \sum_{j=1}^q S_{n,j}^- \right\} + o_{L_2}(1),$$

where $S_{n,j}^\pm := \sqrt{q/(nr)} \sum_{s \in J_j^\pm} \{g(Z_{r,s}) - \mathbb{E}[g(Z_{r,s})]\}$. Note that $(S_{n,j}^\pm)_j$ is stationary for both of the sampling schemes (S1) and (S2).

We will next argue that the contribution of the ‘small blocks’ is negligible. Since $\mathbb{E}[S_{n,j}^-] = 0$, this follows if the variance is shown to converge to 0. We have

$$\begin{aligned} \text{Var} \left(\frac{1}{\sqrt{q}} \sum_{j=1}^q S_{n,j}^- \right) &= \text{Var}(S_{n,1}^-) + \frac{2}{q} \sum_{h=1}^{q-1} (q-h) \text{Cov}(S_{n,1}^-, S_{n,1+h}^-) \\ (B.4) \quad &\leq 3 \text{Var}(S_{n,1}^-) + 2 \sum_{h=2}^{q-1} \left(1 - \frac{h}{q}\right) |\text{Cov}(S_{n,1}^-, S_{n,1+h}^-)| \end{aligned}$$

by Cauchy-Schwarz. By stationarity across blocks, we may write

$$(B.5) \quad \begin{aligned} \text{Var}(S_{n,1}^-) &= \left\| \sqrt{\frac{q}{nr}} \sum_{s \in J_1^-} (g(Z_{r,s}) - \mathbb{E}[g(Z_{r,s})]) \right\|_2^2 \\ &\leq \frac{q}{nr} \left(\sum_{s \in J_1^-} \|g(Z_{r,s}) - \mathbb{E}[g(Z_{r,s})]\|_2 \right)^2 \leq 4 \frac{qr}{n} \left(\frac{1}{r} \sum_{s=1}^r \text{Var}(g(Z_{r,s})) \right)^2. \end{aligned}$$

Since $qr/n = q/m = O(1/m^*) = o(1)$, we obtain that $\text{Var}(S_{n,1}^-) = o(1)$ by Condition 3.3 (sampling scheme (S1)) or Condition 3.3 and Lemma B.7 (sampling scheme (S2)). It remains to consider the sum on the right-hand side of (B.4), which is equal to zero under sampling scheme (S2). For sampling scheme (S1), we may apply Lemma 3.11 in [6] with $1/p = 1/q = 1/(2 + \nu)$ to obtain

$$|\text{Cov}(S_{n,1}^-, S_{n,h}^-)| \leq 10 \|S_{n,1}^-\|_{2+\nu}^2 \alpha(\sigma(S_{n,1}^-), \sigma(S_{n,h}^-))^{\frac{\nu}{2+\nu}} \leq 10 \|S_{n,1}^-\|_{2+\nu}^2 \alpha(r)^{\frac{\nu}{2+\nu}}$$

for $h \geq 3$. Therefore

$$\sum_{h=2}^{q-1} |\text{Cov}(S_{n,1}^-, S_{n,1+h}^-)| \lesssim q \|S_{n,1}^-\|_{2+\nu}^2 \alpha(r)^{\frac{\nu}{2+\nu}} \lesssim q \frac{qr}{n} \alpha(r)^{\frac{\nu}{2+\nu}} = \frac{m}{(m^*)^2} \alpha(r)^{\frac{\nu}{2+\nu}},$$

which converges to zero since $m\alpha(r)^{\frac{\nu}{2+\nu}} = o(1)$ by Condition 2.5(ii) and the choice of ν in Condition 3.3.

The sum over the small blocks being negligible, it remains to show that $q^{-1/2} \sum_{j=1}^q S_{n,j}^+$ converges in distribution to a centered normal distribution with variance σ^2 as in (B.2). For sampling scheme (S2), $(S_{n,j}^+)_j$ is a rowwise independent triangular array, and a standard argument based on characteristic functions shows that we may assume the same for sampling scheme (S1). As a consequence, we may apply Ljapunov's central limit theorem, for which we need to check Lyapunov's Condition:

$$(B.6) \quad \exists \delta > 0: \quad \lim_{n \rightarrow \infty} \frac{\sum_{j=1}^q \mathbb{E}[|S_{n,j}^+|^{2+\delta}]}{\left\{ \sum_{j=1}^q \mathbb{E}[|S_{n,j}^+|^2] \right\}^{1+\frac{\delta}{2}}} = 0.$$

Now, by Condition 3.3 (sampling scheme (S1)) or Condition 3.3 and Lemma B.7 (sampling scheme (S2)),

$$\begin{aligned} \|S_{n,j}^+\|_{2+\delta} &\leq \sqrt{\frac{q}{nr}} (m^* - 2) \sum_{s \in I_1} \|g(Z_{r,s}) - \mathbb{E}[g(Z_{r,s})]\|_{2+\delta} \\ &\lesssim \sqrt{m^*} \frac{1}{r} \sum_{s=1}^r \|g(Z_{r,s}) - \mathbb{E}[g(Z_{r,s})]\|_{2+\delta} = O(\sqrt{m^*}). \end{aligned}$$

As a consequence, provided that $\mathbb{E}[|S_{n,j}^+|^2]$ is converging to a non-zero constant, the fraction in (B.6) is of the order $O(q^{-\delta/2} (m^*)^{1+\delta/2}) = O(m^{-\delta/2} (m^*)^{1+\delta}) = o(1)$ by the choice of δ and m^* in the paragraph below (B.2).

Finally, $\mathbb{E}[|S_{n,j}^+|^2] = \text{Var}(S_{n,j}^+) = \text{Var}(q^{-1/2} \sum_{j=1}^q S_{n,j}^+)$ converges to σ^2 since

$$\lim_{n \rightarrow \infty} \text{Var} \left(q^{-1/2} \sum_{j=1}^q S_{n,j}^+ \right) = \lim_{n \rightarrow \infty} \text{Var}(\mathbb{G}_n^{(\text{sb})} g)$$

by (B.3) and since we have shown that $\|q^{-1/2} \sum_{j=1}^q S_{n,j}^-\|_2 = o(1)$. The right-hand side of the previous display is equal to σ^2 by Lemma B.8 (sampling scheme (S1)) and Lemma B.9 (sampling scheme (S2)).

Finally, for the function class \mathcal{G}' of indicator functions, the previous proof remains valid with only minor modifications: the right hand-side of (B.5) converges to zero since finite linear combinations of indicators are bounded. For the arguments that follow, one may apply (the simpler) Lemma 3.9 rather than Lemma 3.11 in [6]. \square

The next result serves the purpose of proving Equation (A.5) in the proof of Theorem 3.5.

PROPOSITION B.2. *Suppose one of the sampling schemes from Condition 2.2 holds with $\gamma < 1/2$. If Conditions 2.5 and 3.3 are met, then, for $k \in \{0, 1, 2\}$,*

$$(B.7) \quad \sqrt{\frac{n}{r}} \left(\tilde{\beta}_{r,k}^{(\text{mb})} - \beta_{\gamma,k} \right) = \mathbb{G}_n^{(\text{mb})} f_k + B_{n,k}^{(\text{mb},\text{S})} + o_{\mathbb{P}}(1),$$

with $\tilde{\beta}_{r,k}^{(\text{mb})}$, $\mathbb{G}_n^{(\text{mb})}$, f_k and $B_{n,k}^{(\text{mb},\text{S})}$ as defined in (A.1), (A.2) and (A.3), (12) and Condition 3.2, respectively.

PROOF OF PROPOSITION B.2. We start by getting rid of the order statistics and claim that, for $k \in \{0, 1, 2\}$,

$$(B.8) \quad \tilde{\beta}_{r,k}^{(\text{mb})} = \bar{\beta}_{r,k}^{(\text{mb})} + O_{\mathbb{P}}(r/n),$$

where

$$\begin{aligned} \bar{\beta}_{r,k}^{(\text{db})} &= \frac{1}{m} \sum_{j=1}^m Z_{r,j}^{(\text{db})} \hat{H}_r^k(Z_{r,j}^{(\text{db})}), & \bar{\beta}_{r,k}^{(\text{sb})} &= \frac{1}{n-r+1} \sum_{j=1}^{n-r+1} Z_{r,j}^{(\text{sb})} \hat{H}_r^k(Z_{r,j}^{(\text{sb})}), \\ \hat{H}_r^{(\text{db})}(z) &= \frac{1}{m} \sum_{j=1}^m \mathbf{1}(Z_{r,j}^{(\text{db})} \leq z), & \hat{H}_r^{(\text{sb})}(z) &= \frac{1}{n-r+1} \sum_{j=1}^{n-r+1} \mathbf{1}(Z_{r,j}^{(\text{sb})} \leq z) \end{aligned}$$

The assertion is obvious for $k = 0$. For $k \in \{1, 2\}$, consider the disjoint and sliding case separately.

(i) First, let $\text{mb} = \text{db}$, and omit the upper index db for the ease of notation. Due to the no-tie assumption in Condition 2.2, we have $(\hat{H}_r(Z_{r,(1)}), \dots, \hat{H}_r(Z_{r,(m)})) = (1/m, \dots, m/m)$, and therefore

$$\begin{aligned} \tilde{\beta}_{r,1} &= \frac{1}{m} \sum_{j=1}^m Z_{r,(j)} \frac{m \hat{H}_r(Z_{r,(j)}) - 1}{m-1}, \\ \tilde{\beta}_{r,2} &= \frac{1}{m} \sum_{j=1}^m Z_{r,(j)} \frac{m \hat{H}_r(Z_{r,(j)}) - 1}{m-1} \frac{m \hat{H}_r(Z_{r,(j)}) - 2}{m-2}. \end{aligned}$$

As a consequence,

$$\tilde{\beta}_{r,1} - \bar{\beta}_{r,1} = \frac{1}{m(m-1)} \sum_{j=1}^m Z_{r,j} \{ \hat{H}_r(Z_{r,j}) - 1 \} = \frac{1}{m-1} (\bar{\beta}_{r,1} - \bar{\beta}_{r,0}).$$

The arguments to follow imply that the expression on the right-hand side is of the order $O_{\mathbb{P}}(m^{-1}) = O_{\mathbb{P}}(r/n)$. The case $k = 2$ can be treated similarly, and (B.8) is shown.

(ii) Now let $mb = sb$ and again, suppress the upper index sb . Write

$$\begin{aligned}\tilde{\beta}_{r,1} - \bar{\beta}_{r,1} &= \frac{1}{(n-r+1)(n-r)} \sum_{j=1}^{n-r+1} Z_{r,(j)} \left\{ \hat{H}_r(Z_{r,(j)}) - 1 \right\} + R_{n,1} \\ &= \frac{1}{n-r} (\bar{\beta}_{r,1} - \bar{\beta}_{r,0}) + R_{n,1},\end{aligned}$$

where

$$(B.9) \quad R_{n,1} = \frac{1}{n-r+1} \sum_{j=1}^{n-r+1} Z_{r,(j)} \frac{j - (n-r+1) \hat{H}_r(Z_{r,(j)})}{n-r}.$$

Again, the arguments to follow imply that $(\bar{\beta}_{r,1} - \bar{\beta}_{r,0})/(n-r)$ is of order $O_{\mathbb{P}}(1/n)$. For the treatment of $R_{n,1}$, denote the T different and ordered values of the scaled sliding block maxima by $\tilde{Z}_{r,(1)} < \tilde{Z}_{r,(2)} < \dots < \tilde{Z}_{r,(T)}$. Because of the no-tie assumption, we have $T \geq n/r$, which can easily be seen from the fact that the n/r pairwise different disjoint block maxima appear in the sequence of sliding block maxima as well. Now set

$$V_t := \left\{ j \in \{1, \dots, n-r+1\} : Z_{r,j} = \tilde{Z}_{r,(t)} \right\}, \quad t \in \{1, \dots, T\},$$

which defines a partition of $\{1, \dots, n-r+1\}$. We have $\alpha_t = |V_t| \leq r$, because otherwise the no-tie assumption would be violated. The empirical c.d.f. \hat{H}_r is a step function that jumps up by $\alpha_t/(n-r+1)$ in the points $\tilde{Z}_{r,(t)}$, so we have $\hat{H}_r(\tilde{Z}_{r,(t)}) = \sum_{s=1}^t \alpha_s / (n-r+1)$. Further, for each element $Z_{r,(j)}$ of the ordered sample $Z_{r,(1)}, \dots, Z_{r,(n-r+1)}$, we can find an index t_j such that $Z_{r,(j)} = \tilde{Z}_{r,(t_j)}$. As a consequence, $\sum_{s=1}^{t_j-1} \alpha_s < j \leq \sum_{s=1}^{t_j} \alpha_s$, which in turn implies

$$(n-r+1) \hat{H}_r(Z_{r,(j)}) - \alpha_{t_j} < j \leq (n-r+1) \hat{H}_r(Z_{r,(j)}).$$

Hence, by the definition of $R_{n,1}$ in (B.9), we have

$$|R_{n,1}| < \frac{1}{n-r+1} \sum_{j=1}^{n-r+1} |Z_{r,(j)}| \frac{\alpha_{t_j}}{n-r} \leq \frac{r}{n-r} O_{\mathbb{P}}(1) = O_{\mathbb{P}}(r/n),$$

where the $O_{\mathbb{P}}(1)$ -term follows from $\mathbb{E}[|Z_{r,j}|] = \mathbb{E}[|Z|] + o(1)$ by Lemma B.6 and $\text{Var}\{(n-r+1)^{-1} \sum_{j=1}^{n-r+1} |Z_{r,j}|\} = o(1)$ by Lemma B.8 and B.9. This proves (B.8) for $k=1$, and the case $k=2$ can be treated similarly with slightly more effort.

As a consequence, (B.8) is shown, and hence, for proving the proposition, it suffices to show (B.7) with $\beta_{r,k}$ replaced by $\bar{\beta}_{r,k}$. The assertion is immediate for $k=0$. For $k \in \{1, 2\}$, decompose

$$\sqrt{n/r} \left(\bar{\beta}_{r,k}^{(mb)} - \beta_{\gamma,k} \right) = \mathbb{X}_{n,k}^{(mb)} + \mathbb{Y}_{n,k}^{(mb)} + B_{n,k}^{(mb,S)}$$

where

$$\begin{aligned}\mathbb{X}_{n,k}^{(db)} &= \sqrt{\frac{n}{r}} \frac{1}{m} \sum_{j=1}^m Z_{r,j}^{(db)} \left\{ (\hat{H}_r^{(db)}(Z_{r,j}^{(db)}))^k - H_r^k(Z_{r,j}^{(db)}) \right\}, \\ \mathbb{Y}_{n,k}^{(db)} &= \sqrt{\frac{n}{r}} \left\{ \frac{1}{m} \sum_{j=1}^m Z_{r,j}^{(db)} H_r^k(Z_{r,j}^{(db)}) - \mathbb{E} \left[Z_{r,j}^{(db)} H_r^k(Z_{r,j}^{(db)}) \right] \right\},\end{aligned}$$

$$\begin{aligned}\mathbb{X}_{n,k}^{(\text{sb})} &= \sqrt{\frac{n}{r}} \frac{1}{n-r+1} \sum_{j=1}^{n-r+1} Z_{r,j}^{(\text{sb})} \left\{ (\hat{H}_r^{(\text{sb})}(Z_{r,j}^{(\text{sb})}))^k - \bar{H}_r^k(Z_{r,j}^{(\text{sb})}) \right\}, \\ \mathbb{Y}_{n,k}^{(\text{sb})} &= \sqrt{\frac{n}{r}} \left\{ \frac{1}{n-r+1} \sum_{j=1}^{n-r+1} Z_{r,j}^{(\text{sb})} \bar{H}_r^k(Z_{r,j}^{(\text{sb})}) - \mathbb{E} \left[Z_{r,j}^{(\text{sb})} \bar{H}_r^k(Z_{r,j}^{(\text{sb})}) \right] \right\},\end{aligned}$$

Recall that $\bar{H}_r = H_r$ for $(\text{mb}, \text{S}) \neq (\text{sb}, \text{S2})$, which we will occasionally use to rewrite the above expressions. Observing that, for $k \in \{1, 2\}$, f_k from (12) may be written as $f_k = f_{k,1} + f_{k,2}$ with

$$(B.10) \quad f_{k,1}(x) = x G_\gamma^k(x), \quad f_{k,2}(x) = \int_x^\infty y \nu_k'(G_\gamma(y)) dG_\gamma(y)$$

and $\nu_k(x) = x^k$, the proposition is shown once we show that, for $\text{mb} \in \{\text{db}, \text{sb}\}$,

$$(B.11) \quad \mathbb{X}_{n,k}^{(\text{mb})} = \mathbb{G}_n^{(\text{mb})} f_{k,2} + o_{\mathbb{P}}(1), \quad \mathbb{Y}_{n,k}^{(\text{mb})} = \mathbb{G}_n^{(\text{mb})} f_{k,1} + o_{\mathbb{P}}(1).$$

For the second assertion, it is sufficient to show that $\text{Var}(\mathbb{Y}_{n,k}^{(\text{mb})} - \mathbb{G}_n^{(\text{mb})} f_{k,1}) = o(1)$, by centeredness. For that purpose, write

$$Y_{n,j}^{(\text{mb})} = Z_{r,j}^{(\text{mb})} \{ \bar{H}_r^k(Z_{r,j}^{(\text{mb})}) - G_\gamma^k(Z_{r,j}^{(\text{mb})}) \},$$

and consider $\text{mb} \in \{\text{db}, \text{sb}\}$ separately.

First, let $\text{mb} = \text{db}$, and omit the upper index db for notational convenience. Then, by stationarity and assuming $m = n/r \in \mathbb{N}$ for simplicity (otherwise, a negligible remainder shows up),

$$(B.12) \quad \begin{aligned}\text{Var}(\mathbb{Y}_{n,k} - \mathbb{G}_n f_{k,1}) &= \text{Var}(Y_{n,1}) + 2 \sum_{h=1}^{m-1} \frac{m-h}{m} \text{Cov}(Y_{n,1}, Y_{n,1+h}) \\ &\leq 3 \text{Var}(Y_{n,1}) + 20 \|Y_{n,1}\|_{2+\delta}^2 \sum_{h=2}^{m-1} \alpha(\sigma(Y_{n,1}), \sigma(Y_{n,1+h}))^{\frac{\delta}{2+\delta}},\end{aligned}$$

where $\|Y\|_p = \mathbb{E}[|Y|^p]^{1/p}$ and where the last inequality follows from the Cauchy-Schwarz inequality and Lemma 3.11 in [6], with $\frac{1}{p} = \frac{1}{q} = \frac{1}{2+\delta}$ and $\delta \in [2/\omega, \nu]$ and ν from Condition 3.3. Since the sum starts at $h=2$, the variables generating the sigma fields depend on observations which are separated by a time lag of at least r , so each summand is smaller than or equal to $\alpha(r)^{\delta/(2+\delta)}$. Further, noting that $\bar{H}_r = H_r$,

$$\text{Var}(Y_{n,1}) \leq \|Y_{n,1}\|_{2+\delta}^2 \leq \|H_r^k - G_\gamma^k\|_\infty^2 \mathbb{E}[|Z_{r,1}|^{2+\delta}]^{2/(2+\delta)} = o(1)$$

by Conditions 2.1 and 3.3, where $\|F\|_\infty = \sup_{x \in \mathbb{R}} |F(x)|$. As a consequence,

$$\text{Var}(\mathbb{Y}_{n,k} - \mathbb{G}_n f_{k,1}) \leq o(1) \{3 + 20 \cdot m \alpha(r)^{\frac{\delta}{2+\delta}}\}$$

which converges to zero by Condition 2.5(ii), observing that $\delta(1+\omega) \geq 2+\delta$ by the choice of δ . Hence, the second assertion in (B.11) is shown.

Next, consider $\text{mb} = \text{sb}$ and, again assuming $n/r \in \mathbb{N}$, let

$$I_h := \{(h-1)r+1, \dots, hr\}, \quad h \in \{1, \dots, n/r\},$$

denote the set of indices making up the h -th disjoint block of observations. Then

$$(B.13) \quad \frac{\sqrt{n/r}}{n-r+1} \sum_{j=1}^{n-r+1} Y_{n,j} = (1+o(1)) \frac{1}{\sqrt{nr}} \left\{ \sum_{h=1}^{n/r-1} A_h + Y_{n,n-r+1} \right\},$$

where $A_h := \sum_{s \in I_h} Y_{n,s}$. It is sufficient to show that $\text{Var}((nr)^{-1/2} \sum_{h=1}^{n/r-1} A_h) = o(1)$, since the last summand in (B.13) is asymptotically negligible. By stationarity of $(A_h)_h$, we get

$$(B.14) \quad v_n = \text{Var} \left(\frac{1}{\sqrt{nr}} \sum_{h=1}^{n/r-1} A_h \right) = \frac{1}{nr} \left\{ \left(\frac{n}{r} - 1 \right) \text{Var}(A_1) + 2 \left(\frac{n}{r} - 2 \right) \text{Cov}(A_1, A_2) \right. \\ \left. + 2 \left(\frac{n}{r} - 3 \right) \text{Cov}(A_1, A_3) + 2 \sum_{h=3}^{n/r-2} \left(\frac{n}{r} - 1 - h \right) \text{Cov}(A_1, A_{1+h}) \right\}.$$

Now, by the Cauchy-Schwarz and Minkowski inequality,

$$(B.15) \quad |\text{Cov}(A_1, A_{1+h})| \leq \|A_1\|_2^2 = \left\| \sum_{s=1}^r Y_{n,s} \right\|_2^2 \leq r^2 \left(\|\bar{H}_r^k - G_\gamma^k\|_\infty \frac{1}{r} \sum_{s=1}^r \|Z_{r,s}\|_2 \right)^2,$$

where the right-hand side is equal to $r^2 \|H_r^k - G_\gamma^k\|_\infty^2 \|Z_{r,1}\|_2^2$ under sampling scheme (S1). Likewise, for $\delta \in [2/\omega, \nu)$ with ω and ν as in Conditions 2.5 and 3.3, we have

$$(B.16) \quad |\text{Cov}(A_1, A_{1+h})| \leq 10 \|A_1\|_{2+\delta}^2 \alpha(\sigma(A_1), \sigma(A_{1+h}))^{\frac{\delta}{2+\delta}} \leq 10 \|A_1\|_{2+\delta}^2 \alpha((h-2)r)^{\frac{\delta}{2+\delta}}$$

by Lemma 3.11 in [6]. Combining (B.15), (B.16) and the fact that the sum on the right-hand side of (B.14) starts at $h=3$, we get

$$v_n \lesssim \frac{1}{nr} r^2 \|\bar{H}_r^k - G_\gamma^k\|_\infty^2 \left\{ 5 \frac{n}{r} \left(\frac{1}{r} \sum_{s=1}^r \|Z_{r,s}\|_2 \right)^2 + 20 \left(\frac{n}{r} \right)^2 \left(\frac{1}{r} \sum_{s=1}^r \|Z_{r,s}\|_{2+\delta} \right)^2 \alpha(r)^{\frac{\delta}{2+\delta}} \right\} \\ = \|\bar{H}_r^k - G_\gamma^k\|_\infty^2 \left\{ 5 \left(\frac{1}{r} \sum_{s=1}^r \|Z_{r,s}\|_2 \right)^2 + 20 \left(\frac{1}{r} \sum_{s=1}^r \|Z_{r,s}\|_{2+\delta} \right)^2 \frac{n}{r} \alpha(r)^{\frac{\delta}{2+\delta}} \right\} \\ = o(1) \{O(1) + O(1)o(1)\} = o(1)$$

where the orders of the terms in brackets follow from Lemma B.5 (sampling scheme (S2)) or $\bar{H}_r = H_r$ and Condition 2.1 (sampling scheme (S1)), and from Conditions 2.5(ii) and 3.3 in combination with Lemma B.6.

Having treated the cases $\text{mb} \in \{\text{db}, \text{sb}\}$, the second assertion in (B.11) is shown, and it remains to treat the first one. Its proof will be split into two parts:

$$(B.17) \quad \mathbb{X}_{n,k}^{(\text{mb})} = \mathbb{X}'_{n,k}^{(\text{mb})} + o_{\mathbb{P}}(1), \quad \mathbb{X}'_{n,k}^{(\text{mb})} = \mathbb{G}_n^{(\text{mb})} f_{k,2} + o_{\mathbb{P}}(1),$$

where

$$\mathbb{X}'_{n,k}^{(\text{mb})} := \sqrt{\frac{n}{r}} \int_{\mathbb{R}} y \nu'_k(\bar{H}_r(y)) \left\{ \hat{H}_r^{(\text{mb})}(y) - \bar{H}_r(y) \right\} d\hat{H}_r^{(\text{mb})}(y).$$

The first assertion in (B.17) is immediate for $k=1$; in that case, even $\mathbb{X}_{n,1}^{(\text{mb})} = \mathbb{X}'_{n,1}^{(\text{mb})}$.

Treating the case $k = 2$ is more difficult, and for that purpose, let $\mathcal{P}(\mathbb{R})$ denote the set of all probability measures on \mathbb{R} and let

$$(B.18) \quad A = \{f : \mathbb{R} \rightarrow \mathbb{R} : \|f\|_\infty < \infty \text{ and } f \text{ is Borel-measurable}\} \subset \ell^\infty(\mathbb{R}),$$

equipped with the uniform metric. Further, let

$$(B.19) \quad W_1 = W_1(\mathbb{R}) = \{\mu \in \mathcal{P}(\mathbb{R}) : \int_{\mathbb{R}} |x| d\mu(x) < \infty\}$$

denote the Wasserstein space of order 1, equipped with the Wasserstein metric

$$d_{W_1}(\mu, \nu) = \sup_{h \in \text{Lip}_1} \left| \int_{\mathbb{R}} h(x) d\mu(x) - \int_{\mathbb{R}} h(x) d\nu(x) \right|,$$

where Lip_1 is the set of all Lipschitz-continuous functions with Lipschitz-constant 1. Recall that a sequence $(\mu_n)_n$ of probability measures in W_1 is said to converge weakly in W_1 to another probability measure $\mu \in W_1$, if

$$(B.20) \quad \mu_n \rightarrow \mu \text{ weakly} \quad \text{and} \quad \int_{\mathbb{R}} |x| d\mu_n(x) \rightarrow \int_{\mathbb{R}} |x| d\mu(x),$$

for $n \rightarrow \infty$. The Wasserstein metric metrizes weak convergence in W_1 (Theorem 6.8 in [10]):

$$(B.21) \quad \mu_n \rightarrow \mu \text{ weakly in } W_1 \Leftrightarrow d_{W_1}(\mu_n, \mu) \rightarrow 0,$$

Another equivalent property for weak convergence in W_1 is as follows (Definition 6.7 in [10]): for all continuous functions φ such that $|\varphi(x)| \leq C(1+|x|)$ for all x and some $C = C_\varphi$, it holds that

$$(B.22) \quad \int_{\mathbb{R}} \varphi(x) d\mu_n(x) \xrightarrow{n \rightarrow \infty} \int_{\mathbb{R}} \varphi(x) d\mu(x).$$

Now, consider the first assertion in (B.17) with $k = 2$. Observing that $H_r = \bar{H}_r$ for $\text{mb} = \text{db}$, we have

$$\mathbb{X}_{n,2}^{(\text{mb})} = \sqrt{\frac{n}{r}} \int_{\mathbb{R}} y \left((\hat{H}_r^{(\text{mb})}(y))^2 - \bar{H}_r^2(y) \right) d\hat{H}_r^{(\text{mb})}(y)$$

$$(B.23) \quad = \sqrt{\frac{n}{r}} \int_{\mathbb{R}} y \left(\hat{H}_r^{(\text{mb})}(y) - \bar{H}_r(y) \right) \left(\hat{H}_r^{(\text{mb})}(y) + \bar{H}_r(y) \right) d\hat{H}_r^{(\text{mb})}(y) = \mathbb{X}_{n,2}'^{(\text{mb})} + R_{n,2}^{(\text{mb})},$$

where

$$R_{n,2}^{(\text{mb})} = \sqrt{\frac{n}{r}} \int_{\mathbb{R}} y \left(\hat{H}_r^{(\text{mb})}(y) - \bar{H}_r(y) \right)^2 d\hat{H}_r^{(\text{mb})}(y).$$

Now, $|R_{n,2}^{(\text{mb})}| \leq \|\hat{H}_r^{(\text{mb})} - \bar{H}_r\|_\infty \psi(\mathbb{H}_r^{(\text{mb})}, 1, \hat{H}_r^{(\text{mb})})$, where $\psi(a, g, \mu) = \int_{\mathbb{R}} |y|g(y)a(y) d\mu(y)$ and where

$$\mathbb{H}_r^{(\text{mb})}(y) = \sqrt{\frac{n}{r}} \left(\hat{H}_r^{(\text{mb})}(y) - \bar{H}_r(y) \right), \quad y \in \mathbb{R}.$$

It then follows from continuity of ψ (Lemma B.11), $d_{W_1}(\hat{H}_r^{(\text{mb})}, G_\gamma) = o_{\mathbb{P}}(1)$ (Lemma B.12) and weak convergence of $\mathbb{H}_r^{(\text{mb})}$ in $\ell^\infty(\mathbb{R})$ to a process with continuous and bounded sample paths almost surely (Theorem 2.6) that $R_{n,2}^{(\text{mb})} = O_{\mathbb{P}}((r/n)^{-1/2}) = o_{\mathbb{P}}(1)$, which implies the assertion by (B.23).

It remains to show the second assertion in (B.17), for $k \in \{1, 2\}$. For that purpose, write

$$\mathbb{X}_{n,k}^{(\text{mb})} = \phi(\mathbb{H}_r^{(\text{mb})}, \nu'_k \circ \bar{H}_r, \hat{H}_r^{(\text{mb})}),$$

where $\phi : A \times C_b(\mathbb{R}) \times W_1 \rightarrow \mathbb{R}$, $(a, g, \mu) \mapsto \int_{\mathbb{R}} yg(y)a(y) d\mu(y)$. Likewise, a simple calculation shows that we may write

$$\mathbb{G}_n^{(\text{mb})} f_{k,2} = \sqrt{\frac{n}{r}} \int_{-\infty}^{\infty} y \nu'_k(G_\gamma(y)) \left\{ \hat{H}_r^{(\text{mb})}(y) - \bar{H}_r(y) \right\} dG_\gamma(y) = \phi(\mathbb{H}_r^{(\text{mb})}, \nu'_k \circ G_\gamma, G_\gamma).$$

The second assertion in (B.17) is then again an immediate consequence of continuity of ϕ (Lemma B.11), since $d_{W_1}(\hat{H}_r^{(\text{mb})}, G_\gamma) = o_{\mathbb{P}}(1)$ (Lemma B.12), $\|\bar{H}_r - G_\gamma\|_\infty = o(1)$ (Condition 2.1 or, for sampling scheme (S2), Lemma B.5) and since $\mathbb{H}_r^{(\text{mb})}$ converges weakly in $\ell^\infty(\mathbb{R})$ to a continuous limit (Theorem 2.6). The proof of Proposition B.2 is finished. \square

B.2. Weak convergence and moment convergence of sliding block maxima.

LEMMA B.3 (Joint weak convergence of sliding block maxima under (S1)). *Consider sampling scheme (S1) from Condition 2.2, let $r_n \rightarrow \infty$ with $r_n = o(n)$ and suppose there exists a sequence of integers $(\ell_n)_n$ such that $\ell_n = o(r_n)$, $\alpha(\ell_n) = o(\ell_n/r_n)$ for $n \rightarrow \infty$. Then, for any $\xi \geq 0$ and $x, y \in \mathbb{R}$,*

$$\lim_{n \rightarrow \infty} \mathbb{P}(Z_{r,1}^{(\text{sb})} \leq x, Z_{r,1+\lfloor r\xi \rfloor}^{(\text{sb})} \leq y) = G_{\gamma,\xi}(x, y),$$

with $G_{\gamma,\xi}$ as in (13) for $\xi \in [0, 1]$ and $G_{\gamma,\xi}(x, y) = G_\gamma(x)G_\gamma(y)$ for $\xi > 1$.

PROOF. We omit the upper index sb. The case $\xi = 0$ is trivial. For $j, k \in \mathbb{N}$ with $j \leq k$, let $M_{j:k} := \max(X_j, \dots, X_k)$. By similar arguments as in the proof of Lemma 5.1 in [3] (see below for details), we have, for $\xi \in (0, 1)$,

$$\begin{aligned} & \mathbb{P}(Z_{r,1} \leq x, Z_{r,1+\lfloor r\xi \rfloor} \leq y) \\ &= \mathbb{P}(M_{1:\lfloor r\xi \rfloor} \leq a_r x + b_r, M_{\lfloor r\xi \rfloor+1:r} \leq a_r(x \wedge y) + b_r, M_{r+1:r+\lfloor r\xi \rfloor} \leq a_r y + b_r) \\ &= \mathbb{P}(M_{1:\lfloor r\xi \rfloor-\ell} \leq a_r x + b_r, M_{\lfloor r\xi \rfloor+1:r-\ell} \leq a_r(x \wedge y) + b_r, M_{r+1:r+\lfloor r\xi \rfloor} \leq a_r y + b_r) + o(1) \\ &= \mathbb{P}(M_{1:\lfloor r\xi \rfloor-\ell} \leq a_r x + b_r) \mathbb{P}(M_{\lfloor r\xi \rfloor+1:r-\ell} \leq a_r(x \wedge y) + b_r) \\ & \quad \mathbb{P}(M_{r+1:r+\lfloor r\xi \rfloor} \leq a_r y + b_r) + o(1) \\ &= \mathbb{P}(M_{1:\lfloor r\xi \rfloor} \leq a_r x + b_r) \mathbb{P}(M_{\lfloor r\xi \rfloor+1:r} \leq a_r(x \wedge y) + b_r) \\ & \quad \mathbb{P}(M_{r+1:r+\lfloor r\xi \rfloor} \leq a_r y + b_r) + o(1). \end{aligned} \tag{B.24}$$

From the last expression we can then follow the claimed limit, since Condition 2.1 implies

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}(M_{1:\lfloor r\xi \rfloor} \leq a_r x + b_r) &= \mathbb{P}\left(Z_{\lfloor r\xi \rfloor,1} \leq \frac{a_r}{a_{\lfloor r\xi \rfloor}} x + \frac{b_r - b_{\lfloor r\xi \rfloor}}{a_{\lfloor r\xi \rfloor}}\right) \\ &= G_\gamma\left(\xi^{-\gamma} x + \frac{\xi^{-\gamma} - 1}{\gamma}\right), \end{aligned}$$

and analogously

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}(M_{r+1:r+\lfloor r\xi \rfloor} \leq a_r y + b_r) &= G_\gamma\left(\xi^{-\gamma} y + \frac{\xi^{-\gamma} - 1}{\gamma}\right), \\ \lim_{n \rightarrow \infty} \mathbb{P}(M_{\lfloor r\xi \rfloor+1:r} \leq a_r(x \wedge y) + b_r) &= G_\gamma\left((1 - \xi)^{-\gamma}(x \wedge y) + \frac{(1 - \xi)^{-\gamma} - 1}{\gamma}\right). \end{aligned}$$

Multiply the latter three limits to arrive at $G_{\gamma, \xi}$.

Explanation of (B.24): the first equality is obvious. For the second equality, note that $\mathbb{P}(A_n \cap B_n) = \mathbb{P}(A_n) + o(1)$ provided that $\lim_{n \rightarrow \infty} \mathbb{P}(A_n \cap B_n^c) = 0$. Therefore,

$$\begin{aligned} \mathbb{P}(\{M_{1: \lfloor r\xi \rfloor} \leq a_r x + b_r\}) &= \mathbb{P}(\{M_{1: \lfloor r\xi \rfloor - \ell} \leq a_r x + b_r\} \cap \{M_{\lfloor r\xi \rfloor - \ell + 1: \lfloor r\xi \rfloor} \leq a_r x + b_r\}) \\ &= \mathbb{P}(\{M_{1: \lfloor r\xi \rfloor - \ell} \leq a_r x + b_r\}) + o(1), \end{aligned}$$

in view of

$$\begin{aligned} &\mathbb{P}(\{M_{1: \lfloor r\xi \rfloor - \ell} \leq a_r x + b_r\} \cap \{M_{\lfloor r\xi \rfloor - \ell + 1: \lfloor r\xi \rfloor} \leq a_r x + b_r\}^c) \\ &\leq \mathbb{P}(M_{1: \lfloor r\xi \rfloor - \ell} < M_{\lfloor r\xi \rfloor - \ell + 1: \lfloor r\xi \rfloor}) \\ &= \mathbb{P}(M_{1: \lfloor r\xi \rfloor - \ell} < M_{1: \lfloor r\xi \rfloor}), \end{aligned}$$

and the last expression is of order $o(1)$ by Lemma B.15. With the same argument we can cut off the last ℓ observations in $M_{\lfloor r\xi \rfloor + 1: r}$ to treat $\mathbb{P}(M_{\lfloor r\xi \rfloor + 1: r} \leq a_r(x \wedge y) + b_r)$. Combining this we get the second equality. The third equality follows because $\alpha(\ell) = o(1)$, and the observations from one considered set to another consist of observations which are at least ℓ apart. The last equality can be proven in the manner of the second one, just reversely.

Finally, the case $\xi > 1$ can be proven in a similar way and is even easier, since in this case the blocks under consideration do not overlap. \square

LEMMA B.4 (Joint weak convergence of sliding block maxima under (S2)). *Consider sampling scheme (S2) from Condition 2.2, let $r_n \rightarrow \infty$ with $r_n = o(n)$ and suppose there exists a sequence of integers $(\ell_n)_n$ such that $\ell_n = o(r_n)$, $\alpha(\ell_n) = o(\ell_n/r_n)$ for $n \rightarrow \infty$. Then, for any $\xi, \xi' \geq 0$ and any $x, y \in S_\gamma$*

$$\lim_{n \rightarrow \infty} \mathbb{P}(Z_{r, 1 + \lfloor r\xi \rfloor}^{(\text{sb})} \leq x, Z_{r, 1 + \lfloor r\xi' \rfloor}^{(\text{sb})} \leq y) = G_{\gamma, |\xi - \xi'|}(x, y),$$

with $G_{\gamma, \xi}$ as in (13) for $\xi \in [0, 1]$ and $G_{\gamma, \xi}(x, y) = G_\gamma(x)G_\gamma(y)$ for $\xi > 1$.

PROOF. If $|\xi - \xi'| \geq 1$, the respective block maxima are independent, whence their joint c.d.f is the product of their marginal c.d.f.s and the result follows from Lemma 2.4. For $|\xi - \xi'| < 1$ the proof is a slight adaptation of the proof of Lemma B.3. \square

LEMMA B.5 (Convergence of average cdfs under (S2)). *Consider sampling scheme (S2) from Condition 2.2. Then, with \bar{H}_r as defined in Condition 3.2,*

$$\lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |\bar{H}_r(x) - G_\gamma(x)| = 0.$$

PROOF. Recalling $F_{r, j}$ from (4), we may write

$$\begin{aligned} H_{r, j+1}(x) &= F_{r, j+1}(a_r x + b_r) \\ &= \mathbb{P}(\max(X_{j+1}, \dots, X_r) \leq a_r x + b_r, \max(X_{r+1}, \dots, X_{r+j}) \leq a_r x + b_r) \\ \text{(B.25)} \quad &= F_{r-j}(a_r x + b_r) F_j(a_r x + b_r), \end{aligned}$$

with $F_0 \equiv 1$. We may thus write

$$\begin{aligned} \left\| \frac{1}{r} \sum_{j=1}^r H_{r, j} - G_\gamma \right\|_\infty &\leq \frac{1}{r} \sum_{j=1}^r \|H_{r, j} - G_\gamma\|_\infty \\ &= \int_0^1 \|H_{r, \lfloor r\xi \rfloor + 1} - G_\gamma\|_\infty d\xi \\ \text{(B.26)} \quad &= \int_0^1 \|F_{r - \lfloor r\xi \rfloor}(a_r \cdot + b_r) F_{\lfloor r\xi \rfloor}(a_r \cdot + b_r) - G_\gamma(\cdot)\|_\infty d\xi. \end{aligned}$$

Recalling identity (25) and invoking the triangular inequality after adding and subtracting $F_{r-\lfloor r\xi\rfloor}(a_r \cdot + b_r)G_\gamma(\xi^{-\gamma} \cdot + \frac{\xi^{-\gamma}-1}{\gamma})$, we obtain the bound

$$\|F_{r-\lfloor r\xi\rfloor}(a_r \cdot + b_r)F_{\lfloor r\xi\rfloor}(a_r \cdot + b_r) - G_\gamma(\cdot)\|_\infty \leq A_{r1}(\xi) + A_{r2}(\xi),$$

where

$$\begin{aligned} A_{r1}(\xi) &= \|F_{r-\lfloor r\xi\rfloor}(a_r \cdot + b_r) - G_\gamma((1-\xi)^{-\gamma} \cdot + ((1-\xi)^{-\gamma} - 1)/\gamma)\|_\infty \\ A_{r2}(\xi) &= \|F_{\lfloor r\xi\rfloor}(a_r \cdot + b_r) - G_\gamma(\xi^{-\gamma} \cdot + (\xi^{-\gamma} - 1)/\gamma)\|_\infty. \end{aligned}$$

Using the fact that $F_{\lfloor r\xi\rfloor}(x) = H_{\lfloor r\xi\rfloor}((x - b_{\lfloor r\xi\rfloor})/a_{\lfloor r\xi\rfloor})$, we have the bound

$$A_{r2}(\xi) \leq \|H_{\lfloor r\xi\rfloor} - G_\gamma\|_\infty + R_r(\xi),$$

where

$$R_r(\xi) := \left\| G_\gamma\left(\frac{a_r}{a_{\lfloor r\xi\rfloor}} \cdot + \frac{b_r - b_{\lfloor r\xi\rfloor}}{a_{\lfloor r\xi\rfloor}}\right) - G_\gamma\left(\xi^{-\gamma} \cdot + \frac{\xi^{-\gamma} - 1}{\gamma}\right) \right\|_\infty.$$

Likewise,

$$A_{r1}(\xi) \leq \|H_{r-\lfloor r\xi\rfloor} - G_\gamma\|_\infty + R_r(1-\xi),$$

We can thus conclude that the right-hand side of (B.26) may be bounded by

$$\begin{aligned} \int_0^1 \|H_{\lfloor r\xi\rfloor} - G_\gamma\|_\infty + \|H_{r-\lfloor r\xi\rfloor} - G_\gamma\|_\infty + R_r(\xi) + R_r(1-\xi) \, d\xi \\ = 2 \int_0^1 \|H_{\lfloor r\xi\rfloor} - G_\gamma\|_\infty \, d\xi + 2 \int_0^1 R_r(\xi) \, d\xi. \end{aligned}$$

The two integrals on the right-hand side converge to zero by dominated convergence and Equations (2) and (1) from Condition 2.1, respectively. \square

LEMMA B.6 (Moment convergence of block maxima). *Consider one of the sampling schemes from Condition 2.2 with $\gamma < 1/2$. Suppose there exists some $\nu > 0$ such that*

$$\limsup_{r \rightarrow \infty} \mathbb{E}[|Z_r|^{2+\nu}] < \infty,$$

and let f be a real-valued function for which there exist constants $c, d \in [0, \infty)$ and $0 \leq \mu < 2 + \nu$ with $|f(x)| \leq c|x|^\mu + d$ for all $x \in \mathbb{R}$. Then, with $Z \sim G_\gamma$,

$$\lim_{r \rightarrow \infty} \mathbb{E}[f(Z_r)] = \mathbb{E}[f(Z)], \quad \lim_{r \rightarrow \infty} \mathbb{E}\left[\frac{1}{r} \sum_{j=1}^r f(Z_{r,j}^{(\text{sb})})\right] = \mathbb{E}[f(Z)].$$

PROOF. The first assertion is an immediate consequence of weak convergence (Condition 2.1) and uniform integrability. This readily implies the second assertion under sampling scheme (S1). For sampling scheme (S2), we may write

$$\mathbb{E}\left[\frac{1}{r} \sum_{j=1}^r f(Z_{r,j}^{(\text{sb})})\right] = \int_0^1 \mathbb{E}[f(Z_{r,1+\lfloor r\xi\rfloor}^{(\text{sb})})] \, d\xi.$$

The expression inside the integral converges to $\mathbb{E}[f(Z)]$ by weak convergence (Lemma 2.4) and uniform integrability (Lemma B.7). Since the upper bound in the latter lemma holds uniformly in ξ , the assertion follows from dominated convergence. \square

LEMMA B.7 (Uniform integrability under (S2)). *Consider sampling scheme (S2) from Condition 2.2 with $\gamma < 1/2$ and suppose that*

$$(B.27) \quad \limsup_{r \rightarrow \infty} \mathbb{E}[|Z_r|^{2+\nu}] < \infty,$$

for some $\nu > 0$. Then

$$\limsup_{r \rightarrow \infty} \sup_{\xi \in [0,1]} \mathbb{E}\left[|Z_{r,1+\lfloor r\xi \rfloor}^{(\text{sb})}|^{2+\nu}\right] < \infty.$$

PROOF. Throughout, we omit the upper index sb and assume $r/2 \in \mathbb{N}$ for simplicity. By (B.25), the random variables $Z_{r,1+\lfloor r\xi \rfloor}$ and $Z_{r,r-\lfloor r\xi \rfloor+1}$ have the same distribution, whence it is sufficient to restrict the supremum to $\xi \in [0, 1/2]$. Next, note that

$$\max(X_{1+r/2}, \dots, X_r) \leq \max(X_{1+\lfloor r\xi \rfloor}, \dots, X_{r+\lfloor r\xi \rfloor}) \leq \max(X_1, \dots, X_{2r}),$$

which may be written as $M_{r/2,1+r/2} \leq M_{r,1+\lfloor r\xi \rfloor} \leq M_{2r,1} = M_{r,1} \vee M_{r,r+1}$. As a consequence,

$$\begin{aligned} |Z_{r,1+\lfloor r\xi \rfloor}| &= \left| \frac{M_{r,1+\lfloor r\xi \rfloor} - b_r}{a_r} \right| \leq \left| \frac{M_{r,1} - b_r}{a_r} \vee \frac{M_{r,r+1} - b_r}{a_r} \right| + \left| \frac{M_{r/2,1+r/2} - b_r}{a_r} \right| \\ &\leq |Z_{r,1}| + |Z_{r,1+r}| + \left| Z_{r/2,1+r/2} \frac{a_{r/2}}{a_r} + \frac{b_{r/2} - b_r}{a_r} \right|, \end{aligned}$$

which implies

$$\|Z_{r,1+\lfloor r\xi \rfloor}\|_{2+\nu} \leq 2\|Z_{r,1}\|_{2+\nu} + \|Z_{r/2,1}\|_{2+\nu} \left| \frac{a_{r/2}}{a_r} \right| + \left| \frac{b_{r/2} - b_r}{a_r} \right|.$$

This implies the assertion by (B.27) and (1). \square

B.3. Asymptotic covariances for empirical moments of block maxima.

LEMMA B.8. *Consider sampling scheme (S1) from Condition 2.2 with $\gamma < 1/2$ and suppose further that Conditions 2.5 and 3.3 hold. Then, for $g, g' \in \mathcal{G}$ with \mathcal{G} as defined in (B.1),*

$$\lim_{n \rightarrow \infty} \text{Cov}(\mathbb{G}_n^{(\text{sb})} g, \mathbb{G}_n^{(\text{sb})} g') = 2 \int_0^1 \text{Cov}(g(Z_{1\xi}), g'(Z_{2\xi})) \, d\xi.$$

where $\mathbb{G}_n^{(\text{sb})}$ is defined in (A.3) and where $(Z_{1\xi}, Z_{2\xi}) \sim G_{\gamma,\xi}$ with $G_{\gamma,\xi}$ from (13). The same result holds with \mathcal{G} replaced by $\mathcal{G}' = \{\mathbf{1}_{(-\infty, t]} : t \in \mathbb{R}\}$; in that case, one may dispense with Condition 3.3.

PROOF. We only give a proof for $g, g' \in \mathcal{G}$, as the case $g, g' \in \mathcal{G}'$ is similar but simpler. Without making further assumptions, the sequence ℓ_n that satisfies the condition from Lemma B.3 can be chosen as $\ell_n = \max\{s_n, \lfloor r_n \sqrt{\alpha(s_n)} \rfloor\}$, where $s_n = \lfloor \sqrt{r_n} \rfloor$ (see [4]), so we can apply that Lemma. We proceed similar as in [4]: for $h \in \{1, \dots, m\}$, let $I_h = \{(h-1)r_n + 1, \dots, hr_n\}$ denote the set of indices which make up the h -th disjoint block of observations. For simplicity, assume $n/r \in \mathbb{N}$. Then,

$$\sum_{j=1}^{n-r+1} g(Z_{r,j}) = g(Z_{r,n-r+1}) + \sum_{h=1}^{m-1} A_h, \quad \sum_{j=1}^{n-r+1} g'(Z_{r,j}) = g'(Z_{r,n-r+1}) + \sum_{h=1}^{m-1} B_h$$

where

$$A_h = \sum_{s \in I_h} g(Z_{r,s}), \quad B_h = \sum_{s \in I_h} g'(Z_{r,s}).$$

Note that, by stationarity, the sequences $(A_h)_h, (B_h)_h$ are stationary as well. By uniform integrability (Condition 3.3), the contribution of $g(Z_{r,n-r+1})$ and $g'(Z_{r,n-r+1})$ to the asymptotic covariance is negligible. Further, since

$$\sqrt{\frac{n}{r}} \frac{1}{n-r+1} = \frac{1}{\sqrt{nr}} \{1 + o(1)\},$$

it is sufficient to show that

$$v_n = \frac{1}{nr} \text{Cov} \left(\sum_{h=1}^{m-1} A_h, \sum_{j=1}^{m-1} B_j \right) \rightarrow 2 \int_0^1 \text{Cov} (g(Z_{1\xi}), g'(Z_{2\xi})) d\xi = v.$$

For that purpose, write

$$\begin{aligned} v_n &= \frac{1}{nr} \left\{ (m-1) \text{Cov}(A_1, B_1) + \sum_{h=1}^{m-2} (m-1-h) (\text{Cov}(A_1, B_{1+h}) + \text{Cov}(A_{1+h}, B_1)) \right\} \\ &= \frac{1}{nr} \left\{ (m-1) \text{Cov}(A_1, B_1) + (m-2) \text{Cov}(A_2, B_1 + B_3) \right\} \\ &\quad + \frac{1}{nr} \sum_{h=2}^{m-2} (m-1-h) \{ \text{Cov}(A_1, B_{1+h}) + \text{Cov}(A_{1+h}, B_1) \} \end{aligned}$$

(B.28)

$$= v_{n1} + v_{n2} + v_{n3},$$

where

$$\begin{aligned} v_{n1} &= \frac{1}{r^2} \text{Cov}(A_2, B_1 + B_2 + B_3), \quad v_{n2} = -\frac{1}{nr} \text{Cov}(A_2, 2B_1 + B_2 + 2B_3) \\ v_{n3} &= \frac{1}{nr} \left\{ \sum_{h=2}^{m-2} (m-1-h) \{ \text{Cov}(A_1, B_{1+h}) + \text{Cov}(A_{1+h}, B_1) \} \right\}. \end{aligned}$$

Next, for $\xi \geq 0$, define

$$g_{n1}(\xi) := \text{Cov} (g(Z_{r,1}), g'(Z_{r,1+\lceil r\xi \rceil})), \quad g_{n2}(\xi) := \text{Cov} (g(Z_{r,1+\lceil r\xi \rceil}), g'(Z_{r,1})),$$

such that, by stationarity,

$$\begin{aligned} \frac{1}{r^2} \text{Cov}(A_2, B_2) &= \frac{1}{r^2} \sum_{s=1}^{r_n} \sum_{t=1}^{r_n} \text{Cov}(g(Z_{r,s}), g'(Z_{r,t})) \\ &= \frac{1}{r} g_{n1}(0) + \frac{1}{r} \sum_{h=1}^{r-1} \left(1 - \frac{h}{r} \right) \{ g_{n1} \left(\frac{h}{r} \right) + g_{n2} \left(\frac{h}{r} \right) \}. \end{aligned}$$

Similarly, we obtain

$$\begin{aligned} \frac{1}{r^2} \text{Cov}(A_2, B_3) &= \frac{1}{r} \sum_{h=1}^{r-1} \frac{h}{r} g_{n1} \left(\frac{h}{r} \right) + \frac{1}{r} \sum_{h=r}^{2r-1} \left(2 - \frac{h}{r} \right) g_{n1} \left(\frac{h}{r} \right), \\ \frac{1}{r^2} \text{Cov}(A_2, B_1) &= \frac{1}{r} \sum_{h=1}^{r-1} \frac{h}{r} g_{n2} \left(\frac{h}{r} \right) + \frac{1}{r} \sum_{h=r}^{2r-1} \left(2 - \frac{h}{r} \right) g_{n2} \left(\frac{h}{r} \right). \end{aligned}$$

Combining the previous three equations, we get

$$\begin{aligned} v_{n1} &= \frac{1}{r} \sum_{h=0}^{r_n-1} \left\{ g_{n1}\left(\frac{h}{r}\right) + g_{n2}\left(\frac{h}{r}\right) \right\} - \frac{1}{r} g_{n2}(0) + \frac{1}{r} \sum_{h=r}^{2r-1} \left(2 - \frac{h}{r} \right) \left\{ g_{n1}\left(\frac{h}{r}\right) + g_{n2}\left(\frac{h}{r}\right) \right\} \\ &= \int_0^1 g_{n1}(\xi) + g_{n2}(\xi) \, d\xi + R_n, \end{aligned}$$

where

$$\begin{aligned} |R_n| &\leq \frac{1}{r_n} |g_{n2}(0)| + \int_1^2 \left| 2 - \frac{\lfloor r\xi \rfloor}{r} \right| |g_{n1}(\xi) + g_{n2}(\xi)| \, d\xi \\ &\leq \frac{1}{r_n} |g_{n2}(0)| + 2 \int_1^2 |g_{n1}(\xi) + g_{n2}(\xi)| \, d\xi. \end{aligned}$$

Now, weak convergence (Lemma B.3) and uniform integrability (Condition 3.3) implies that $\lim_{n \rightarrow \infty} g_{nj}(\xi) = \text{Cov}(g(Z_{1\xi}), g'(Z_{2\xi}))$ for $j \in \{1, 2\}$ and $\xi \geq 0$; in particular, the limit is zero for $\xi > 1$. As a consequence, by dominated convergence, $R_n = o(1)$ and then $\lim_{n \rightarrow \infty} v_{n1} = v$.

It remains to prove that v_{n2} and v_{n3} in Equation (B.28) converge to zero. It can be shown by similar arguments as for v_{n1} that $v_{n2} = O(r/n) = o(1)$. Considering v_{n3} , we start by treating the sum over those summands for which $h \geq 3$. Lemma 3.11 in [6] yields

$$\begin{aligned} |\text{Cov}(A_1, B_{1+h})| &\leq 10 \|A_1\|_{2+\nu} \|B_1\|_{2+\nu} \alpha(\sigma(A_1), \sigma(B_{1+h}))^{\frac{\nu}{2+\nu}} \\ &\leq 10r^2 \|g(Z_{r,1})\|_{2+\nu} \|g'(Z_{r,1})\|_{2+\nu} \alpha((h-2)r)^{\frac{\nu}{2+\nu}}, \end{aligned}$$

where ν is taken from Condition 3.3, so that the norms are uniformly bounded by some constant C . $\text{Cov}(A_{1+h}, B_1)$ can be bounded in the same way, whence the sum over the summands with $h \geq 3$ in v_{n3} may be bounded by

$$\frac{1}{r^2} \sum_{h=3}^{m-2} |\text{Cov}(A_1, B_{1+h})| + |\text{Cov}(A_{1+h}, B_1)| \leq 20C^2 \sum_{h=1}^{m-4} \alpha(hr)^{\frac{\nu}{2+\nu}},$$

which converges to zero by Condition 2.5(ii) and the choice of ν in Condition 3.3. The summand for $h = 2$ can be written as

$$\int_2^3 \left(3 - \frac{\lfloor r\xi \rfloor}{r} \right) (g_{n1}(\xi) + g_{n2}(\xi)) \, d\xi$$

and this converges to 0 by the same arguments as used in the treatment of R_n . \square

LEMMA B.9. *Consider sampling scheme (S2) from Condition 2.2 with $\gamma < 1/2$ and suppose further that Conditions 2.5 and 3.3 hold. Then, for $g, g' \in \mathcal{G}$ with \mathcal{G} as defined in (B.1),*

$$\lim_{n \rightarrow \infty} \text{Cov}(\mathbb{G}_n^{(\text{sb})} g, \mathbb{G}_n^{(\text{sb})} g') = 2 \int_0^1 \text{Cov}(g(Z_{1\xi}), g'(Z_{2\xi})) \, d\xi.$$

where $\mathbb{G}_n^{(\text{sb})}$ is defined in (A.3) and where $(Z_{1\xi}, Z_{2\xi}) \sim G_{\gamma, \xi}$ with $G_{\gamma, \xi}$ from (13). The same result holds with \mathcal{G} replaced by $\mathcal{G}' = \{\mathbf{1}_{(-\infty, t]} : t \in \mathbb{R}\}$; in that case, one may dispense with Condition 3.3.

PROOF. As in the previous proof, we only consider the case $g, g' \in \mathcal{G}$. Let $g_n(\xi, \xi') = \text{Cov}(g(Z_{r,1+\lfloor r\xi \rfloor}, g'(Z_{r,1+\lfloor r\xi' \rfloor})))$. With the same arguments and notations as in the proof of Lemma B.8 for sampling scheme (S1) we obtain that the leading term in the covariance under consideration is

$$(B.29) \quad \frac{1}{r^2} \{ \text{Cov}(A_1, B_1) + \text{Cov}(A_1, B_2) + \text{Cov}(A_2, B_1) \} \\ = \int_0^1 \int_0^1 g_n(\xi, \xi') \, d\xi \, d\xi' + \int_0^1 \int_1^2 g_n(\xi, \xi') \, d\xi \, d\xi' + \int_1^2 \int_0^1 g_n(\xi, \xi') \, d\xi \, d\xi'.$$

Weak convergence (Lemma B.4) and uniform integrability (Lemma B.7) implies that $g_n(\xi, \xi')$ converges to $\text{Cov}(g(Z_{1,|\xi-\xi'|}), g'(Z_{2,|\xi-\xi'|}))$, where $(Z_{1,|\xi-\xi'|}, Z_{2,|\xi-\xi'|}) \sim G_{\gamma,|\xi-\xi'|}$. By dominated convergence, the integrals in (B.29) converge as well, the limit being

$$(B.30) \quad \int_0^1 \int_0^1 \text{Cov}(g(Z_{1,|\xi-\xi'|}), g'(Z_{2,|\xi-\xi'|})) \, d\xi \, d\xi' \\ + 2 \int_0^1 \int_1^{1+\xi'} \text{Cov}(g(Z_{1,|\xi-\xi'|}), g'(Z_{2,|\xi-\xi'|})) \, d\xi \, d\xi',$$

where we have used symmetry and the fact that $Z_{1,|\xi-\xi'|}$ and $Z_{2,|\xi-\xi'|}$ are independent if $|\xi - \xi'| > 1$.

It remains to show that the last expression is equal to $2 \int_0^1 \text{Cov}(g(Z_{1\xi}), g'(Z_{2\xi})) \, d\xi$. For that purpose, note that the function $\xi \mapsto u(\xi) = \text{Cov}(g(Z_{1\xi}), g'(Z_{2\xi}))$ is bounded by some constant independent of ξ , as can be seen by applying the Cauchy-Schwarz inequality. Therefore, $u(\cdot)$ is integrable on every closed interval $[a, b] \subset \mathbb{R}$ with $\int_a^b u(\xi) \, d\xi = U(b) - U(a)$, where U denotes an antiderivative of u . We need to show that (B.30) may be written as $2 \{U(1) - U(0)\}$. By changing variables we obtain

$$(B.30) = \int_0^1 \left\{ \int_0^1 u(|\xi - \xi'|) \, d\xi + 2 \int_1^{1+\xi'} u(|\xi - \xi'|) \, d\xi \right\} d\xi' \\ = \int_0^1 \left\{ \int_0^{\xi'} u(\xi' - \xi) \, d\xi + \int_{\xi'}^1 u(\xi - \xi') \, d\xi + 2 \int_1^{1+\xi'} u(\xi - \xi') \, d\xi \right\} d\xi' \\ = \int_0^1 \left\{ \int_0^{\xi'} u(v) \, dv + \int_0^{1-\xi'} u(v) \, dv + 2 \int_{1-\xi'}^1 u(v) \, dv \right\} d\xi' \\ = \int_0^1 \left\{ \int_0^{\xi'} u(v) \, dv + \int_{1-\xi'}^1 u(v) \, dv + \int_0^1 u(v) \, dv \right\} d\xi' \\ = \int_0^1 \{U(\xi') - U(0) + U(1) - U(1 - \xi')\} \, d\xi' + U(1) - U(0) \\ = 2 \{U(1) - U(0)\},$$

since $\int_0^1 U(\xi') \, d\xi' = \int_0^1 U(1 - \xi') \, d\xi'$. □

LEMMA B.10. Let $\Lambda^{(\text{mb})}$ be defined as in Theorem B.1. Then $\Lambda^{(\text{sb})} \leq_L \Lambda^{(\text{db})}$.

PROOF. By the definition of the Loewner-order, we have to show that $\text{Var}(c' \mathbf{Y}^{(\text{db})}) \geq \text{Var}(c' \mathbf{Y}^{(\text{sb})})$ for any $c \in \mathbb{R}^p$, where $\mathbf{Y}^{(\text{mb})}$ denotes the limit variable of Theorem B.1. Choos-

ing an iid sequence satisfying the conditions from that theorem, we have

$$\text{Var}(c'Y^{(\text{mb})}) = \lim_{n \rightarrow \infty} \text{Var}\left(\mathbb{G}_n^{(\text{mb})}g\right), \quad \text{mb} \in \{\text{db}, \text{sb}\},$$

for some function $g \in \mathcal{G}$ (note that \mathcal{G} is closed under taking linear combinations), see Lemma B.8 for the case $\text{mb} = \text{sb}$. In view of the fact that $(g(Z_{r,i}^{(\text{sb})}))_i$ is r_n -dependent, the assertion follows from Lemma A.10 in [11]. \square

B.4. Further auxiliary results.

LEMMA B.11. *Recall the definition of A and W_1 in (B.18) and (B.19), respectively, and let*

$$\begin{aligned} \psi : A \times C_b(\mathbb{R}) \times W_1 &\rightarrow \mathbb{R}, & (a, g, \mu) &\mapsto \int_{\mathbb{R}} |y|g(y)a(y) \, d\mu(y), \\ \phi : A \times C_b(\mathbb{R}) \times W_1 &\rightarrow \mathbb{R}, & (a, g, \mu) &\mapsto \int_{\mathbb{R}} yg(y)a(y) \, d\mu(y). \end{aligned}$$

The maps ψ and ϕ are continuous in every $(a, g, \mu) \in C_b(\mathbb{R}) \times C_b(\mathbb{R}) \times W_1$.

PROOF. We only consider ϕ . For $(a, g, \mu) \in C_b(\mathbb{R}) \times C_b(\mathbb{R}) \times W_1$, let $(a_n, g_n, \mu_n)_n \subset A \times C_b(\mathbb{R}) \times W_1$ such that $\lim_{n \rightarrow \infty} (a_n, g_n, \mu_n) = (a, g, \mu)$, i.e., $\lim_{n \rightarrow \infty} \|a_n - a\|_{\infty} = 0$, $\lim_{n \rightarrow \infty} \|g_n - g\|_{\infty} = 0$ and $\lim_{n \rightarrow \infty} d_{W_1}(\mu_n, \mu) = 0$. Then,

$$\begin{aligned} &|\phi(a_n, g_n, \mu_n) - \phi(a, g, \mu)| \\ &\leq |\phi(a_n, g_n, \mu_n) - \phi(a, g_n, \mu_n)| + |\phi(a, g_n, \mu_n) - \phi(a, g, \mu_n)| \\ &\quad + |\phi(a, g, \mu_n) - \phi(a, g, \mu)| \\ &= \left| \int_{\mathbb{R}} yg_n(y)\{a_n(y) - a(y)\} \, d\mu_n(y) \right| + \left| \int_{\mathbb{R}} y\{g_n(y) - g(y)\}a(y) \, d\mu_n(y) \right| \\ &\quad + \left| \int_{\mathbb{R}} yg(y)a(y) \, d\{\mu_n(y) - \mu(y)\} \right| \\ &\leq \{\|a_n - a\|_{\infty} \|g_n\|_{\infty} + \|g_n - g\|_{\infty} \|a\|_{\infty}\} \int_{\mathbb{R}} |y| \, d\mu_n(y) + \left| \int_{\mathbb{R}} \varphi_{a,g}(y) \, d\{\mu_n(y) - \mu(y)\} \right|, \end{aligned}$$

where $\varphi_{a,g}(y) = yg(y)a(y)$. The first term on the right-hand side of the previous display converges to 0, since $d_{W_1}(\mu_n, \mu) \rightarrow 0$ implies $\int_{\mathbb{R}} |y| \, d\mu_n(y) \rightarrow \int_{\mathbb{R}} |y| \, d\mu(y) < \infty$ by (B.20). Since a and g are continuous, $\varphi_{a,g}$ is continuous as well and satisfies $|\varphi_{a,g}(y)| \leq \|g\|_{\infty} \|a\|_{\infty} |y| \leq \|g\|_{\infty} \|a\|_{\infty} (1 + |y|)$. Hence, the second term converges to 0 by (B.22). \square

LEMMA B.12 (Wasserstein consistency). *Consider one of the sampling schemes from Condition 2.2 with $\gamma < 1/2$. If Conditions 2.5 and 3.3 are met, then $d_{W_1}(\hat{H}_r^{(\text{mb})}, G_{\gamma}) = o_{\mathbb{P}}(1)$.*

PROOF. The result follows from application of Lemma B.13. First of all, for every n , $\hat{H}_r^{(\text{mb})}$ is a discrete probability measure and hence an element of W_1 . Next, note that

$$\|\hat{H}_r^{(\text{mb})} - G_{\gamma}\|_{\infty} \leq \|\hat{H}_r^{(\text{mb})} - \bar{H}_r\|_{\infty} + \|\bar{H}_r - G_{\gamma}\|_{\infty} = O_{\mathbb{P}}((r/n)^{1/2}) + o(1) = o_{\mathbb{P}}(1)$$

as a consequence of Theorem 2.6. It remains to be shown that $M_n = \int |z| \, d\hat{H}_r^{(\text{mb})}(z) = \mathbb{E}[|Z|] + o_{\mathbb{P}}(1)$, where $Z \sim G_{\gamma}$. First, $\mathbb{E}[M_n] \rightarrow \mathbb{E}[|Z|]$ by Lemma B.6. It thus suffices to

show that $\text{Var}(M_n) = o(1)$. This follows by the same arguments as for the treatment of $\text{Var}(\mathbb{Y}_{n,k}^{(\text{mb})} - \mathbb{G}_n^{(\text{mb})} f_{k,1})$ in the proof of Proposition B.2, see in particular (B.12) for disjoint blocks and (B.14) for sliding blocks. \square

LEMMA B.13. *If, for a sequence of random probability measures $(\hat{\mu}_n)_n$, $\hat{\mu}_n : \Omega \rightarrow W_1$ with distribution functions \hat{F}_n and some $\mu \in W_1$ with continuous distribution function F , the conditions*

$$\|\hat{F}_n - F\|_\infty \xrightarrow{\mathbb{P}} 0 \text{ and } \int_{\mathbb{R}} |x| d\hat{\mu}_n(x) \xrightarrow{\mathbb{P}} \int_{\mathbb{R}} |x| d\mu(x)$$

hold, then

$$d_{W_1}(\hat{\mu}_n, \mu) \xrightarrow{\mathbb{P}} 0.$$

PROOF. Weak convergence of probability measures μ_n on the real line to a limit μ_0 with continuous distribution function F_{μ_0} is well-known to be equivalent to uniform convergence of the respective distribution functions F_{μ_n} . As a consequence, by (B.21),

$$d_{W_1}(\mu_n, \mu_0) \rightarrow 0 \iff \|F_{\mu_n} - F_{\mu_0}\|_\infty + \left| \int |x| d\mu_n(x) - \int |x| d\mu_0(x) \right| \rightarrow 0.$$

The imposed assumptions imply

$$\|\hat{F}_n - F\|_\infty + \left| \int |x| d\hat{\mu}_n(x) - \int |x| d\mu(x) \right| \xrightarrow{\mathbb{P}} 0.$$

The assertion then follows from standard arguments based on passing to almost surely convergent subsequences. \square

The following two lemmas are simple adaptations of Lemma A.7 and A.8 in [3].

LEMMA B.14. *Assume Condition 2.1 and let $M_k = \max(X_1, \dots, X_k)$. If $r \rightarrow \infty$, $r = o(n)$, $\ell \rightarrow \infty$, $\ell = o(r)$ and $\alpha(\ell) = o(\ell/r)$ for $n \rightarrow \infty$, then, for all $y \in S_\gamma$,*

$$\mathbb{P}(M_\ell \geq a_r y + b_r) = O(\ell/r), \quad n \rightarrow \infty.$$

PROOF. From [2], Lemma 7.1 we know that, for all $u > 0$,

$$\mathbb{P}(F_r(M_\ell) > u) = O(\ell/r), \quad n \rightarrow \infty.$$

Since for all $y \in S_\gamma$ we have $\lim_{n \rightarrow \infty} F_r(a_r y + b_r) = G_\gamma(y)$, we have $F_r(a_r y + b_r) > G_\gamma(y)/2 > 0$ for sufficiently large n . Hence, the previous display implies

$$\begin{aligned} \mathbb{P}(M_\ell \geq a_r y + b_r) &\leq \mathbb{P}(F_r(M_\ell) \geq F_r(a_r y + b_r)) \\ &\leq \mathbb{P}(F_r(M_\ell) > G_\gamma(y)/2) = O(\ell/r). \end{aligned} \quad \square$$

LEMMA B.15. *Under the same conditions as in Lemma B.14, we have*

$$\lim_{n \rightarrow \infty} \mathbb{P}(M_r > M_{r-\ell}) = 0.$$

PROOF. For any $y \in S_\gamma$, we have

$$\begin{aligned} &\mathbb{P}(M_r > M_{r-\ell}) \\ &= \mathbb{P}(M_r > M_{r-\ell}, M_{r-\ell} \leq a_r y + b_r) + \mathbb{P}(M_r > M_{r-\ell}, M_{r-\ell} > a_r y + b_r) \\ &\leq \mathbb{P}(M_{r-\ell} \leq a_r y + b_r) + \mathbb{P}(\max\{X_{r-\ell+1}, \dots, X_r\} > a_r y + b_r). \end{aligned}$$

The first summand converges to $G_\gamma(y)$ because of Condition 2.1, invoking local uniform convergence in (1). The second summand is equal to $\mathbb{P}(M_\ell > a_r y + b_r)$ by stationarity, which converges to 0 by Lemma B.14. Now let $y \downarrow x_L$, the left endpoint of G_γ , to obtain the assertion. \square

APPENDIX C: EXACT FORMULAS FOR THE ASYMPTOTIC COVARIANCE MATRIX

LEMMA C.1. *Let $k, k' \in \{0, 1, 2\}$ and $\gamma < 1/2$. The asymptotic covariance from Theorem 3.5 can be written as*

$$\Omega_{k,k'}^{(\text{sb})} = 2C_\gamma \int_0^{1/2} \frac{h_{\gamma,k,k'}(w) + h_{\gamma,k',k}(w)}{\{w(1-w)\}^{1+\gamma}} dw,$$

where

$$C_\gamma = \begin{cases} \Gamma(2|\gamma|), & \gamma < 0, \\ 1, & \gamma = 0, \\ -\Gamma(1-2\gamma)/2\gamma, & \gamma > 0 \end{cases}$$

and, writing $c_{k,k'}(w) = kw + k'(1-w)$,

$$h_{\gamma,k,k'}(w) = \frac{\{c_{k,k'}(w) + 1\}^{2\gamma+1} - \{c_{k,k'}(w) + 1 - w\}^{2\gamma+1}}{w(2\gamma+1)} - \{c_{k,k'}(w) + 1\}^{2\gamma}$$

for $\gamma \notin \{0, -1/2\}$,

$$h_{-\frac{1}{2},k,k'}(w) = \frac{\log(c_{k,k'}(w) + 1) - \log(c_{k,k'}(w) + 1 - w)}{w} - \{c_{k,k'}(w) + 1\}^{-1}$$

and

$$h_{0,k,k'}(w) = 1 - \frac{c_{k,k'}(w) + 1 - w}{w} \log\left(\frac{c_{k,k'}(w) + 1}{c_{k,k'}(w) + 1 - w}\right).$$

Moreover,

$$\Omega_{k,k'}^{(\text{db})} = H_{k,k',\gamma} + H_{k',k,\gamma},$$

where

$$H_{k,k',\gamma} = \int_0^1 u^{k-1} (1-u) (-\log u)^{-1-\gamma} \int_0^u s^{k'} (-\log s)^{-1-\gamma} ds du.$$

The above integrals cannot be solved explicitly, but can be approximated to an arbitrary precision based on numerical integration.

PROOF. We only treat the sliding blocks case, the disjoint block case follows from a simple calculation and has, for instance, been worked out in [7], see their formula (12).

Recall that, for fixed $k \in \mathbb{N}$, we may write $f_k = f_{k,1} + f_{k,2}$ with

$$f_{k,1}(x) = xG_\gamma^k(x), \quad f_{k,2}(x) = \int_x^\infty y \nu'_k(G_\gamma(y)) dG_\gamma(y)$$

and $\nu_k(x) = x^k$. Therefore, supposing for the moment that $\gamma \neq 0$ and choosing $x \in S_\gamma$ and $k \in \mathbb{N}_{\geq 1}$, we may invoke

$$\begin{aligned} xG_\gamma^k(x) &= xe^{-k(1+\gamma x)^{-\frac{1}{\gamma}}} = \int_{-\infty}^x \frac{d}{dy} ye^{-k(1+\gamma y)^{-\frac{1}{\gamma}}} \mathbf{1}_{S_\gamma}(y) dy \\ &= \int_{-\infty}^x \{1 + ky(1 + \gamma y)^{-\frac{1}{\gamma}-1}\} e^{-k(1+\gamma y)^{-\frac{1}{\gamma}}} \mathbf{1}_{S_\gamma}(y) dy, \end{aligned}$$

to obtain

$$\begin{aligned} f_k(x) &= xG_\gamma^k(x) + \int_x^\infty kye^{-(k-1)(1+\gamma y)^{-\frac{1}{\gamma}}} (1 + \gamma y)^{-\frac{1}{\gamma}-1} e^{-(1+\gamma y)^{-\frac{1}{\gamma}}} \mathbf{1}_{S_\gamma}(y) dy \\ &= \int_{-\infty}^x e^{-k(1+\gamma y)^{-\frac{1}{\gamma}}} \mathbf{1}_{S_\gamma}(y) dy + \int_{-\infty}^\infty ky(1 + \gamma y)^{-\frac{1}{\gamma}-1} e^{-k(1+\gamma y)^{-\frac{1}{\gamma}}} \mathbf{1}_{S_\gamma}(y) dy \\ &= \int_{(1+\gamma x)^{-\frac{1}{\gamma}}}^\infty e^{-kt} t^{-\gamma-1} dt + k\beta_{\gamma, k-1}, \end{aligned}$$

where we made use of the substitution $t = (1 + \gamma y)^{-\frac{1}{\gamma}}$, $-t^{-\gamma-1} dt = dy$, and where $\beta_{\gamma, k}$ is defined in (9). Some thoughts reveal that the same equation is true for $\gamma = 0$, i.e.,

$$f_k(x) = \int_{e^{-x}}^\infty e^{-kt} t^{-1} dt + k\beta_{0, k-1}, \quad k \in \mathbb{N}_{\geq 1}.$$

Now, if $Z \sim G_\gamma$, the transformation $S = (1 + \gamma Z)^{-\frac{1}{\gamma}}$ (with $S = \exp(-Z)$ for $\gamma = 0$) is exponentially distributed with rate $\lambda = 1$. Indeed, for any $x > 0$, we have:

1. Case $\gamma > 0$:

$$\mathbb{P}(S \geq x) = \mathbb{P}\left((1 + \gamma Z)^{-\frac{1}{\gamma}} \geq x\right) = \mathbb{P}\left(Z \leq \frac{x^{-\gamma} - 1}{\gamma}\right) = \exp(-x).$$

2. Case $\gamma < 0$:

$$\mathbb{P}(S \geq x) = \mathbb{P}\left((1 - |\gamma|Z)^{\frac{1}{|\gamma|}} \geq x\right) = \mathbb{P}\left(Z \leq \frac{1 - x^{|\gamma|}}{|\gamma|}\right) = \exp(-x).$$

3. Case $\gamma = 0$:

$$\mathbb{P}(S \geq x) = \mathbb{P}(\exp(-Z) \geq x) = \mathbb{P}(Z \leq -\log(x)) = \exp(-x).$$

Therefore we may write

$$(C.1) \quad f_k(Z) = \int_S^\infty e^{-kt} t^{-\gamma-1} dt + k\beta_{\gamma, k-1},$$

so that, when taking the expectation, Fubini's Theorem yields

$$\mathbb{E}[f_k(Z)] = \mathbb{E}\left[\int_0^\infty \mathbf{1}(S \leq t) e^{-kt} t^{-\gamma-1} dt\right] + k\beta_{\gamma, k-1}$$

$$\begin{aligned}
&= \int_0^{\infty} \mathbb{P}(S \leq t) e^{-kt} t^{-\gamma-1} dt + k\beta_{\gamma, k-1} \\
\text{(C.2)} \quad &= \int_0^{\infty} \left(e^{-kt} - e^{-(k+1)t} \right) t^{-\gamma-1} dt + k\beta_{\gamma, k-1}.
\end{aligned}$$

To calculate integrals of that type, we distinguish cases according to the sign of γ :

1. Case $\gamma < 0$. For every $z > 0$, we have

$$\text{(C.3)} \quad \int_0^{\infty} e^{-zt} t^{-\gamma-1} dt = z^{\gamma} \int_0^{\infty} e^{-s} s^{-\gamma-1} ds = z^{\gamma} \Gamma(|\gamma|).$$

2. Case: $0 < \gamma < 1$. First notice that, by partial integration and since $\gamma < 1$,

$$\int_0^{\infty} (1 - e^{-s}) s^{-\gamma-1} ds = \frac{-s^{-\gamma}}{\gamma} (1 - e^{-s}) \Big|_{s=0}^{\infty} + \frac{1}{\gamma} \int_0^{\infty} e^{-s} s^{-\gamma} ds = \frac{\Gamma(1-\gamma)}{\gamma}$$

and with the same substitution as in the first case we get for every $z > 0$

$$\text{(C.4)} \quad \int_0^{\infty} (1 - e^{-zs}) s^{-\gamma-1} ds = z^{\gamma} \frac{\Gamma(1-\gamma)}{\gamma}.$$

3. Case: $\gamma = 0$. Recall the Exponential Integral

$$E_1(x) = \int_x^{\infty} e^{-t} t^{-1} dt = -\gamma - \log(x) + \int_0^x \frac{1 - e^{-t}}{t} dt.$$

Then we may write, for $0 < z_1 < z_2$,

$$\begin{aligned}
\int_0^{\infty} (e^{-z_1 t} - e^{-z_2 t}) t^{-1} dt &= \lim_{a \downarrow 0} \left\{ \int_{az_1}^{\infty} e^{-t} t^{-1} dt - \int_{az_2}^{\infty} e^{-t} t^{-1} dt \right\} \\
&= \lim_{a \downarrow 0} \left\{ \log(az_2) - \log(az_1) - \int_{az_1}^{az_2} \frac{1 - e^{-t}}{t} dt \right\} \\
\text{(C.5)} \quad &= \log\left(\frac{z_2}{z_1}\right).
\end{aligned}$$

Next, for $(Z_{1\xi}, Z_{2\xi}) \sim G_{\gamma, \xi}$ from Theorem 3.5, let

$$(S_{1\xi}, S_{2\xi}) = ((1 + \gamma Z_{1\xi})^{-\frac{1}{\gamma}}, (1 + \gamma Z_{2\xi})^{-\frac{1}{\gamma}})$$

be the random vector arising from the transformation of the marginal distributions to standard exponentially distributed random variables (with $(S_{1\xi}, S_{2\xi}) = (\exp(-Z_{1\xi}), \exp(-Z_{2\xi}))$ in case $\gamma = 0$). Note that, recalling $A_{\xi}(w) = \xi + (1 - \xi)\{w \vee (1 - w)\}$, we have

$$\mathbb{P}(S_{1\xi} \leq s, S_{2\xi} \leq t) = 1 - e^{-s} - e^{-t} + e^{-(s+t)A_{\xi}\left(\frac{t}{t+s}\right)}, \quad s, t > 0,$$

by a simple calculation. Invoking (C.1) and (C.2), we get, for $k, k' \in \mathbb{N}_{\geq 1}$,

$$\begin{aligned}
& \text{Cov}(f_k(Z_{1\xi}), f_{k'}(Z_{2\xi})) \\
&= \mathbb{E} \left[\int_{S_{1\xi}}^{\infty} \int_{S_{2\xi}}^{\infty} e^{-kt-k's}(ts)^{-\gamma-1} ds dt \right] \\
&\quad - \left(\int_0^{\infty} (e^{-kt} - e^{-(k+1)t}) t^{-\gamma-1} dt \right) \left(\int_0^{\infty} (e^{-k's} - e^{-(k'+1)s}) s^{-\gamma-1} ds \right) \\
&= \int_{(0,\infty)^2} \mathbb{P}(S_{1\xi} \leq s, S_{2\xi} \leq t) e^{-kt-k's}(ts)^{-\gamma-1} d(s, t) \\
&\quad - \int_{(0,\infty)^2} (1 - e^{-t})(1 - e^{-s}) e^{-kt-k's}(ts)^{-\gamma-1} d(s, t) \\
&= \int_{(0,\infty)^2} \left(e^{-(t+s)A_\xi(\frac{t}{t+s})} - e^{-(t+s)} \right) e^{-kt-k's}(ts)^{-\gamma-1} d(s, t) \\
&\tag{C.6} \\
&= \int_0^1 \int_0^\infty \frac{e^{-u\{A_\xi(w)+kw+k'(1-w)\}} - e^{-u\{kw+k'(1-w)+1\}}}{\{w(1-w)\}^{\gamma+1}} u^{-2\gamma-1} dudw,
\end{aligned}$$

where we used the change of variables $u = t + s$, $w = t/(t + s)$, i.e., $uw = t$, $u(1 - w) = s$ with Jacobian determinant u . As the function f_k is the identity for $k = 0$, applying Hoeffding's formula for $k = k' = 0$ yields

$$\begin{aligned}
\text{Cov}(f_0(Z_{1,\xi}), f_0(Z_{2,\xi})) &= \text{Cov}(Z_{1,\xi}, Z_{2,\xi}) \\
&= \int_{\mathbb{R}} (\mathbb{P}(Z_{1\xi} \geq x, Z_{2\xi} \geq y) - \mathbb{P}(Z_{1\xi} \geq x)\mathbb{P}(Z_{2\xi} \geq y)) \mathbf{1}_{S_\gamma^2}(x, y) dx(x, y) \\
&= \int_0^\infty \int_0^\infty \left(e^{-(t+s)A_\xi(\frac{t}{t+s})} - e^{-(t+s)} \right) (ts)^{-\gamma-1} ds dt,
\end{aligned}$$

which implies (C.6) with $k = k' = 0$. Finally, for the case $k \in \mathbb{N}_{\geq 1}$ and $k' = 0$, we may apply a generalized Hoeffding formula ([9], Theorem 3.1), which yields, with

$$f'_k(x) = \frac{d}{dx} \left\{ \int_{(1+\gamma x)^{-\frac{1}{\gamma}}}^{\infty} e^{-kt} t^{-\gamma-1} dt + k\beta_{\gamma, k-1} \right\} = e^{-k(1+\gamma x)^{-\frac{1}{\gamma}}},$$

(defined as $e^{-e^{-x}}$ for $\gamma = 0$) that (C.6) is also valid if only one of k, k' equals 0. As a summary, the equation holds for all $k, k' \in \mathbb{N}_{\geq 0}$.

We proceed by first restricting attention to the case $\gamma < 0$. By (C.3), for every $z > 0$,

$$\int_0^\infty e^{-uz} u^{-2\gamma-1} du = \Gamma(2|\gamma|) z^{2\gamma},$$

so that (C.6) equals

$$\Gamma(2|\gamma|) \int_0^1 \frac{\{kw + k'(1-w) + A_\xi(w)\}^{2\gamma} - \{kw + k'(1-w) + 1\}^{2\gamma}}{\{w(1-w)\}^{\gamma+1}} dw.$$

By symmetry and the definition of A_ξ , this expression may be written as

$$\Gamma(2|\gamma|) \{J_{\gamma,k,k'}(\xi) + J_{\gamma,k',k}(\xi)\}$$

where

$$J_{\gamma,k,k'}(\xi) = \int_0^{1/2} \frac{\{(k+\xi)w + (k'+1)(1-w)\}^{2\gamma} - \{kw + k'(1-w) + 1\}^{2\gamma}}{\{w(1-w)\}^{\gamma+1}} dw,$$

As a summary,

$$\Omega_{k,k'}^{(sb)} = 2 \int_0^1 \text{Cov}(f_k(Z_{1\xi}), f_{k'}(Z_{2\xi})) d\xi = 2\Gamma(2|\gamma|) \int_0^1 J_{\gamma,k,k'}(\xi) + J_{\gamma,k',k}(\xi) d\xi.$$

Finally, note that, for $\beta \neq -1$ and $a \neq 0$,

$$\int_0^1 (a\xi + b)^\beta d\xi = \frac{(a+b)^{\beta+1} - b^{\beta+1}}{(\beta+1)a},$$

while for $\beta = -1$ and $a \neq 0$,

$$\int_0^1 (a\xi + b)^{-1} d\xi = \frac{\log(a+b) - \log(b)}{a},$$

which readily implies

$$\int_0^1 J_{\gamma,k,k'}(\xi) d\xi = \int_0^{1/2} \frac{h_{\gamma,k,k'}(w)}{\{w(1-w)\}^{1+\gamma}} dw$$

after changing the order of integration, and hence yields the asserted formula after assembling terms.

For the case $\gamma > 0$, we add ± 1 in the numerator in (C.6) and with (C.4) we see that,

$$\text{Cov}(f_k(Z_{1\xi}), f_{k'}(Z_{2\xi})) = -\frac{\Gamma(1-2\gamma)}{2\gamma} \{J_{\gamma,k,k'}(\xi) + J_{\gamma,k',k}(\xi)\}.$$

The calculations for the case $\gamma < 0$ imply the asserted formula.

Finally, for the case $\gamma = 0$, (C.5) yields that the expression in (C.6) may be rewritten as

$$\int_0^1 \log\left(\frac{c_{k,k'}(w) + 1}{A_\xi(w) + c_{k,k'}(w)}\right) \{w(1-w)\}^{-1} dw = J_{0,k,k'}(\xi) + J_{0,k',k}(\xi),$$

where

$$J_{0,k,k'}(\xi) = \int_0^{1/2} \log\left(\frac{c_{k,k'}(w) + 1}{\xi w + c_{k,k'}(w) + 1 - w}\right) \{w(1-w)\}^{-1} dw.$$

Since

$$\int_0^1 \log\left(\frac{c}{\xi a + b}\right) d\xi = 1 + \frac{1}{a} \left((a+b) \log\left(\frac{c}{a+b}\right) - b \log\left(\frac{c}{b}\right) \right)$$

for $a, b, c > 0$, we get

$$\int_0^1 J_{0,k',k}(\xi) d\xi = \int_0^{1/2} \frac{h_{0,k,k'}(w)}{w(1-w)} dw,$$

which implies the final formula. \square

LEMMA C.2. *The entries $(c_{j,k})_{j,k}$ of the Jacobian matrix C of ϕ from Corollary 3.6 are given by*

$$\begin{aligned} c_{11} &= \left(\frac{3^\gamma - 1}{2^\gamma - 1} - 1 \right) \tilde{c}_{\gamma,1}, & c_{12} &= -2 \frac{3^\gamma - 1}{2^\gamma - 1} \tilde{c}_{\gamma,1}, & c_{13} &= 3 \tilde{c}_{\gamma,1}, \\ c_{21} &= -\tilde{c}_{\gamma,2} + c_{11} \tilde{c}_{\gamma,3}, & c_{22} &= 2 \tilde{c}_{\gamma,2} + c_{12} \tilde{c}_{\gamma,3}, & c_{23} &= c_{1,3} \tilde{c}_{\gamma,3}, \\ c_{31} &= 1 + \tilde{c}_{\gamma,4} c_{11} + \tilde{c}_{\gamma,5} c_{21}, & c_{32} &= \tilde{c}_{\gamma,4} c_{12} + \tilde{c}_{\gamma,5} c_{22}, & c_{33} &= \tilde{c}_{\gamma,4} c_{13} + \tilde{c}_{\gamma,5} c_{23}, \end{aligned}$$

where

$$\begin{aligned} \tilde{c}_{\gamma,1} &= \frac{\gamma}{\Gamma(1-\gamma)(2^\gamma-1)} \left\{ \frac{3^\gamma \log(3)}{2^\gamma-1} - \frac{2^\gamma(3^\gamma-1)\log(2)}{(2^\gamma-1)^2} \right\}^{-1}, \\ \tilde{c}_{\gamma,2} &= \frac{\gamma}{\Gamma(1-\gamma)(2^\gamma-1)}, \\ \tilde{c}_{\gamma,3} &= \frac{1}{\gamma} - \frac{2^\gamma \log(2)}{2^\gamma-1} + \frac{\Gamma'(1-\gamma)}{\Gamma(1-\gamma)}, \\ \tilde{c}_{\gamma,4} &= \frac{1}{\gamma^2} \{ \gamma \Gamma'(1-\gamma) - 1 + \Gamma(1-\gamma) \}, \\ \tilde{c}_{\gamma,5} &= \frac{1}{\gamma} (1 - \Gamma(1-\gamma)). \end{aligned}$$

For $\gamma = 0$, the expressions are interpreted as the limit for $\gamma \rightarrow 0$, yielding

$$\begin{aligned} c_{11} &= \left(\frac{1}{\log(2)} - \frac{1}{\log(3)} \right) \tilde{c}_{0,1}, & c_{12} &= -\frac{2}{\log(2)} \tilde{c}_{0,1}, & c_{13} &= \frac{3}{\log(3)} \tilde{c}_{0,1}, \\ c_{21} &= -\frac{1}{\log(2)} + c_{11} \tilde{c}_{0,3}, & c_{22} &= \frac{2}{\log(2)} + c_{12} \tilde{c}_{0,3}, & c_{23} &= c_{1,3} \tilde{c}_{0,3}, \\ c_{31} &= 1 + \Gamma''(1) c_{11} + \Gamma'(1) c_{21}, & c_{32} &= \Gamma''(1) c_{12} + \Gamma'(1) c_{22}, & c_{33} &= \Gamma''(1) c_{13} + \Gamma'(1) c_{23}, \end{aligned}$$

where

$$\tilde{c}_{0,1} = \left\{ \frac{\log(3)}{2} - \frac{\log(2)}{2} \right\}^{-1}, \quad \tilde{c}_{0,3} = \frac{\log(2)}{2} + \Gamma'(1).$$

PROOF. This follows from straightforward calculations. \square

APPENDIX D: A FINITE SAMPLE COMPARISON OF SLIDING VS. DISJOINT BLOCK MAXIMA SAMPLES

Lemma 2.4 shows that, under suitable natural assumptions, each sliding block maximum is asymptotically GEV-distributed with the ‘correct’ asymptotic parameters. While this is

natural under sampling scheme (S1), it is rather surprising under (S2). Throughout this section, we discuss the distributional approximation in finite-sample situations for the ARMAX-GPD-model, for which explicit calculations are possible. The results will provide theoretical justification for some statements from the main paper where it was argued (based on simulation evidence) that the bias induced by sampling scheme (S2) does not play a significant role in practical applications, where the temporal dependence is typically at most moderate and the block size is rather large ($r \geq 90$).

We start by calculating the c.d.f. of disjoint and sliding block maxima in the ARMAX-GPD model. Based on Formula (10.5) in [1], a straightforward calculation implies that a (disjoint) block maximum of size r has cdf F_r given by

$$\begin{aligned} F_r(x) &= \mathbb{P}\left(\max_{i=1}^r Y_i \leq \frac{1}{-\log F_\gamma(x)}\right) \\ &= \exp\left\{-\left(1 + (1-b)(r-1)\right)(-\log F_\gamma(x))\right\} \\ &= F_\gamma(x)^{b+(1-b)r}, \end{aligned}$$

with F_γ the c.d.f. of the GPD(0,1, γ)-distribution (recall that the extremal index is given by $\theta = 1 - b$). Under sampling scheme (S2), the distribution of a sliding block maximum made up from j observations in one season and $r - j$ observations in another ($j \in \{1, \dots, r-1\}$) has cdf

$$F_{r,j}(x) = F_j(x)F_{r-j}(x) = F_\gamma(x)^{2b+(1-b)r}.$$

Remarkably, the distribution does not depend on j (which is not the case in general, but a specific feature of the ARMAX-dynamics). Moreover, since $F_{r,j}(x) = F_\gamma(x)^b \times F_r(x)$, the difference between $F_{r,j}$ and F_r is clearly (and unsurprisingly) increasing in b .

Now, a sliding block maxima sample based on m seasons of size r consists of exactly m observations drawn from F_r , and $n - r - m + 1$ observations from $F_{r,j}$. Hence, a randomly chosen sliding block maxima has cdf

$$F_{r,m}^* = \frac{m}{n-r+1}F_r + \frac{n-r-m+1}{n-r+1}F_{r,j} = \left\{\frac{1}{r}F_r + \left(1 - \frac{1}{r}\right)F_{r,j}\right\} + O(m^{-1})$$

for $n \rightarrow \infty$, provided that $r = o(n)$. Subsequently, the c.d.f. in curly brackets on the right-hand side will be denoted as F_r^* . In view of the fact that the return level $\text{RL}(T, r)$ (as a central target quantity) is a function of F_r (and not of F_r^*), one may indeed expect less biased estimates when applying the disjoint block maxima method (note that all disjoint block maxima are drawn from F_r).

In the following, we carry out a graphical comparison between the sliding block maxima distribution F_r^* and the disjoint block maxima distribution F_r , for various values of γ , b and r . Note that the quantiles of F_r^* must be calculated numerically, which, however, is straightforward. Respective qq-plots for quantile levels $\{0.01, 0.02, \dots, 0.98, 0.99\}$ are depicted in Figure D.1. In Figure D.2, we depict, as a function of b , the relative difference

$$D_r = \frac{(F_r^*)^{-1}(0.99)}{F_r^{-1}(0.99)} - 1 = \frac{(F_r^*)^{-1}(0.99) - F_r^{-1}(0.99)}{F_r^{-1}(0.99)}$$

for block sizes $r \in \{90, 180, 360\}$. Note that $F_r^{-1}(0.99) = \text{RL}(100, r)$ in the paper's notation for the return level.

It can be seen that, for each fixed value of γ and $b = 1 - \theta$, the difference is decreasing in r , which illustrates the asymptotic validity of the sliding block maxima method under (S2), see in particular Lemma 2.4. At the same time, for fixed block size r (note that the block

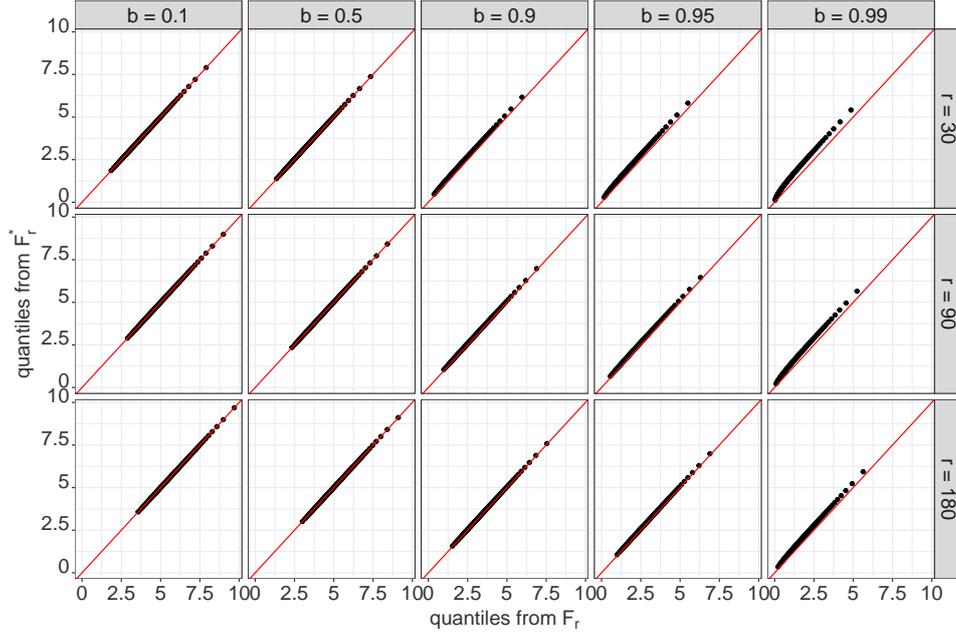


FIG D.1. *QQ-plots of the theoretical disjoint and sliding BM distributions in the ARMAX(b) model, with $b \in \{0.1, 0.5, 0.9, 0.95, 0.99\}$, blocksize $r \in \{30, 90, 180\}$ and marginal $GPD(0, 1, 0)$ distribution.*

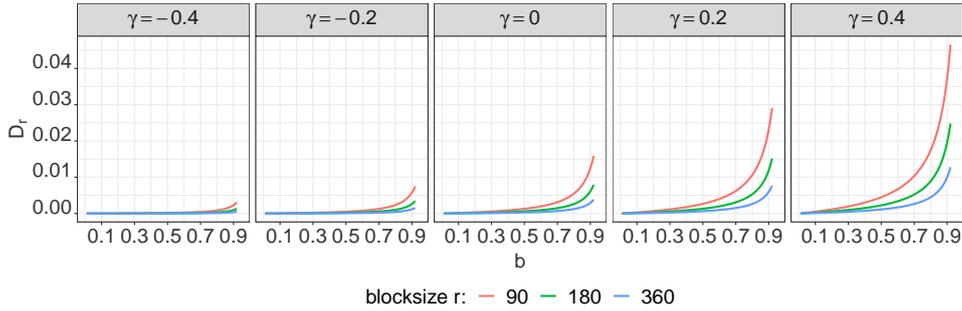


FIG D.2. *The relative difference D_r as a function of b , for block sizes $r \in \{90, 180, 360\}$ and shape parameters $\gamma \in \{0.4, -0.2, 0, 0.2, 0.4\}$.*

size is typically given in applications), it can be seen that larger values of b (i.e., smaller extremal indices) yield a larger discrepancy between the two quantiles, which will eventually show up as a bias when applying the sliding block maxima method for estimation. However, a significant deviation requires quite small extremal indices, in particular for non-positive shape parameters (for which we advocate the use of the PWM estimator). Further note that the extremal index is typically around 45-50% for daily temperatures and above 80-90% for cumulative daily precipitation, so we hardly get into the critical range in practice when using $r = 90$.

Furthermore, it is worthwhile to mention that the relative difference D_r is not invariant with respect to affine linear transformations of the underlying distributions. Indeed, if the initial $GPD(0, 1, \gamma)$ distribution (with associated c.d.f.s F_r and F_r^*) is replaced with a $GPD(\mu, \sigma, \gamma)$ distribution, which corresponds to the linear transformation $y \mapsto \sigma y + \mu$, this

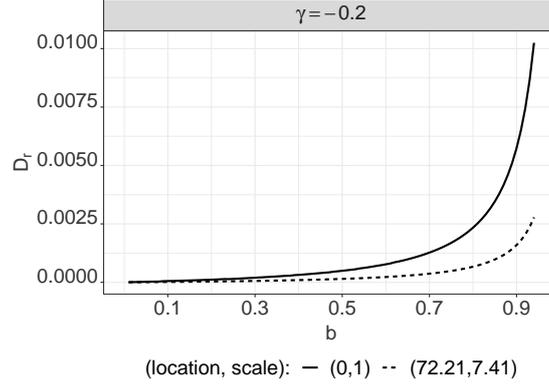


FIG D.3. The relative difference D_r as a function of b , for blocksize $r = 90$ and initial marginal distribution $F_{(\mu, \sigma, \gamma)} \in \{\text{GPD}(0, 1, -0.2), \text{GPD}(72.21, 7.41, -0.2)\}$.

transformation propagates to the quantile function, with the relative difference becoming

$$D_r = \frac{\sigma(F_r^*)^{-1}(0.99) + \mu}{\sigma F_r^{-1}(0.99) + \mu} - 1 = \frac{(F_r^*)^{-1}(0.99) - F_r^{-1}(0.99)}{F_r^{-1}(0.99) + \mu/\sigma}.$$

A large ratio μ/σ of location and scale parameter therefore contributes to a small relative difference D_r . An example is illustrated in Figure D.3, where relative differences for $\text{GPD}(0, 1, -0.2)$ and $\text{GPD}(72.21, 7.41, -0.2)$ distributions are depicted, both for blocksize $r = 90$ and again as a function of b . The explicit values were chosen following the simulations for temperature data as described in Section 4.3 (without innerseasonal non-stationarities) of the main paper; they correspond to realistic values for temperature extremes measured in Fahrenheit. It can be seen that the relative difference is negligible even for high level of serial dependence; for instance, the difference is smaller than 0.2% for $\theta = 0.9$.

APPENDIX E: ADDITIONAL SIMULATION RESULTS

E.1. Additional results for fixed block size. This section contains additional results comparable to those in Figures 3 and 4 in a situation where $r = 90$ is fixed. For each time series model, we show results for both shape estimation and $\text{RL}(T, 90)$ estimation under sampling schemes (S1) and (S2). Results for the AR-GPD-model are shown in Figure E.1 and Figure E.2, those for the CAR-GPD-model are shown in Figure E.3 and Figure E.4, and those for the ARMAX-GPD-model can be found in Figure E.5 and Figure E.6.

Remarkably, under sampling scheme (S1) and for almost all dependence structures, it is only the $\gamma = 0.4$ case for which the sliding blocks version does not provide an improvement over the disjoint blocks counterpart. Under sampling scheme (S2), this observation does not hold universally anymore, but depends on the combination of shape parameter and magnitude of serial dependence. For strong serial dependence (here: extremal index smaller than or equal to 0.25) and non-negatives shapes, there is no improvement of the sliding over the disjoint version (in particular for return level estimation), while it can be observed in all scenarios with either moderate serial dependence or negative shapes. This issue is discussed further in Section 4.5 from the main paper.

TABLE E.1
Population return levels $RL(T, 90)$ for $T = 50$ ($T = 100$)

Model	$\gamma = -0.4$	$\gamma = -0.2$	$\gamma = 0$	$\gamma = 0.2$	$\gamma = 0.4$
IID	2.41 (2.43)	4.07 (4.19)	8.40 (9.09)	21.85 (25.84)	69.18 (92.18)
AR 0.5	2.41 (2.43)	4.06 (4.18)	8.37 (9.06)	21.74 (25.77)	68.62 (92.01)
ARMAX 0.5	2.39 (2.41)	3.93 (4.07)	7.72 (8.43)	18.41 (21.94)	52.27 (69.99)
CAR 0.5	2.39 (2.41)	3.93 (4.07)	7.73 (8.43)	18.36 (21.86)	52.36 (70.27)

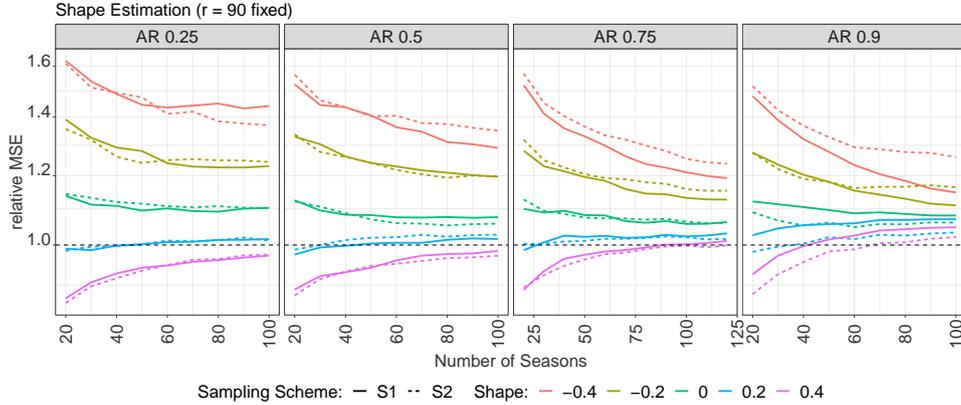


FIG E.1. Relative MSE (MSE of disjoint blocks estimator divided by MSE of sliding blocks estimator) of shape estimation in a transformed AR(1) model with GPD-margins under sampling scheme (S1) (straight line) and (S2) (dashed line) for fixed block size $r = 90$.

E.2. Results for fixed sample size. In this experiment, instead of fixing the block length parameter r , we fixed the overall sample length and investigated the estimators' performance for varying values of r . We considered each combination of the time series models and the marginal distribution functions described in Section 4 for fixed sample length $n = 1000$ and sampling scheme (S1). The setting aims at evaluating the common bias-variance tradeoff in extreme value statistics, which becomes visible when treating the block length as a hyperparameter to be chosen by the statistician with the ultimate goal of maximizing the estimation accuracy (which is comparable to the choice of the number of upper order statistics in the peaks-over-threshold approach). Note that treating the blocksize as a hyperparameter is only valid for sampling scheme (S1) (it is given when considering sampling scheme (S2)) and for estimating the shape parameter (as return levels depend on the blocksize). For the experiment, the block length has been chosen as $r \in \{4, 5, 6, 7, 8, 9, 10, 12, 14, 16, 18, 20, 25, 30, 40\}$, yielding between 25 and 250 disjoint blocks. All estimators (disjoint, sliding, and sliding bias reduced) have been evaluated in terms of their empirical MSE, variance and squared bias based on $N = 1000$ simulation repetitions.

Results for fixed sample size $n = 1000$ for shape estimation under sampling scheme (S1) can be found in Figure E.7 for the AR-GPD-model, in Figure E.8 for the CAR-GPD-model, and in Figure E.9 for the ARMAX-GPD-model. In these figures, the x-axis corresponds to the effective sample size $\lfloor n/r \rfloor$, i.e., the number of disjoint blocks. The general shape of the curves is mostly (with the exception of $\gamma = 0.4$) as follows: we observe a decreasing variance curve that is universally smaller for the sliding blocks method (as expected from the theoretic results) and an (eventually) increasing bias curve that is mostly comparable between the two methods. As a result, the MSE curve is mostly u-shaped, representing the typical bias-variance tradeoff. The improvement of the sliding blocks method over the disjoint blocks method is largest for negative shape parameters, while no significant improvement is visible for positive shape parameters. The time series model does not have a significant effect

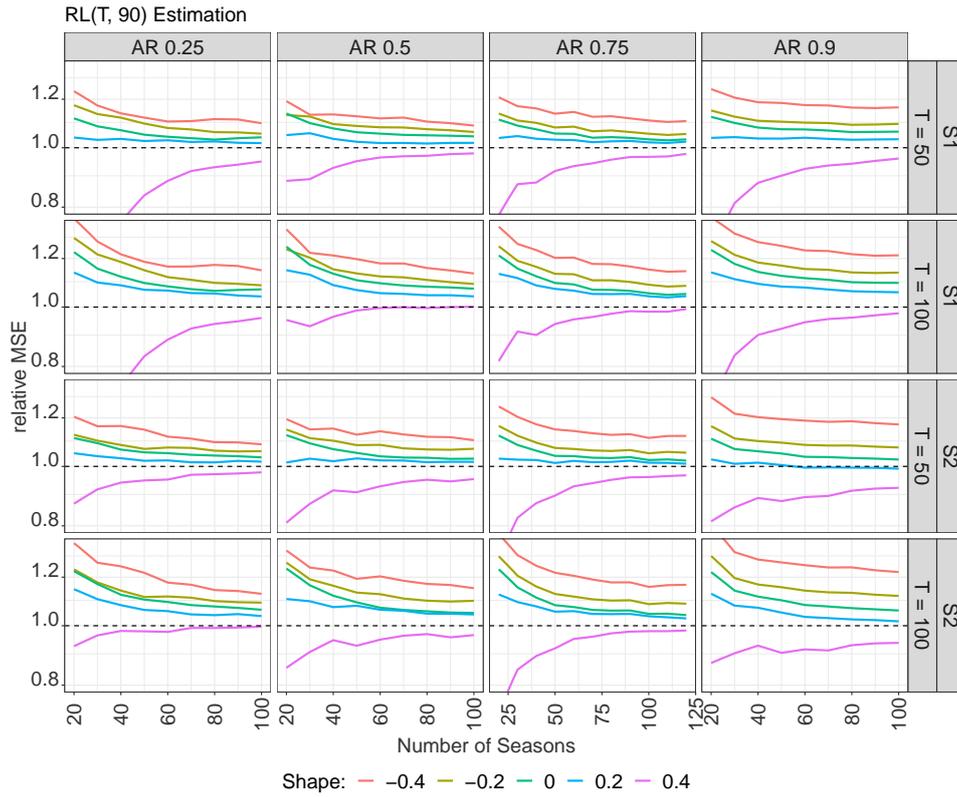


FIG E.2. *Relative MSE (MSE of disjoint blocks estimator divided by MSE of sliding blocks estimator) of RL(T,r)-estimation where $T = 50$ (rows 1 and 3) or $T = 100$ (rows 2 and 4) in a transformed AR(1) model with GPD-margins under sampling schemes (S1) (top two rows) and (S2) (bottom two rows) for fixed block size $r = 90$.*

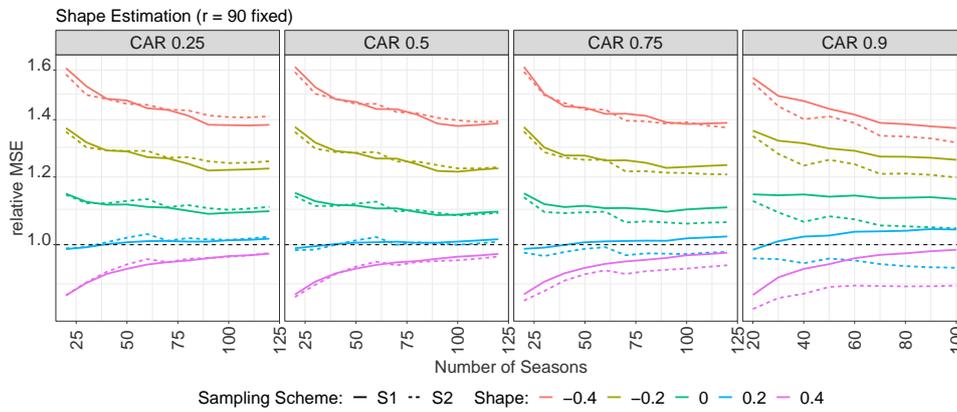


FIG E.3. *Relative MSE (MSE of disjoint blocks estimator divided by MSE of sliding blocks estimator) of shape estimation in CAR-GPD-models under sampling scheme (S1) (straight line) and (S2) (dashed line) for fixed block size $r = 90$.*

on the qualitative performance, despite that a remarkably strong bias is visible for strong temporal dependence, small block sizes and negative shape parameters, for both disjoint and sliding blocks estimators. Overall, for small effective sample sizes (i.e., large block sizes), we

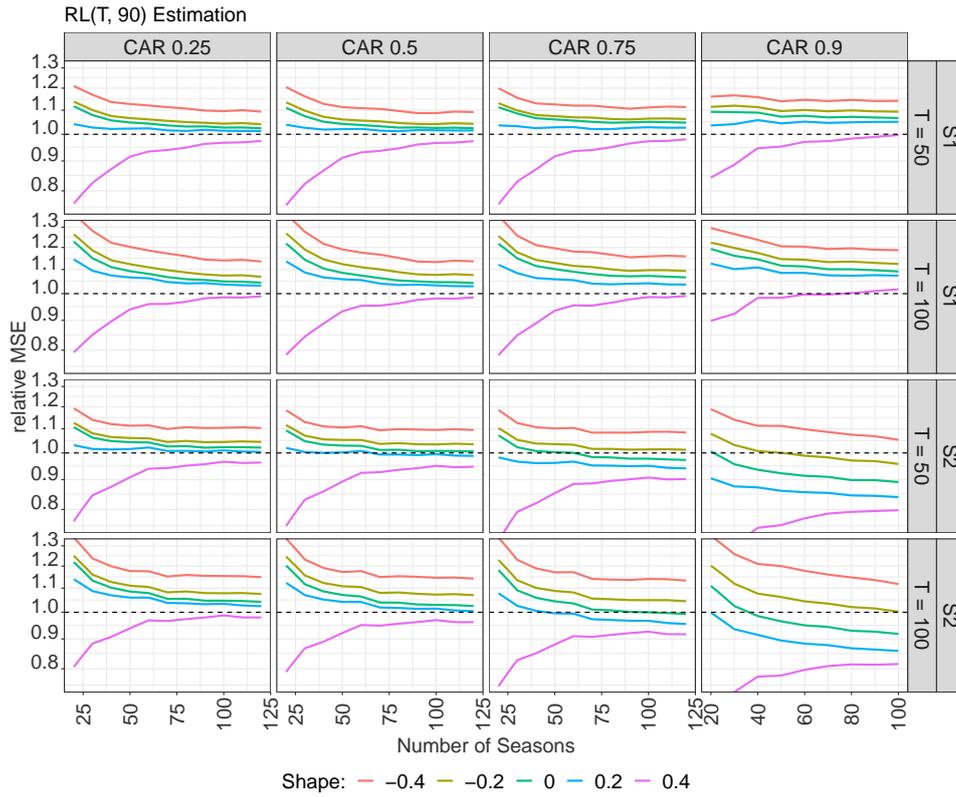


FIG E.4. Relative MSE (MSE of disjoint blocks estimator divided by MSE of sliding blocks estimator) of $RL(T, r)$ -estimation where $T = 50$ (rows 1 and 3) or $T = 100$ (rows 2 and 4) in CAR-GPD-models under sampling schemes (S1) (top two rows) and (S2) (bottom two rows) for fixed block size $r = 90$.

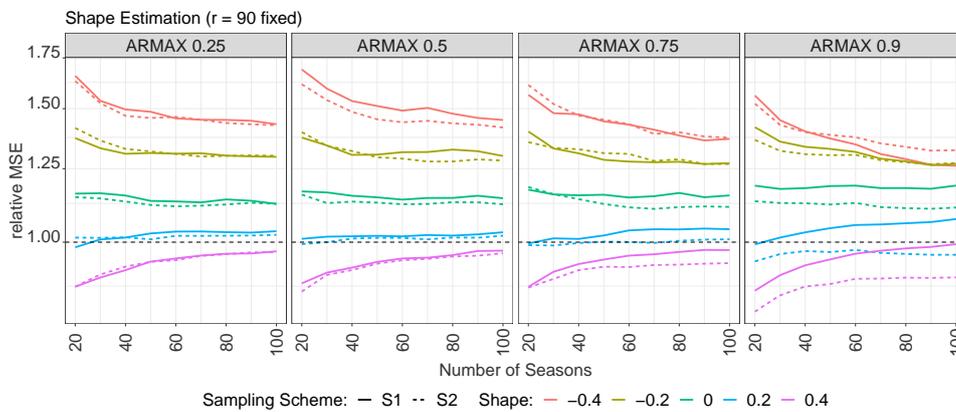


FIG E.5. Relative MSE (MSE of disjoint blocks estimator divided by MSE of sliding blocks estimator) of shape estimation in ARMAX-GPD-models under sampling scheme (S1) (straight line) and (S2) (dashed line) for fixed block size $r = 90$.

sometimes observe a significantly higher bias for the sliding blocks method, which may be explained by the dependency bias discussed in Remark 3.1; see also the results in Section E.3 for further discussions.

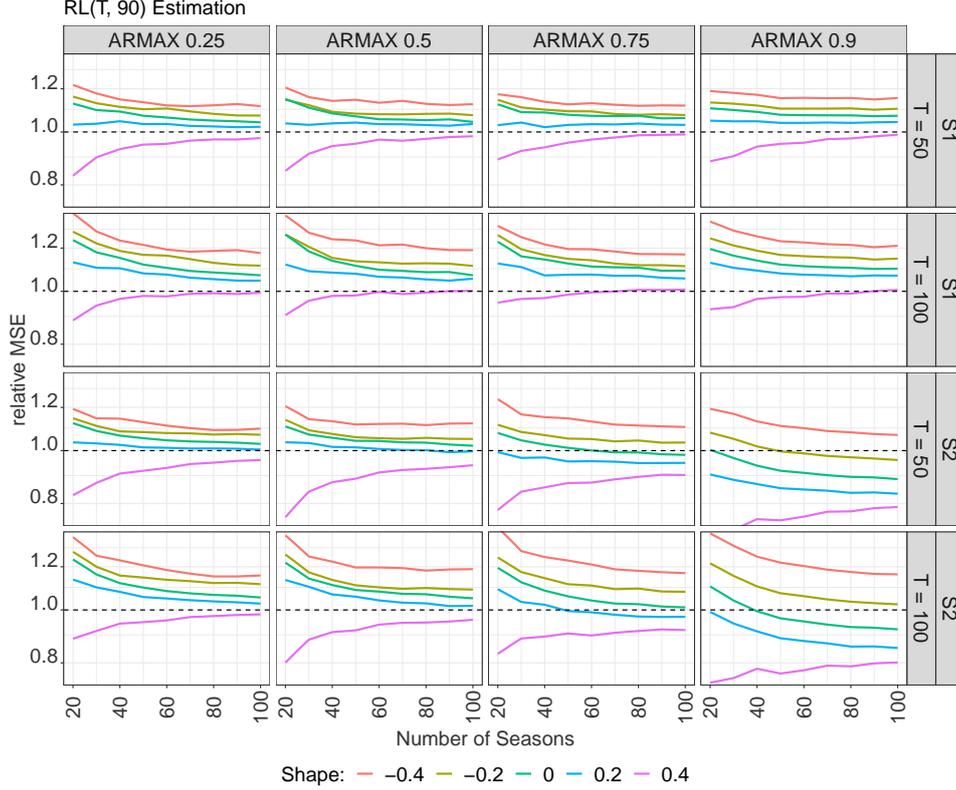


FIG E.6. *Relative MSE (MSE of disjoint blocks estimator divided by MSE of sliding blocks estimator) of RL(T, r)-estimation where $T = 50$ (rows 1 and 3) or $T = 100$ (rows 2 and 4) in ARMAX-GPD-models under sampling schemes (S1) (top two rows) and (S2) (bottom two rows) for fixed block size $r = 90$.*

As discussed before, the blocksize r should not be considered a hyperparameter when dealing with sampling scheme (S2). Nevertheless, we show the Bias, MSE and variance of shape parameter estimation for fixed sample size $n = 1000$ under sampling scheme (S2) for a selection of time series models in Figure E.10. The results are overall very similar to the respective scenarios in sampling scheme (S1). However, note that the curves cannot be interpreted as bias-variance-tradeoff induced by the choice of r in this case.

E.3. Results for comparing the plain and bias-reduced sliding blocks estimator. As mentioned in the main paper, the bias-reduced sliding blocks estimator from Remark 3.1 is computationally costly for situations involving overall sample sizes of $n = 9000$ or larger. Therefore, when comparing results for fixed blocksize $r = 90$, we restrict attention to sampling scheme (S2) and a selection of 20 models that are made up of 4 different time series models (i.i.d., AR 0.5, CAR 0.5 and ARMAX 0.5) and the 5 different GPD-margins (GPD(γ) with $\gamma \in \{-0.4, -0.2, 0, 0.2, 0.4\}$). The bias and MSE of shape estimation as obtained for the disjoint, sliding and bias-reduced sliding blocks methods are shown in Figures E.11 and E.12, respectively. The bias of the bias-reduced sliding blocks estimator matches the bias of the disjoint blocks estimator almost perfectly. For positive shape parameters that results in equal performance in terms of MSE (with a tiny advantage for the sliding version and small sample sizes) for those two estimators. For negative shapes, the plain sliding version still has the smallest MSE, which can be explained by its smaller variance.

For fixed sample size, the sliding blocks estimator is compared with its bias-reduced version in Figure E.13 (shape estimation, AR-GPD-model, sampling scheme (S1), $n = 1000$),

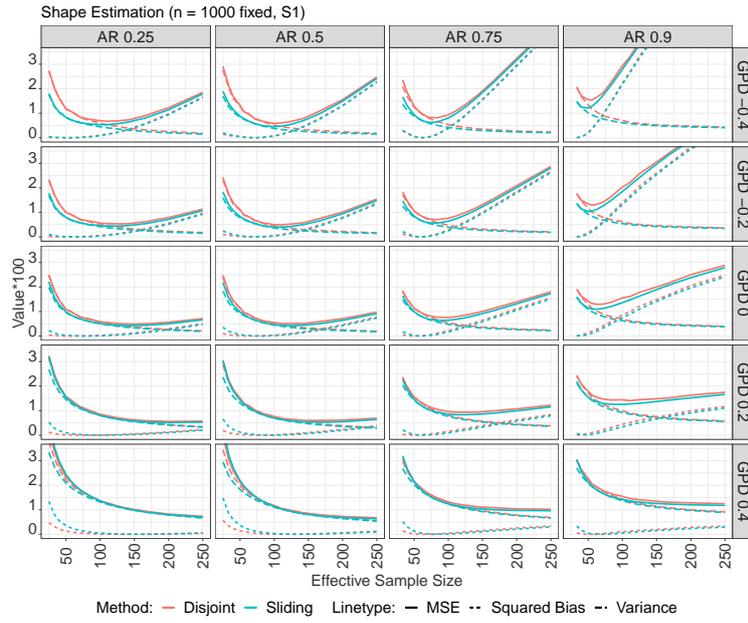


FIG E.7. MSE, squared bias and variance as a function of the effective sample size for the estimation of the shape parameter γ in a transformed AR(1) model with GPD-margins under sampling scheme (S1) for fixed sample size $n = 1000$.

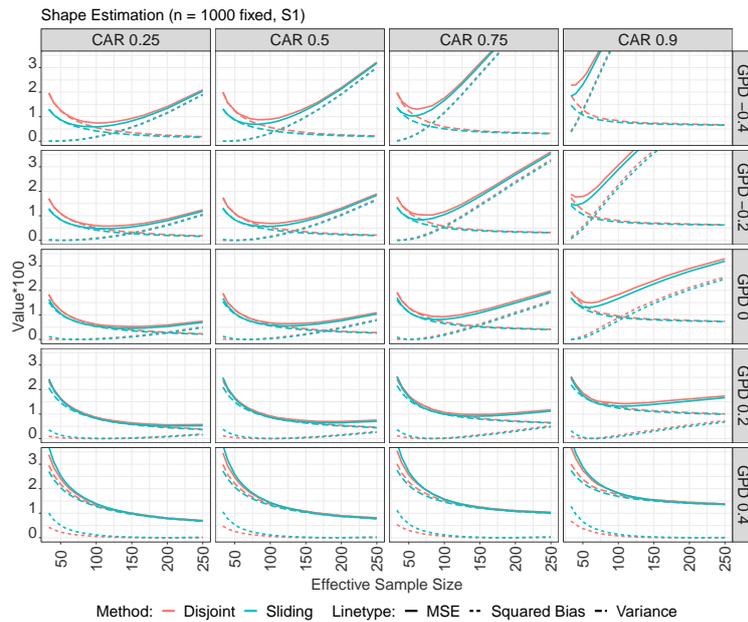


FIG E.8. MSE, squared bias and variance as a function of the effective sample size for the estimation of the shape parameter γ in a transformed Cauchy AR(1) model with GPD-margins under sampling scheme (S1) for fixed sample size $n = 1000$.

Figure E.14 (the same for the CAR-GPD-model) and Figure E.15 (the same for the ARMAX-GPD-model). Considering only the squared bias, it can be seen that the bias-reduced version may outperform its counterpart for small block sizes, in particular in scenarios involving non-

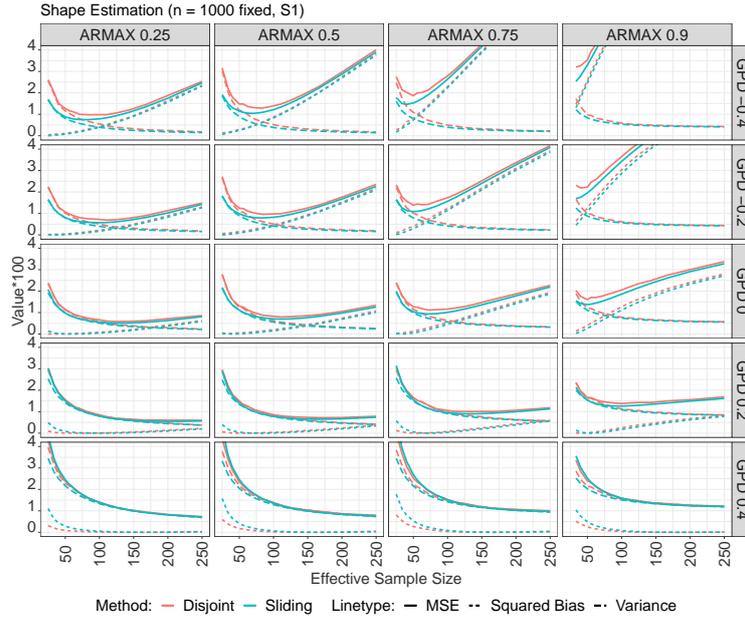


FIG E.9. *MSE, squared bias and variance as a function of the effective sample size for the estimation of the shape parameter γ in a transformed ARMAX(1) model with GPD-margins under sampling scheme (S1) for fixed sample size $n = 1000$.*

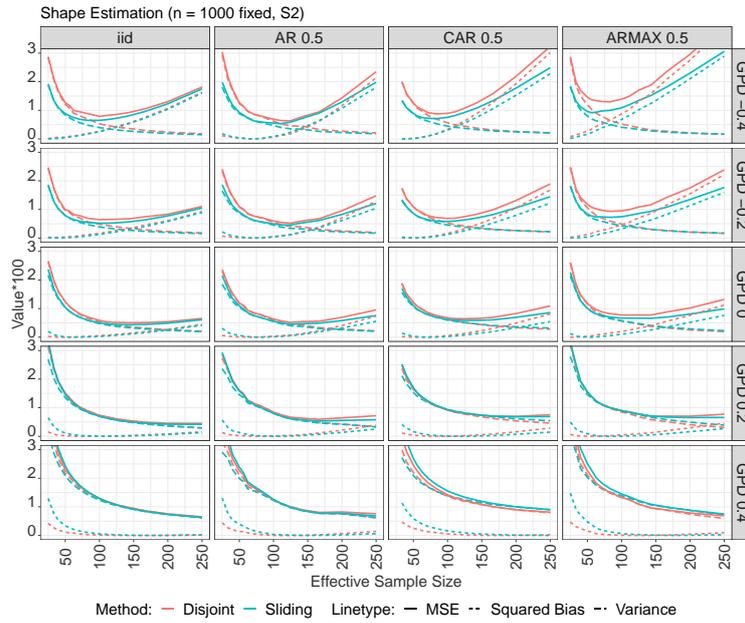


FIG E.10. *MSE, squared bias and variance as a function of the effective sample size for the estimation of the shape parameter γ in a selection of time series models with GPD-margins under sampling scheme (S2) for fixed sample size $n = 1000$.*

negative shapes and positive AR parameters. However, under strong temporal dependence, the bias-reduced estimator may also exhibit a uniformly larger squared bias. In terms of variance, the plain estimator mostly has a slight edge. Summarizing the findings is rather difficult,

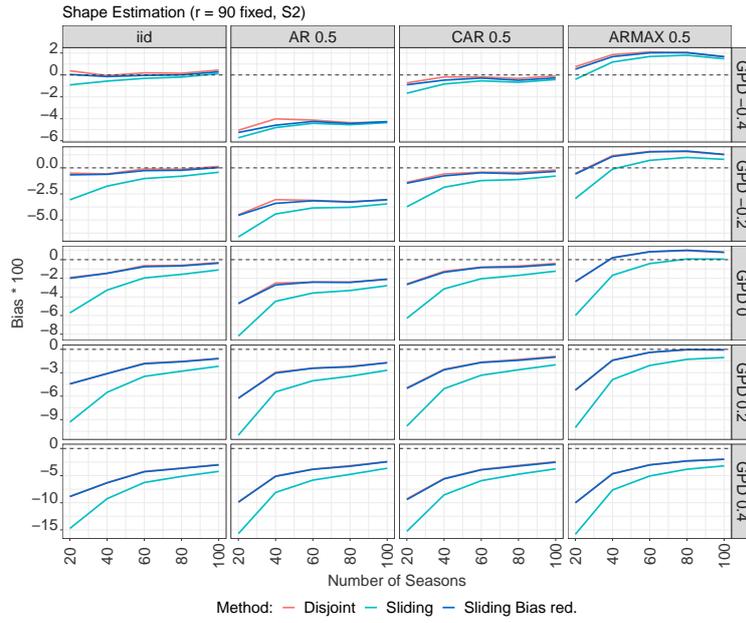


FIG E.11. Bias of shape estimation in a selection of transformed time series models with GPD margins for fixed $r = 90$ and growing number of seasons under sampling scheme (S2).

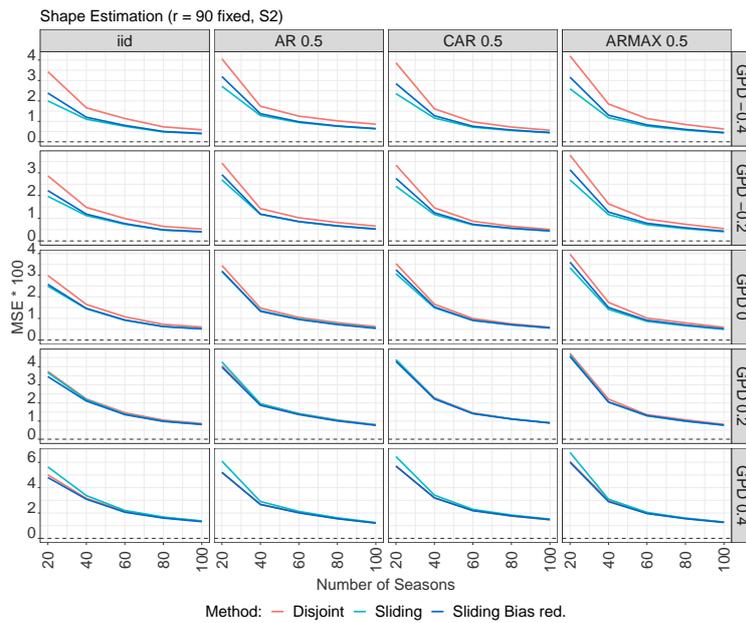


FIG E.12. MSE of shape estimation in a selection of transformed time series models with GPD margins for fixed $r = 90$ and growing number of seasons under sampling scheme (S2).

whence we tend to recommend the use of the plain version merely for computational reasons (in particular for non-negative shapes).

E.4. Results for comparing sampling schemes (S1) and (S2) for fixed sample size n . Results for the comparison of sampling schemes (S1) and (S2) in situations of fixed sample

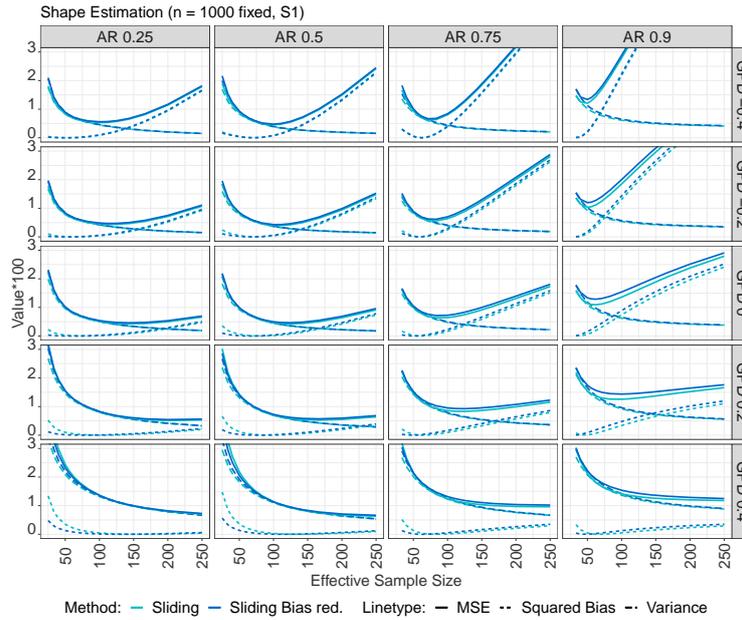


FIG E.13. Comparison of MSE, squared Bias and Variance of shape estimation as a function of the effective sample size for the plain and bias reduced sliding blocks estimators in the AR-GPD-models under sampling scheme (S1), for fixed sample size $n = 1000$.

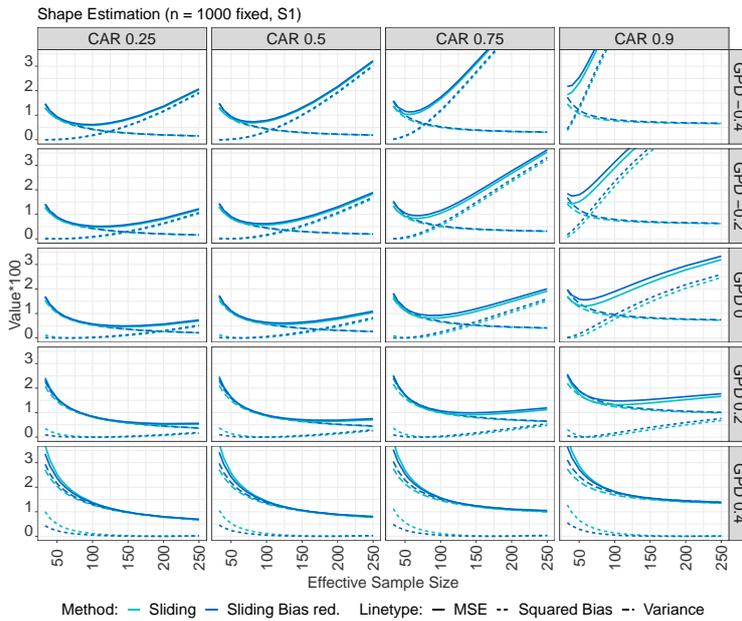


FIG E.14. Comparison of MSE, squared Bias and Variance of shape estimation as a function of the effective sample size for the plain and bias reduced sliding blocks estimators in the CAR-GPD-models under sampling scheme (S1), for fixed sample size $n = 1000$.

size $n = 1000$ can be found in Figure E.16 (shape estimation within the AR-GPD-model), Figure E.17 (shape estimation within the CAR-GPD-model) and Figure E.18 (shape estimation within the ARMAX-GPD-model). In most cases the behavior between the two sampling

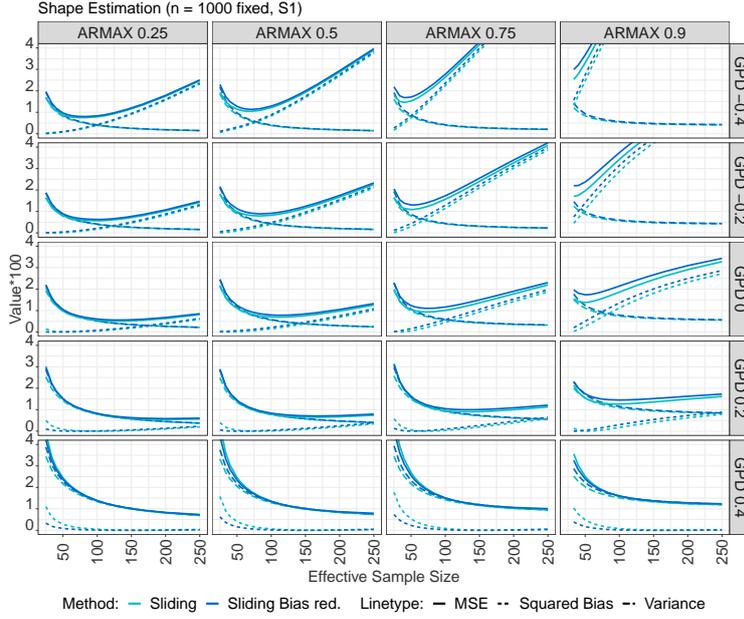


FIG E.15. Comparison of MSE, squared Bias and Variance of shape estimation as a function of the effective sample size for the plain and bias reduced sliding blocks estimators in the ARMAX-GPD-models under sampling scheme (S1), for fixed sample size $n = 1000$.

schemes is similar, as was to be expected from the theoretic results. A notable exception concerns high level of serial dependence, non-positive shape parameters and small block sizes, where the MSE for the sliding blocks estimator is smaller in scenario (S2) than in (S1). This difference may be explained by the fact that, heuristically, the non-constancy of $j \mapsto H_{r,j}$ (see Condition 3.2) is increasing in the strength of serial dependence and decreasing in the block size. As a consequence, the bias $B_{n,k}^{(mb,S)}$ in Condition 3.2 shows a similar behavior, eventually impacting the MSE in the observed way.

E.5. Additional results for different marginal distributions. Block maxima obtained from an i.i.d. GPD sample are known to converge comparably fast to their limiting GEV distribution. The speed of convergence may be measured with the second order parameter, say $\rho = \rho_{BM}$, which takes its values in $[-\infty, 0]$. The smaller ρ is, the higher is the speed of convergence. For the GPD distribution, we have $\rho = -1$, see Section 2 in [5]. Slower convergence is thus obtained if the second order parameter is larger than -1 , whence we chose to (partially) repeat our simulation study for distributions such that $\rho = -1/2$.

More precisely, for positive γ , we chose to consider a member from the Hall-and-Welsh (HW) distribution family, defined by its cumulative distribution function $F_\gamma(x) = 1 - x^{-1/\gamma}(1 + x^{-1/(2\gamma)})/2, x \geq 1$. It can be shown that this distribution is in the maximum domain of attraction of G_γ with second order parameter $\rho = -1/2$ (see Table 1 in [5]). For negative γ , we chose the distribution of the random variable $-1/Z$ where $Z \sim F_{|\gamma|}$, whose second order parameter is again $-1/2$ (model RHW in Table 3 in [5]). Finally, for $\gamma = 0$, we chose to consider the distribution defined by its inverse $F^{-1}(p) = \log\{1/(1-p)\} \times \{1 + (1-p)^{1/2}\}$. Further, in order to avoid division by values close to zero when evaluating the (relative) performance of return level estimators, all distributions were shifted by adding 1 to the simulated values.

Simulations were carried out for all dependence structures of Section 4 in the main paper and marginal distributions as described above, with $\gamma \in \{-0.4, -0.2, 0, 0.2, 0.4\}$. Again, the

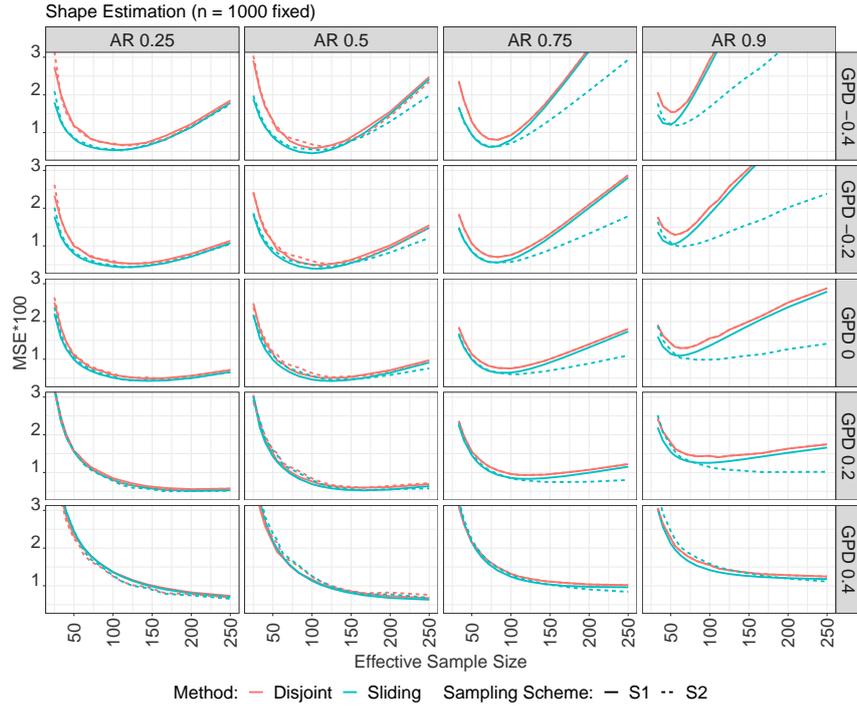


FIG E.16. *MSE of shape estimation for observations from sampling scheme (S1) and (S2) based on a transformed AR(1) model with GPD-margins for fixed sample size $n = 1000$.*

quantile transformation method was applied for sampling from the respective models. For the ease of presentation and because findings were similar, we restrict attention to models with medium (AR 0.5, CAR 0.5, ARMAX 0.5) or no (i.i.d.) dependence under sampling scheme (S2).

The resulting MSE curves are shown in Figure E.19 (shape estimation) and Figure E.20 (RL(100, 90) estimation). Regarding the latter, the MSE is computed from $\{\widehat{\text{RL}}(T, r) - \text{RL}(T, r)\} / \text{RL}(T, r)$, with $T = 100$, $r = 90$ and $\text{RL}(T, r)$ computed from a preliminary simulation involving $N = 10^6$ blockmaxima of independent blocks of size r .

For estimation of the shape, the MSE curves for the two second order parameters are nearly identical, while some differences are visible for return level estimation. For the latter however, a direct comparison is not quite sensible, as the true values deviate from each other. Overall, the qualitative behaviour is not significantly influenced by the second order parameter, in particular when comparing disjoint and sliding blocks.

E.6. Results for comparing Maximum Likelihood and PWM Estimation. A (reduced) simulation study was performed to compare the PWM estimator to its most popular competitor, the (pseudo) Maximum Likelihood estimator. Attention was restricted to 20 selected models that are made up from 4 different time series models (i.i.d., AR 0.5, CAR 0.5 and ARMAX 0.5) and the 5 different GPD-margins (GPD(γ) with $\gamma \in \{-0.4, -0.2, 0, 0.2, 0.4\}$). The sliding blocks maximum likelihood estimator was obtained by maximizing the likelihood function that results from treating the sliding blocks as independent, see [4] for respective theoretical results in the heavy tailed case.

The respective results for the estimation of γ and RL(100, 90) are summarized in Figure E.21 and Figure E.22. Both figures are slightly manipulated in favor of the maximum likelihood estimator as all presented results are conditional on the event that $|\hat{\gamma}_{\text{ML}} - \gamma| \leq 1$.

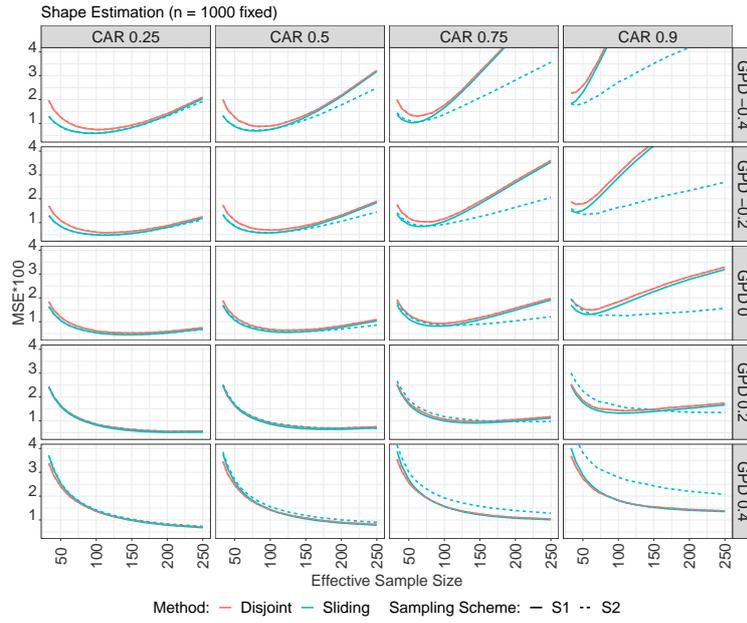


FIG E.17. *MSE of shape estimation for observations from sampling scheme (S1) and (S2) based on a transformed Cauchy AR(1) model with GPD-margins for fixed sample size $n = 1000$.*

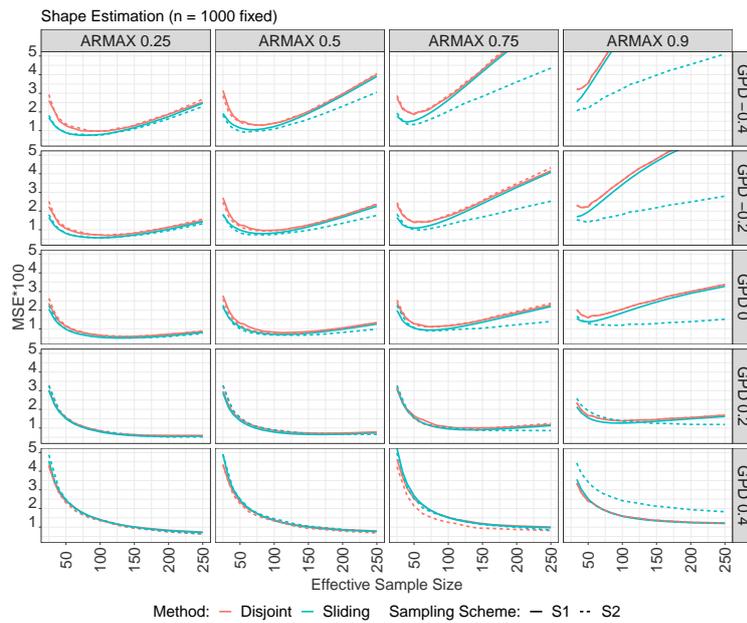


FIG E.18. *MSE of shape estimation for observations from sampling scheme (S1) and (S2) based on a transformed ARMAX(1) model with GPD-margins for fixed sample size $n = 1000$.*

The latter happens to be the case in approx. 95% of the simulation runs for $m = 10$ and in up to 99.5% for $m \geq 20$; not omitting the remaining (unrealistic) cases yields quite unstable curves for the ML estimator when $|\gamma| = 0.4$.

The results reveal that the PWM estimator has a tendency to be superior for small sample sizes while the maximum likelihood estimator is superior for large sample sizes; to the best

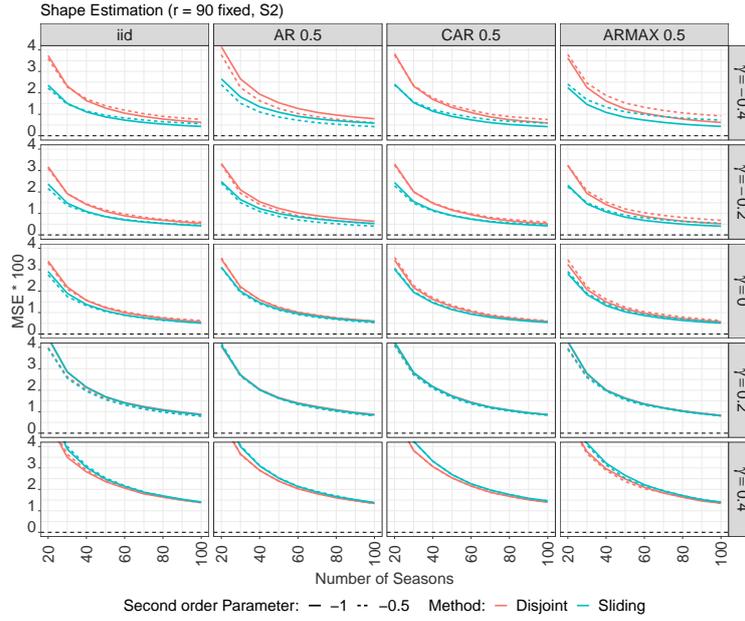


FIG E.19. *MSE of shape estimation as a function of the number of seasons (blocksize $r = 90$ is fixed) under sampling scheme (S2) and a selection of different dependence structures. The marginal distributions are attracted to G_γ , where γ varies across rows and the second order parameter is either -1 (solid line) or -0.5 (dashed line).*

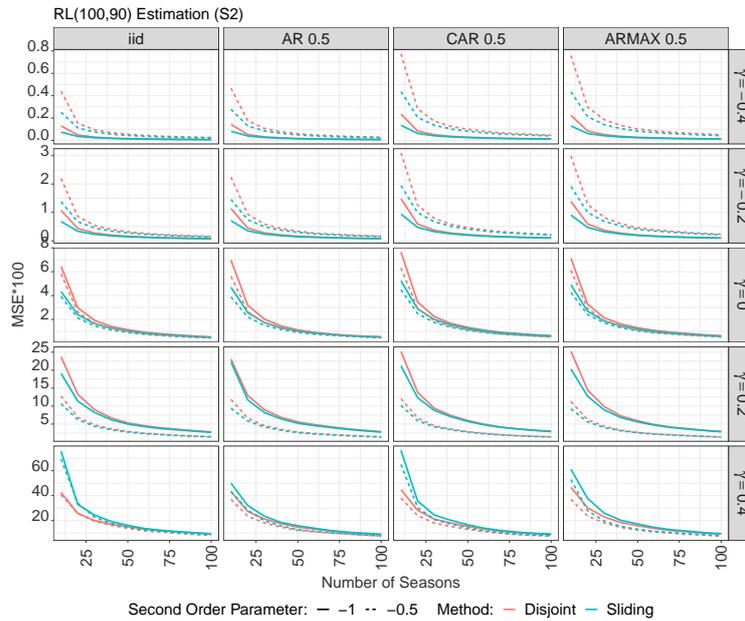


FIG E.20. *Rescaled MSE of RL(100,90) estimation under the same specifications as in Figure E.19.*

of our knowledge this is a usual view of the two estimators among applied statisticians. For shape estimation, smaller shapes yield better results for the PWM estimator, while for return level estimation, the picture is almost reversed. This seems to be an interesting aspect that could be confirmed in an extensive simulation study in future research.

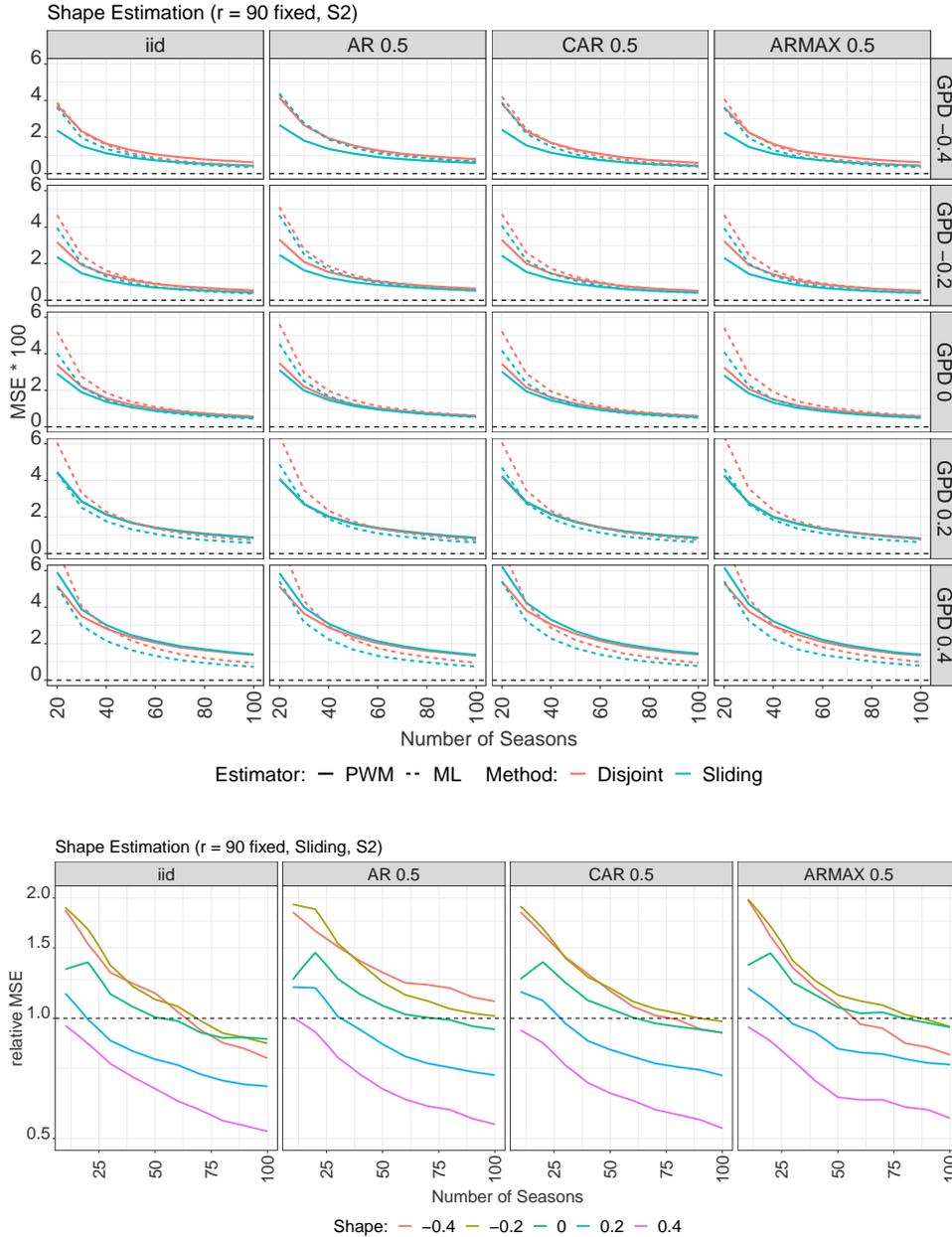


FIG E.21. Top: MSE obtained from PWM (solid line) and ML (dashed line) shape estimation based on disjoint and sliding blocks under sampling scheme (S_2) for fixed $r = 90$. Bottom: Relative MSE of sliding blocks shape estimation (MSE of ML estimation divided by MSE of PWM estimation) under sampling scheme (S_2), also for $r = 90$ fixed.

APPENDIX F: GEV-FIT EXAMINATION FOR THE CASE STUDY

To assess whether the fitted GEV distributions in the case study are plausible, we generated QQ-plots, which can be found in Figure F.1 and reveal a remarkably good fit.

Acknowledgements. This work has been supported by the integrated project “Climate Change and Extreme Events - ClimXtreme Module B - Statistics (subproject B3.3)” funded by the German Bundesministerium für Bildung und Forschung, which is gratefully acknowl-

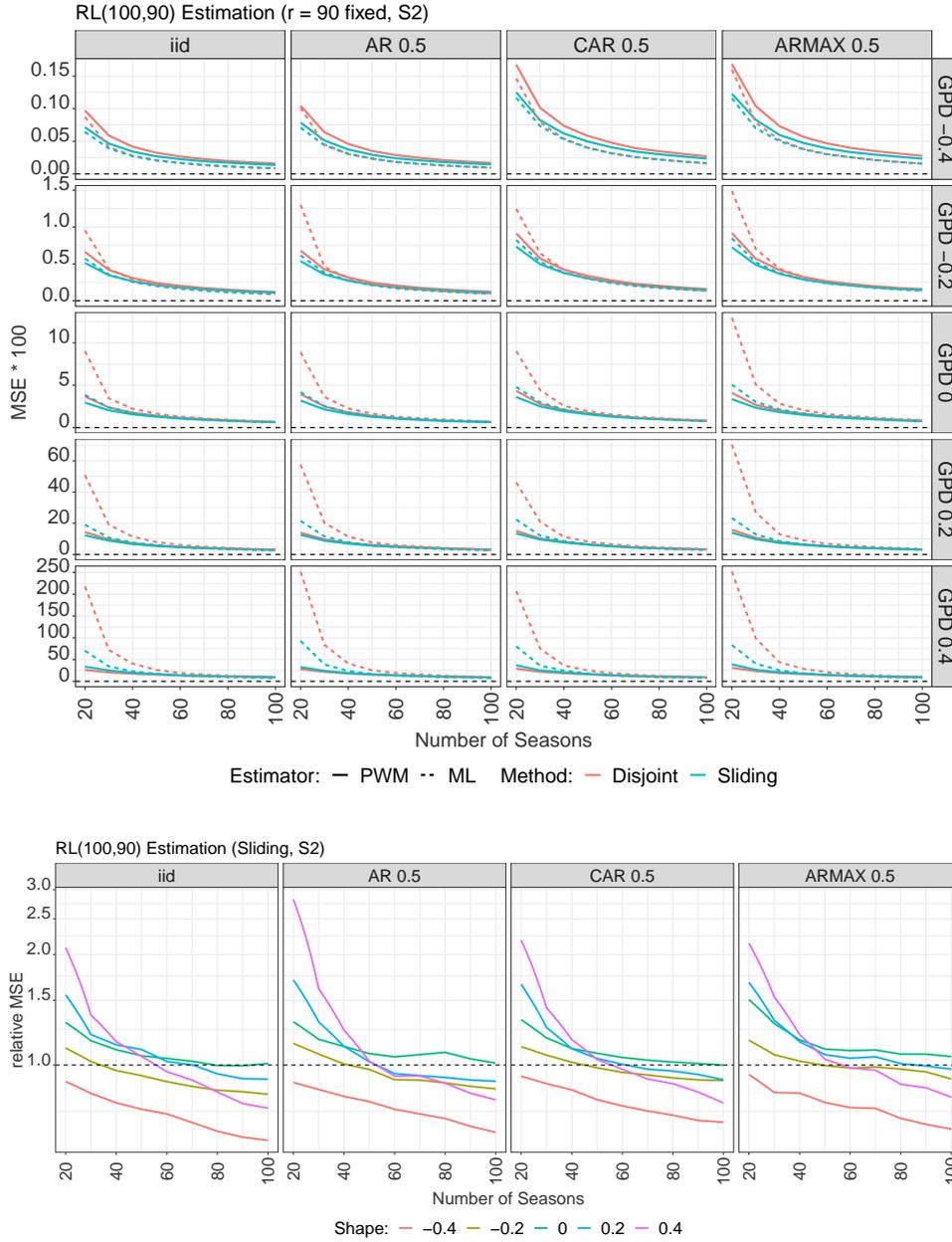


FIG E.22. Same as Figure E.21, but for the estimation of RL(100, 90).

edged. Computational infrastructure and support were provided by the Centre for Information and Media Technology at Heinrich Heine University Düsseldorf.

The authors are grateful to three unknown referees and an associate editor for their constructive comments that helped to improve the presentation substantially. The authors also appreciate valuable comments by Petra Friederichs, Marco Oesting and multiple participants of the Extreme Value Analysis (EVA) Conference 2021 in Edinburgh.

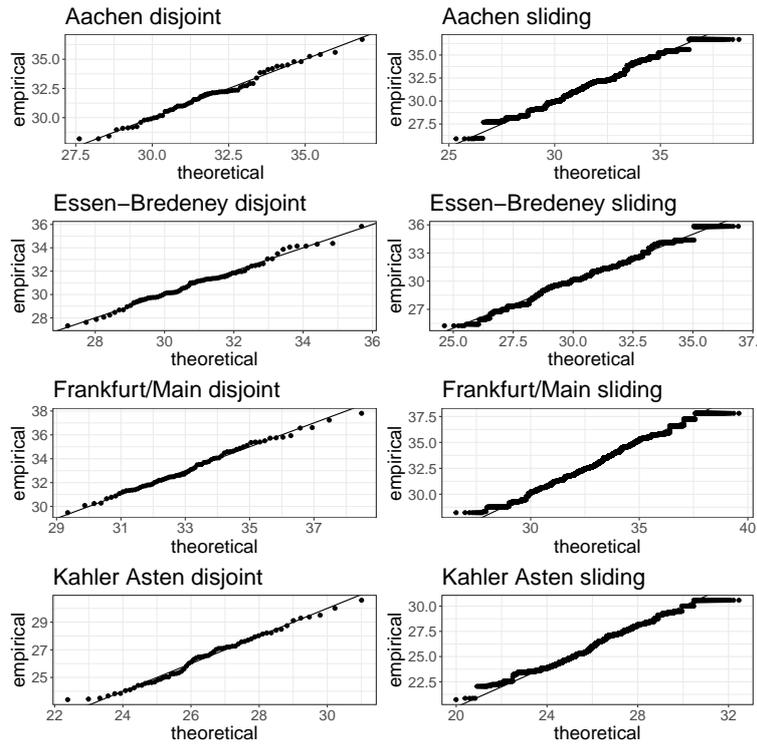


FIG F.1. *QQ-Plots for the fitted models based on disjoint BM (first column) and sliding BM (second column).*

REFERENCES

- [1] BEIRLANT, J., GOEGEBEUR, Y., SEGERS, J. and TEUGELS, J. (2004). *Statistics of extremes: Theory and Applications*. *Wiley Series in Probability and Statistics*. John Wiley & Sons Ltd., Chichester.
- [2] BÜCHER, A. and SEGERS, J. (2014). Extreme value copula estimation based on block maxima of a multivariate stationary time series. *Extremes* **17** 495–528.
- [3] BÜCHER, A. and SEGERS, J. (2018). Maximum likelihood estimation for the Fréchet distribution based on block maxima extracted from a time series. *Bernoulli* **24** 1427–1462.
- [4] BÜCHER, A. and SEGERS, J. (2018). Inference for heavy tailed stationary time series based on sliding blocks. *Electron. J. Statist.* **12** 1098–1125.
- [5] BÜCHER, A. and ZHOU, C. (2021). A Horse Race between the Block Maxima Method and the Peak-over-Threshold Approach. *Statistical Science* **36** 360–378.
- [6] DEHLING, H. and PHILIPP, W. (2002). Empirical process techniques for dependent data. In *Empirical process techniques for dependent data* 3–113. Birkhäuser Boston, Boston, MA.
- [7] FERREIRA, A. and DE HAAN, L. (2015). On the block maxima method in extreme value theory: PWM estimators. *Ann. Statist.* **43** 276–298.
- [8] KOJADINOVIC, I. and NAVEAU, P. (2017). Detecting distributional changes in samples of independent block maxima using probability weighted moments. *Extremes* **20** 417–450.
- [9] LO, A. (2017). Functional generalizations of Hoeffding’s covariance lemma and a formula for Kendall’s tau. *Statistics & Probability Letters* **122** 218–226.
- [10] VILLANI, C. (2008). *Optimal transport: old and new* **338**. Springer Science & Business Media.
- [11] ZOU, N., VOLGUSHEV, S. and BÜCHER, A. (2021). Multiple block sizes and overlapping blocks for multivariate time series extremes. *Ann. Statist.* **49** 295–320.

A note on statistical tests for homogeneities in multivariate extreme value models for block maxima

Jona Lilienthal¹ | Leandra Zanger² | Axel Bücher²  | Roland Fried¹ 

¹Department of Statistics, TU Dortmund University, Dortmund, Germany

²Mathematisches Institut, Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany

Correspondence

Axel Bücher, Mathematisches Institut, Heinrich-Heine-Universität Düsseldorf, Universitätsstr. 1, 40225 Düsseldorf, Germany.

Email: axel.buecher@hhu.de

Funding information

Bundesministerium für Bildung und Forschung

Abstract

Mathematical theory suggests to model annual or seasonal maxima by the generalized extreme value distribution. In environmental applications like hydrology, record lengths are typically small, whence respective parameter estimators typically exhibit a large variance. The variance may be decreased by pooling observations from different sites or variables, but this requires to check the validity of the inherent homogeneity assumption. The present paper provides an overview of (partly new) respective asymptotic significance tests. It is found that the tests' levels are often violated in typical finite-sample situations, whence a parametric bootstrap approach based on max-stable process models is proposed to obtain more accurate critical values. As a side product, we present an overview of asymptotic results on a variety of common estimators for GEV parameters in a multisample situation of varying record lengths.

KEYWORDS

flood extremes, index-flood assumption, max-stable process, parametric bootstrap, probability weighted moment estimator, pseudo-maximum likelihood

1 | INTRODUCTION

Environmental extreme value analysis is concerned with assessing the risk of rare extreme events like flooding or heat waves, see Beirlant et al. (2004) or the recent special issue in *Extremes Journal* (Cooley & Naveau, 2021). One of the most common statistical approaches in the field is the celebrated block maxima method popularized by Gumbel (1958): based on, say, daily observations of one or more variables of interest collected at one or more locations, calculate the coordinate wise annual or seasonal maximum for further investigation. Mathematical theory then suggests to model such data by max-stable processes, as the latter are the only processes that can arise, after proper affine transformation, as the limit of maxima of independent and identically distributed random fields (Davison et al., 2012). We also refer the reader to the latter paper for a description of a variety of respective parametric models.

Fitting a multivariate max-stable (process) model is typically done in two steps: first, estimate each marginal distribution, which is known to necessarily be a three parametric generalized extreme value (GEV) distribution (see Section 2); for instance by maximum likelihood or some moment method. Second, for estimating the dependence, one typically transforms the margins to some common scale (often the Fréchet(1)-distribution) and then fits a max-stable model with prescribed margins to the transformed data, for instance by pairwise maximum likelihood (Padoan et al., 2010). In view of the fact that the data transformation often involves the estimates obtained from the first step, there is quite some risk of error propagation, showing the need for estimates as accurate as possible on the marginal models.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Environmetrics* published by John Wiley & Sons Ltd.

The latter however may be a delicate task, as environmental sample sizes are often quite small (say, $n \leq 100$ or even $n \leq 50$ for hydrological data or meteorological station or reanalysis data) and GEV parameter estimates are known to be volatile in such situations. It has therefore been advertised to incorporate suitable homogeneity constraints on the marginal model parameters, thereby eventually reducing estimation uncertainty. For instance, the popular Index Flood model (Dalrymple, 1960, see also the monograph Hosking and Wallis, 1997, and Section 4.1 below) builds upon the assumption that the distribution at each station is the same except for some local scale parameter. If such an assumption is met, one may ultimately estimate the marginal model at each specific site using observations from other homogeneous sites. A second example concerns model assumptions that link intensity and duration of extreme precipitation events to their frequency, see (Boukhelifa et al., 2018; Jurado et al., 2020; Koutsoyiannis et al., 1998), and Section 4.2 below.

Before relying on a homogeneity assumption on marginal GEV parameters, it is advisable to check the assumption's validity. Respective methods have been proposed particularly for the afore-mentioned index-flood assumption, see Fill & Stedinger (1995), Hosking & Wallis (1993), Lu & Stedinger (1992), Viglione et al. (2007), among others. Among those tests, the Hosking–Wallis (HW) procedure (Hosking & Wallis, 1993) based on L-moments has become a common standard. The method has been modified and extended in several directions: Chebana & Ouarda (2007), Chebana & Ouarda (2009), Šimková (2018) consider extensions to the multivariate case, while certain nonparametric versions can be found in Masetlot et al. (2017; Šimková, 2017). The validity of the afore-mentioned procedures has typically been checked by simulation experiments.

The present paper expands this line of research, and is concerned with formal statistical homogeneity tests derived from limit theory for suitable parameter estimates. Quite surprisingly, the respective asymptotic theory has never been collected and worked out in full detail. It is the first major contribution of this paper to work out asymptotic results on multiple sample versions of (1) the probability weighted moment estimator (Hosking et al., 1985), (2) truncated L-moment estimators (Hosking, 2007), and (3) (pseudo) maximum likelihood estimators (Bücher & Segers, 2017), alongside with consistency results on estimators for their asymptotic covariance matrices. As a second major contribution, the results are used to design and investigate asymptotic tests for the aforementioned homogeneity constraints. Within extensive simulation experiments concerning the index flood assumption in situations of moderate dimension, the tests were found to require unpleasantly high sample sizes to keep their nominal level. It is further illustrated that this nuisance may be avoided by relying on a suitable parametric bootstrap device which was found to yield accurate level approximations and decent power properties in typical small-sample situations. The results are illustrated in a case study involving flood extremes.

We would like to point out that Bayesian (MCMC) methods provide yet another method to obtain parameter estimates. The approach has been widely applied in environmental extremes, see Boukhelifa et al. (2018), Gaume et al. (2010), Nguyen et al. (2014), Renard et al. (2006, Viglione et al. (2013), among others, and yields the advantage of providing credibility intervals for the parameters when following the Bayesian paradigm. However, throughout this paper we follow the frequentist paradigm, and deriving respective limit results would require a frequentist evaluation of Bayesian methods (see, e.g., Section 10 in van der Vaart, 1998), which is beyond the scope of this paper.

The remainder of this article is organized as follows. After introducing the underlying asymptotic sampling scheme at the beginning of Section 2, we present general limit results on empirical probability weighted moments and truncated L-moments (Section 2.1; the results built upon a previous discussion paper co-authored by a subset of the authors, see Lilienthal et al., 2016). The results are then transferred to limit results for respective moment estimators for GEV parameters, which are accompanied by results on a pseudo maximum likelihood estimator (Section 2.2). Section 3 is devoted to estimation of the estimators' asymptotic covariance matrices. Formal asymptotic homogeneity tests are discussed in Section 4, and the particular case of testing the index flood assumption is examined in a finite-sample Monte Carlo simulation study in Section 5. A case study on flood extremes is worked out in Section 6. The main findings are summarized in the concluding Section 7. Proofs and technical details are deferred to the Supplementary Material.

2 | PARAMETRIC ESTIMATORS FOR MULTIVARIATE GEV-MODELS

Recall the generalized extreme value distribution defined by its cumulative distribution function (c.d.f.)

$$G_{\mu,\sigma,\xi}(x) = \exp\left(-\left[1 + \xi \frac{x - \mu}{\sigma}\right]^{-1/\xi}\right), \quad 1 + \xi \frac{x - \mu}{\sigma} > 0,$$

$$\hat{\boldsymbol{\beta}}_j^{(R)} = (\hat{\beta}_{0,j}, \dots, \hat{\beta}_{R-1,j})' \quad \text{and} \quad \hat{\boldsymbol{\beta}}^{(R)} = \left((\hat{\boldsymbol{\beta}}_1^{(R)})', \dots, (\hat{\boldsymbol{\beta}}_d^{(R)})' \right)',$$

Theorem 1 (Multivariate asymptotic normality of empirical PWMs). *Consider observations from sampling scheme (1) with continuous marginal c.d.f. F_j for $j \in \{1, \dots, d\}$ such that the second moment of F_j exists. Suppose further that $\sup_{x \in \mathbb{R}} |x \{F_j(x)(1 - F_j(x))\}^w| < \infty$ for all $j = 1, \dots, d$ and some $w \in [0, 1/2)$. Then, for any fixed $R \in \mathbb{N}$ and as $n \rightarrow \infty$,*

$$\sqrt{n} \left(\hat{\boldsymbol{\beta}}^{(R)} - \boldsymbol{\beta}^{(R)} \right) \xrightarrow{D} \mathcal{N}_{dR} \left(\mathbf{0}, \boldsymbol{\Xi}^{(R)} \right),$$

where the limiting covariance matrix $\boldsymbol{\Xi}^{(R)} \in \mathbb{R}^{dR \times dR}$ is given by $\boldsymbol{\Xi}^{(R)} = (\boldsymbol{\Xi}_{j,\ell}^{(R)})_{j,\ell=1}^d$ with

$$\boldsymbol{\Xi}_{j,\ell}^{(R)} = \frac{1}{\max(c_j, c_\ell)} \text{Cov}(\mathbf{Z}_j, \mathbf{Z}_\ell) \in \mathbb{R}^{R \times R}$$

and where $\mathbf{Z}_j = \mathbf{Z}_j^{(R)} = (Z_{0,j}, Z_{1,j}, \dots, Z_{R-1,j})'$ has coordinates, for $r \in \{0, \dots, R-1\}$,

$$Z_{r,j} = X_j \cdot F_j^r(X_j) + r \cdot \int_{\mathbb{R}} x \cdot F_j^{r-1}(x) \cdot \mathbf{1}(X_j \leq x) dF_j(x). \quad (3)$$

Next, consider TL-moments and assume that F_j has a finite mean. For $m \in \mathbb{N}$ and $s, t \in \mathbb{N}_0$, the m th TL-moment of F_j with left- and right-trimming parameters s and t , respectively, is defined as

$$\lambda_{m,j}^{(s,t)} = \frac{1}{m} \sum_{i=0}^{m-1} (-1)^i \binom{m-1}{i} E(X_{m+s-i:m+s+t,j}),$$

where $X_{1:n,j} \leq \dots \leq X_{n:n,j}$ denote order statistics of a random sample of size n drawn from F_j (Elamir & Seheult, 2003). Recall that TL-moments generalize L-moments which are obtained for $s = t = 0$. By definition, both are built on linear combinations of expected values of order statistics. TL-moments with $s, t > 0$ avoid using the most extreme order statistics which may suffer from high variability and other effects. In extreme value analysis of heavy tailed maxima this particularly applies to the upper tail so that choosing $t > 0$ is often worthwhile. The TL-moment is known to satisfy (see Hosking, 2007)

$$\lambda_{m+1,j}^{(s,t)} = \sum_{i=s}^{m+s+t} z_{m,i}^{(s,t)} \beta_{ij} = (\mathbf{z}_m^{(s,t)})' \boldsymbol{\beta}_j^{(m+s+t+1)}, \quad m \in \mathbb{N}_0, \quad (4)$$

with $\mathbf{z}_m^{(s,t)} = (z_{m,0}^{(s,t)}, \dots, z_{m,m+s+t}^{(s,t)})' \in \mathbb{R}^{m+s+t+1}$ being a coefficient vector with components $z_{m,i}^{(s,t)} = 0$ for $i \in \{0, \dots, s-1\}$ and

$$z_{m,i}^{(s,t)} = \frac{m!(m+s+t+1)!}{(m+1)(m+s)!(m+t)!} (-1)^{s+m+i} \binom{m+t}{i+s} \binom{m+i}{m},$$

for $i \in \{s, \dots, m+s+t\}$. It is easy to see that, for fixed $M \in \mathbb{N}$, Equation (4) allows to write $\lambda_j^{(M,s,t)} = (\lambda_{1,j}^{(s,t)}, \dots, \lambda_{M,j}^{(s,t)})'$ as $\lambda_j^{(M,s,t)} = \Delta^{(M,s,t)} \boldsymbol{\beta}_j^{(M+s+t)}$ for some matrix $\Delta^{(M,s,t)} \in \mathbb{R}^{M \times (M+s+t)}$. The first M sample TL(s, t)-moments $\hat{\lambda}_j^{(M,s,t)} = (\hat{\lambda}_{1,j}^{(s,t)}, \dots, \hat{\lambda}_{M,j}^{(s,t)})'$ of the j th coordinate are therefore defined as $\hat{\lambda}_j^{(M,s,t)} = \Delta^{(M,s,t)} \hat{\boldsymbol{\beta}}_j^{(M+s+t)}$. Finally, write

$$\hat{\boldsymbol{\lambda}}^{(M,s,t)} = \left((\hat{\lambda}_1^{(M,s,t)})', \dots, (\hat{\lambda}_d^{(M,s,t)})' \right)', \quad \boldsymbol{\lambda}^{(M,s,t)} = \left((\lambda_1^{(s,t)})', \dots, (\lambda_d^{(s,t)})' \right)'$$

Corollary 1 (Multivariate asymptotic normality of empirical TLMs). *Under the conditions of Theorem 1 we have, for any $M \in \mathbb{N}$, $s, t \in \mathbb{N}_0$ and as $n \rightarrow \infty$,*

$$\sqrt{n} \left(\hat{\boldsymbol{\lambda}}^{(M,s,t)} - \boldsymbol{\lambda}^{(M,s,t)} \right) \xrightarrow{D} \mathcal{N}_{dM} \left(\mathbf{0}, \boldsymbol{\Xi}^{(M,s,t)} \right),$$

where $\Xi^{(M,s,t)} = \tilde{\Delta}^{(M,s,t)} \Xi^{(M+s+t)} (\tilde{\Delta}^{(M,s,t)})'$ with $\Xi^{(M+s+t)}$ from Theorem 1 and with block-diagonal matrix $\tilde{\Delta}^{(M,s,t)} = \text{diag}(\Delta^{(M,s,t)}, \dots, \Delta^{(M,s,t)}) \in \mathbb{R}^{dM \times d(M+s+t)}$, where the matrix $\Delta^{(M,s,t)}$ is repeated d times.

2.2 | Three parametric estimators in multivariate GEV-models

The PWM-estimator: Recall the relationship between the GEV-parameters and the PWMs in Equation (2). As shown in Hosking et al. (1985), the first three PWMs $\beta_j^{(3)} = (\beta_{0,j}, \beta_{1,j}, \beta_{2,j})$ uniquely determine the GEV-parameters (μ_j, σ_j, ξ_j) :

$$\xi_j = g_1^{-1} \left(\frac{3\beta_{2,j} - \beta_{0,j}}{2\beta_{1,j} - \beta_{0,j}} \right), \quad \sigma_j = g_2(\xi_j) (2\beta_{1,j} - \beta_{0,j}), \quad \mu_j = \beta_{0,j} + \sigma_j g_3(\xi_j), \quad (5)$$

where

$$g_1(\xi) = \frac{3^\xi - 1}{2^\xi - 1}, \quad g_2(\xi) = \frac{\xi}{\Gamma(1 - \xi)(2^\xi - 1)}, \quad g_3(\xi) = \frac{1 - \Gamma(1 - \xi)}{\xi},$$

with $g_1(0) = \frac{\log 3}{\log 2}$, $g_2(0) = \frac{1}{\log 2}$ and $g_3(0) = -\gamma$ defined by continuity and with the Euler–Mascheroni constant γ . Writing the equation system from (5) as $\vartheta_j = \varphi(\beta_j^{(3)})$, we arrive at the PWM-estimator for the GEV-parameters,

$$\hat{\vartheta}_j = \hat{\vartheta}_j^{(\text{pwm})} = \varphi(\hat{\beta}_j^{(3)}) \quad \hat{\vartheta} = \hat{\vartheta}^{(\text{pwm})} = ((\hat{\vartheta}_1^{(\text{pwm})})', \dots, (\hat{\vartheta}_d^{(\text{pwm})})')'.$$

Corollary 2. Consider sampling scheme (1) with $F_j = G_{\vartheta_j}$ and $\xi_j < 1/2$ for all $j \in \{1, \dots, d\}$. Then,

$$\sqrt{n} \left(\hat{\vartheta}^{(\text{pwm})} - \vartheta \right) \xrightarrow{D} \mathcal{N}_{3d} \left(\mathbf{0}, \Sigma^{(\text{pwm})} \right)$$

as $n \rightarrow \infty$, where $\Sigma^{(\text{pwm})} = \Delta \Xi^{(3)} \Delta'$ with $\Xi^{(3)}$ from Theorem 1 and where

$$\Delta = \Delta^{(\text{pwm})} = \text{diag} \left(\frac{\partial \varphi(\beta_1)}{\partial \beta}, \dots, \frac{\partial \varphi(\beta_d)}{\partial \beta} \right).$$

The TLM-estimator: Just as for PW-moments, GEV parameters may be uniquely expressed through the first three TL-moments for $\xi_j < 1$, that is, $\vartheta_j = \varphi^{(s,t)}(\lambda_j^{(3,s,t)})$ for some known function $\varphi^{(s,t)}$ (Lilienthal, 2019a).

Details for the case of TL(0,0)-moments (also known as L-moments) and TL(0,1)-moments are summarized in Appendix A. Let

$$\hat{\vartheta}^{(\text{tl}), (s,t)} = \left(\left(\hat{\vartheta}_1^{(\text{tl}), (s,t)} \right)', \dots, \left(\hat{\vartheta}_d^{(\text{tl}), (s,t)} \right)' \right)', \quad \hat{\vartheta}_j^{(\text{tl}), (s,t)} = \varphi^{(s,t)} \left(\hat{\lambda}_j^{(3,s,t)} \right).$$

Corollary 3. Consider sampling scheme (1) with $F_j = G_{\vartheta_j}$ and $\xi_j < 1/2$ for all $j \in \{1, \dots, d\}$. Then,

$$\sqrt{n} \left(\hat{\vartheta}^{(\text{tl}), (s,t)} - \vartheta \right) \xrightarrow{D} \mathcal{N}_{3d} \left(\mathbf{0}, \Sigma^{(\text{tl}), (s,t)} \right),$$

as $n \rightarrow \infty$, where $\Sigma^{(\text{tl}), (s,t)} = \Delta^{(s,t)} \Xi^{(3,s,t)} (\Delta^{(s,t)})'$ with $\Xi^{(3,s,t)} \in \mathbb{R}^{3d \times 3d}$ from Theorem 1 and where

$$\Delta^{(s,t)} = \text{diag} \left(\frac{\partial \varphi^{(s,t)}(\lambda_1)}{\partial \lambda}, \dots, \frac{\partial \varphi^{(s,t)}(\lambda_d)}{\partial \lambda} \right).$$

Remark 1. Note that the first m TL(0,0)-moments are linear combinations of the first m PWMs (let $s = t = 0$ in (4) to see this). In particular, the first three TL(0,0)- and PW-moments are in one-to-one correspondence, whence the GEV-estimators based on these methods coincide.

The PML-estimator: The PML-estimator is defined as any coordinate-wise maximum of the independence log-likelihood function, which may be rewritten as

$$\hat{\boldsymbol{\vartheta}}^{(\text{pml})} \in \operatorname{argmax}_{(\boldsymbol{\vartheta}_1, \dots, \boldsymbol{\vartheta}_d) \in \Theta^d} \sum_{j=1}^d \sum_{i=a_j+1}^n \ell_{\boldsymbol{\vartheta}_j}(X_{i,j}),$$

where $\ell_{\boldsymbol{\vartheta}} : \mathbb{R} \rightarrow [-\infty, \infty)$, $x \mapsto \log p_{\boldsymbol{\vartheta}}(x)$ with the density $p_{\boldsymbol{\vartheta}}$ of the GEV-distribution. Here, Θ denotes a subset of $\mathbb{R} \times (0, \infty) \times \mathbb{R}$ that is to be considered the coordinate-wise parameter space. As is well-known, consistency can only be guaranteed if $\Theta \subset \mathbb{R} \times (0, \infty) \times (-1, \infty)$ (Dombry, 2015) and asymptotic normality requires $\Theta \subset \mathbb{R} \times (0, \infty) \times (-1/2, \infty)$ (Bücher & Segers, 2017).

Theorem 2. Consider sampling scheme (1) with $F_j = G_{\boldsymbol{\vartheta}_j}$ and $\xi_j > -1/2$ for all $j \in \{1, \dots, d\}$. Then, for any compact parameter set $\Theta \subset \mathbb{R} \times (0, \infty) \times (-\frac{1}{2}, \infty)$ containing $\boldsymbol{\vartheta}_j$ in its interior for all $j \in \{1, \dots, d\}$, any sequence $\hat{\boldsymbol{\vartheta}}^{(\text{pml})} = \hat{\boldsymbol{\vartheta}}_n^{(\text{pml})}$ of coordinate-wise maximum likelihood estimators over Θ^d , such maximizers always existing, is strongly consistent and satisfies

$$\sqrt{n} \left(\hat{\boldsymbol{\vartheta}}^{(\text{pml})} - \boldsymbol{\vartheta} \right) \xrightarrow{D} \mathcal{N}_{3d} \left(\mathbf{0}, \boldsymbol{\Sigma}^{(\text{pml})} \right),$$

where $\boldsymbol{\Sigma}^{(\text{pml})} = (\boldsymbol{\Sigma}_{j,\ell}^{(\text{pml})})_{j,\ell=1}^d$ with

$$\boldsymbol{\Sigma}_{j,\ell}^{(\text{pml})} = \frac{1}{\max(c_j, c_{\ell})} I_{\boldsymbol{\vartheta}_j}^{-1} E \left[\dot{\ell}_{\boldsymbol{\vartheta}_j}(X_{1,j}) (\dot{\ell}_{\boldsymbol{\vartheta}_{\ell}}(X_{1,\ell}))' \right] I_{\boldsymbol{\vartheta}_{\ell}}^{-1} \quad (6)$$

and where $I_{\boldsymbol{\vartheta}}$ denotes the Fisher information matrix of the three parametric GEV-family and where $\dot{\ell}_{\boldsymbol{\vartheta}}(x) = \frac{\partial}{\partial \boldsymbol{\vartheta}} \ell_{\boldsymbol{\vartheta}}(x)$ is the corresponding score vector.

3 | ESTIMATION OF LIMITING COVARIANCE MATRICES

In this section, details regarding the estimation of the limiting covariance matrices in the previous section are given. We start with an empirical estimator for $\boldsymbol{\Xi}^{(R)}$ from Section 2.1, which may readily be turned into an estimator for $\boldsymbol{\Xi}^{(M,s,t)}$, $\boldsymbol{\Sigma}^{(\text{pwm})}$ and $\boldsymbol{\Sigma}^{(\text{tl),(s,t)}}$. Likewise, an empirical version for $\boldsymbol{\Sigma}^{(\text{pml})}$ suggests itself.

3.1 | Covariance estimation for empirical PW- and TL-moments

Suppose that we have collected observations based on the sampling scheme in (1), with continuous marginal c.d.f.s. Recall that the asymptotic covariance matrix of the first R PWMs, $\boldsymbol{\Xi}_{j,\ell}^{(R)}$ from Theorem 1, may be expressed in terms of the covariance $\operatorname{Cov}(\mathbf{Z}_j, \mathbf{Z}_{\ell})$ with \mathbf{Z}_j as defined in (3). The latter covariance may be consistently estimated by the following sample analogues: for $j \in \{1, \dots, d\}$, $i \in \{a_j + 1, \dots, n\}$ and $r \in \mathbb{N}_0$, let

$$\hat{Z}_{i,r,j} = X_{i,j} \cdot F_{a_j+1:n_j}^r(X_{i,j}) + \frac{1}{n_j} \sum_{i'=a_j+1}^n X_{i',j} \cdot r \cdot F_{a_j+1:n_j}^{r-1}(X_{i',j}) \cdot \mathbb{1}(X_{i,j} \leq X_{i',j})$$

and $\hat{\mathbf{Z}}_{i,j} = (Z_{i,0,j}, Z_{i,1,j}, \dots, Z_{i,R-1,j})'$ be observable counterparts of \mathbf{Z}_j . Then, let

$$\hat{\boldsymbol{\Xi}}^{(R)} = (\hat{\boldsymbol{\Xi}}_{j,\ell}^{(R)})_{j,\ell=1}^d, \quad \hat{\boldsymbol{\Xi}}_{j,\ell}^{(R)} = \frac{n}{\max(n_j, n_{\ell})} \widehat{\operatorname{Cov}}(\hat{\mathbf{Z}}_j, \hat{\mathbf{Z}}_{\ell}),$$

where $\widehat{\operatorname{Cov}}(\hat{\mathbf{Z}}_j, \hat{\mathbf{Z}}_{\ell})$ is the empirical covariance matrix of the observable sample

$$\hat{\mathbf{Z}}_{j,\ell} = (\hat{\mathbf{Z}}_{\max(a_j, a_{\ell})+1, j}, \hat{\mathbf{Z}}_{\max(a_j, a_{\ell})+1, \ell}, \dots, (\hat{\mathbf{Z}}_{n_j}, \hat{\mathbf{Z}}_{n_{\ell}})$$

of sample size $\min(n_j, n_{\ell})$.

Based on the fact that $\Xi^{(M,s,t)}$ from Corollary 1 is a known functional of $\Xi^{(R)}$, we readily obtain an estimator for the latter matrix, say $\hat{\Xi}^{(M,s,t)}$, as well.

Both estimators are consistent.

Corollary 4. *Under the assumptions of Theorem 1 and for $n \rightarrow \infty$, we have*

$$\hat{\Xi}^{(R)} = \Xi^{(R)} + o_p(1), \quad \hat{\Xi}^{(M,s,t)} = \Xi^{(M,s,t)} + o_p(1).$$

3.2 | Covariance estimation in GEV-models

The asymptotic covariance matrices $\Sigma^{(\text{pwm})}$ and $\Sigma^{(\text{tl},(s,t))}$ from Corollary 2 and 3 are known functionals of $(\Xi^{(R)}, \boldsymbol{\vartheta})$ and $(\Xi^{(M,s,t)}, \boldsymbol{\vartheta})$, respectively. Replacing both objects by their respective estimators (where the PWM estimator $\hat{\boldsymbol{\vartheta}}^{(\text{pwm})}$ is used for estimation of $\Sigma^{(\text{pwm})}$ and likewise for the TLM estimator), we obtain what we call empirical estimators and denote them by $\hat{\Sigma}^{(\text{pwm})}$ and $\hat{\Sigma}^{(\text{tl},(s,t))}$, respectively. Consistency is a direct consequence of consistency of $\hat{\Xi}^{(c)}$ and $\hat{\boldsymbol{\vartheta}}^{(c)}$. In view of the fact that the j th element of the block diagonal, say $\Sigma_{jj}^{(c)}$, is a known, smooth functional of $\boldsymbol{\vartheta}_j$, a second set of consistent estimators may be obtained by replacing $\boldsymbol{\vartheta}_j$ by the respective estimator in that functional. However, throughout extensive simulation experiments, these modified estimators were found to frequently produce invalid covariance matrices (unless $d = 1$), whence they are unsuitable for testing issues.

Finally, regarding the PML-estimator, the covariance matrix $\Sigma^{(\text{pml})}$ from Theorem 2 may be estimated by $\hat{\Sigma}^{(\text{pml})} = (\hat{\Sigma}_{j,\ell}^{(\text{pml})})_{j,\ell=1}^d$ where

$$\hat{\Sigma}_{j,\ell}^{(\text{pml})} = \frac{n}{\max(n_j, n_\ell)} \hat{\Gamma}_{\boldsymbol{\vartheta}_j}^{-1} \left[\frac{1}{\min(n_j, n_\ell)} \sum_{i=\max(a_j, a_\ell)+1}^n (Y_{i,j} - \bar{Y}_{j,n_j}) (Y_{i,\ell} - \bar{Y}_{\ell,n_\ell})' \right] \hat{\Gamma}_{\boldsymbol{\vartheta}_\ell}^{-1},$$

with $Y_{i,j} = \ell'_{\boldsymbol{\vartheta}_j}(X_{i,j})$, $\bar{Y}_{j,n_j} = \frac{1}{n_j} \sum_{i=a_j+1}^n Y_{i,j}$ and

$$\hat{\Gamma}_{\boldsymbol{\vartheta}_j} = \frac{1}{n_j} \sum_{i=a_j+1}^n (Y_{i,j} - \bar{Y}_{j,n_j}) (Y_{i,j} - \bar{Y}_{j,n_j})'.$$

4 | ASYMPTOTIC TESTING PROCEDURES

From Corollaries 2 and 3 and Theorem 2 we conclude the asymptotic normality of PWM-, TLM-, and PML-estimators of the GEV-parameters, respectively; subsequently generically denoted by $\sqrt{n}(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0) \xrightarrow{D} \mathcal{N}(0, \Sigma)$. To construct asymptotically valid tests, that is, tests that hold the respective significance level for $n \rightarrow \infty$, the principle of the Wald test can be used. More precisely, for hypotheses of the form

$$H_0 : h(\boldsymbol{\vartheta}) = 0 \text{ versus } H_1 : h(\boldsymbol{\vartheta}) \neq 0 \quad (7)$$

with some known continuously differentiable function $h : \mathbb{R}^{3d} \rightarrow \mathbb{R}^q$, the test statistic

$$T_n = n (h(\hat{\boldsymbol{\vartheta}}))' \left(\hat{h}(\hat{\boldsymbol{\vartheta}}) \hat{\Sigma} (\hat{h}(\hat{\boldsymbol{\vartheta}}))' \right)^{-1} h(\hat{\boldsymbol{\vartheta}}), \quad (8)$$

converges in distribution to the χ_q^2 -distribution with q degrees of freedom, under H_0 . Here, $\hat{\Sigma}$ denotes a consistent estimator of Σ and \hat{h} is the Jacobi matrix of h . The convergence result is an immediate consequence of the delta method, Slutsky's theorem and the continuous mapping theorem. As a consequence, rejecting H_0 if the observed value of T_n exceeds the $(1 - \alpha)$ -quantile of the χ_q^2 -distribution defines an asymptotic test of significance level α . The $(1 - \alpha)$ -quantile may also be replaced by the empirical $(1 - \alpha)$ -quantile of a suitable bootstrap sample of T_n , see Section 5 for details.

Within the next two subsections, we discuss two practical examples of primal importance for which hypotheses of the general form in (7) play a major role.

4.1 | Testing the index flood assumption

A common application in hydrology is regional flood frequency analysis (RFFA), where one observes flood flows at d sites of a region with site-specific distributions F_j for $j \in \{1, \dots, d\}$ and in which it is of interest to combine regional information in order to decrease the variability of estimation at specific sites. Such an approach, referred to as a pooling method, is typically based on certain assumptions that ensure “regional homogeneity”. As an important example, the so-called index flood (IF) approach (Dalrymple, 1960) considers the homogeneity hypothesis

$$\mathcal{H}_{0,IF} : \begin{cases} \exists \text{ c.d.f. } G \text{ and constants } s_j = s(F_j) > 0 \text{ such that} \\ F_j^{-1} = s_j \cdot G^{-1} \text{ for all } j = 1, \dots, d. \end{cases}$$

Here, $s_j = s(F_j)$ denotes a site specific scaling factor (often a population mean or some other location parameter) and G is typically assumed to follow some parametric family, that is, $G = G_{\boldsymbol{\vartheta}}$ with $\boldsymbol{\vartheta}$ unknown (e.g., the GEV distribution).

While a moderate amount of heterogeneity in the group of sites may still lead to an overall improvement compared to individual local estimation (because of a reduction of variance, see Lettenmaier et al., 1987), strong heterogeneity typically leads to a severe bias that may render the potential improvement in terms of estimation variance useless. It is thus important to identify serious sources of heterogeneity. The given framework allows to construct an asymptotic testing procedure under a GEV assumption.

More precisely, suppose that we have observations meeting sampling scheme (1) with site-specific distribution functions $F_j = G_{\boldsymbol{\vartheta}_j}$, $j = 1, \dots, d$, the c.d.f. of the GEV(μ_j, σ_j, ξ_j)-distribution with parameter vector $\boldsymbol{\vartheta}_j = (\mu_j, \sigma_j, \xi_j)'$. In this case, a simple calculation shows that the Index Flood assumption $\mathcal{H}_{0,IF}$ is equivalent to

$$\mathcal{H}'_{0,IF} : \frac{\mu_1}{\sigma_1} = \dots = \frac{\mu_d}{\sigma_d} \quad \text{and} \quad \xi_1 = \dots = \xi_d. \quad (9)$$

Let $\hat{\boldsymbol{\vartheta}} = (\hat{\mu}_1, \hat{\sigma}_1, \hat{\xi}_1, \dots, \hat{\mu}_d, \hat{\sigma}_d, \hat{\xi}_d)'$ denote a generic estimator of the parameter vector $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}_1, \dots, \boldsymbol{\vartheta}_d)'$ and let $\hat{\boldsymbol{\Sigma}}$ denote a corresponding estimator (see Section 3) for the covariance matrix of the limiting distribution. Defining $h : \mathbb{R}^{3d} \rightarrow \mathbb{R}^{2(d-1)}$ as

$$h(\boldsymbol{\vartheta}) = \left(\frac{\mu_1}{\sigma_1} - \frac{\mu_2}{\sigma_2}, \dots, \frac{\mu_{d-1}}{\sigma_{d-1}} - \frac{\mu_d}{\sigma_d}, \xi_1 - \xi_2, \dots, \xi_{d-1} - \xi_d \right)'$$

the hypothesis in (9) can be expressed as $h(\boldsymbol{\vartheta}) = 0$, which is of the required form in (7). Further note that $h(\boldsymbol{\vartheta}) = \tilde{h}(g(\boldsymbol{\vartheta}))$ with $g(\boldsymbol{\vartheta}) = (\mu_1/\sigma_1, \dots, \mu_d/\sigma_d, \xi_1, \dots, \xi_d)$ and $\tilde{h}(\delta_1, \dots, \delta_d, \xi_1, \dots, \xi_d) = (\delta_1 - \delta_2, \dots, \delta_{d-1} - \delta_d, \xi_1 - \xi_2, \dots, \xi_{d-1} - \xi_d)$, showing that we are dealing with a linear hypothesis in $g(\boldsymbol{\vartheta})$, whence the Wald test statistic is invariant with respect to permutations of the coordinates.

Section 5.2 contains a thorough study concerning this testing procedure. Power and size are analyzed in finite sample cases and subsequently a bootstrap procedure for size correction is given.

4.2 | Testing assumptions for intensity-duration-frequency curves

Another potential application for the general testing procedure arises from intensity-duration-frequency (IDF) curves, a commonly used tool for linking intensity and duration of extreme precipitation events to their frequency, that is, their return period. More precisely, suppose that yearly maxima of the time series of average precipitation amounts (e.g., measured in mm/h) collected over successive time periods of duration d_1, \dots, d_k (for instance, $d_1 = 1, d_2 = 24, d_3 = 48$, measured in hours) at a fixed location is modeled by a c.d.f. $F = F_j$.

According to Koutsoyiannis et al. (1998), the intensity $i_j = i(d_j, p) = F_j^{-1}(p)$ related to non-exceedance probability $p \in (0, 1)$ (i.e., frequency) and duration d_j is typically assumed to satisfy

$$i_j = \frac{\omega}{(d_j^\nu + \theta)^\eta}, \quad (10)$$

where $\omega, \nu, \theta, \eta$ are certain non-negative parameters satisfying $\nu\eta \geq 1$. If $d_j \geq 1$ hour for all j , one often applies a simplified or reduced model with $\nu = 1$ and $\theta = 0$, which may be motivated by simple and multifractal scaling properties (Boukhelifa et al., 2018; Jurado et al., 2020). In that case, if $F_j = G_{\boldsymbol{\theta}_j}$ is the c.d.f. of the GEV-distribution with parameter vector $\boldsymbol{\theta}_j = (\mu_j, \sigma_j, \xi_j)'$, relationship (10) is typically guaranteed by assuming that

$$\mu_j = \tilde{\mu} \sigma_j, \quad \sigma_j = \tilde{\sigma} d_j^{-\eta_0}, \quad \xi_j = \xi_0,$$

for some parameters $\tilde{\mu}, \tilde{\sigma}, \xi_0, \eta_0$ that do not depend on the duration. A simple calculation shows that the parametric constraints of this simplified model are equivalent to

$$\begin{aligned} \frac{\mu_1}{\sigma_1} &= \dots = \frac{\mu_k}{\sigma_k}, \\ \xi_1 &= \dots = \xi_k, \\ \frac{\ln(\sigma_2/\sigma_1)}{\ln(d_2/d_1)} &= \dots = \frac{\ln(\sigma_k/\sigma_1)}{\ln(d_k/d_1)}. \end{aligned} \quad (11)$$

Similar to the application in Section 4.1, a function h can be constructed such that the constraints in (11) are equivalent to $h(\boldsymbol{\theta}) = 0$, where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_k)'$. This, again, is of the required form in (7).

5 | FINITE-SAMPLE PROPERTIES

The finite-sample properties of the homogeneity tests for the Index Flood Assumption (Section 4) are evaluated in a large scale simulation study. The underlying data generating processes are explained in Section 5.1 and the homogeneity test is illustrated in Sections 5.2–5.4. Throughout, results on the PWM estimator are omitted in view of Remark 1, as they can be directly inferred from the L-moment results.

5.1 | Data generating process

Many environmental applications involve block maxima collected at various spatial stations. Mathematical theory suggests to model such data by max-stable processes, as the latter are the only processes that can arise, after proper affine transformation, as the limit of maxima of independent and identically distributed random fields $\{Y_i(x) : x \in \mathbb{R}^p\}$ (Davison et al., 2012). In particular, the marginal distribution at each specific location is necessarily GEV($\boldsymbol{\theta}$), possibly with location-specific parameter vector $\boldsymbol{\theta} = \boldsymbol{\theta}(x)$.

The data generating process used throughout this simulation study builds upon the max-stable Schlather model (Schlather, 2002). The latter is a stationary process $\{Z(x) : x \in \mathbb{R}^2\}$ with unit Fréchet margins whose bivariate distribution function is given by

$$P(Z(0) \leq z_1, Z(x) \leq z_2) = \exp \left[-\frac{1}{2} (z_1^{-1} + z_2^{-1}) \left\{ 1 + \left(1 - 2 \frac{(\rho(\|x\|) + 1) z_1 z_2}{(z_1 + z_2)^2} \right)^{\frac{1}{2}} \right\} \right],$$

where ρ is a correlation function and $\|x\|$ denotes the Euclidean norm of x . Throughout, we employ the Whittle-Matérn correlation function

$$\rho_{\eta, \nu}(h) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{h}{\eta} \right)^\nu K_\nu \left(\frac{h}{\eta} \right), \quad h \geq 0,$$

where K_ν denotes the modified Bessel function of order ν and where $\nu > 0$ and $\eta > 0$ are called smoothness and range parameter, respectively; see Figure 1.

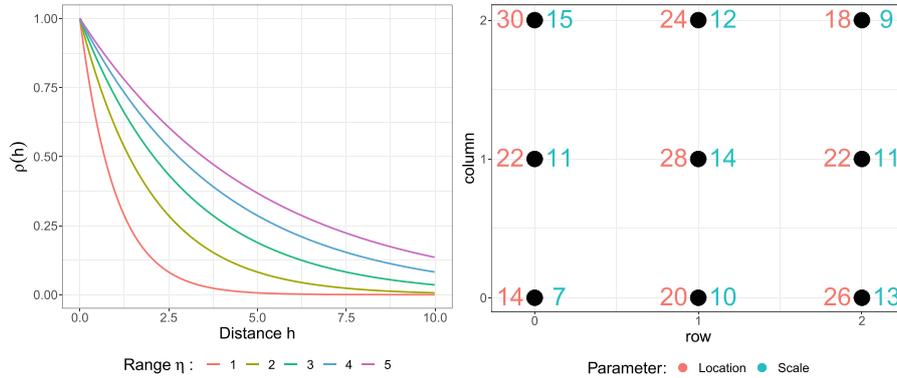


FIGURE 1 Left: Whittle-Matérn correlation function for fixed smoothing parameter $\nu = 0.5$ and range parameter $\eta \in \{1, 2, 3, 4, 5\}$. Right: Stations on a grid with respective location and scale parameters

The max-stable Schlather model can be transformed into a max-stable process with $GEV(\mu(\cdot), \sigma(\cdot), \xi(\cdot))$ margins by letting

$$X(x) = \mu(x) + \sigma(x) \frac{Z(x)^{\xi(x)} - 1}{\xi(x)}. \tag{12}$$

A simulated sample from this model at a given finite set of locations $\mathcal{X} = \{x_1, \dots, x_d\}$ can be obtained by the function ‘rmaxstab’ from the R-package SPATIALEXTREMES (Ribatet, 2020). Throughout, \mathcal{X} is chosen as a $\sqrt{d} \times \sqrt{d}$ -grid $\{0, 1, \dots, \sqrt{d}\}^2$, where $d \in \{4, 9\}$. Moreover, we fix $\nu = 0.5$, as suggested by a data analysis of annual maxima of daily maximum rainfall in Davison et al. (2012). We further allow for different $\eta \in \{1, 3, 5\}$ to control the strength of spatial dependence. Finally, $\mu(x)$, $\sigma(x)$ and $\xi(x)$ are specified in the next subsections.

5.2 | Testing the index flood assumption

In this section, we empirically examine the homogeneity test from Section 4.1. The data generating process is as in Section 5.1, with results only reported for $d = 9$ locations $x \in \mathcal{X} = \{0, 1, 2\}^2$. Under the null hypothesis, we set $\xi(x) = \xi \in \{-0.4, -0.2, 0, 0.2, 0.4\}$ constant, $\mu(x) = 2\sigma(x)$ and

$$\sigma(x) = 7 + 3x_1 + 4x_2 - 3x_1x_2 \cdot \mathbb{1}(x_1 + x_2 \geq 3),$$

see the right panel in Figure 1 for an illustration. Moreover, we consider as an alternative a scenario with two nearly homogeneous groups of stations containing 6 and 3 stations, respectively. The larger group is generated as under the null hypothesis, while the smaller group deviates from the other group as follows: the shape parameter in the deviating group is $\xi_{\text{dev}} = \xi + c_\xi$, with $c_\xi \in \{-0.15, -0.10, -0.05, 0, 0.05, 0.10, 0.15\}$. The scale parameters are uniformly manipulated by adding $c_\sigma \in \{-2.25, -1.125, 0, 1.125, 2.25\}$ to the local scale parameter of the other group, while the location parameters stay fixed. More precisely, we have scale parameters $\sigma_{\text{dev}}(x) = \sigma(x) + c_\sigma$ and location parameters $\mu_{\text{dev}} = \mu(x)$ as before. The dispersion parameters in the deviating group are thus $\mu(x)/\sigma_{\text{dev}}(x) \neq 2$. The constants c_ξ and c_σ will be referred to as the ‘manipulation’ of shape and scale, respectively. Members of the deviating group are the stations located at (2, 0), (2, 1) and (2, 2). The null hypothesis is investigated for $n_1 = n \in \{50, 75, 100, 150, 200, 250, 300, 500, 1000, 1750, 2500\}$, and with $n_1 = n_2 = n_8 = n_9 = n$, $n_3 = n_7 = \lfloor 0.8n \rfloor$, $n_4 = n_6 = \lfloor 0.85n \rfloor$ and $n_5 = \lfloor 0.9n \rfloor$. These values, many of which seem unrealistically large for real world record lengths, are chosen mainly for illustrative purposes and shall emphasize that the derived tests are only asymptotic level α tests, that is, the significance level can only be guaranteed to hold in the limit. The same setting is evaluated under the alternative, but for $n_1 = n \in \{100, 300, 500\}$ only. The level of the test is fixed to $\alpha = 0.05$. During simulation, the strength of spatial correlation, that is, the value of the range parameter η , was found to have very little influence on the rejection rates of the tests, so we only present results for $\eta = 3$.

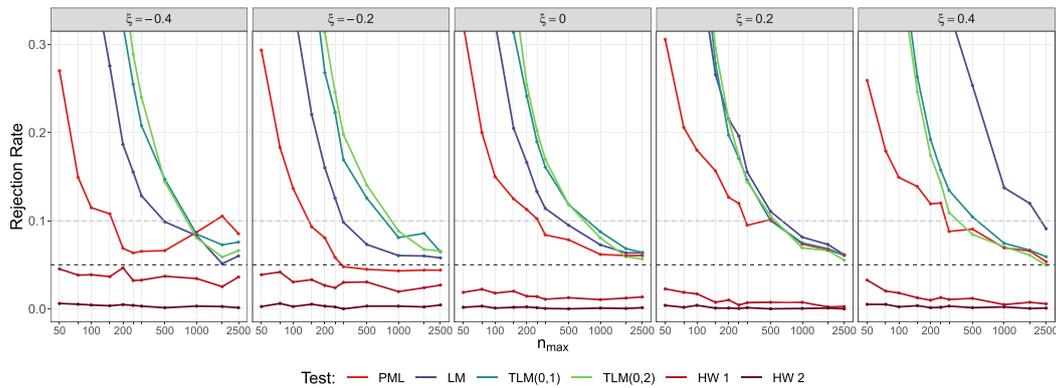


FIGURE 2 Error rates of the proposed test for increasing sample length

The test statistic (8) is calculated with $\hat{\vartheta}$ being either the LM, TLM(0,1), TLM(0,2) or PML estimator, and with $\hat{\Sigma}$ the respective covariance estimator. Stationwise GEV-parameter estimation based on the first three estimators is carried out with functions provided in the R-package `TLMoments` (Lilienthal, 2019b), while for ML estimation the function `fgev()` from the `evd`-package (Stephenson, 2018) is used. For the latter, when true shape parameters were very small ($\xi_{\text{dev}} \leq -0.4$) and starting values were computed with the standard routine implemented in `fgev()`, we noticed some instabilities regarding convergence toward the global maximum, resulting in unrealistic estimates. Therefore, we choose L-Moment based GEV-parameter estimates as starting values whenever the covariance matrix of the resulting estimate can be computed and is not singular, otherwise we stick to the default method.

Next to our formal tests, we also compute the heterogeneity measure HW of Hosking and Wallis (1993) as implemented in the R-Package `homtest` (Viglione, 2012), which compares the observed variability of L-moment ratios to that expected under homogeneity. Following Hosking and Wallis (1993), the sample is called “acceptably homogeneous”, “possibly heterogeneous” or “definitely heterogeneous”, whenever $HW < 1$, $1 \leq HW < 2$ or $HW \geq 2$, respectively. To make the results comparable, the test that is obtained when rejecting the null hypothesis when $HW \geq 1$ is called HW 1, while the one obtained when rejecting if $HW \geq 2$ is called HW 2. However, it is important to keep in mind that these tests are not classical level- α -tests.

The proportion of test rejections is interpreted as empirical error in the homogeneous case and as empirical power in the heterogeneous case. The results under the null hypothesis are illustrated in Figure 2. It can be seen that most tests (except for HW) approach the nominal level of 5% for increasing maximal sample size n . However, all tests (except for HW) are very liberal (i.e., the null hypothesis is rejected more frequently than suggested by the significance level) for small or even medium sample lengths ($n \leq 200$), with decent level approximations often requiring $n \geq 1000$. This is an important message for applications: the asymptotic tests are not reliable for realistic real world record lengths, and modifications such as those worked out in Section 5.3 below are necessary.

Comparing across (non HW-)methods, the PML method overall provides the most accurate test, except for $\xi = -0.4$. Finally, the tests derived from the Hosking Wallis criterion are quite conservative at the 5% level, in particular for larger shapes.

The empirical power of the tests is illustrated in Figure 3, where we restrict attention to $n = 300$ ($n = 100$ and $n = 500$ lead to qualitatively similar results, with slightly higher error rates and lower power for $n = 100$, and vice versa for $n = 500$). The plot is to be read as follows: the columns indicate the shape parameter of the GEV distribution of the reference group, while the rows indicate the estimator that the test was based on. The central square of each grid represents the rejection rate under the null hypothesis (only one group, no station was manipulated), which is to be compared to the theoretical significance level $\alpha = 5\%$. As explained at the beginning of this section, deviations from the null hypothesis are generated by uniformly manipulating the scale and shape parameters of three stations, which is done by adding constants c_ξ and c_σ to the scale and shape parameters of the reference stations. Moving along the x -axis therefore corresponds to the amount c_σ that was added to the initial scale parameters $\sigma(x)$ of the reference group, while moving along the y -axis corresponds to the amount c_ξ that the shape parameters between the groups differ. Hence, the alternative is met for all but the central square, and the corresponding rejection rates provide an approximation of the power of the test. In accordance

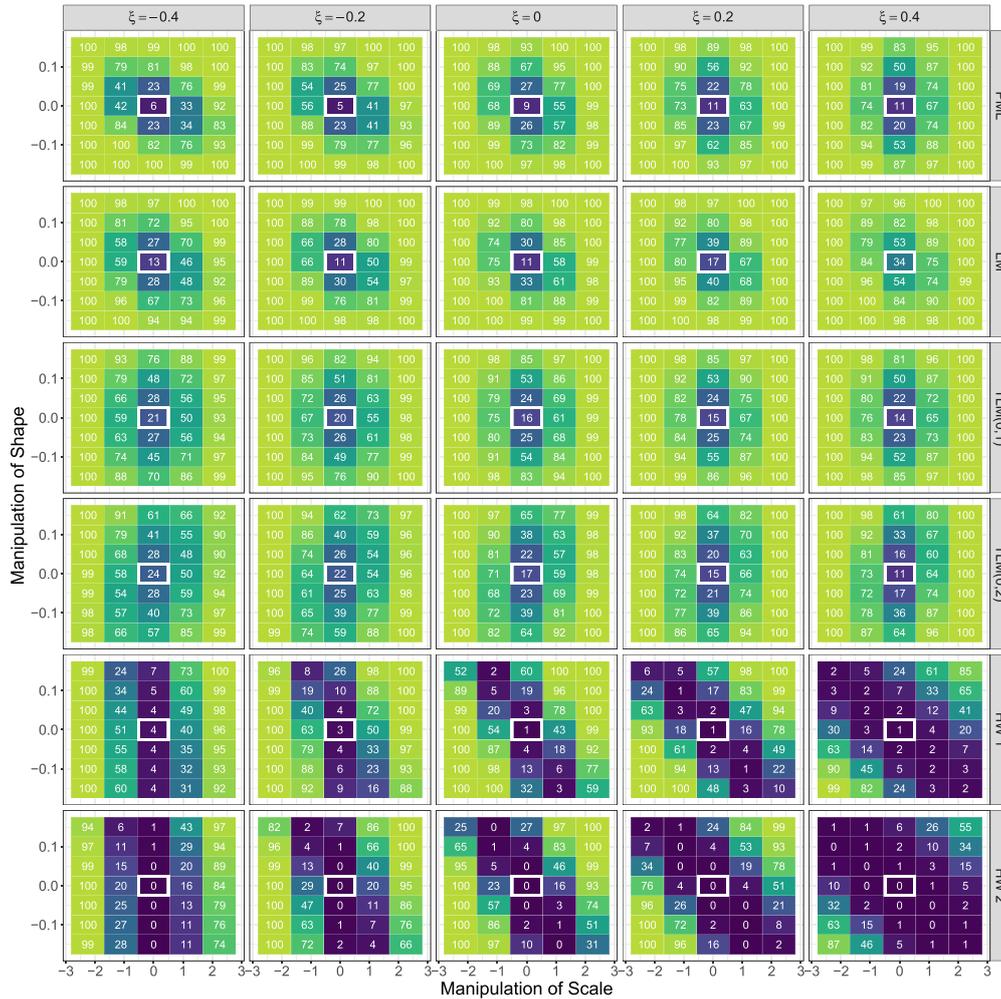


FIGURE 3 Rejection rates (%) of the proposed test with $n = 300$ and $\eta = 3$

with the results under H_0 , the error rate (centre point) exceeds the level of 5% for all of the non-HW versions, with barely acceptable rates for all but the PML test (for which error rates are between 5% and 11%).

The HW-tests' level is well below 5%, and accordingly the tests exhibit quite little power.

5.3 | Bootstrap procedure

In view of the disappointing test behavior for small to moderate sample sizes, we next investigate a modification of the above tests that is based on a parametric bootstrap technique. More precisely, given an observed sample as in (1), we start by computing the value of the test statistic T_n from (8). Next, we calculate estimators of the GEV parameters under the homogeneity constraints from (9). For that purpose, we apply different methods depending on the method of parameter estimation.

First, for the TL-moment versions, we rely on the following relationship between GEV parameters and TL-moment ratios: let $\tau_2^{(s,t)} = \lambda_2^{(s,t)} / \lambda_1^{(s,t)}$, $\tau_3^{(s,t)} = \lambda_3^{(s,t)} / \lambda_2^{(s,t)}$ denote the first two TL(s, t)-moment ratios. Then

$$\begin{aligned}\xi &= g_\xi \left(\tau_3^{(s,t)} \right), \\ \delta &= g_\delta \left(\tau_2^{(s,t)}, \tau_3^{(s,t)} \right), \\ \mu &= g_\mu \left(\lambda_1^{(s,t)}, \tau_2^{(s,t)}, \tau_3^{(s,t)} \right),\end{aligned}$$

where the explicit forms of g_ξ, g_δ, g_μ can be derived from the relation between τ_2, τ_3 and $\lambda_1, \lambda_2, \lambda_3$ and Equations (A.1), (A.1) from the Appendix, see also Lilienthal (2019a). The shape and dispersion parameter satisfying the homogeneity assumption are then estimated by plugging the weighted sample means $\bar{\tau}_3^{(s,t)}$ and $\bar{\tau}_2^{(s,t)}$ of $\tau_3^{(s,t)}$ and $\tau_2^{(s,t)}$, respectively, into g_ξ and g_δ , where $\bar{\tau}_i^{(s,t)} = \left(\sum_{j=1}^d n_j \hat{\tau}_{i,j}^{(s,t)} \right) / \left(\sum_{j=1}^d n_j \right)$. Finally, for estimation of the location parameters, the local estimates of $\lambda_1^{(s,t)}$ are plugged into g_μ along with $\bar{\tau}_3^{(s,t)}$ and $\bar{\tau}_2^{(s,t)}$. Altogether, the GEV parameters at site j are estimated by

$$\hat{\boldsymbol{\theta}}_j^{(tl),(s,t)} = (\hat{\mu}_j, \hat{\sigma}_j, \hat{\xi}_j)' = \left(g_\mu(\lambda_{1,j}^{(s,t)}, \bar{\tau}_2^{(s,t)}, \bar{\tau}_3^{(s,t)}), \frac{g_\mu(\lambda_{1,j}^{(s,t)}, \bar{\tau}_2^{(s,t)}, \bar{\tau}_3^{(s,t)})}{g_\delta(\bar{\tau}_2^{(s,t)}, \bar{\tau}_3^{(s,t)})}, g_\xi(\bar{\tau}_3^{(s,t)}) \right)'$$

For PML-estimation under the homogeneity constraints, we choose to minimize

$$(\delta, \sigma_1, \dots, \sigma_d, \xi) \mapsto \ell_{\delta, \sigma_1, \dots, \sigma_d, \xi}(\mathbf{x}) = \sum_{j=1}^d \sum_{i=a_j+1}^n \left\{ \log(\sigma_j) + \left(\frac{1}{\xi} + 1 \right) \log(y_{ij}) + y_{ij}^{-\frac{1}{\xi}} \right\},$$

where $y_{ij} = \max\{1 + \xi(\frac{x_{ij}}{\sigma_j} - \delta), 0\}$ and where the summands are defined by continuous extension for $\xi = 0$. From the obtained estimates $(\hat{\delta}, \hat{\sigma}_1, \dots, \hat{\sigma}_d, \hat{\xi})$, local location parameter estimates are derived by $\hat{\mu}_j = \hat{\delta} \hat{\sigma}_j$.

In order to obtain bootstrap samples that exhibit the (approximately) correct dependence structure, we proceed as follows: once the local GEV parameters are estimated, we transform the observations stationwise to approximately unit Fréchet distributed data (under H_0) by inverting (12):

$$Z_{i,j} = \left\{ 1 + \frac{\hat{\xi}_j}{\hat{\sigma}_j} \frac{X_{i,j} - \hat{\mu}_j}{\hat{\sigma}_j} \right\}_+^{\frac{1}{\hat{\xi}_j}}, \quad i = a_j + 1, \dots, n.$$

The transformed sample may then be fitted to a max-stable process model. Since the true model of the max-stable process is unknown, we fit several ones and select the one with minimal value of the composite likelihood information criterion (CLIC), which corresponds to the Takeuchi model selection criterion adapted to the pairwise likelihood setting (Davison and Gholamrezaee, 2012). Models fitted are the Schlather model with Whittle Matérn and Powered Exponential correlation function as well as a Smith model. For more details on the models, see for example, Davison et al. (2012).

Now that both the margins and the dependence structure are specified, we repeat the following bootstrap step for $b = 1, \dots, B$. First, we generate new data by sampling from the selected max-stable process model with respective estimated spatial dependence parameters. Then we transform the unit Fréchet to GEV margins by applying transformation (12) stationwise with local GEV parameters $\hat{\mu}_j, \hat{\sigma}_j, \hat{\xi}_j, j = 1, \dots, d$. Note that the obtained sample fulfills the null hypothesis of homogeneity. Then the value of the test statistic $T_{n,b}^*$ is computed and from the series of test statistics we calculate the bootstrapped p -value by

$$p_{corr} = \frac{1}{B+1} \sum_{b=1}^B \mathbb{1}(T_n \leq T_{n,b}^*).$$

The bootstrap test was carried out for the heterogeneous data setting described in the previous section. Attention was restricted to the LM-, TLM(0,1)- and PML-based estimation, as TLM(0,2) was found to perform similarly, but slightly worse than TLM(0,1) (see Figures 2 and 3).

We chose $B = 300$ and again every scenario is replicated 2000 times. Results are shown in Figure 4 for $n_{\max} = 100$. For all bootstrap tests (top three rows), the empirical error is between 3% and 6%, most often even between 4% and 5%, therefore accurately achieving the intended asymptotic level of the test. The LM and PML versions outperform the TLM(0,1)

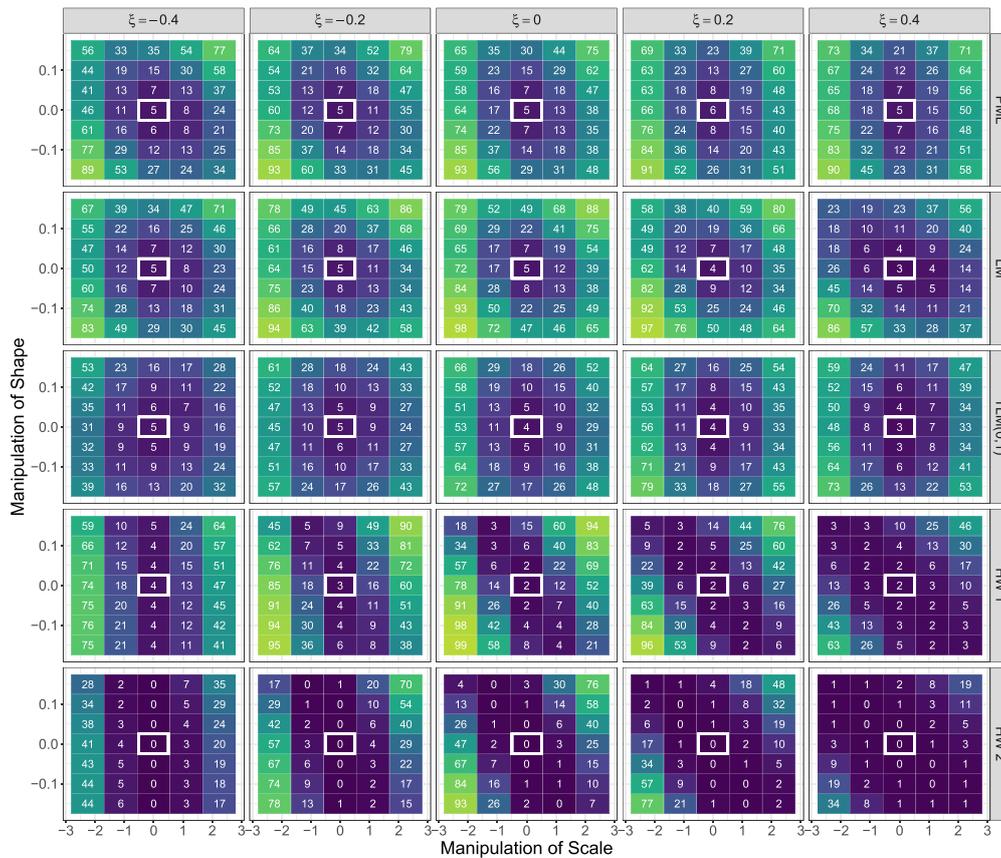


FIGURE 4 Empirical power (in %) of the proposed test for $n_{max} = 100$, $d = 9$ and $\eta = 3$, based on the bootstrap procedure (top three rows) and the heterogeneity measure of Hosking and Wallis (bottom two rows)

version in terms of power, with a slightly better overall behavior of the LM version, except for scenarios with very high shape parameter ($\xi = 0.4$, right column of Figure 4). Rejections rates based on the Hosking-Wallis criteria are shown in the bottom two rows. Quite remarkably, the PML and LM bootstrap tests outperform or can at least compete with the Hosking-Wallis criteria, since for the former tests the power is higher or at least comparable. Especially in scenarios where both the shape and dispersion parameters of the deviating group deviate toward the same direction (i.e., when c_ξ and c_σ have opposite signs; note that decreasing c_σ increases the dispersion).

5.4 | Dimension versus record length

In practice, homogeneity tests will typically be applied in situations where one is a priori expecting some form of homogeneity. However, when the number of potentially homogenous stations d increases, it is more likely that the signal-to-noise ratio for detecting possible heterogeneities decreases, such that it is harder for any test to detect those heterogeneities. The tests will hence be less likely to be significant, resulting in nonrejection of the null-hypothesis and thus in treating the data as homogeneous.

To illustrate this issue, we performed some simulations specifically addressing the detection of heterogeneity when the number of stations d grows. To make the magnitudes of heterogeneity and thus, the rejections rates for different d , comparable, we proceed in the following way. For each value of d , we generate data that consists of either

- (a) two equally sized heterogeneous groups that are homogeneous within the groups;

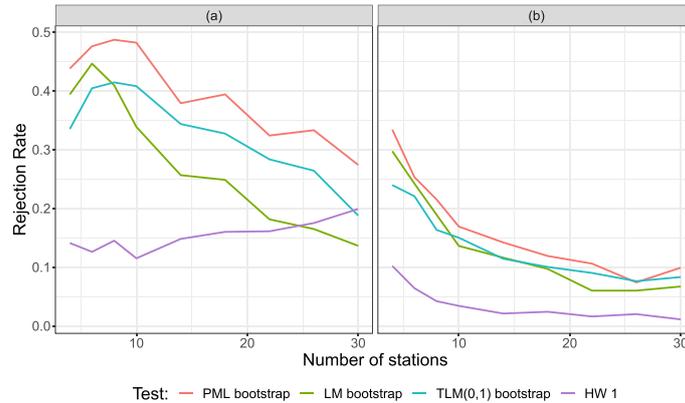


FIGURE 5 Rejection rates as a function of the dimension d , that is, the number of stations considered for the test, for scenarios (a) and (b) as described in the text

(b) one large homogeneous group with $d - 1$ stations and one single station whose distribution deviates from the large group.

More precisely, we started by simulating d -dimensional data of length $n = 75$ from a max-stable process with max-stable process parameters $\eta = 3$ and $\nu = 0.5$ and with unit Fréchet margins. For simplicity, we restrict attention to full record lengths, that is, no values are missing. Margins are then transformed to GEV margins, with constant location, scale and shape parameters within the groups. In the interest of mimicking a real-data scenario, the GEV parameters that were chosen for the latter step are based on clustering the stationwise dispersion and shape parameters of the winter maxima used in the case study of Section 6 into two clusters, and then computing the mean location and mean scale parameters within the found clusters. The two parameter vectors are $(50.28, 23.69, 0.27)$ and $(14.15, 8.29, 0.14)$ (i.e., the ratio of location and scale is approximately 1.7 in the first and 2.12 in the second group). Then the PML, LM, and TLM(0,1) bootstrap tests are applied, as well as the test based on HW's heterogeneity measure.

The resulting rejection rates as a function of dimension d are shown in Figure 5, with the left-hand side giving the results for scenario (a) and the right-hand side the results for scenario (b). It can be seen that the curves of the rejection rates are falling quite monotonically in scenario (b). Compared to scenario (a), the rejection rates are smaller overall, which also seems natural. However, we also find an approximately decreasing rejection rate in scenario (a) for the bootstrap based tests. This could be explained by the circumstance that the number of parameters that need to be estimated grows as $O(d^2)$. To sum up, the results indicate that the proposed tests should only be applied when the ratio of dimension and record length and hence the potential signal-to-noise ratio is moderate.

6 | CASE STUDY

The derived index flood tests were applied to flood peak data (maximal water discharge in m^3/s) observed at several gauging stations located in the Elbe river basin in Saxony, Germany. In total there are 26 gauging stations at 15 distinct waters, with record lengths varying from 64 to 103 years and catchment areas between 36 and 5433 km^2 . Since the basin is located at the north side of the Ore Mountains, the sites differ a lot in mean elevation, from a minimum of 168 m to a maximum of 1244 m above sea level. We perform a group-building process prior to applying the homogeneity tests. This has been done for two reasons: first, the index flood assumption is not expected to hold over the entire region, as we have quite heterogeneous site characteristics. Second, as argued in Section 5.4, the tests do not have high power in scenarios where the record length is small compared to the dimension. In fact, we will see below that even group-wise homogeneity is rejected in some of the groups, which naturally implies that the entire set of stations is heterogeneous as well.

For the group building, we chose the k -means clustering algorithm based on five standardized site characteristics: mean elevation, proportion of forest area, stream density, length of main watercourse and length of stream network. Note

TABLE 1 Cluster means of the different clustering variables (standardized)

	Group size	Mean elevation	Forest area prop.	Length main watercourse	Length stream network	Stream density
Group 1	9	-0.66	-0.63	-0.62	-0.50	0.22
Group 2	8	-0.32	-0.52	1.18	1.04	0.42
Group 3	9	0.94	1.09	-0.43	-0.43	-0.59

Note: Length of main watercourse, length of stream network and stream density are given as logarithmic values.

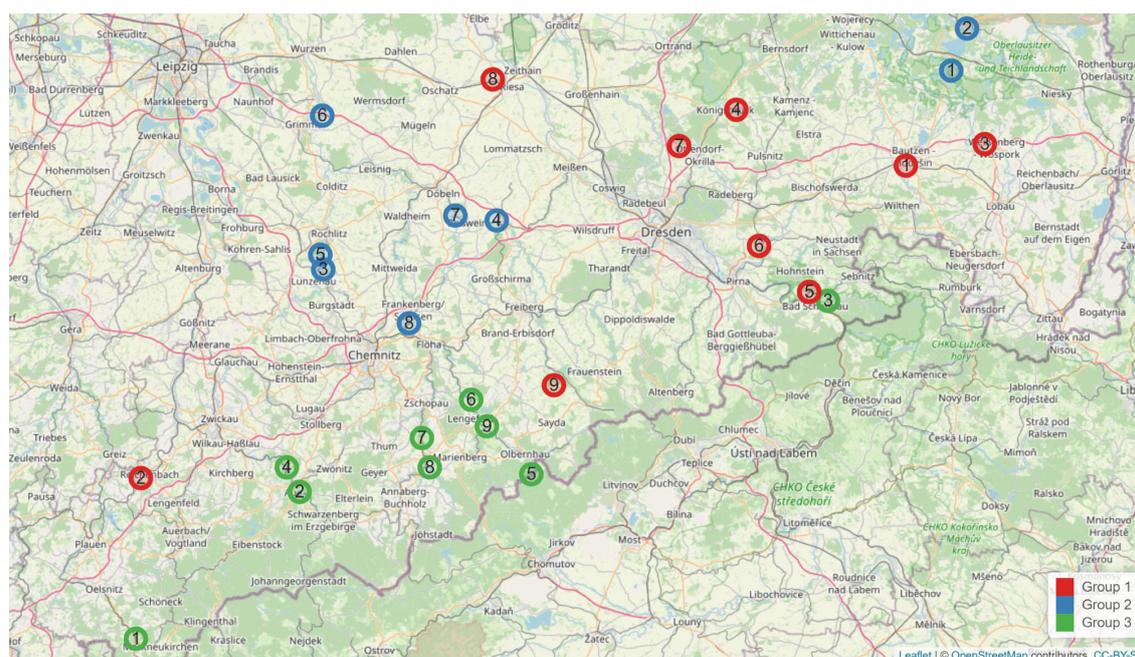


FIGURE 6 Locations of the 26 gauging stations colored by groups found by k-means clustering. Numbers within the circles indicate the numbering within groups, as treated in the parameter estimates in Tables 4 and 3

that both longitude and latitude were found to be highly correlated to mean elevation ($\rho = -0.66$ and $\rho = -0.85$, respectively), and were hence excluded from the group building process in order to guarantee (approximate) independence of the chosen explanatory variables. Based on visual inspection of the “elbow plot”, $k = 3$ clusters were chosen, with resulting cluster sizes and cluster means (of the standardized values) presented in Table 1. Figure 6 shows the locations of the stations and the belonging to groups.

Next, tests on homogeneity were performed within each of the three groups, and with candidate dependence models as in Section 5.3. Since monthly flood peaks are available, we split the analysis into the maxima of summer months (May–October) and winter months (November–April) of the (german) hydrological year. Since record lengths are small and the asymptotic tests did not achieve their theoretical level in the simulation studies of Section 5, we decided to only apply the bootstrap procedures. Since there was no scenario where the TLM(0, 1)-bootstrap test had more power than one of the LM- or PML-bootstrap test, we only compute p -values of the latter two. In every scenario, the bootstrap tests were carried out with $B = 500$ bootstrap repetitions.

The results are shown in Table 2. The PML-based test does not reject the hypothesis of homogeneity for any of the groups regarding winter maxima, while the LM-based test rejects the homogeneity assumption at the 5% level in groups 2 and 3. The Hosking Wallis heterogeneity measure comes to the same conclusion as the LM-based test (version HW 1 from Section 5.2). For summer maxima, both bootstrap tests agree that, at the 5% level, the index flood assumption does not hold in group 1, while it cannot be rejected in group 2. In group 3 the PML-based test rejects homogeneity at the 5%

TABLE 2 Row 2 and 3: p -values of the bootstrap-based homogeneity test for winter (summer) maximal flood peaks based on 500 bootstrap replications, for the PML- and the LM-based method, with significant values at the 5% level in boldface. Row 4: values of the Hosking Wallis homogeneity statistic; values above 1 in boldface

test	Group 1	Group 2	Group 3
boot-PML	0.685 (0.018)	0.134 (0.327)	0.271 (0.034)
boot-LM	0.186 (0.045)	0.010 (0.110)	0.024 (0.073)
HW	-2.92 (-0.03)	1.18 (0.42)	2.94 (0.27)

TABLE 3 Obtained local parameter estimates for **winter** maxima of flood peaks in the Elbe basin data, along with parameter estimation under the constraint of homogeneity (column H_0 est.)

Group	Parameter	est.	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	Site 9	H_0 est.	wss
1	Dispersion	PML	2.05	1.99	1.82	2.11	2.23	1.98	1.96	2.10	2.13	2.03	0.10
		LM	1.96	1.97	1.71	1.97	2.11	1.91	1.83	2.15	2.18	2.00	0.16
	Loc	PML	7.73	19.65	15.55	293.08	111.54	40.44	3.96	46.83	7.93		
		LM	7.81	19.59	15.80	298.68	113.14	40.61	4.04	46.24	7.85		
	Scale	PML	3.76	9.87	8.55	138.76	50.10	20.39	2.02	22.33	3.72		
		LM	3.98	9.93	9.22	151.45	53.54	21.23	2.20	21.47	3.61		
Shape	PML	0.23	0.11	0.04	0.20	0.29	0.11	0.20	0.18	0.20	0.18	0.04	
	LM	0.17	0.11	-0.03	0.11	0.22	0.08	0.10	0.22	0.23	0.14	0.05	
2	Dispersion	PML	1.61	1.77	2.29	2.34	2.34	2.03	2.23	1.68		1.95	0.66
		LM	1.48	1.66	2.41	2.39	2.23	1.97	2.08	1.59		2.07	0.91
	Loc	PML	13.10	6.23	52.92	31.90	18.03	16.49	47.66	13.30			
		LM	13.76	6.29	52.09	31.58	18.13	16.50	48.62	13.45			
	Scale	PML	8.14	3.52	23.15	13.65	7.70	8.12	21.42	7.92			
		LM	9.29	3.78	21.65	13.23	8.15	8.37	23.34	8.46			
Shape	PML	0.26	0.02	0.29	0.33	0.07	0.09	0.36	0.05		0.19	0.12	
	LM	0.09	-0.03	0.36	0.36	0.03	0.07	0.25	-0.01		0.24	0.23	
3	Dispersion	PML	1.64	2.09	1.98	2.09	2.33	2.19	1.80	1.59	1.84	1.90	0.42
		LM	1.56	1.97	1.93	1.98	2.29	2.10	1.71	1.50	1.70	1.88	0.43
	Loc	PML	5.12	37.68	115.86	35.21	7.74	10.41	20.00	25.92	13.99		
		LM	5.24	38.22	116.66	35.84	7.73	10.51	20.17	26.06	14.44		
	Scale	PML	3.13	18.04	58.46	16.81	3.33	4.76	11.13	16.30	7.62		
		LM	3.36	19.42	60.57	18.10	3.37	5.01	11.78	17.38	8.49		
Shape	PML	0.42	0.12	0.28	0.32	0.23	0.28	0.11	-0.01	0.24	0.23	0.10	
	LM	0.32	0.05	0.23	0.23	0.21	0.22	0.06	-0.05	0.11	0.18	0.09	

Note: The last column shows the sum of weighted squared deviations of local parameter estimates from the estimated value under the homogeneity assumption, where the weights are the ratios of local sample lengths and maximal sample length of the group.

TABLE 4 Obtained local parameter estimates for **summer** maxima of flood peaks in the Elbe basin data, along with parameter estimation under the constraint of homogeneity (column H_0 est.)

Group	Estimated.par	Estimator	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	Site 9	H_0 est.	wss
1	Dispersion	PML	1.59	1.54	1.30	1.48	1.74	1.79	1.76	1.55	1.57	1.56	0.18
		LM	1.62	1.54	1.22	1.41	1.67	1.87	1.79	1.58	1.54	1.53	0.31
	loc	PML	5.30	12.65	9.44	185.96	95.29	36.97	2.61	31.48	6.87		
		LM	5.24	12.61	9.87	190.31	96.76	36.34	2.57	30.97	6.86		
	scale	PML	3.32	8.22	7.27	125.47	54.84	20.71	1.48	20.36	4.39		
		LM	3.24	8.20	8.08	134.89	57.78	19.41	1.44	19.55	4.44		
	shape	PML	0.47	0.29	0.48	0.47	0.44	0.22	0.27	0.31	0.35	0.36	0.08
		LM	0.48	0.29	0.34	0.37	0.37	0.28	0.32	0.36	0.34	0.36	0.03
2	Dispersion	PML	1.37	1.38	1.56	1.52	1.42	1.64	1.52	1.15		1.42	0.15
		LM	1.27	1.28	1.48	1.43	1.37	1.58	1.43	1.09		1.41	0.16
	Loc	PML	7.26	4.40	43.89	26.19	10.78	12.71	35.64	8.20			
		LM	7.67	4.65	45.43	27.04	11.01	12.96	36.98	8.67			
	Scale	PML	5.30	3.18	28.11	17.26	7.60	7.76	23.47	7.13			
		LM	6.03	3.64	30.78	18.87	8.01	8.22	25.92	7.98			
	Shape	PML	0.54	0.30	0.51	0.49	0.35	0.40	0.45	0.55		0.43	0.06
		LM	0.37	0.14	0.38	0.37	0.27	0.32	0.33	0.39		0.35	0.05
3	Dispersion	PML	1.72	1.42	1.44	1.64	1.98	1.74	1.29	1.26	1.50	1.52	0.36
		LM	1.76	1.40	1.36	1.52	1.96	1.71	1.20	1.16	1.50	1.40	0.55
	Loc	PML	2.77	21.55	73.00	25.76	6.88	12.85	12.76	18.00	11.09		
		LM	2.75	21.75	75.80	26.85	6.89	12.91	13.76	19.57	11.03		
	Scale	PML	1.61	15.15	50.53	15.73	3.48	7.40	9.85	14.31	7.41		
		LM	1.56	15.52	55.55	17.63	3.51	7.54	11.45	16.95	7.37		
	Shape	PML	0.41	0.57	0.49	0.48	0.41	0.40	0.59	0.40	0.37	0.44	0.04
		LM	0.44	0.52	0.37	0.33	0.38	0.36	0.37	0.18	0.36	0.38	0.05

Note: The last column shows the sum of weighted squared deviations of local parameter estimates from the estimated value under the homogeneity assumption, where the weights are the ratios of local sample lengths and maximal sample length of the group.

level, while the LM-based test's p -value slightly exceeds 0.05. The Hosking Wallis heterogeneity measure, on the other hand, does not find any evidence against homogeneity.

The obtained results may be interpreted in light of the findings of the simulation experiments. First of all, the Hosking Wallis procedure was overall found to be most conservative, and, indeed, any of its rejections is also detected by the LM-based test. The different findings between the LM- and PML-based tests may partly be explained by the different power properties uncovered in Section 5.3; namely, that the LM-based bootstrap test outperforms the PML-based one when shapes are moderately high. Taking a closer look at the local parameter estimates for winter maxima (Table 3) reveals that the data might fall into this scenario. On the other hand, for summer maxima and group 3, the PML-based bootstrap is the only test that detects heterogeneity. Again, the parameter estimates in Table 4 provide a possible explanation: the estimated local shape parameters are rather high, exceeding 0.4 almost everywhere. This is exactly the scenario in which the PML-bootstrap detected heterogeneity more often during the simulation study than the LM-bootstrap did. Last but not least, the fact that HW 1 detects heterogeneity in group 2 for winter maxima while PML does not can be explained by the simulations as well, as this behavior was found particularly in cases where the reference shape parameter is 0 or 0.2 (compare the bottom left and top right corners of the 5×7 grids in the first and fourth rows in Figure 4 for the columns were $\xi = 0$ and $\xi = 0.2$). Only the PML-based bootstrap not detecting

heterogeneity in group 3 for winter maxima lacks justification, since here also HW 2 would reject the homogeneity assumption.

7 | CONCLUSION

Many applications of GEV models in environmental statistics are based on small data sets where parameter estimates suffer from large variance. Under appropriate homogeneity assumptions, the variance can be reduced by pooling observations from different variables. Such assumptions should be checked beforehand, and in this article, we have addressed respective statistical tests that are based on large sample theory for a variety of classical parametric estimators. The large sample theory was summarized in several mathematical theorems, which were augmented by consistency results on estimators for the estimation variance. The main findings regarding practical applications can be summarized as follows:

- The tests based on critical values derived from large sample theory were found to be unreasonably liberal in simulation experiments involving small to moderate sample sizes. Their practical use cannot be recommended for such sample sizes.
- Fortunately, applying the tests with modified critical values based on a parametric bootstrap device was found to provide accurate approximations of the nominal level, even for small sample sizes.
- Quite naturally, the tests' power was found to be decreasing in the number of variables, and for typical sample lengths of $n \leq 100$ it is not advisable to apply the modified tests to more than 10 variables.
- Within a case study on flood peak data, the number of variables was first reduced by a clustering algorithm applied to site characteristics such that the requirement from the previous bullet point was met. After that reduction, the tests were able to detect heterogeneity within some of the groups.

Additional information and supporting material for this article is available online at the journal's website.

ACKNOWLEDGMENTS

This work has been supported by the integrated project "Climate Change and Extreme Events - ClimXtreme Module B - Statistics (subproject B3.3)" funded by the German Federal Ministry of Education and Research (BMBF), which is gratefully acknowledged. Computational infrastructure and support were provided by the Centre for Information and Media Technology at Heinrich Heine University Düsseldorf. The authors are grateful to Paul Kinsvater for helpful discussions, and to two unknown referees, an associate editor and an editor for their constructive comments on an earlier version of this article which helped to improve the presentation substantially. Open Access funding enabled and organized by Projekt DEAL.

DATA AVAILABILITY STATEMENT

The R-Code used for the simulation study is available online at the journal's website. The data used in the case study is not open access but can be made available upon request.

ORCID

Axel Bücher  <https://orcid.org/0000-0002-1947-1617>

Roland Fried  <https://orcid.org/0000-0002-9830-9713>

REFERENCES

- Beirlant, J., Goegebeur, Y., Teugels, J., & Segers, J. (2004). *Statistics of extremes*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester. Theory and applications, With contributions from Daniel De Waal and Chris Ferro. <https://doi.org/10.1002/0470012382>
- Blöschl, G., Hall, J., Parajka, J., Perdigão, R. A. P., Merz, B., Arheimer, B., Aronica, G. T., Bilibashi, A., Bonacci, O., Borga, M., Čanjevac, I., Castellarin, A., Chirico, G. B., Claps, P., Fiala, K., Frolova, N., Gorbachova, L., Gül, A., Hannaford, J., ... Živković, N. (2017). Changing climate shifts timing of european floods. *Science*, 357(6351), 588–590. <https://doi.org/10.1126/science.aan2506>
- Boukhefifa, M., Meddi, M., & Gaume, E. (2018). Integrated Bayesian estimation of intensity-duration- frequency curves: Consolidation and extensive testing of a method. *Water Resources Research*, 54, 7459–7477.

- Bücher, A., & Segers, J. (2017). On the maximum likelihood estimator for the generalized extreme-value distribution. *Extremes*, 20(4), 1–34.
- Chebana, F., & Ouarda, T. B. (2007). Multivariate l-moment homogeneity test. *Water Resources Research*, 43(8).
- Chebana, F., & Ouarda, T. B. (2009). Index flood-based multivariate regional frequency analysis. *Water Resources Research*, 45(10).
- Cooley, D., & Naveau, P. (2021). Editorial to the special issue: Statistical modeling of environmental extremes. *Extremes*, 24(2), 197–198.
- Dalrymple, T. (1960). *Flood-frequency analyses, manual of hydrology: Part 3* (Technical report), USGPO.
- Davison, A. C., & Gholamrezaee, M. M. (2012). Geostatistics of extremes. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 468(2138), 581–608.
- Davison, A. C., Padoan, S. A., & Ribatet, M. (2012). Statistical modeling of spatial extremes. *Statistical Science*, 27(2), 161–186.
- Dombry, C. (2015). Existence and consistency of the maximum likelihood estimators for the extreme value index within the block maxima framework. *Bernoulli*, 21(1), 420–436. <https://doi.org/10.3150/13-BEJ573>
- Elamir, E. A., & Seheult, A. H. (2003). Trimmed L-moments. *Computational Statistics & Data Analysis*, 43(3), 299–314. [https://doi.org/10.1016/S0167-9473\(02\)00250-5](https://doi.org/10.1016/S0167-9473(02)00250-5)
- Fill, H. D., & Stedinger, J. R. (1995). Homogeneity tests based upon gumbel distribution and a critical appraisal of dalrymple's test. *Journal of Hydrology*, 166(1–2), 81–105.
- Gaume, E., Gaál, L., Viglione, A., Szolgay, J., Kohnová, S., & Blöschl, G. (2010). Bayesian MCMC approach to regional flood frequency analyses involving extraordinary flood events at ungauged sites. *Journal of Hydrology*, 394(1), 101–117. Flash Floods: Observations and Analysis of Hydrometeorological Controls <https://doi.org/10.1016/j.jhydrol.2010.01.008>
- Gumbel, E. J. (1958). *Statistics of extremes*. Columbia University Press.
- Hosking, J., & Wallis, J. (1993). Some statistics useful in regional frequency analysis. *Water Resources Research*, 29(2), 271–281. <https://doi.org/10.1029/92WR01980>
- Hosking, J. R. M. (2007). Some theory and practical uses of trimmed L-moments. *Journal of Statistical Planning and Inference*, 137(9), 3024–3039.
- Hosking, J. R. M., & Wallis, J. R. (1997). *Regional frequency analysis*. Cambridge.
- Hosking, J. R. M., Wallis, J. R., & Wood, E. F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, 27(3), 251–261. <https://doi.org/10.1080/00401706.1985.10488049>
- Jurado, O. E., Ulrich, J., Scheibel, M., & Rust, H. W. (2020). Evaluating the performance of a max-stable process for estimating intensity-duration-frequency curves. *Water*, 12(12), 3314. <https://doi.org/10.3390/w12123314>
- Koutsoyiannis, D., Kozonis, D., & Manetas, A. (1998). A mathematical framework for studying rainfall intensity-duration-frequency relationships. *Journal of Hydrology*, 206(1–2), 118–135.
- Landwehr, J. M., Matalas, N. C., & Wallis, J. R. (1979). Probability weighted moments compared with some traditional techniques in estimating Gumbel Parameters and quantiles. *Water Resources Research*, 15(5), 1055–1064. <https://doi.org/10.1029/WR015i005p01055>
- Lettenmaier, D. P., Wallis, J. R., & Wood, E. F. (1987). Effect of regional heterogeneity on flood frequency estimation. *Water Resources Research*, 23(2), 313–323. <https://doi.org/10.1029/WR023i002p00313>
- Lilienthal, J. (2019a). *Group-based regionalization in flood frequency analysis considering heterogeneity*. PhD thesis, Universitätsbibliothek Dortmund.
- Lilienthal, J. (2019b). *TLMoments: Calculate TL-moments and convert them to distribution parameters*. R package version 0.7.5.
- Lilienthal, J., Kinsvater, P., & Fried, R. (2016). *On the method of probability weighted moments in regional frequency analysis* (SFB 823 Discussion Paper, 63/2016).
- Lu, L. H., & Stedinger, J. R. (1992). Variance of two- and three-parameter GEV/PWM quantile estimators: Formulae, confidence intervals, and a comparison. *Journal of Hydrology*, 138(1–2), 247–267. [https://doi.org/10.1016/0022-1694\(92\)90167-T](https://doi.org/10.1016/0022-1694(92)90167-T)
- Masselot, P., Chebana, F., & Ouarda, T. B. (2017). Fast and direct nonparametric procedures in the l-moment homogeneity test. *Stochastic Environmental Research and Risk Assessment*, 31(2), 509–522.
- Nguyen, C., Gaume, E., & Payrastre, O. (2014). Regional flood frequency analyses involving extraordinary flood events at ungauged sites: further developments and validations. *Journal of Hydrology*, 508, 385–396. <https://doi.org/10.1016/j.jhydrol.2013.09.058>
- Padoan, S. A., Ribatet, M., & Sisson, S. A. (2010). Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association*, 105(489), 263–277. <https://doi.org/10.1198/jasa.2009.tm08577>
- Renard, B., Garreta, V., & Lang, M. (2006). An application of Bayesian analysis and Markov chain Monte Carlo methods to the estimation of a regional trend in annual maxima. *Water Resources Research*, 42(12).
- Ribatet, M. (2020). *SpatialExtremes: Modelling spatial extremes*. R package version 2.0-9.
- Schlather, M. (2002). Models for stationary max-stable random fields. *Extremes*, 5(1), 33–44. <https://doi.org/10.1023/A:1020977924878>
- Šimková, T. (2017). Homogeneity testing for spatially correlated data in multivariate regional frequency analysis. *Water Resources Research*, 53(8), 7012–7028.
- Šimková, T. (2018). L-moment homogeneity test in trivariate regional frequency analysis of extreme precipitation events. *Meteorological Applications*, 25(1), 11–22.
- Stephenson, A. G. (2018). *evd: Extreme value distributions*. R package version, 2.3-3.
- van der Vaart, A. W. (1998). *Asymptotic statistics* Cambridge Series in Statistical and Probabilistic Mathematics (Vol. 3). Cambridge University Press.
- Viglione, A. (2012). *homtest: Homogeneity tests for regional frequency analysis*. R package version 1.0-5.

- Viglione, A., Laio, F., & Claps, P. (2007). A comparison of homogeneity tests for regional frequency analysis. *Water Resources Research*, 43(3).
- Viglione, A., Merz, R., Salinas, J. L., & Blöschl, G. (2013). Flood frequency hydrology: 3. A Bayesian analysis. *Water Resources Research*, 49(2), 675–692.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Lilienthal, J., Zanger, L., Bücher, A., & Fried, R. (2022). A note on statistical tests for homogeneities in multivariate extreme value models for block maxima. *Environmetrics*, 33(7), e2746. <https://doi.org/10.1002/env.2746>

Supplement to the Paper: A note on statistical tests for homogeneities in multivariate extreme value models for block maxima

Jona Lilienthal, Leandra Zanger,
Roland Fried, Axel Bücher

May 25, 2022

A Re-parametrization of the GEV distribution by TL-moments

This section recaps the equation systems used to calculate GEV parameters from TL(0,0)- and TL(0,1)-moments, respectively. We also present the corresponding Jacobi matrices involved in the limiting covariance matrix of Corollary 2.4.

TL(0,0)

Let $\boldsymbol{\vartheta} = (\mu, \sigma, \xi)'$ with $\xi < 1$ and $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)'$ denote parameters and untrimmed L-moments of a GEV distribution, respectively. Hosking, Wallis, and Wood (1985) proved

that $\boldsymbol{\vartheta} = \varphi^{(0,0)}(\boldsymbol{\lambda})$, where $\varphi^{(0,0)}$ is implicitly defined by equation system

$$\begin{cases} \frac{2 \cdot 3^\xi - 3 \cdot 2^\xi + 1}{2^\xi - 1} = \frac{\lambda_3}{\lambda_2} \\ \sigma = \frac{\lambda_2 \xi}{\Gamma(1-\xi)(2^\xi - 1)} \\ \mu = \lambda_1 + \frac{\sigma}{\xi}(1 - \Gamma(1 - \xi)) \end{cases}, \quad (\text{A.1})$$

with Γ denoting the gamma function. However, there is no explicit expression for $\varphi^{(0,0)}$ as a function of $\boldsymbol{\lambda}$. Practitioners thus commonly replace the first line by

$$\xi = -7.859z - 2.9554z^2, \quad z = \frac{2}{3 + \lambda_3/\lambda_2} - \frac{\log 2}{\log 3}$$

based on a second order polynomial approximation in order to obtain an explicit solution. Slightly abusing notation, we denote the resulting function by $\varphi^{(0,0)}$ as well. Accordingly the Jacobi matrix $\frac{\partial}{\partial \boldsymbol{\lambda}} \varphi^{(0,0)}(\boldsymbol{\lambda})$ involved in the asymptotic distribution of L-moment estimators is approximated by that of the explicit solution. For the latter we obtain

$$A = \begin{pmatrix} 1 & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{pmatrix} \quad (\text{A.2})$$

with

$$\begin{aligned} a_{12} &= \frac{\log(2) \lambda_2 (\Gamma(1-\pi) - 1) 2^\pi \rho \theta}{\Gamma(1-\pi) (1-2^\pi)^2} + \frac{\lambda_2 \psi_0 (1-\pi) (\Gamma(1-\pi) - 1) \rho \theta}{\Gamma(1-\pi) (1-2^\pi)} - \\ &\quad \frac{\lambda_2 \psi_0 (1-\pi) \rho \theta}{1-2^\pi} + \frac{\Gamma(1-\pi) - 1}{\Gamma(1-\pi) (1-2^\pi)} \\ a_{13} &= - \frac{\log(2) \lambda_2^2 (\Gamma(1-\pi) - 1) 2^{\pi+1} \rho \zeta^2}{\Gamma(1-\pi) (1-2^\pi)^2} - \\ &\quad \frac{2 \lambda_2^2 \psi_0 (1-\pi) (\Gamma(1-\pi) - 1) \rho \zeta^2}{\Gamma(1-\pi) (1-2^\pi)} + \frac{2 \lambda_2^2 \psi_0 (1-\pi) \rho \zeta^2}{1-2^\pi} \\ a_{22} &= - \frac{\log(2) \lambda_2 \pi 2^\pi \rho \theta}{\Gamma(1-\pi) (1-2^\pi)^2} - \frac{\lambda_2 \rho \theta (\psi_0 (1-\pi) \pi + 1)}{\Gamma(1-\pi) (1-2^\pi)} - \frac{\pi}{\Gamma(1-\pi) (1-2^\pi)} \\ a_{23} &= \frac{\log(2) \lambda_2^2 \pi 2^{\pi+1} \rho \zeta^2}{\Gamma(1-\pi) (1-2^\pi)^2} + \frac{2 \lambda_2^2 \rho \zeta^2 (\psi_0 (1-\pi) \pi + 1)}{\Gamma(1-\pi) (1-2^\pi)} \\ a_{32} &= -2 \lambda_3 (2b \kappa \lambda_3 - a \lambda_3 + 6b \kappa \lambda_2 - 4b \lambda_2 - 3a \lambda_2) \zeta^3 \end{aligned}$$

$$a_{33} = 2\lambda_2 (2b\kappa\lambda_3 - a\lambda_3 + 6b\kappa\lambda_2 - 4b\lambda_2 - 3a\lambda_2) \zeta^3,$$

where

$$\begin{aligned} a &= -7.859, & b &= -2.9554, & \kappa &= \log 2 / \log 3, \\ \eta &= (2\lambda_2\zeta - \kappa), & \pi &= b\eta^2 + a\eta, & \zeta &= 1/(\lambda_3 + 3\lambda_2), \\ \theta &= (2\zeta - 6\lambda_2\zeta^2), & \rho &= 2b\eta + a, & \psi_0(x) &= \Gamma'(x)/\Gamma(x). \end{aligned}$$

TL(0,1)

Considering trimmed L-moments $\boldsymbol{\lambda}^{(0,1)}$ of a GEV distribution with parameters $\boldsymbol{\vartheta}$ we have $\boldsymbol{\vartheta} = \varphi^{(0,1)}(\boldsymbol{\lambda}^{(0,1)})$, where $\varphi^{(0,1)}$ is implicitly defined by

$$\left\{ \begin{array}{l} \frac{5.4\xi - 12.3\xi + 9.2\xi - 2}{3\xi - 2\xi^{+1} + 1} = \frac{9\lambda_3^{(0,1)}}{4\lambda_2^{(0,1)}} \\ \sigma = \frac{2 \cdot \lambda_2^{(0,1)}}{3\Gamma(-\xi) \cdot (3\xi - 2\xi^{+1} + 1)} \\ \mu = \lambda_1^{(0,1)} + \frac{\sigma}{\xi} - \frac{\sigma \cdot \Gamma(-\xi)}{(2\xi - 2)^{-1}} \end{array} \right. \quad (\text{A.1})$$

In order to obtain an explicit solution, the first line can be replaced by a second order polynomial approximation

$$\xi = -8.5674z + 0.6760z^2, \quad z = \frac{10}{9} \frac{\lambda_2^{(0,1)}}{2\lambda_2^{(0,1)} + \lambda_3^{(0,1)}} - \frac{2 \log 2 - \log 3}{3 \log 3 - 2 \log 4}.$$

Slightly abusing notation and denoting the resulting function by $\varphi^{(0,1)}$ as well, the Jacobi matrix $\frac{\partial}{\partial \lambda^{(0,1)}} \varphi^{(0,1)}(\boldsymbol{\lambda}^{(0,1)})$ is approximated by

$$A = \begin{pmatrix} 1 & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{pmatrix}, \quad (\text{A.2})$$

where

$$\begin{aligned}
a_{12} &= -\frac{2\lambda_2^{(0,1)}\psi_0(-\pi)\left(-2b(\zeta-\eta)\left(\lambda_2^{(0,1)}\zeta-\kappa\right)-a(\zeta-\eta)\right)+2}{3\pi(-2^{\pi+1}+3^\pi+1)\gamma(-\pi)} \\
&\quad -\frac{2\lambda_2^{(0,1)}(\log(3)\iota 3^\pi-\log(2)\iota 2^{\pi+1})(1-(2^\pi-2)\pi\gamma(-\pi))}{3\pi(-2^{\pi+1}+3^\pi+1)^2\gamma(-\pi)} \\
&\quad -\frac{2\iota\lambda_2^{(0,1)}}{3\pi^2(-2^{\pi+1}+3^\pi+1)\gamma(-\pi)}-\frac{\log(2)\iota\lambda_2^{(0,1)}2^{\pi+1}+2(2^\pi-2)}{3(-2^{\pi+1}+3^\pi+1)} \\
a_{13} &= -\frac{2\lambda_2^{(0,1)}(\log(2)2^{\pi+1}\rho-\log(3)3^\pi\rho)(1-(2^\pi-2)\pi\gamma(-\pi))}{3\pi(-2^{\pi+1}+3^\pi+1)^2\gamma(-\pi)} \\
&\quad -\frac{2\lambda_2^{(0,1)}\rho(\psi_0(-\pi)\pi-2^\pi\log(2)-1)}{3\pi^2(-2^{\pi+1}+3^\pi+1)\gamma(-\pi)} \\
a_{22} &= -\frac{2\lambda_2^{(0,1)}\psi_0(-\pi)\left(-2b(\zeta-\eta)\left(\lambda_2^{(0,1)}\zeta-\kappa\right)-a(\zeta-\eta)\right)+2}{3(-2^{\pi+1}+3^\pi+1)\gamma(-\pi)} \\
&\quad -\frac{2\lambda_2^{(0,1)}(\log(3)\iota 3^\pi-\log(2)\iota 2^{\pi+1})}{3(-2^{\pi+1}+3^\pi+1)^2\gamma(-\pi)} \\
a_{23} &= -\frac{2\lambda_2^{(0,1)}(\log(2)2^{\pi+1}\rho-\log(3)3^\pi\rho)}{3(-2^{\pi+1}+3^\pi+1)^2\gamma(-\pi)}-\frac{2\lambda_2^{(0,1)}\psi_0(-\pi)\rho}{3(-2^{\pi+1}+3^\pi+1)\gamma(-\pi)} \\
a_{32} &= \iota \\
a_{33} &= -\rho
\end{aligned}$$

with

$$\begin{aligned}
a &= -8.5674, & b &= 0.6760, \\
\kappa &= \frac{2\log 2 - \log 3}{3\log 3 - 2\log 4}, & \theta &= 3(\lambda_3^{(0,1)} + 2\lambda_2^{(0,1)}), \\
\zeta &= 10/(3\theta), & \eta &= 20\lambda_2^{(0,1)}/\theta^2, \\
\pi &= b(\lambda_2^{(0,1)}\zeta - \kappa)^2 + a(\lambda_2^{(0,1)}\zeta - \kappa), & \rho &= -b\eta(\lambda_2^{(0,1)}\zeta - \kappa), \\
\iota &= 2b(\zeta - \eta)(\lambda_2^{(0,1)}\zeta - \kappa) + a(\zeta - \eta), & \psi_0(x) &= \Gamma'(x)/\Gamma(x).
\end{aligned}$$

B Proofs

For sake of readability the proofs are given for $d = 2$ only. The derivation for arbitrary dimensions $d \geq 2$ can be established at the cost of a more complex notation but without additional technical difficulties.

Furthermore, we may and will assume the same beginnings and different end points, that is, we compute the statistics of interest from the variables $X_1, \dots, X_{\lfloor nc_1 \rfloor}$ and $Y_1, \dots, Y_{\lfloor nc_2 \rfloor}$, where (X_i, Y_i) , $i \geq 1$, is a sequence of independent and identically distributed bivariate vectors with margins $F(x) = P(X_i \leq x)$ and $G(y) = P(Y_i \leq y)$, respectively.

Proof of Theorem 2.1

The first R probability weighted moments of F and G are denoted by $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \dots, \alpha_{R-1})'$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_{R-1})'$, respectively, and we let $\boldsymbol{\gamma} = (\boldsymbol{\alpha}', \boldsymbol{\beta}')' \in \mathbb{R}^{2R}$. We set

$$\bar{\alpha}_{r,c_1,n} = \frac{1}{\lfloor nc_1 \rfloor} \sum_{i=1}^{\lfloor nc_1 \rfloor} X_i \cdot F_{\lfloor nc_1 \rfloor}^r(X_i), \quad \bar{\beta}_{r,c_2,n} = \frac{1}{\lfloor nc_2 \rfloor} \sum_{i=1}^{\lfloor nc_2 \rfloor} Y_i \cdot G_{\lfloor nc_2 \rfloor}^r(Y_i) \quad (\text{B.1})$$

with F_{n_1} (resp. G_{n_2}) denoting the empirical distribution function of the sample X_1, \dots, X_{n_1} (resp. Y_1, \dots, Y_{n_2}). All these components are collected in $\bar{\boldsymbol{\alpha}}_{c_1,n}, \bar{\boldsymbol{\beta}}_{c_2,n} \in \mathbb{R}^R$ and $\bar{\boldsymbol{\gamma}}_{c,n} = (\bar{\boldsymbol{\alpha}}'_{c_1,n}, \bar{\boldsymbol{\beta}}'_{c_2,n})' \in \mathbb{R}^{2R}$. The estimators of interest are denoted by

$$\hat{\alpha}_{r,c_1,n} = \frac{1}{\lfloor nc_1 \rfloor} \sum_{i=1}^{\lfloor nc_1 \rfloor} \left(\prod_{\ell=1}^r \frac{i-\ell}{\lfloor nc_1 \rfloor - \ell} \right) X_{(i)}, \quad \hat{\beta}_{r,c_2,n} = \frac{1}{\lfloor nc_2 \rfloor} \sum_{i=1}^{\lfloor nc_2 \rfloor} \left(\prod_{\ell=1}^r \frac{i-\ell}{\lfloor nc_2 \rfloor - \ell} \right) Y_{(i)}$$

and are collected in a vector $\hat{\boldsymbol{\gamma}}_{c,n} = (\hat{\boldsymbol{\alpha}}'_{c_1,n}, \hat{\boldsymbol{\beta}}'_{c_2,n})'$.

As a first step, we prove that

$$\hat{\boldsymbol{\gamma}}_{c,n} = \bar{\boldsymbol{\gamma}}_{c,n} + O_P(n^{-1}), \quad (\text{B.2})$$

which readily implies

$$\sqrt{n}(\hat{\gamma}_{\mathbf{c},n} - \gamma) = \sqrt{n}(\tilde{\gamma}_{\mathbf{c},n} - \gamma) + o_P(1).$$

For that purpose, it is sufficient to show that $\hat{\alpha}_{r,c_1,n} = \bar{\alpha}_{r,c_1,n} + O_P(n^{-1})$ for fixed $r \in \{0, \dots, R-1\}$. Writing $n_1 = \lfloor nc_1 \rfloor$ and observing that $F_{n_1}(X_{(i)}) = i/n_1$, we have the representation

$$\begin{aligned} \hat{\alpha}_{r,c_1,n} - \bar{\alpha}_{r,c_1,n} &= \frac{1}{n_1} \sum_{i=1}^{n_1} X_{(i)} \left\{ \left(\prod_{\ell=1}^r \frac{i-\ell}{n_1-\ell} \right) - F_{n_1}^r(X_{(i)}) \right\} \\ &= \left(\prod_{\ell=1}^r (n_1-\ell) \right)^{-1} \frac{1}{n_1} \sum_{i=1}^{n_1} X_{(i)} \left\{ \prod_{\ell=1}^r (i-\ell) - \left(\frac{i}{n_1} \right)^r \prod_{\ell=1}^r (n_1-\ell) \right\}, \end{aligned} \quad (\text{B.3})$$

We now show that the last formula can be expressed as a linear combination of the components in $\bar{\alpha}_{c_1,n}$, multiplied by a deterministic term of order $O(n^{-1})$. Since we show $\bar{\alpha}_{c_1,n} = O_P(1)$ later in the proof, this yields the assertion. Note that the two summands within the curly brackets on the right-hand side of (B.3) are polynomials of order r in i , both with leading term i^r . More precisely, we have, for certain coefficients $b_\ell \in \mathbb{N}, \ell \in \{1, \dots, r-1\}$, the representation

$$\begin{aligned} \prod_{\ell=1}^r (i-\ell) - \left(\frac{i}{n_1} \right)^r \prod_{\ell=1}^r (n_1-\ell) &= i^r + \sum_{\ell=1}^{r-1} b_\ell i^\ell - \frac{i^r}{n_1^r} \left(n_1^r + \sum_{\ell=1}^{r-1} b_\ell n_1^\ell \right) \\ &= \sum_{\ell=1}^{r-1} b_\ell i^\ell - i^r \sum_{\ell=1}^{r-1} b_\ell n_1^{\ell-r}. \end{aligned}$$

As a consequence, recalling that $\bar{\alpha}_{\ell,c_1,n} = \frac{1}{n_1} \sum_{i=1}^{n_1} X_{(i)} (i/n_1)^\ell$,

$$\hat{\alpha}_{r,c_1,n} - \bar{\alpha}_{r,c_1,n} = \frac{n_1^r}{\prod_{\ell=1}^r (n_1-\ell)} \left\{ \sum_{\ell=1}^{r-1} b_\ell n_1^{\ell-r} \bar{\alpha}_{\ell,c_1,n} - \left(\sum_{\ell=1}^{r-1} b_\ell n_1^{\ell-r} \right) \bar{\alpha}_{r,c_1,n} \right\},$$

which is of the desired form, and the proof of (B.2) is finished.

Next, let $\tilde{\alpha}_{c_1,n}, \tilde{\beta}_{c_2,n}$ and $\tilde{\gamma}_{\mathbf{c},n} = (\tilde{\alpha}'_{c_1,n}, \tilde{\beta}'_{c_2,n})'$ be defined analogously to the bar-versions in (B.1) but with $F_{\lfloor nc_1 \rfloor}$ and $G_{\lfloor nc_2 \rfloor}$ replaced by their true counterparts F and G , respectively.

We write

$$\sqrt{n}(\bar{\gamma}_{c,n} - \gamma) = \mathbf{Q}_{c,n} + \mathbf{\Delta}_{c,n}, \quad (\text{B.4})$$

where $\mathbf{Q}_{c,n} = \sqrt{n}(\tilde{\gamma}_{c,n} - \gamma)$ and $\mathbf{\Delta}_{c,n} = \sqrt{n}(\bar{\gamma}_{c,n} - \tilde{\gamma}_{c,n})$. The remainder of the proof is organized in the following three steps:

- a) Verify that $\mathbf{Q}_{c,n} \xrightarrow{D} \mathbf{Q}_c$, where the limit is a zero mean normally distributed random vector and show that the convergence holds jointly with that of the weighted empirical processes $\mathbb{U}_{c_1,n}$ and $\mathbb{V}_{c_2,n}$ defined below.
- b) Show that $\mathbf{\Delta}_{c,n} = \mathbf{R}_{c,n} + o_P(1)$ for $n \rightarrow \infty$, where all components of $\mathbf{R}_{c,n}$ can be represented as continuous functionals of either $\mathbb{U}_{c_1,n}$ or $\mathbb{V}_{c_2,n}$. Verify that $\mathbf{R}_{c,n}$ converges weakly towards a zero mean normally distributed random vector \mathbf{R}_c .
- c) Conclude that (B.4) is asymptotically normal with mean zero and compute the limiting variance matrix $\mathbf{\Xi}_c = \text{Var}(\mathbf{Q}_c + \mathbf{R}_c)$.

Step a) Let $\mathbb{U}_{c_1,n}$ and $\mathbb{V}_{c_2,n}$ be $\ell^\infty([0, 1])$ -valued processes defined by

$$\begin{aligned} \mathbb{U}_{c_1,n}(u) &= \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nc_1 \rfloor} \{\mathbb{1}(F(X_i) \leq u) - u\}}{\{u(1-u)\}^w}, \\ \mathbb{V}_{c_2,n}(v) &= \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nc_2 \rfloor} \{\mathbb{1}(G(Y_i) \leq v) - v\}}{\{v(1-v)\}^w} \end{aligned}$$

for $u, v \in [0, 1]$. These are called weighted empirical processes and their weak convergence is studied, e.g., in Genest and Segers (2009, Appendix G) and Kojadinovic and Naveau (2017, Appendix B) in a more general context. The weighting is needed for step b) of the proof in order to be able to express the components of $\mathbf{R}_{c,n}$ as continuous functionals of the empirical processes. Without loss of generality let $c_1 \leq c_2$ and note that

$$\mathbb{W}_{r,n} = (\mathbb{U}_{c_1,n}, \mathbb{V}_{c_2,n}) = (\mathbb{U}_{c_1,n}, \mathbb{V}_{c_1,n}) + (0, \mathbb{V}_{c_2,n} - \mathbb{V}_{c_1,n})$$

is a sum of two independent processes with $\mathbb{V}_{c_2,n} - \mathbb{V}_{c_1,n} \stackrel{D}{=} \mathbb{V}_{c_2-c_1,n} + o_P(1)$, where the remainder is due to the fact that $\lfloor nc_2 \rfloor - \lfloor nc_1 \rfloor \neq \lfloor n(c_2 - c_1) \rfloor$ in general. By the continuous mapping theorem and by Genest and Segers (2009, Th. G.1), both summands on the right-hand side of the previous equation converge weakly in $(\ell^\infty([0, 1]))^2$ towards centered Gaussian processes and, by independence of the summands, so does $\mathbb{W}_{c,n}$. Let \mathbb{W}_c denote the limiting process.

In almost the same manner we can write $\sqrt{n}(\tilde{\gamma}_{c,n} - \gamma)$ as a sum of two independent random vectors, where weak convergence of both summands towards centered normal distributions easily follows from the central limit theorem for sums of i.i.d. random vectors. The limit is denoted by $\sqrt{n}(\tilde{\gamma}_{c,n} - \gamma) \xrightarrow{D} \mathbf{Q}_c$. In fact, weak convergence of $\mathbb{W}_{c,n}$ and that of $\sqrt{n}(\tilde{\gamma}_{c,n} - \gamma)$ holds jointly as a random element in $(\ell^\infty([0, 1]))^2 \times \mathbb{R}^{2R}$. The only thing left to verify is that the finite dimensional convergence holds, which again follows from the central limit theorem for sums of i.i.d. random vectors.

Step b) Let $\mathbf{R}_{c,n} = (\mathbf{S}'_{c_1,n}, \mathbf{T}'_{c_2,n})'$ with $\mathbf{S}_{c_1,n} = (S_{0,c_1,n}, \dots, S_{R-1,c_1,n})'$,

$$S_{r,c_1,n} = \frac{1}{c_1} \int_{\mathbb{R}} x \cdot r \cdot F^{r-1}(x) \cdot \mathbb{U}_{c_1,n}(F(x)) \cdot \{F(x)(1 - F(x))\}^w dF(x)$$

and analogously define $\mathbf{T}_{c_2,n}$ but with (c_1, F, \mathbb{U}) replaced by (c_2, G, \mathbb{V}) . In order to show that $\mathbf{\Delta}_{c,n} = \mathbf{R}_{c,n} + o_P(1)$ for $n \rightarrow \infty$, it suffices to consider each component separately by proving

$$\sqrt{n}(\bar{\alpha}_{r,c_1,n} - \tilde{\alpha}_{r,c_1,n}) = S_{r,c_1,n} + o_P(1)$$

for each $r = 0, \dots, R - 1$ and analogously for the β -components. But this follows from (C.9) in the proof of Proposition C.2 in Kojadinovic and Naveau (2017).

Let $\varphi_r : \ell^\infty([0, 1]) \rightarrow \mathbb{R}$, $r = 0, \dots, R - 1$, be defined by

$$\varphi_r(g) = \int_{\mathbb{R}} x \cdot r \cdot F^{r-1}(x) \{F(x)(1 - F(x))\}^w \cdot g(F(x)) dF(x)$$

and note that $S_{r,c_1,n} = \varphi_r(\mathbb{U}_{c_1,n})$. Since $\sup_{x \in \mathbb{R}} |x \cdot r \cdot F^{r-1}(x) \{F(x)(1-F(x))\}^w| < \infty$ by assumption, it follows that φ_r is a continuous map. Similarly we can define continuous maps ψ_r , $r = 0, \dots, R-1$, such that $T_{r,c_2,n} = \psi_r(\mathbb{V}_{c_2,n})$. Bringing things together we conclude that $\mathbf{R}_{c,n} = \Psi(\mathbb{W}_{c,n}) \xrightarrow{D} \Psi(\mathbb{W}_{c,n}) = \mathbf{R}_c$, where $\Psi : (\ell^\infty[0,1])^2 \rightarrow \mathbb{R}^{2R}$ with

$$\Psi(f, g) = (\varphi_0(f), \dots, \varphi_{R-1}(f), \psi_0(g), \dots, \psi_{R-1}(g))'$$

is continuous. Since each component of $\mathbf{R}_{c,n}$ is a sum of i.i.d. zero-mean random variables with existing second moments, we conclude that the limit is a zero-mean normal distribution.

Step c) From steps a) and b) we obviously obtain the joint asymptotic normality of $\mathbf{Q}_{c,n}$ and $\mathbf{R}_{c,n}$. By the continuous mapping theorem we conclude that

$$\sqrt{n}(\bar{\gamma}_{c,n} - \gamma) \xrightarrow{D} \mathcal{N}(0, \mathbf{\Xi}_c) \quad \text{for } n \rightarrow \infty,$$

where $\mathbf{\Xi}_c = \text{Var}(\mathbf{Q}_c + \mathbf{R}_c)$. The calculation of the variance matrix is a simple exercise since each component of the random vector $\mathbf{Q}_{c,n} + \mathbf{R}_{c,n}$ is a sum of i.i.d. random variables and $\mathbf{\Xi}_c = \lim_{n \rightarrow \infty} \text{Var}(\mathbf{Q}_{c,n} + \mathbf{R}_{c,n})$. \square

Proof of Theorem 2.6.

Throughout the proof, the true parameter vector will be denoted by $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}'_1, \boldsymbol{\vartheta}'_2)'$. For simpler notation we further write $n_1 = \lfloor nc_1 \rfloor$ and $n_2 = \lfloor nc_2 \rfloor$ for $c_1, c_2 \in (0, 1]$, and without loss of generality we assume $c_1 < c_2$.

First note that our assumptions allow for an application of Proposition 3.3 in Bücher and Segers (2017) for each component of $\hat{\boldsymbol{\vartheta}}^{(\text{pml})}$. Therefore, consistency follows immediately from the fact that each $\hat{\boldsymbol{\vartheta}}_j^{(\text{pml})}$ is strongly consistent. Furthermore, the same proposition

yields the representation

$$\sqrt{n} \begin{pmatrix} \hat{\boldsymbol{\vartheta}}_1^{(\text{pml})} - \boldsymbol{\vartheta}_1 \\ \hat{\boldsymbol{\vartheta}}_2^{(\text{pml})} - \boldsymbol{\vartheta}_2 \end{pmatrix} = \begin{pmatrix} \sqrt{\frac{n}{n_1}} \sqrt{n_1} (\hat{\boldsymbol{\vartheta}}_1^{(\text{pml})} - \boldsymbol{\vartheta}_1) \\ \sqrt{\frac{n}{n_2}} \sqrt{n_2} (\hat{\boldsymbol{\vartheta}}_2^{(\text{pml})} - \boldsymbol{\vartheta}_2) \end{pmatrix} = A_{n1} + A_{n2} + o_P(1),$$

where

$$A_{n1} = \frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} \begin{pmatrix} \sqrt{\frac{n}{n_1}} g_{\boldsymbol{\vartheta}_1}(X_i) \\ \sqrt{\frac{nn_1}{n_2^2}} g_{\boldsymbol{\vartheta}_2}(Y_i) \end{pmatrix}, \quad A_{n2} = \frac{1}{\sqrt{n_2 - n_1}} \sum_{i=n_1+1}^{n_2} \begin{pmatrix} 0 \\ \sqrt{\frac{n(n_2 - n_1)}{n_2^2}} g_{\boldsymbol{\vartheta}_2}(Y_i) \end{pmatrix}$$

with $g_{\boldsymbol{\vartheta}_j}(\cdot) := I_{\boldsymbol{\vartheta}_j}^{-1} \dot{\ell}_{\boldsymbol{\vartheta}_j}(\cdot)$. Since $\dot{\ell}_{\boldsymbol{\vartheta}_j}$ is a score vector, we have $E[\dot{\ell}_{\boldsymbol{\vartheta}_1}(X_1)] = E[\dot{\ell}_{\boldsymbol{\vartheta}_2}(Y_1)] = \mathbf{0}_3$, whence A_{n1} is a sum of independent and identically distributed, centered random vectors.

The multivariate central limit theorem yields

$$A_{n1} \xrightarrow{D} \begin{pmatrix} \sqrt{\frac{1}{c_1}} Z_1 \\ \sqrt{\frac{c_1}{c_2^2}} Z_2 \end{pmatrix},$$

where $(Z_1, Z_2) \sim \mathcal{N}_6(0, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_{j,l})_{j,l=1}^2$ defined as

$$\begin{aligned} \boldsymbol{\Sigma}_{j,j} &= I_{\boldsymbol{\vartheta}_j}^{-1}, \quad j = 1, 2, \\ \boldsymbol{\Sigma}_{1,2} &= \boldsymbol{\Sigma}_{2,1} = I_{\boldsymbol{\vartheta}_1}^{-1} E \left[\dot{\ell}_{\boldsymbol{\vartheta}_1}(X_1) \left(\dot{\ell}_{\boldsymbol{\vartheta}_2}(Y_1) \right)' \right] I_{\boldsymbol{\vartheta}_2}^{-1}. \end{aligned} \tag{B.5}$$

By the same arguments,

$$A_{n2} \xrightarrow{D} \begin{pmatrix} 0 \\ \sqrt{\frac{c_2 - c_1}{c_2^2}} \tilde{Z}_2 \end{pmatrix},$$

where $\tilde{Z}_2 \sim \mathcal{N}_3(0, \boldsymbol{\Sigma}_{2,2})$ with $\boldsymbol{\Sigma}_{2,2}$ from (B.5). Since A_{n1} and A_{n2} are stochastically independent, so are the limits. Putting everything together yields the claimed limit distribution after a straightforward calculation. \square

Proof of Corollary 3.1.

Let

$$Z_{i,r,1} = X_i \cdot F^r(X_i) + \int_{\mathbb{R}} xrF^{r-1}(x)\mathbb{1}(X_i \leq x) dF(x),$$

$$Z_{i,\ell,2} = Y_i \cdot G^\ell(Y_i) + \int_{\mathbb{R}} y\ell G^{\ell-1}(y)\mathbb{1}(Y_i \leq y) dG(y)$$

for $r, \ell \in \mathbb{N}_0$, $i = 1, \dots, m$ and $m = \min\{\lfloor nc_1 \rfloor, \lfloor nc_2 \rfloor\}$. Slightly abusing notation, we further let $\hat{Z}_{i,r,1}$ (resp. $\hat{Z}_{i,\ell,2}$) be defined analogously with F (resp. G) replaced by its empirical counterpart $F_{\lfloor nc_1 \rfloor}$ (resp. $G_{\lfloor nc_2 \rfloor}$). We denote by $\tilde{\sigma}_{r,\ell,m}$ (resp. $\hat{\sigma}_{r,\ell,m}$) the empirical covariance of the bivariate sample $(Z_{i,r,1}, Z_{i,\ell,2})$, $i = 1, \dots, m$ (resp. $(\hat{Z}_{i,r,1}, \hat{Z}_{i,\ell,2})$, $i = 1, \dots, m$). From the strong law of large numbers we immediately obtain that $\tilde{\sigma}_{r,\ell,m} \xrightarrow{a.s.} \text{Cov}(Z_{1,r,1}, Z_{1,\ell,2})$ for $n \rightarrow \infty$. It thus remains to show that

$$|\hat{\sigma}_{r,\ell,m} - \tilde{\sigma}_{r,\ell,m}| \xrightarrow{P} 0 \text{ for } n \rightarrow \infty. \quad (\text{B.6})$$

To make a long story short, (B.6) follows from the consistency of probability weighted moments proven in Theorem 2.1, (C.12) in Kojadinovic and Naveau (2017) and from the consistency of the empirical process $\mathbb{W}_{c,n}$ defined in the proof of Theorem 2.1. A detailed presentation is omitted for the sake of brevity. \square

References

- Bücher, A., & Segers, J. (2017). On the maximum likelihood estimator for the Generalized Extreme-Value distribution. *Extremes*, 20(4), 1–34. <https://doi.org/10.1007/s10687-017-0292-6>
- Genest, C., & Segers, J. (2009). Rank-based inference for bivariate extreme-value copulas. *The Annals of Statistics*, 37(5B), 2990–3022. <https://doi.org/10.1214/08-AOS672>

- Hosking, J. R. M., Wallis, J. R., & Wood, E. F. (1985). Estimation of the Generalized Extreme-Value Distribution by the Method of Probability-Weighted Moments. *Technometrics*, *27*(3), 251–261. <https://doi.org/10.1080/00401706.1985.10488049>
- Kojadinovic, I., & Naveau, P. (2017). Detecting distributional changes in samples of independent block maxima using probability weighted moments. *Extremes*, *20*, 417–450. <https://doi.org/10.1007/s10687-016-0273-1>

Regional Pooling in Extreme Event Attribution Studies: an Approach Based on Multiple Statistical Testing

Leandra Zanger¹, Axel Bücher ^{*1}, Frank Kreienkamp², Philip Lorenz², and
Jordis Tradowsky²

¹Heinrich-Heine Universität Düsseldorf, Mathematisches Institut,
Düsseldorf, Germany

²Deutscher Wetterdienst, Regionales Klimabüro Potsdam, Potsdam,
Germany

January 5, 2023

Abstract

Statistical methods are proposed to select homogeneous locations when analyzing spatial block maxima data, such as in extreme event attribution studies. The methods are based on classical hypothesis testing using Wald-type test statistics, with critical values obtained from suitable parametric bootstrap procedures and corrected for multiplicity. A large-scale Monte Carlo simulation study finds that the methods are able to accurately identify homogeneous locations, and that pooling the selected locations improves the accuracy of subsequent statistical analyses. The approach is illustrated with a case study on precipitation extremes in Western Europe. The methods are implemented in an R package that allows easy application in future extreme event attribution studies.

Key words: Extreme event attribution; Extreme Value Statistics; Homogeneity Tests; Multiple Comparison Problem; Parametric Bootstrap; Max-Stable Processes.

1 Introduction

Extreme event attribution studies on precipitation extremes are typically motivated by the occurrence of an extreme event which causes major impacts such as damages to infrastructure and agriculture, or even fatalities, see, for instance, van der Wiel *et al.* (2017); van Oldenborgh *et al.* (2017); Otto *et al.* (2018); Kreienkamp *et al.* (2021). A key task for attributing the event to anthropogenic climate change consists of a statistical analysis of available observational data products at the location

*Corresponding author: axel.buecher@hhu.de

or region of interest (Philip *et al.*, 2020). Typically, the observed time period is short, often less than 100 years, which ultimately leads to large statistical uncertainties. One possibility to reduce those uncertainties is to incorporate observations from nearby locations/regions, given that their meteorological characteristics are sufficiently similar and governed by the same underlying processes to those from the region affected by an extreme event. The selection of surrounding areas for which these criteria are met can be based on expert knowledge of the meteorological characteristics and dynamics, for instance provided by experts from the national meteorological and hydrological service of the affected country, like the Deutsche Wetterdienst in Germany. The expert knowledge-based suggestion may next be assessed statistically, which, to the best of our knowledge, has been done based on ad hoc methods in the past. In this paper, we propose profound statistical methods that can complement the expert’s knowledge and which is based on statistically evaluating observational data from the past. Once regions with sufficiently similar characteristics of the analysed variable, e.g., the yearly maximum of daily rainfall, have been identified, the time series of all identified regions can be combined, thereby extending the available time series for the benefit of a more efficient statistical analysis.

The building blocks for the new approach are classical Wald-type tests statistics (Lehmann and Romano, 2021) for testing the null hypothesis that the temporal dynamics at multiple locations of interest are the same. Unlike in the classical textbook case, and motivated by the fact that standard likelihood-based inference for extreme value distributions requires unreasonably large sample sizes for sufficient finite-sample accuracy, we employ a parametric bootstrap device to approximate the distribution of the test statistics under the null hypothesis. This approach is motivated by results in Lilienthal *et al.*, 2022 for respective stationary extreme value models. Based on suitable decompositions of a global null hypothesis, we then propose to test for carefully selected sub-hypotheses, possibly after correcting the individual tests’ level for multiple comparisons.

The new methods are illustrated by a large-scale Monte Carlo simulation study and by an application to the severe flooding event in Western Europe during July 2021 for which spatial pooling was applied in an attribution study following the event (Kreienkamp *et al.*, 2021; Tradowsky *et al.*, 2022). For the benefit of researchers who would like to use this spatial pooling approach, an implementation of the method in the statistical programming environment R (R Core Team, 2022) is publicly available as an R package called `findpoolreg` on GitHub (Zanger, 2022).

Attribution analysis of precipitation extremes is especially challenging due to short observational time series as well as their often limited spatial extend, which further complicates the detection of a trend and estimation of return periods based on the limited time series (see Tradowsky *et al.*, 2022, for a discussion on this). Therefore, we will in the following present the suggested approach for a heavy rainfall event, however, the method could equally be applied to other parameters.

The remaining parts of this paper are organized as follows. Section 2 explains

the mathematical concept of the proposed methods, starting with a detailed description of the underlying model assumptions and a strategy for the detection of a possible pooling region in Section 2.1. It is recommended to all readers. In Sections 2.2 and 2.3, mathematical details on the applied estimators and test statistics are given, and they may be skipped by non-expert readers. Next, the ideas of the bootstrap procedures that allow to draw samples from the distribution under the null hypothesis are explained. Again, this may be skipped by non-statisticians. Section 2.5 goes into detail about the detection strategy of possible pooling regions and the treatment of the related multiple testing problem. This section is of interest to all readers who want to apply the methods, since it provides details on how to process a set of p-values obtained from testing multiple hypotheses. Next, Section 3 gives the results of the simulation study that was performed in order to evaluate the performance of the proposed methods. These results are of interest to all readers, and they serve as a basis for the case study conducted in Section 4. Section 5 then discusses several extensions of the proposed methods. In 5.1, we provide a method for estimating *region-wise* return periods once a pooling region has been chosen. Here, a region-wise return period of a given event is defined as the number of years that one has to wait on average until an event of the same or even greater magnitude is observed anywhere in the pooling region. Extensions to different model assumptions that suit e.g. other variables such as temperature are discussed in Section 5. Last but not least, we come to a conclusion in Section 6. Some mathematical details and further illustrations on the simulation study and the case study are postponed to a supplement.

2 Assessing spatial homogeneities for precipitation extremes

2.1 A homogeneous model for precipitation extremes

The observational data of interest consists of annual or seasonal maximal precipitation amounts (over some fixed time duration, e.g., a day) collected over various years and at various locations (in practice, each *location* may correspond to a spatial *region*; we separate these two terms from the outset to avoid misunderstandings: subsequently, a *region* shall be a set of locations). More precisely, we denote by $m_d^{(t)}$ the observed maximal precipitation amount in season t and at location d , with $t = 1, \dots, n$ and $d = 1, \dots, D$. The location of primary interest shall be the one with index $d = 1$. Note that the choice of $d = 1$ is made for illustrative purposes only and can be replaced by any index $d \in \{1, \dots, D\}$.

In view of the stochastic nature, we assume that $m_d^{(t)}$ is an observed value of some random variable $M_d^{(t)}$. Since $M_d^{(t)}$ is generated by a maxima operation, standard extreme value theory (Coles, 2001) suggests to assume that $M_d^{(t)}$ follows the generalized extreme value (GEV) distribution, i.e.,

$$M_d^{(t)} \sim \text{GEV}(\mu_d(t), \sigma_d(t), \gamma_d(t))$$

for some $\mu_d(t), \sigma_d(t) > 0, \gamma_d(t) \in \mathbb{R}$, where the $\text{GEV}(\mu, \sigma, \gamma)$ distribution with location parameter $\mu > 0$, scale parameter $\sigma > 0$ and shape parameter $\gamma \in \mathbb{R}$ is defined by its cumulative distribution function

$$G_{(\mu, \sigma, \gamma)}(x) = \exp \left\{ - \left(1 + \gamma \frac{x - \mu}{\sigma} \right)^{-1/\gamma} \right\} \quad (1)$$

for x such that $1 + \gamma \frac{x - \mu}{\sigma} > 0$. Due to climate change, the temporal dynamics at location d , which are primarily driven by the function $t \mapsto (\mu_d(t), \sigma_d(t), \gamma_d(t))$, are typically non-constant. Any proxy for climate change qualifies as a suitable temporal covariate, and a standard assumption in extreme event attribution studies, motivated by the Clausius–Clapeyron relation, postulates that

$$\mu_d(t) = \mu_d \exp \left(\frac{\alpha_d \text{GMST}'(t)}{\mu_d} \right), \quad \sigma_d(t) = \sigma_d \exp \left(\frac{\alpha_d \text{GMST}'(t)}{\mu_d} \right), \quad \gamma_d(t) = \gamma_d \quad (2)$$

for certain parameters $\alpha_d, \gamma_d \in \mathbb{R}, \mu_d, \sigma_d > 0$. Here, $\text{GMST}'(t)$ denotes the smoothed global mean surface temperature anomaly, see Philip *et al.* (2020). Note that (2) implies

$$\text{GEV}(\mu_d(t), \sigma_d(t), \gamma_d(t)) = \exp \left(\frac{\alpha_d \text{GMST}'(t)}{\mu_d} \right) \text{GEV}(\mu_d, \sigma_d, \gamma_d),$$

hence the model may be identified as a temporal scaling model. It is further assumed that any temporal dependence at location d is completely due to $\text{GMST}'(t)$, which we treat as deterministic and which implies that $M_d^{(1)}, \dots, M_d^{(n)}$ are stochastically independent, for each $d = 1, \dots, D$. For the moment, the spatial dependence will be left unspecified.

Recall that the location of interest is the one with $d = 1$, which is characterised by the four parameters $\mu_1, \sigma_1, \gamma_1, \alpha_1$. As described before, estimating those parameters based on the observations from location $d = 1$ only may be unpleasantly inaccurate, which is why one commonly assumes that the D locations have been carefully selected by experts to meet the following space-time homogeneity assumption:

$$H_0^{\text{ED}} : \quad \exists \boldsymbol{\vartheta} \in \Theta \quad \forall d \in \{1, \dots, D\} : \quad \boldsymbol{\vartheta}_d = \boldsymbol{\vartheta}, \quad (3)$$

where $\Theta := (0, \infty)^2 \times \mathbb{R}^2$ and $\boldsymbol{\vartheta} = (\mu, \sigma, \gamma, \alpha)^\top$, $\boldsymbol{\vartheta}_d = (\mu_d, \sigma_d, \gamma_d, \alpha_d)^\top$, and where the upper index ED stands for ‘equal distribution’, since, in short, Equation (3) states that the location-wise GEV parameters coincide for the D locations.

In the subsequent sections, we aim at testing the validity of the expert’s hypothesis H_0^{ED} . Here, it is not only of interest to test the hypothesis for the whole set $\{1, \dots, D\}$, but also to find a (maximal) subset $A \subset \{1, \dots, D\}$ with $1 \in A$ and $|A| = k \geq 2$ on which the space-time homogeneity assumption holds. Here, for an arbitrary index set A , the latter assumption may be expressed through

$$H_0^{\text{ED}}(A) : \quad \exists \boldsymbol{\vartheta}_A \in \Theta \quad \forall d \in A : \quad \boldsymbol{\vartheta}_d = \boldsymbol{\vartheta}_A, \quad (4)$$

with Θ as in Equation (3) and $\boldsymbol{\vartheta}_A = (\mu_A, \sigma_A, \gamma_A, \alpha_A)^\top$, meaning that the location-wise GEV parameters coincide for all locations with index in the set A , making the respective locations a possible pooling region.

Now, a maximal subset A for which Equation (4) holds may be determined with the following strategy: Since we are interested in finding all locations that ‘match’ the location of primary interest with index $d = 1$, we test for each pair $A_d = \{1, d\}, d = 2, \dots, D$, whether the null hypothesis $H_0^{\text{ED}}(A_d)$ holds. This will provide us with a set of p-values based on which we can decide which locations to reject and which not to reject. Those locations that are not rejected can then be assumed to be sufficiently homogeneous and are thus included in the suggestion of a pooling region of maximal extent. For further details on this strategy and the impact of the induced multiple testing problem, see Section 2.5.

2.2 Coordinate-wise maximum likelihood estimation

The starting point for the subsequent test statistics are the coordinate-wise maximum likelihood estimators for the model specified in (2). Writing $c^{(t)} = \text{GMST}'(t)$ for brevity, the log-likelihood contribution of observation $(M_d^{(t)}, c^{(t)})$ is given by $\ell_{\boldsymbol{\vartheta}_d}(M_d^{(t)}, c^{(t)})$, where

$$\ell_{\boldsymbol{\vartheta}_d}(x, c) = \log g_{(\mu_d \exp(\alpha_d c / \mu_d), \sigma_d \exp(\alpha_d c / \mu_d), \gamma_d)}(x) \quad (5)$$

with $g_{(\gamma, \mu, \sigma)}(x) = \frac{\partial}{\partial x} G_{(\mu, \sigma, \gamma)}(x)$ the probability density function of the GEV(μ, σ, γ)-distribution. The maximum likelihood estimator for $\boldsymbol{\vartheta}_d$ at location d is then defined as

$$\hat{\boldsymbol{\vartheta}}_d \in \operatorname{argmax}_{\boldsymbol{\vartheta}_d \in \Theta} \sum_{t=1}^n \ell_{\boldsymbol{\vartheta}_d}(M_d^{(t)}, c^{(t)}). \quad (6)$$

The arg-maximum cannot be calculated explicitly, but may be found by suitable numerical optimization routines. We denote the gradient and the Hessian matrix of $\boldsymbol{\vartheta} \mapsto \ell_{\boldsymbol{\vartheta}}(x, c)$ by $\dot{\ell}_{\boldsymbol{\vartheta}}(x, c) \in \mathbb{R}^4$ and $\ddot{\ell}_{\boldsymbol{\vartheta}}(x, c) \in \mathbb{R}^{4 \times 4}$, respectively. Under appropriate regularity assumptions, standard asymptotic expansions (van der Vaart, 1998, see also Bücher and Segers, 2017 for the stationary GEV family) imply that $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\vartheta}}_1^\top, \dots, \hat{\boldsymbol{\vartheta}}_D^\top)^\top \in \Theta^D$ is approximately Gaussian with mean $\boldsymbol{\theta} = (\boldsymbol{\vartheta}_1^\top, \dots, \boldsymbol{\vartheta}_D^\top)^\top$ and covariance $n^{-1} \boldsymbol{\Sigma}_n$, where $\boldsymbol{\Sigma}_n = (\boldsymbol{\Sigma}_{n;j,k})_{j,k=1}^D \in \mathbb{R}^{4D \times 4D}$ is defined as

$$\boldsymbol{\Sigma}_{n;j,k} = J_{n,j,\boldsymbol{\vartheta}_j}^{-1} \left(\frac{1}{n} \sum_{t=1}^n \operatorname{Cov} [\dot{\ell}_{\boldsymbol{\vartheta}_j}(M_j^{(t)}, c^{(t)}), \dot{\ell}_{\boldsymbol{\vartheta}_k}(M_k^{(t)}, c^{(t)})] \right) J_{n,k,\boldsymbol{\vartheta}_k}^{-1} \in \mathbb{R}^{4 \times 4} \quad (7)$$

with $J_{n,j,\boldsymbol{\vartheta}} = \frac{1}{n} \sum_{t=1}^n \mathbb{E}[\ddot{\ell}_{\boldsymbol{\vartheta}}(M_j^{(t)}, c^{(t)})] \in \mathbb{R}^{4 \times 4}$. See Appendix A.1 for details and Appendix A.2 for a suitable estimator $\hat{\boldsymbol{\Sigma}}_n$ for $\boldsymbol{\Sigma}_n$.

2.3 Wald-type test statistics

We define test statistics which allow to test for the sub-hypotheses $H_0^{\text{ED}}(A)$ of H_0^{ED} from Equation (4), where $A \subset \{1, \dots, D\}$. For that purpose, we propose to use classical Wald-type test statistics; see Section 14.4.2 in Lehmann and Romano (2021) for

a general discussion and Lilienthal *et al.* (2022) for a similar approach in temporally stationary GEV models, i.e., with α_d fixed to $\alpha_d = 0$.

Write $A = \{d_1, \dots, d_k\}$ with $1 \leq d_1 < \dots < d_k \leq D$ and let $h_A : \mathbb{R}^{4D} \rightarrow \mathbb{R}^{4(k-1)}$ be defined by

$$\begin{aligned} h_A(\boldsymbol{\theta}) &= h_A(\boldsymbol{\vartheta}_1, \dots, \boldsymbol{\vartheta}_D) = (\boldsymbol{\vartheta}_{d_1}^\top - \boldsymbol{\vartheta}_{d_2}^\top, \boldsymbol{\vartheta}_{d_2}^\top - \boldsymbol{\vartheta}_{d_3}^\top, \dots, \boldsymbol{\vartheta}_{d_{k-1}}^\top - \boldsymbol{\vartheta}_{d_k}^\top)^\top \\ &= (\mu_{d_1} - \mu_{d_2}, \sigma_{d_1} - \sigma_{d_2}, \gamma_{d_1} - \gamma_{d_2}, \alpha_{d_1} - \alpha_{d_2}, \\ &\quad \dots, \gamma_{d_{k-1}} - \gamma_{d_k}, \alpha_{d_{k-1}} - \alpha_{d_k})^\top. \end{aligned}$$

We may then write $H_0^{\text{ED}}(A)$ equivalently as

$$H_0^{\text{ED}}(A) : h_A(\boldsymbol{\theta}) = 0.$$

Hence, significant deviations of $h_A(\hat{\boldsymbol{\theta}})$ from 0 with $\hat{\boldsymbol{\theta}}$ from Section 2.2 provide evidence against $H_0^{\text{ED}}(A)$. Such deviations may be measured by the Wald-type test statistic

$$T_n(A) = n (h_A(\hat{\boldsymbol{\theta}}))^\top \left(\mathbf{H}_A \hat{\boldsymbol{\Sigma}}_n \mathbf{H}_A^\top \right)^{-1} h_A(\hat{\boldsymbol{\theta}}), \quad (8)$$

where $\mathbf{H}_A = \dot{h}_A(\boldsymbol{\theta}) \in \mathbb{R}^{4(k-1) \times 4D}$ denotes the Jacobian matrix of $\boldsymbol{\theta} \mapsto h_A(\boldsymbol{\theta})$, which is a matrix with entries in $\{-1, 0, 1\}$ that does not depend on $\boldsymbol{\theta}$. In view of the asymptotic normality of $\hat{\boldsymbol{\theta}}$, see Section 2.2, the asymptotic distribution of $T_n(A)$ under the null hypothesis $H_0^{\text{ED}}(A)$ is the chi-square distribution $\chi_{4(k-1)}^2$ with $4(k-1)$ degrees of freedom; see also Section 4 in Lilienthal *et al.* (2022). Hence, rejecting $H_0^{\text{ED}}(A)$ if $T_n(A)$ exceeds the $(1 - \alpha)$ -quantile of the $\chi_{4(k-1)}^2$ -distribution provides a statistical test of asymptotic level $\alpha \in (0, 1)$. The finite-sample performance of the related test in the stationary setting was found to be quite inaccurate (see Lilienthal *et al.*, 2022). To overcome this issue, we propose a suitable bootstrap scheme in the next section.

2.4 Parametric bootstrap devices for deriving p-values

Throughout this section, we propose two bootstrap devices that allow to simulate approximate samples from the $H_0^{\text{ED}}(A)$ -distribution of the test statistic $T_n(A)$ from Equation (8). Based on a suitably large set of such samples, one can compute a reliable p-value for testing $H_0^{\text{ED}}(A)$, even for short sample sizes.

The first method is based on a global fit of a max-stable process model to the entire region under consideration, while the second one is based on fitting multiple pairwise models. The main difference of the two approaches is that the first one can test the hypothesis $H_0^{\text{ED}}(A)$ for arbitrary subsets $A \subset \{1, \dots, D\}$, while the second approach is restricted to testing the null hypothesis on subsets of cardinality two, i.e., it can only test whether a pair of locations is homogeneous. Depending on the question that is asked, applying the one or the other method may be advantageous.

2.4.1 Global bootstrap based on max-stable process models

The subsequent bootstrap device is a modification of the parametric bootstrap procedure described in Section 5.3 of Lilienthal *et al.* (2022). Fix some large number B , say $B = 200$, noting that larger numbers are typically better, but going beyond $B = 1000$ is usually not worth the extra computational effort.

The basic idea is as follows: for each $b = 1, \dots, B$, simulate artificial *bootstrap samples*

$$\mathcal{D}_b^* = \{M_{d,b}^{(t),*} : t \in \{1, \dots, n\}, d \in \{1, \dots, D\}\}$$

that have a sufficiently similar spatial dependence structure as the observed data $\mathcal{D} = \{M_d^{(t)} : t \in \{1, \dots, T\}, d \in \{1, \dots, D\}\}$ and that satisfy the null hypothesis H_0^{ED} . For each fixed $A \subset \{1, \dots, D\}$ with $k = |A| \geq 2$, the test statistics computed on all bootstrap samples, say $(T_{n,b}^*(A))_{b=1, \dots, B}$, are then compared to the observed test statistic $T_n(A)$. Since the bootstrap samples do satisfy $H_0^{\text{ED}}(A)$, the observed test statistic $T_n(A)$ should differ significantly from the bootstrapped test statistics in case $H_0^{\text{ED}}(A)$ is not satisfied on the observed data.

Here, for simulating the bootstrap samples, we assume that the spatial dependence structure of the observed data can be sufficiently captured by a max-stable process model. Max-stable processes provide a natural choice here, since they are the only processes that can arise, after proper affine transformation, as the limit of maxima of independent and identically distributed random fields $\{Y_i(x) : x \in \mathbb{R}^p\}$ (Coles, 2001, Section 9.3). Parametric models for max-stable processes are usually stated for unit Fréchet (i.e., $\text{GEV}(1, 1, 1)$) margins. Therefore, the first steps in our algorithm below aim at transforming the margins of our observed data to be approximately unit Fréchet.

More precisely, the parametric bootstrap algorithm is defined as follows:

Algorithm 1 (Bootstrap based on max-stable processes).

- (1) For each $d \in \{1, \dots, D\}$, calculate $\hat{\boldsymbol{\theta}}_d$ from Section 2.2.
- (2) For each $d \in \{1, \dots, D\}$, transform the observations to approximately i.i.d. Fréchet-distributed data, by letting

$$Y_d^{(t)} = \left\{ 1 + \hat{\gamma}_d \frac{M_d^{(t)} - \hat{\mu}_d \exp\left(\frac{\hat{\alpha}_d \text{GMST}'(t)}{\hat{\mu}_d}\right)}{\hat{\sigma}_d \exp\left(\frac{\hat{\alpha}_d \text{GMST}'(t)}{\hat{\mu}_d}\right)} \right\}_+^{1/\gamma_d} \quad (t \in \{1, \dots, n\}). \quad (9)$$

- (3) Fit a set of candidate max-stable process models with standard Fréchet margins to the observations $(Y_1^{(t)}, \dots, Y_D^{(t)})_{t=1, \dots, n}$ and choose the best fit according to the composite likelihood information criterion (CLIC), which is a model selection criterion that is commonly applied when fitting max-stable process models. Throughout, we chose the following three models:
 - (a) Smith's model (3 parameters);
 - (b) Schlather's model with a powered exponential correlation function (3 parameters);

(c) the Brown-Resnick process (2 parameters).

For further details on max-stable processes, the mentioned models and the CLIC, see Davison *et al.* (2012) and Davison and Gholamrezaee (2012). Respective functions are implemented in the R package `SpatialExtremes` (Ribatet, 2022).

- (4) For $b \in \{1, \dots, B\}$ and $t \in \{1, \dots, n\}$, simulate spatial data with unit Fréchet margins from the chosen max-stable process model, denoted by

$$(Y_{1,b}^{(t),*}, Y_{2,b}^{(t),*}, \dots, Y_{D,b}^{(t),*}).$$

Note that until now we haven't used the particular hypothesis $H_0^{\text{ED}}(A)$. Subsequently, fix $A = \{d_1, \dots, d_k\}$ with $1 \leq d_1 < \dots < d_k \leq D$.

- (5) Assume that $H_0^{\text{ED}}(A)$ from Equation (4) is true, and estimate the four dimensional model parameters $\boldsymbol{\vartheta}_A = (\mu_A, \sigma_A, \gamma_A, \alpha_A)^\top \in \Theta$ by (pseudo) maximum likelihood based on the pooled sample

$$(M_{d_1}^{(1)}, c^{(1)}), \dots, (M_{d_1}^{(n)}, c^{(n)}), (M_{d_2}^{(1)}, c^{(1)}), \dots, (M_{d_2}^{(n)}, c^{(n)}), \dots, \\ \dots, (M_{d_k}^{(1)}, c^{(1)}), \dots, (M_{d_k}^{(n)}, c^{(n)}).$$

Denote the resulting parameter vector as $\hat{\boldsymbol{\vartheta}}_A = (\hat{\mu}_A, \hat{\sigma}_A, \hat{\gamma}_A, \hat{\alpha}_A)^\top$, and note that $\hat{\boldsymbol{\vartheta}}_A$ should be close to $\boldsymbol{\vartheta}_d$ for each $d \in A$, if $H_0^{\text{ED}}(A)$ is met.

- (6) Transform the margins of the bootstrap samples to the ones of a GEV-model satisfying $H_0^{\text{ED}}(A)$, by letting

$$M_{d,b}^{(t),*} = \hat{\mu}_A \exp\left(\frac{\hat{\alpha}_A \text{GMST}'(t)}{\hat{\mu}_A}\right) + \hat{\sigma}_A \exp\left(\frac{\hat{\alpha}_A \text{GMST}'(t)}{\hat{\mu}_A}\right) \frac{(Y_{d,b}^{(t),*})^{\hat{\gamma}_A} - 1}{\hat{\gamma}_A} \quad (10)$$

for $t \in \{1, \dots, n\}, d \in A$ and $b \in \{1, \dots, B\}$. For each resulting bootstrap sample $\mathcal{D}_b^*(A) = \{M_{d,b}^{(t),*} : t \in \{1, \dots, n\}, d \in A\}$, compute the value $t_{n,b}^*(A)$ of the test statistic $T_n(A)$ from Equation (8). Note that $T_n(A)$ only depends on the coordinates with $d \in A$.

- (7) Compute the value $t_n(A)$ of the test statistic $T_n(A)$ from Equation (8) on the observed sample.
(8) Compute the bootstrapped p -value by

$$p(A) = \frac{1}{B+1} \sum_{b=1}^B \mathbf{1}(t_n(A) \leq t_{n,b}^*(A)).$$

In a classical test situation, one may now reject $H_0^{\text{ED}}(A)$ for a fixed set A at significance level $\alpha \in (0, 1)$ if $p(A) \leq \alpha$. In the current pooling situation, we would need to apply the test to multiple pooling regions A , which hence constitutes a multiple testing problem where standard approaches yield inflated levels. We discuss possible remedies in Section 2.5.

2.4.2 Pairwise bootstrap based on bivariate extreme value distributions

Recall that the location of primary interest is the one with index $d = 1$.

As stated in Section 2.1, it is of interest to test for all bivariate hypotheses $H_0^{\text{ED}}(\{1, d\})$ with $d = 2, \dots, D$. For that purpose, we may apply a modification of the bootstrap procedure from the previous section that makes use of bivariate extreme value models only. By doing so, we decrease the model risk implied by imposing a possibly restrictive global max stable process model.

The modification only affects step (3) and (4) from Algorithm 1. More precisely, for testing the hypothesis $H_0^{\text{ED}}(A_d)$ with $A_d = \{1, d\}$ for some fixed value $d = 2, \dots, D$, we make the following modifications:

Algorithm 2 (Pairwise bootstrap based on bivariate extreme value distributions). Perform step (1) and (2) from Algorithm 1 with the set $\{1, \dots, D\}$ replaced by A_d .

- (3a) Fit a set of bivariate extreme value distributions to the bivariate sample $(Y_1^{(t)}, Y_d^{(t)})_{t=1, \dots, n}$, assuming the marginal distributions to be unit Fréchet. Choose the best fit according to the Akaike information criterion (AIC), a model selection criterion that rewards a good fit of a model and penalises the model's complexity at the same time (Akaike, 1973). Possible models are:
- (a) the Hüsler-Reiss model (1 parameter);
 - (b) the logistic model (1 parameter);
 - (c) the asymmetric logistic model (2 parameters).

Note that all models are implemented in Stephenson (2002).

- (4a) For $b \in \{1, \dots, B\}$ and $t \in \{1, \dots, n\}$, simulate bivariate data with unit Fréchet margins from the chosen bivariate extreme value model, denoted by $(Y_{1,b}^{(t),*}, Y_{d,b}^{(t),*})$.

Perform Steps (5)-(8) from Algorithm 1 with $A = A_d$.

Note that Algorithm 2 is computationally more expensive than Algorithm 1 since model selection and fitting of dependence models and its subsequent simulation must be performed separately for each hypothesis $H_0^{\text{ED}}(A_d)$ of interest.

2.5 Combining test statistics

As already addressed at the end of Section 2.1, it is not only of interest to test the global hypothesis H_0^{ED} , since a possible rejection of H_0^{ED} gives no indication about which locations deviate from the one of primary interest. Instead, one might want to test hypotheses on several subsets and then pool those subsets for which no signal of heterogeneity was found. In this subsection, we provide the mathematical framework of testing sub-hypotheses and discuss how to deal with the induced multiple testing problem.

Mathematically, we propose to regard H_0^{ED} as a global hypothesis that is built up from elementary hypotheses of smaller dimension. A particularly useful decomposition is based on pairwise elementary hypotheses: recalling the notation $H_0^{\text{ED}}(A)$

from Equation (4), we clearly have

$$H_0^{\text{ED}} = \bigcap_{d=2}^D H_0^{\text{ED}}(\{1, d\}), \quad (11)$$

i.e., H_0^{ED} holds globally when it holds locally for all pairs $\{1, d\}$ with $d \in \{2, \dots, D\}$. We may now either apply Algorithm 1 or Algorithm 2 to obtain a p-value, say $p_d^{\text{raw}} = p(\{1, d\})$, for testing $H_0^{\text{ED}}(\{1, d\})$, for any $d \in \{2, \dots, D\}$. Each p-value may be interpreted as a signal for heterogeneity between locations 1 and d , with smaller values indicating stronger heterogeneity. The obtained raw list of p-values may hence be regarded as an exploratory tool for identifying possible heterogeneities.

Since we are now dealing with a multiple testing problem, it might be advisable to adjust for multiple comparison in order to control error rates. This can be done by interpreting the raw list based on classical statistical testing routines, in which p-values are compared with suitable critical values to declare a hypothesis significant. Several methods appear to be meaningful, and we discuss three of them in the following. For this, let $\alpha \in (0, 1)$ denote a significance level, e.g., $\alpha = 0.1$.

IM (Ignore multiplicity): reject homogeneity for all pairs $\{1, d\}$ for which $p_d^{\text{raw}} \leq \alpha$. In doing so, we do not have any control over false rejections. In particular, in case D is large, false rejections of some null hypotheses will be very likely. On the other hand, the procedure will have decent power properties, and will likely detect most alternatives. Hence, in a subsequent analysis based on the pooled sample of homogeneous locations, we can expect estimators to exhibit comparably little bias and large variance.

Holm (Control the family-wise error rate): apply Holm's stepdown procedure (Holm, 1979). For that purpose, sort the p-values $p_j = p_{1+j}^{\text{raw}} = p(\{1, 1+j\})$ with $j = 1, \dots, D-1$; denote them by $p_{(1)} \leq \dots \leq p_{(D-1)}$. Starting from $j = 1$, determine the smallest index j such that

$$p_{(j)} > \alpha_j := \alpha / (D - j).$$

If $j = 1$, then reject no hypotheses. If no such index exists, then reject all hypotheses. Otherwise, if $j \in \{2, \dots, D-1\}$, reject the hypotheses that belong to the p-values $p_{(1)}, \dots, p_{(j-1)}$.

The procedure can be equivalently expressed by adjusted p-values. Recursively defining $\tilde{p}_{(1)} = \min\{1, (D-1)p_{(1)}\}$ and

$$\tilde{p}_{(j)} = \min\{1, \max\{\tilde{p}_{(j-1)}, (D-j)p_{(j)}\}\}$$

for $j = 2, \dots, D-1$, we simply reject those hypotheses that belong to the adjusted p-values with $\tilde{p}_{(j)} \leq \alpha$.

Holm's stepdown procedure is known to asymptotically control the family-wise error rate (FWER) at level α , i.e.,

$$\text{FWER} := \Pr(\text{reject any true null hypothesis } H_0^{\text{ED}}(\{1, d\})) \leq \alpha,$$

see Theorem 9.1.2 in Lehmann and Romano (2021).

In general, controlling the family-wise error rate will result in comparably little power, i.e., we might falsely identify some pairs of locations as homogeneous. Hence, in a subsequent analysis based on the pooled sample of homogeneous locations, we can expect estimators to exhibit comparably large bias and little variance.

BH (Control the false discovery rate): apply the Benjamini Hochberg stepup procedure (Benjamini and Hochberg, 1995). For that purpose, sort the p-values $p_j = p_{1+j}^{\text{aw}} = p(\{1, 1+j\})$ with $j = 1, \dots, D-1$; denote them by $p_{(1)} \leq \dots \leq p_{(D-1)}$. Starting from $j = D-1$, determine the largest index j such that

$$p_{(j)} \leq \alpha_j := \frac{j\alpha}{(D-1)}.$$

If no such index exists, then reject no hypotheses. Otherwise, if $j \in \{1, \dots, D-1\}$, reject the hypotheses that belong to the p-values $p_{(1)}, \dots, p_{(j)}$.

Again, one can compute adjusted p -values $\tilde{p}_{(j)}$ such that the procedure is equivalent to rejecting those hypotheses for which $\tilde{p}_{(j)} \leq \alpha$. For that purpose, let $\tilde{p}_{(D-1)} = \min\{1, (D-1)p_{(D-1)}\}$ and recursively define, for $j = D-2, \dots, 1$,

$$\tilde{p}_{(j)} = \min \left\{ 1, \min \left\{ (D-1) \frac{p_{(j)}}{j}, \tilde{p}_{(j+1)} \right\} \right\}.$$

Under an additional assumption on the p-values that belong to the true null hypotheses (they must exhibit some positive dependence), the BH procedure is known to asymptotically control the false discovery rate (FDR) at level α , i.e.,

$$\text{FDR} := \mathbb{E} \left[\frac{\text{Number of false rejections}}{\text{Number of all rejections}} \mathbf{1}(\text{at least one rejections}) \right] \leq \alpha,$$

see Theorem 9.3.3 in Lehmann and Romano (2021). Control of the FDR will be confirmed by the simulation experiments in Section 3.

If one were interested in guaranteed theoretical control of the FDR rate, one might alternatively apply the Benjamini Yekutieli (BY) stepup procedure, see (Benjamini and Yekutieli, 2001) and Theorem 9.3.3 in Lehmann and Romano (2021). In view of the fact that the procedure is much more conservative than BH, we do not recommend its application in the current setting.

Concerning a subsequent analysis, estimators based on a pooled sample obtained from the BH procedure can be expected to exhibit bias and variance to be somewhere between the IM and Holm procedure.

Remark 1. The decomposition of H_0^{ED} into hypotheses of smaller dimensionality is not unique. For instance, we may alternatively write

$$H_0^{\text{ED}} = \bigcap_{k=1}^K H_0^{\text{ED}}(B_k), \quad (12)$$

where $\{1\} \subset B_1 \subset B_2 \cdots \subset B_K = \{1, \dots, d\}$ denotes an increasing sequence of regions with $2 \leq |B_1| < |B_2| < \dots < |B_K| = d$ (for instance, $B_k = \{1, 2, \dots, 1+k\}$)

with $k = 1, \dots, D-1$). In practice, the sequence is supposed to be derived from some expert knowledge of the region of interest; it shall represent a sequence of possible pooling regions where B_k is constructed from B_{k-1} by adding the locations which are a priori ‘most likely’ homogeneous to the locations in B_k . Note that, necessarily, $K \leq D - 1$, which provides an upper bound on the number of hypotheses to be tested.

The derivation of respective testing methods is straightforward. In view of the facts that the choice of the sequence is fairly subjective and that the eventual results crucially depend on that choice, we do not pursue the method any further.

3 Simulation Study

A large-scale Monte Carlo simulation study was conducted to assess the performance of the proposed bootstrap procedures in finite sample situations. We aim at answering the following questions:

- (a) Regarding the test’s *power*: What percentage of locations that are heterogeneous w.r.t. the location of primary interest can be expected to be identified correctly?
- (b) Regarding the test’s *error rates*: What percentage of locations that are homogeneous w.r.t. the location of primary interest can be expected to be wrongly identified as heterogeneous (FDR)? What is the probability of wrongly identifying at least one location that is homogeneous w.r.t. the location of interest as heterogeneous (FWER)?
- (c) Regarding the chosen pooling regions: How does return level (RL) estimation based on the pooling regions proposed by the bootstrap procedures compare to RL estimation based on the location of interest only or the whole (heterogeneous) region?

The data was generated in such a way that the temporal spatial dynamics from the case study in Section 4 are mimicked. To achieve this, we started by fitting the scale-GEV model from Equation (2) to annual block-maxima of observations from 1950–2021 at 16 spatial locations in Western Europe (i.e., $n = 72$ and $D = 16$) that are arranged in a 4×4 grid; see Figure 1 and the additional explanations in Section 4. The locations correspond to the center points of the grid cells; the distance between the center points of two neighbouring grid cells is approximately 140 km. The location-wise GEV parameter estimates $\hat{\boldsymbol{\theta}}_d$ exhibit the following approximate ranges over $d \in \{1, \dots, 16\}$: $\hat{\mu}_d \in (18.1, 30.8)$ with a mean of 20.85, $\hat{\sigma}_d \in (4.185, 7.92)$ with a mean of 5.3, $\hat{\gamma}_d \in (-0.13, 0.36)$ with a mean of 0.08 and $\hat{\alpha}_d \in (-2.3, 5.08)$ with a mean of 1.5. Fitting the scale-GEV model to the full pooled sample of size $n \cdot D = 1152$, we obtained parameter estimates that were close to the means over the location-wise parameter estimates, with 20.37, 5.8, 0.1, 1.5 for location, scale, shape and trend parameter, respectively. Next, we transformed the margins to (approximate) unit Fréchet by applying the transformation from Equation (9), such that we can fit several max-stable process models to the transformed data. The best fit was Smith’s model with approximate dependence parameters $\sigma_{11} = 0.4, \sigma_{12} =$

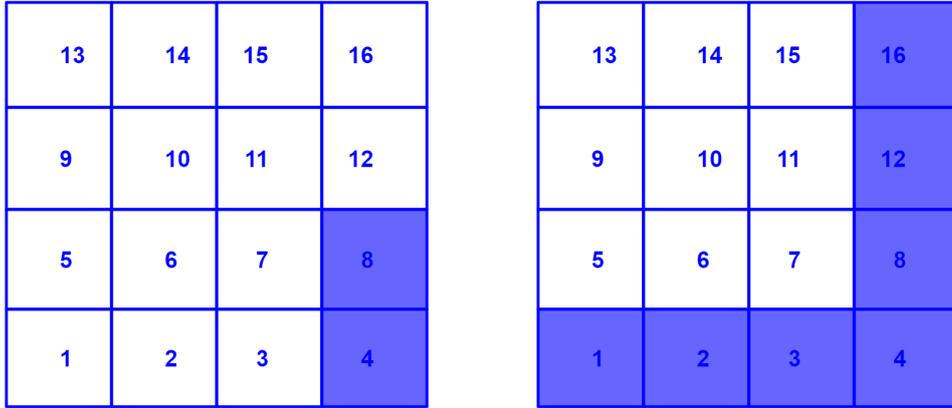


Figure 1: Illustration of the grid used for the simulation. The regions contained in A_{dev} are shaded in blue, with $|A_{\text{dev}}| = 2$ shown in the left plot and $|A_{\text{dev}}| = 7$ shown on the right. The region of interest is the one labelled 10.

0.2, $\sigma_{22} = 0.9$; see Davison *et al.* (2012) for details on the model.

Based on these model fits, we chose to generate data with the following specifications: first, the sample size was fixed to $n = 75$ and the regional 4×4 grid was chosen as described before, i.e., $d = 16$. The grid cell/location labelled ‘10’ is chosen as the one of primary interest. Further, the dependence structure is fixed to that of Smith’s model with (approximately) those parameters that gave the best fit on the observed data, i.e. $\sigma_{11} = 0.4, \sigma_{12} = 0.2, \sigma_{22} = 0.9$. For simulating data, we first simulate from this max-stable process model (Ribatet, 2022) and then transform the margins to scale-GEV distributions, either in a homogeneous or in a heterogeneous manner. Here, the globally homogeneous model is defined by fixing the marginal scale-GEV parameters to approximately the mean values of the location-wise GEV parameters obtained for the real observations, i.e.,

$$\mu_d = 20, \sigma_d = 5.5, \gamma_d = 0.1, \alpha_d = 1.5 \quad (13)$$

for each $d \in \{1, \dots, 16\}$.

Starting from this homogeneous model, we consider two different heterogeneous scenarios. In the first scenario, we fix $\boldsymbol{\vartheta}_d = (\mu_d, \sigma_d, \gamma_d, \alpha_d)^\top$ as in Equation (13) for all $d \in A_{\text{hom}} = \{1, \dots, 16\} \setminus \{4, 8\}$, while

$$\begin{aligned} \mu_d &= 20 + c_\mu, & c_\mu &\in \{-3, -1.5, 0, 1.5, 3\}, \\ \sigma_d &= 5.5 \cdot c_\sigma, & c_\sigma &\in \{0.7, 0.85, 1, 1.15, 1.3\}, \\ \gamma_d &= 0.1 + c_\gamma, & c_\gamma &\in \{-0.1, 0, 0.1\}, \\ \alpha_d &= 1.5 + c_\alpha, & c_\alpha &\in \{-1, 0, 1\}, \end{aligned} \quad (14)$$

for $d \in A_{\text{dev}} = \{4, 8\}$ with $(c_\mu, c_\sigma, c_\gamma, c_\alpha) \neq (0, 0, 0, 0)$. Note that this defines $5 \cdot 5 \cdot 3 \cdot 3 - 1 = 224$ different heterogeneous models. In the second scenario, we consider the same construction with $A_{\text{hom}} = \{5, 6, 7, 9, 10, 11, 13, 14, 15\}$ and

$A_{\text{dev}} = \{1, 2, 3, 4, 8, 12, 16\}$. An illustration of the grid cells and their partition into homogeneous and non-homogeneous areas can be found in Figure 1. Overall, we obtain 448 different heterogeneous models and one homogeneous model.

For each of the 449 models, we now apply the following three bootstrap procedures, each carried out with $B = 300$ bootstrap replications (recall that the grid cell of interest is the one labelled with 10):

- (B1) The bootstrap procedure from Algorithm 1 with $A = \{1, \dots, 16\}$.
- (B2) The bootstrap procedure from Algorithm 1 for all sets $A_d = \{10, d\}$ with $d \in \{1, \dots, 16\} \setminus \{10\}$.
- (B3) The bootstrap procedure from Algorithm 2 for all sets $A_d = \{10, d\}$ with $d \in \{1, \dots, 16\} \setminus \{10\}$.

Note that the second and third method both yield 15 raw p-values. Each procedure was applied to 500 simulated samples from all models under consideration.

Regarding (B1), we compute the percentage of rejections among the 500 replications, which represents the empirical type I error of the test under the homogeneous model and the empirical power under the heterogeneous models. The results can be found in Figure 2. The null hypothesis is met in the central square only, and we observe that the nominal level of $\alpha = 0.1$ is perfectly matched. All non-central squares correspond to different alternatives, and we observe decent power properties in both scenarios. Note that a rejection only implies that the entire region $\{1, \dots, 16\}$ is not homogeneous; there is no information on possible smaller subgroups that are homogeneous to the location of interest.

Regarding (B2) and (B3), rejection decisions were obtained for each hypothesis $H_0^{\text{ED}}(\{10, d\})$ by one of the three methods from Section 2.5. The empirical family-wise error rate is then the percentage of cases (over 500 replications) for which at least one null hypothesis was rejected. Likewise, for the false discovery rate, we calculate, for each replication, the number of false rejections and divide that by the total number of rejections (when the number of total rejections is 0, this ratio is set to 0). The empirical false discovery rate is obtained by taking the mean over all 500 replications. Similarly, for assessing the power properties, we calculate the empirical proportion of correct rejections (i.e., among the 2 or 7 locations that deviate, the proportion of detected heterogeneous locations) over all 500 replications.

Results for the false discovery and family-wise error rate are given in Table 1. We find that the p -value combination methods from Section 2.5 are sufficiently accurate: the BH method controls the false discovery rate, while Holm's method controls the family-wise error rate. This holds exactly for procedures (B3), where the maximal FDR (FWER) of the BH (Holm) method is at 9.4% (8.7%), and approximately for (B2), where the maximal FDR (FWER) is at 12.2% (12.6%). Further, we see that the IM procedure neither controls the FWER nor the FDR.

The power properties for procedure (B2) combined with the BH method are shown in Figure 3. We see that the procedure is able to detect some of the deviations of the null hypothesis, with more correct rejections the stronger the deviation is. The method is particularly powerful when the location and scale parameters deviate into

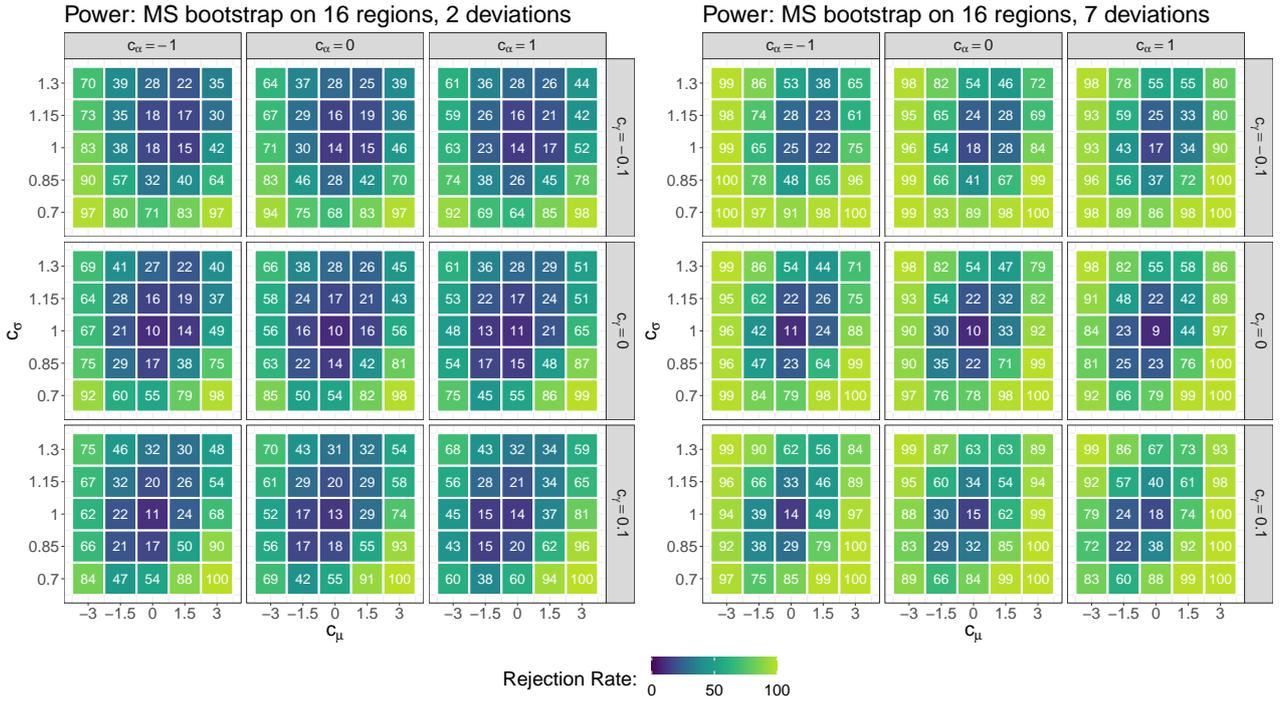


Figure 2: Rejection rates in % obtained for (B1), in the setting where either 2 (left plot) or 7 (right plot) regions deviate from the others. Each coloured square contains the rejection rate for one of the 225 different models, with the central square with $c_\mu = c_\sigma = 0$ corresponding to the null hypothesis. The x - and y -axis and the facets determine the values of the scale-GEV parameter vector of the deviating locations through Equation (14).

opposite directions, i.e. when $c_\mu > 0$ and $c_\sigma < 1$ or $c_\mu < 0$ and $c_\sigma > 1$. There is no obvious pattern regarding the deviations of the shape and trend parameter. Further, we analogously show the power properties of the IM method with bootstrap (B2) in Figure 3. As expected, this method has more power against all alternatives under consideration. However, this comes at the cost of more false discoveries, as can be seen in Table 1.

The results for bootstrap scheme (B3) were very similar and are therefore not shown here, but can be found in Section B of the supplementary material. Likewise, we omit the results for the more conservative Holm procedure, which exhibits, as expected, less power against all alternatives. Further, we repeated the simulation study with an increased location-wise sample size of $n = 100$. As one would expect, the tests have more power in this case.

The results presented so far show that the proposed pooling methods work ‘as intended’, since the theoretical test characteristics are well approximated in finite sample situations, and since we observe decent power properties. In practical applications however, spatial pooling of locations is usually the starting point for subsequent analyses. For instance, one may be interested in estimating return levels

Method	min FDR		max FDR		mean FDR		min FWER		max FWER		mean FWER	
	(B2)	(B3)	(B2)	(B3)	(B2)	(B3)	(B2)	(B3)	(B2)	(B3)	(B2)	(B3)
<i>Scenario 1: $A_{\text{dev}} = 2$</i>												
BH	7.3	5.6	12.2	9.4	9.4	7.5	9.1	6.9	21.2	19.0	14.4	11.8
Holm	3.0	2.3	11.7	8.3	7.1	5.1	6.9	5.0	12.6	8.7	9.5	6.6
IM	25.8	25.3	61.7	60.1	37.7	37.2	53.4	53.4	64.5	62.8	59.1	58.3
<i>Scenario 2: $A_{\text{dev}} = 7$</i>												
BH	3.6	2.4	12.1	8.9	5.6	4.9	6.2	4.2	32.0	29.8	18.6	16.8
Holm	1.0	0.9	11.3	7.9	3.4	2.6	3.8	2.4	11.3	7.9	7.1	5.1
IM	7.9	7.8	61.5	60.1	16.0	15.8	40.9	40.9	61.5	60.1	47.3	46.4

Table 1: False Discovery Rate (FDR) and family-wise Error Rate (FWER) for the three p-value combination methods from Section 2.5 and the two bootstrap methods (B2) and (B3). The stated values are the minimum, maximum and mean across the 224 alternative models from each scenario.

at the location of interest based on the data from all locations that were identified as homogeneous. Moreover, the analysis of alternative data sets like climate model data may be based on the homogeneous locations identified within the analysis of observations.

This suggests that the methods should be examined with regard to their quality in subsequent analyses. For that purpose, we consider, as an example, the problem of return level estimation at the location of interest. The state-of-the-art method would consist of GEV fitting at the location of interest only, which results in (asymptotically) unbiased estimators that suffer from large variance. Basing the estimator on pooled regions will decrease the variance, but at the same time increase its bias if some heterogeneous locations have been wrongly identified as homogeneous.

In particular, pooling based on a conservative testing approach like the BH procedure leads to the acceptance of many locations and thus to a large pooling area and low estimation variance. Most likely, some of the chosen locations will be violating the null hypothesis though, which yields a rather large estimation bias. For pooling based on a more liberal rejection approach like the IM procedure, the estimation bias and variance behave exactly opposite: since the null hypotheses are more likely to be rejected, the resulting pooling sample is smaller (i.e., larger estimation variance) but ‘more accurate’ (i.e., smaller estimation bias).

For our comparison, we consider fitting the scale-GEV model based on pooled locations that have been obtained from one of the following eight methods

$$m \in \{\text{LOI, full, MS IM, MS Holm, MS BH, biv. IM, biv. Holm, biv. BH}\}.$$

Here, LOI refers to considering the location of interest only (no pooling), full refers to pooling all available locations, and the last six methods encode pooling based on any combination of the proposed p-value combination methods and bootstrap approaches.

For each method, we compute the maximum likelihood estimate $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\sigma}, \hat{\gamma}, \hat{\alpha})^\top$ of the scale-GEV model parameters and transform this to an estimate of the T -year

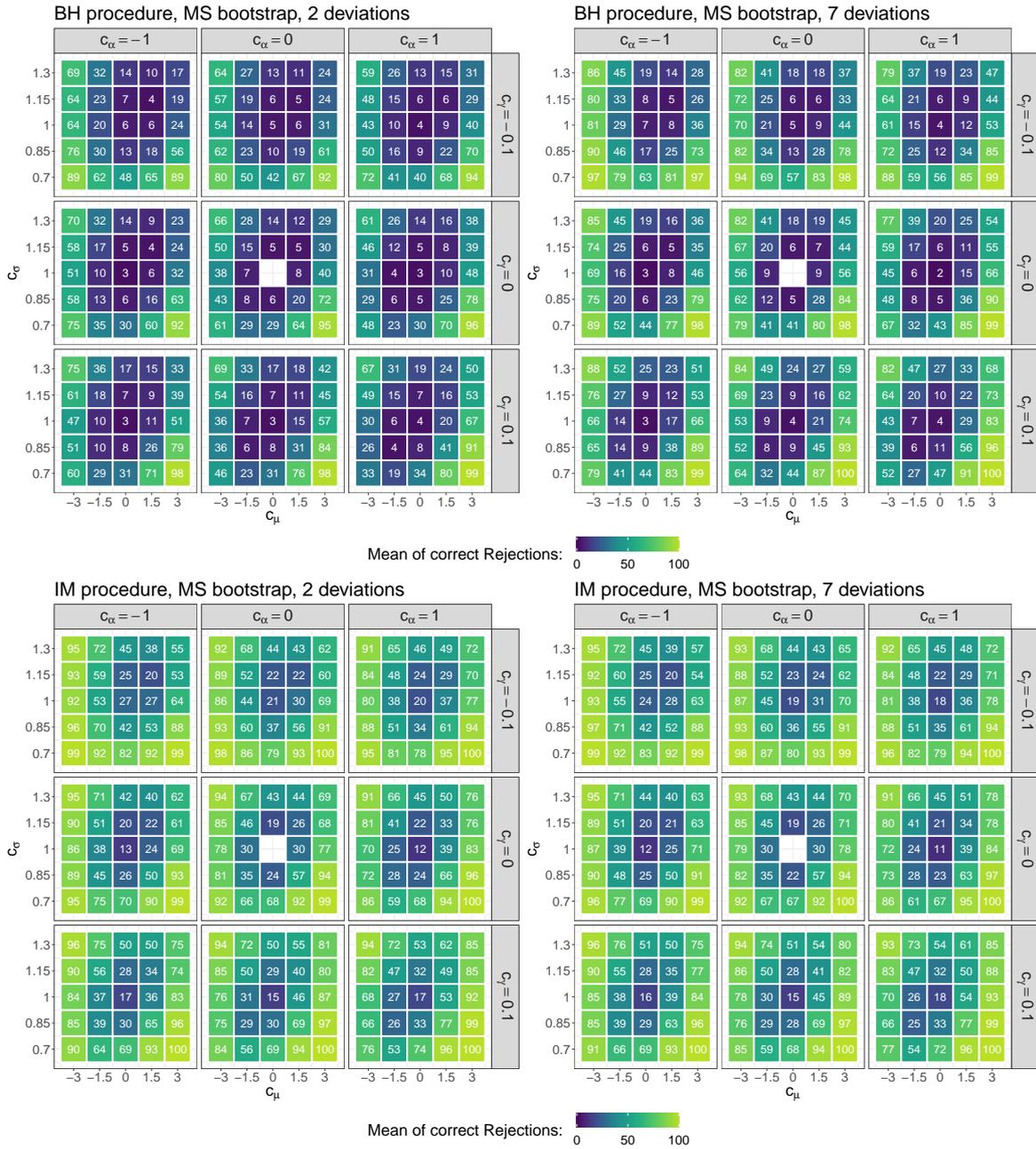


Figure 3: Proportion of correct rejections in % obtained with the BH procedure (upper row) and the IM procedure (lower row) at a level of 0.1, in the setting where two stations deviate from the rest (left column) or 7 locations deviate from the rest (right column), with the bootstrap procedure based on max-stable processes. The axis and facets are as described in Figure 2.

return level (RL) in the reference climate of year t by

$$\widehat{RL}_t(T) = G_{(\hat{\mu}(t), \hat{\sigma}(t), \hat{\gamma})}^{-1}(1 - 1/T),$$

where $\hat{\mu}(t) = \hat{\mu} \exp(\hat{\alpha} \text{GMST}'(t)/\hat{\mu})$ and $\hat{\sigma}(t) = \hat{\sigma} \exp(\hat{\alpha} \text{GMST}'(t)/\hat{\mu})$ and where G

is the cumulative distribution function of the GEV-distribution, see Equation (1). Now, in our simulation study, we know that the true value of the target RL is given by $\text{RL}_t(T) = G_{(\mu_0(t), \sigma_0(t), \gamma_0)}^{-1}(1 - 1/T)$ with

$$\mu_0(t) = 20 \exp\left(\frac{1.5\text{GMST}'(t)}{20}\right), \sigma_0(t) = 5.5 \exp\left(\frac{1.5\text{GMST}'(t)}{20}\right), \gamma_0 = 0.1.$$

From the 500 replications we can therefore compute the empirical Mean Squared Error (MSE) of method m as

$$\text{MSE}(m) = \frac{1}{500} \sum_{j=1}^{500} \left(\widehat{\text{RL}}_t^{(m,j)}(T) - \text{RL}_t(T) \right)^2,$$

where $\widehat{\text{RL}}_t^{(m,j)}(T)$ denotes the estimated RL obtained in the j -th replication with method m . Note that we have suppressed the MSE's dependence on T and t from the notation.

In Figure 4 we compare MSEs of the 100-year RL with reference climate as in year 2021, which is given by $\text{RL}_{2021}(100) = 55.87$, by plotting the difference $\text{MSE}(m_1) - \text{MSE}(m_2)$ with $m_1 \in \{\text{MS BH}, \text{MS IM}\}$ and $m_2 \in \{\text{full}, \text{ROI}\}$ as obtained for the setting where $|A_{\text{dev}}| = 7$. The plots reveal that both the MS BH and the MS IM method are superior to the the LOI fit for almost all scenarios. Comparing the two methods to the full fit reveals that there are certain scenarios for which the full fit performs substantially worse, mostly when the shape and scale parameter deviate towards the same direction for the alternatives. For those scenarios where the full fit outperforms the two methods, the discrepancy is not very large, with the BH method performing slightly better than the IM method.

A comparison between MS BH and MS IM is shown in Figure 5 for $|A_{\text{dev}}| \in \{2, 7\}$. The results reveal that the BH method slightly outperforms the IM method in the case $|A_{\text{dev}}| = 2$ for almost all alternative scenarios. In case $|A_{\text{dev}}| = 7$, the results are quite mixed, with the IM method becoming clearly superior when the shape, scale and location parameters deviate jointly to the top. In all other scenarios, the differences are only moderate, sometimes favoring one method and sometimes the other. Corresponding results for the bootstrap methods based on bivariate extreme value distributions are very similar and therefore not shown. Further, the results were found to be robust against the choices of $t = 2021$ and $T = 100$ that were made here for the return level estimation.

Overall, the results suggest the following practical recommendation: if the full sample is expected to be quite homogeneous a priori (for instance, because it was built based on expert knowledge), then estimation based on BH-based pooling is preferable over the other options (LOI, the full and the IM-based fit). If one expects to have a larger number of heterogeneous locations, it is advisable to apply the IM procedure (or any other liberal procedure), which likely rejects most of the heterogeneous locations and hence reduces the bias. In general, the liberal behavior of IM-based pooling suggests its use when it is of highest practical interest to obtain

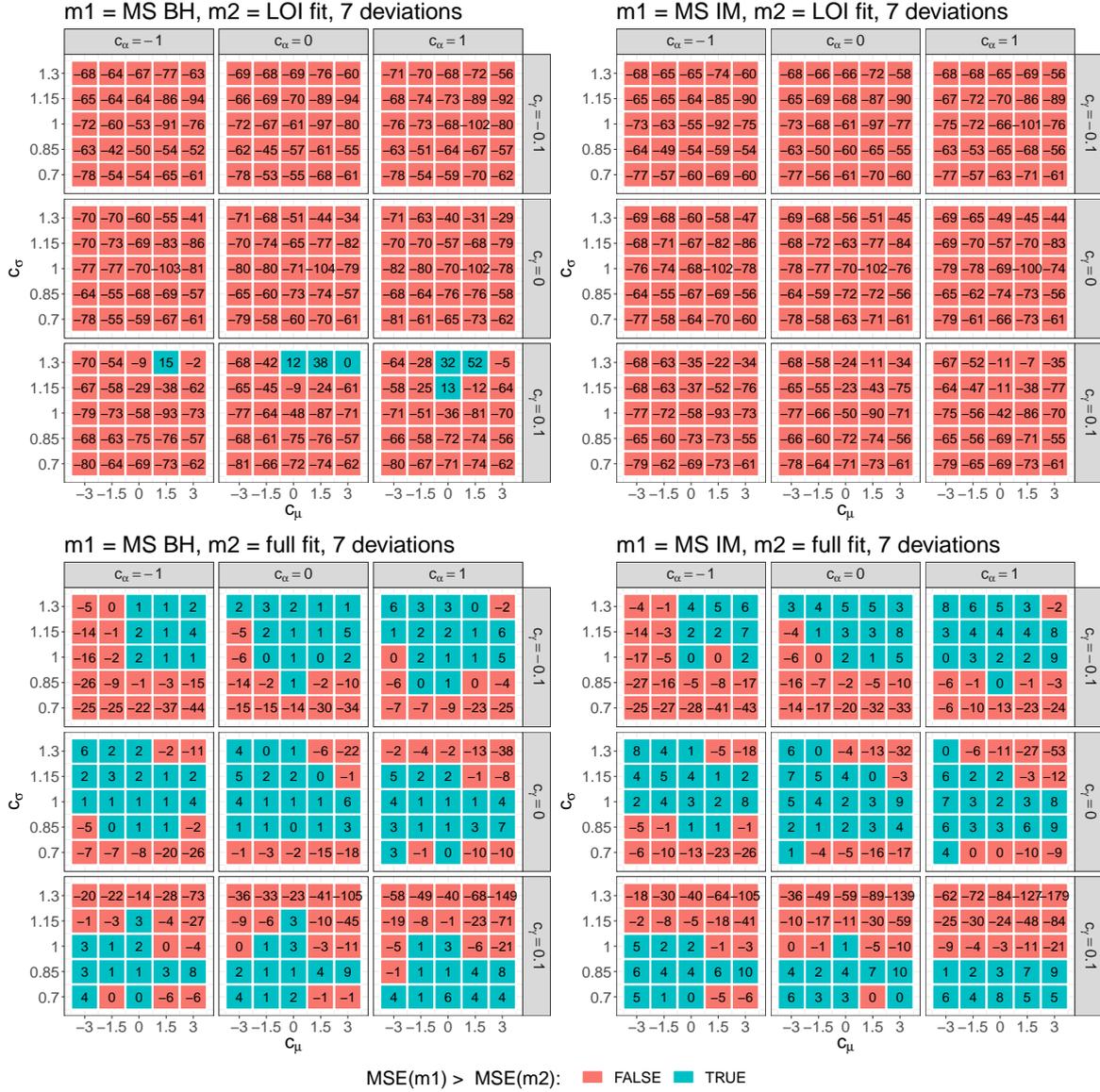


Figure 4: Comparison of MSEs of $RL_{2021}(100)$ obtained for different choices of the method m , in the setting where $|A_{\text{dev}}| = 7$. Shown are the differences $\text{MSE}(m1) - \text{MSE}(m2)$ with $m1$ and $m2$ as indicated in the plot title. Negative values (red) therefore indicate a lower MSE for the method mentioned first, and vice versa for positive values. The axis and facets are as described in Figure 2.

a pooled region that is as homogeneous as possible (as a trade-off, one has to accept that the region is probably much smaller than the initial full region).

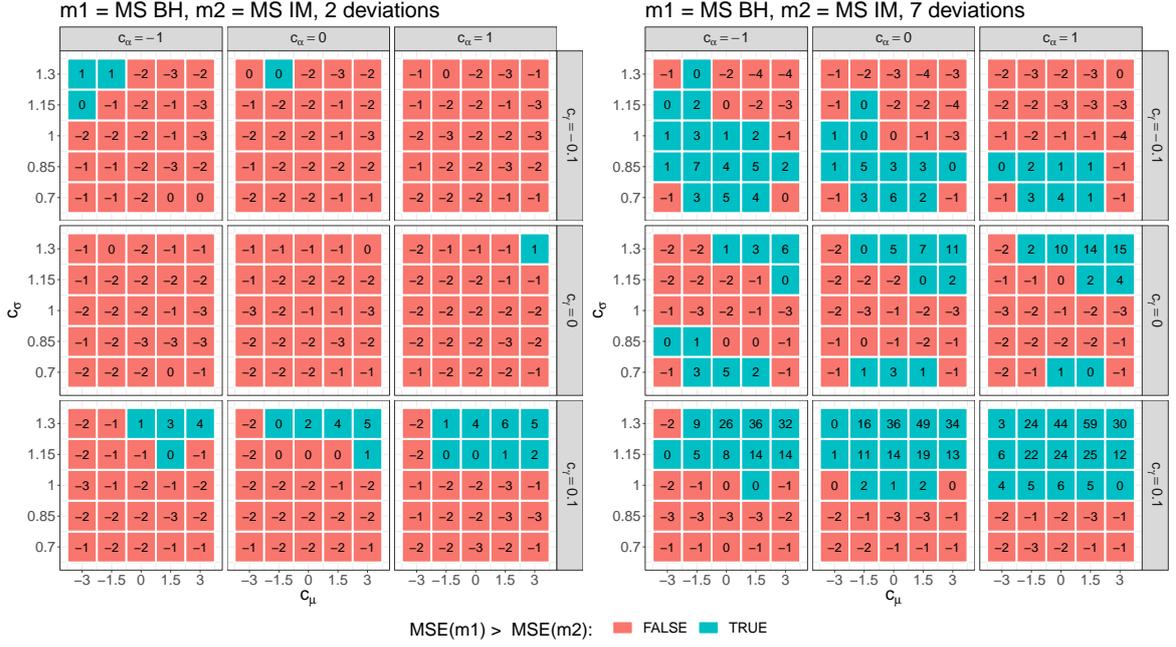


Figure 5: Comparison of MSEs of $RL_{2021}(100)$ in the setting where $|A_{\text{dev}}| = 2$ (left) and $|A_{\text{dev}}| = 7$ (right). Shown are the differences $\text{MSE}(\text{MS BH}) - \text{MSE}(\text{MS IM})$. Negative values (red) therefore indicate a lower MSE for the BH method, while positive values (blue) indicate a lower MSE for the IM method. The axis and facets are as described in Figure 2.

4 Severe flooding in Western Europe during July 2021 revisited

We illustrate the new pooling methods in a case study by revisiting the extreme event attribution study for the heavy precipitation event that led to severe flooding in Western Europe during July 2021, see Kreienkamp *et al.* (2021); Tradowsky *et al.* (2022). In that study, observational data were pooled together based on expert knowledge and on ad hoc tests, with the ultimate goal of assessing the influence of human-made climate change on the likelihood and severity of similar events in Western and Central Europe.

The full region under investigation in Kreienkamp *et al.* (2021); Tradowsky *et al.* (2022) consists of sixteen ($2.0^\circ \times 1.25^\circ$) (i.e. about $(140 \text{ km} \times 140 \text{ km})$) grid cells reaching from the northern Alps to the Netherlands, see Figure 5 in Kreienkamp *et al.* (2021) or the right-hand side of Figure 6. Two of the 16 locations were rejected in that study due to expert knowledge and too large deviations in fitted GEV-parameters (regions 17 and 11 of Figure). Among other things, our illustrative application of the methods explained above will reveal that grid cell 11 has been rightfully dismissed, while grid cell 17 might have been considered homogeneous. Further, there is no clear evidence that any other grid cell that has been declared

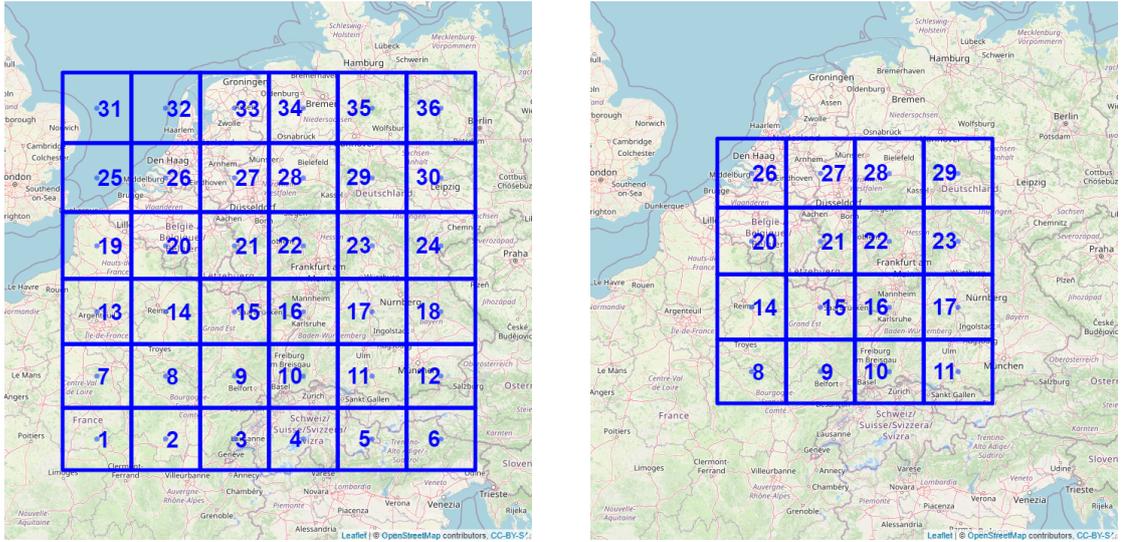


Figure 6: Regions analysed within this case study and the respective numbering used here. The data consists of April-September block maxima of tile-wise averaged daily precipitation sums (RX1day) from 1950-2021.

homogeneous should rather be considered non-homogeneous.

For illustrative purposes, we apply our methods to two different initial areas:

- (A) An area consisting of 6×6 grid cells covering a large part of Western/ Central Europe, as shown in Figure 6 on the left.
- (B) The original 4×4 grid cells from Kreienkamp *et al.* (2021) as shown in Figure 6 on the right.

Note that homogeneity for the 20 grid cells at the boundary of the larger area in (A) has been dismissed based on expert knowledge in Kreienkamp *et al.* (2021); the larger area is included here for illustrative purposes only.

The data used throughout the study consists of April-September block-maxima of tile-wise averaged 1-day accumulated precipitation amounts of the E-OBS data set (Cornes *et al.* (2018), Version 23.1e). In both cases, the grid cell with label 21 is the one of primary interest, since it is the one containing the target location of the study, i.e., the region that accumulated the highest precipitation sum and experienced the largest impacts during the flooding of July 2021. The time series are shown in Figure C.1 in the supplementary material. There, we also plot values of $\hat{\mu}(t) = \hat{\mu} \exp(\hat{\alpha} \text{GMST}'(t)/\hat{\mu})$ obtained from different data sets: once from data of location 21 only, once from data of the respective location only, and once from the pooled data of the respective pair $(21, d)$ for $d \in \{1, \dots, 36\} \setminus \{21\}$.

We apply the two proposed bootstrap procedures to areas (A) and (B). Note that the raw p-values obtained with the bootstrap based on bivariate extreme value distributions should be very similar (or even identical when using the same seed for random number generation) for those grid cells that appear in both areas, while they may differ to a greater extent for the MS bootstrap. This is because the p-value for a given pair obtained with the bivariate bootstrap procedure only depends on the

observations of the pair, while the respective p-value obtained with the MS bootstrap also depends on the spatial model that was fitted to the whole area. However, even if the raw p-value of a given pair obtained for setting (B) coincides with the raw p-value obtained for setting (A), the adjustment for multiple testing can lead to slightly different rejection decisions of the pair at a given level α . The bootstrap procedures are applied with $B = 2000$ bootstrap replications.

We start by discussing the results of the application to the larger grid in (A). Recall that, for a given significance level α , one rejects the null hypothesis for all grid cells whose p-value is smaller than α . To visualise the results, we therefore shade the grid cells according to the magnitude of their (adjusted) p-value. Here, we divide the colour scale into three groups: $[0, 0.05]$, $(0.05, 0.1]$ and $(0.1, 1]$, with a dark red tone assigned to the first group, a brighter red tone for Group 2 and an almost transparent shade for Group 3. This allows us to see the test decisions for significance levels of $\alpha \in \{0.05, 0.1\}$: when the significance level is chosen as $\alpha = 0.1$, all tiles with a reddish shade are rejected, while when working with a level of $\alpha = 0.05$ only tiles shaded in the dark shade are rejected.

Results for both the conservative BH procedure and the liberal IM procedure are shown in Figure 7. For completeness, results on Holm’s method, which is even more conservative than BH, as well as the BH and IM p-values themselves can be found in the supplementary material, Tables C.2 and C.3. One can see that, for a given rejection method (i.e. BH or IM), the MS and bivariate procedures mostly agree on the rejection decisions that would be made at a level of 10% (compare the rows of Figure 7 to see this). The same holds when working with a significance level of 5%.

Further, as expected, the IM method rejects more hypotheses than the BH method. However, according to the results of the simulation study, it is quite likely that at least one of these rejections is a false discovery.

Analogous results for the 4×4 grid in (B) are shown in Figure 8. As discussed above, except for the MS BH method, the results are consistent with the results obtained for the 6×6 grid in the sense that for those locations which are contained in both grids, the locations with p-values of critical magnitude ($< 10\%$) coincide (compare the plot in the upper right corner of Figure 8 to the plot in the upper right corner of Figure 7 to see this for the MS IM method, and similar for the other methods). For the MS BH method, grid cells 10, 14, 15, and 16 are not significant anymore at a level of 10 %, but we recorded an adjusted p-value of 0.106 for those four grid cells, so this is a rather tight decision. The p-values obtained for the 4×4 grid can be found in Table C.1 in the supplementary material.

Let us now move on to the interpretation: considering the larger grid first, some grid cells for which the characteristics of extreme precipitation are different (according to expert opinion) from the grid cell of the target location are detected as heterogeneous. These rejected grid cells are located along the coast and in the mountainous terrain. Comparing the results with Kreienkamp *et al.* (2021); Tradowsky *et al.* (2022), we observe that grid cell 11 has been rejected in their study based on expert knowledge. For grid cell 17, however, we do not detect any statistical evi-

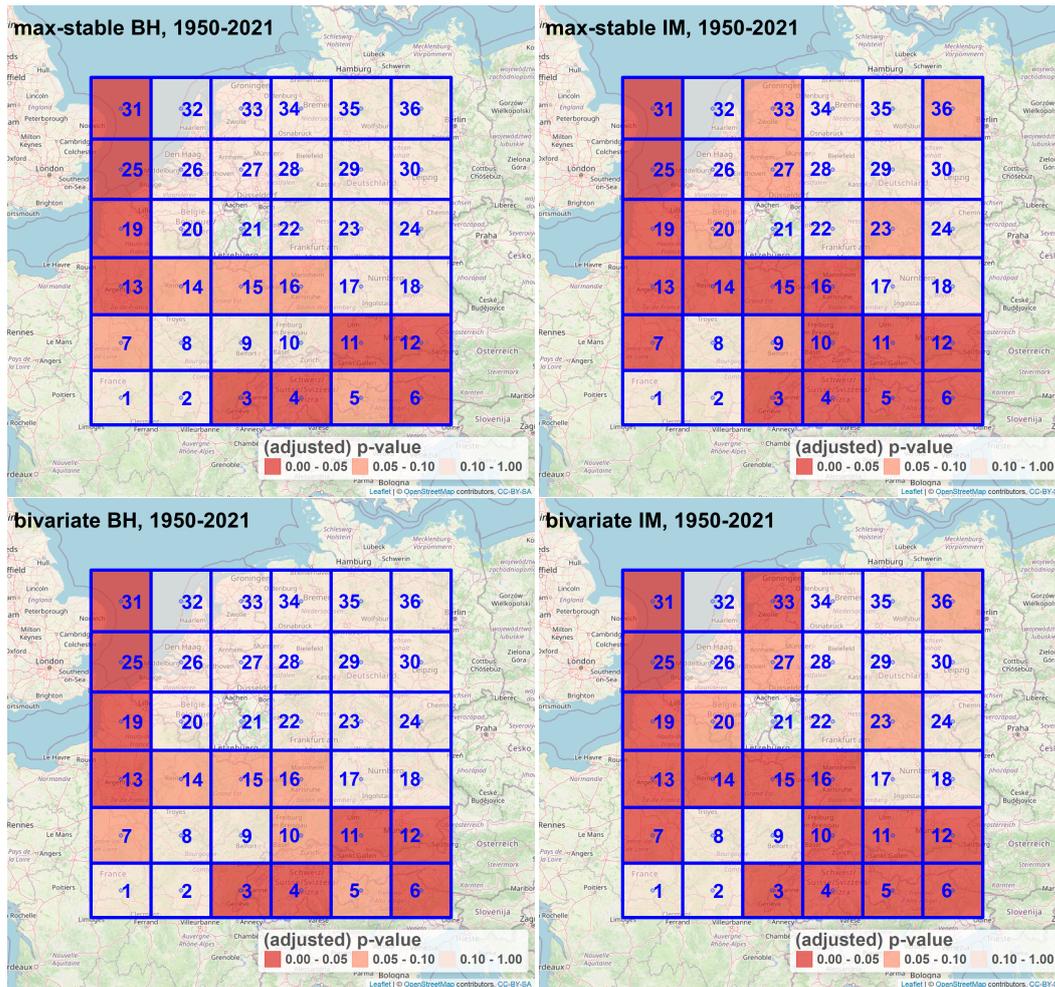


Figure 7: (Adjusted) p -values obtained with the BH (left) and the IM (right) method on the 6×6 grid, with the bootstrap based on max-stable processes (top row) and the bootstrap based on bivariate extreme value distributions (bottom row).

dence that the probabilistic behavior of extreme precipitation is different from the grid cell of the target location, even when applying the liberal IM procedure. We would like to stress though that non-rejection of a null hypothesis does not provide any evidence of the null hypothesis, even when ignoring the multiple testing issue. Hence, expert knowledge that leads to rejection should, in general, outweigh any statistical non-rejection. This particularly applies to the eastern (continental) grid cells in the larger 6×6 -grid, which can be influenced by heavy precipitation caused by different synoptic situations compared to the target region.

Moreover, as the results for locations 10, 14, 15, and 16 showed some discrepancy across the different testing procedures, we suggest that the final decision on the exclusion or inclusion of these locations in a spatial pooling approach should be based on expert knowledge of the meteorological characteristics, and the willingness to trade possible bias for variance (with a possibly larger bias when including the lo-

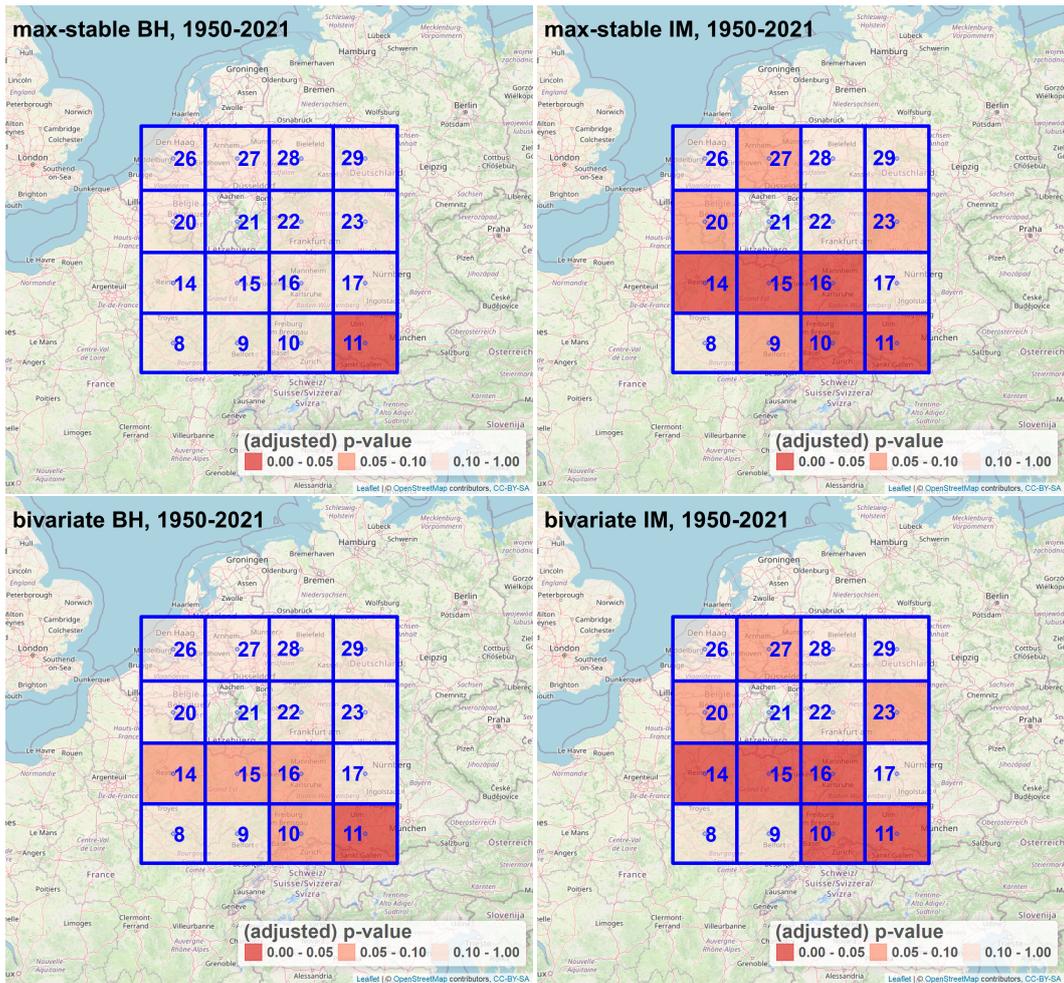


Figure 8: Adjusted p -values obtained with the BH (left) and the IM (right) method on the 4×4 grid, obtained with the bootstrap based on max-stable processes (top row) and the bootstrap based on bivariate extreme value distributions (bottom row).

cations – note that statistical evidence against homogeneity in the bivariate extreme value distribution-based bootstrap is only weak, and wrongly declaring the regions as homogeneous is possibly not too harmful). The same holds for locations 9, 20, 23 and 27 for which only the IM method yielded p -values between 5% and 10%. Again, these rather small p -values could be subject to false discoveries though, and since the heterogeneity signal is also not too strong, there is no clear evidence that these need to be excluded from pooling.

For a last evaluation of results from pairwise tests, we estimated the 100-year RLs in the reference climate of the year 2021, i.e. with reference value $\text{GMST}'(2021) = 0.925^\circ\text{C}$, on five different data sets obtained from the 4×4 grid. Here, we use the data sets consisting of data from

- the location of interest only
- pooling those grid cells suggested by the results of the case study (i.e., all cells

but 11, or all cells but 10, 11, 14, 15, 16) or expert opinion (i.e., all cells but 11, 17)

- pooling all grid cells of the 4×4 grid.

The results can be found in Table 2 and reveal that excluding cell 11 has a clear effect on the estimated RL. Ex- or including grid cell 17 once 11 is excluded does not have a large effect, while excluding cells 10, 14, 15 and 16 additionally to cell 11 has a moderate effect.

cells excluded	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\gamma}$	$\hat{\alpha}$	$\widehat{\text{RL}}_{2021}(100)$
none	20.37	5.80	0.1039	1.50	58.43
11	20.01	5.44	0.0676	1.45	52.74
11, 17	20.01	5.40	0.0760	1.29	52.82
10, 11, 14, 15, 16	19.90	5.41	0.0484	1.79	51.93
all but 21	21.92	6.08	0.0634	-0.00	54.37

Table 2: Estimated parameters and estimate of $\text{RL}_{2021}(100)$ obtained when pooling all grid cells but the ones given in the first column.

Finally, we would like to mention that similar results were obtained when applying the BH test procedures to all triples containing the pair of grid cells (20, 21), i.e., the extended target region considered in the study of Kreienkamp *et al.* (2021); Tradowsky *et al.* (2022), consisting of those regions in Germany and Belgium affected worst by the July 2021 flooding.

5 Extensions

In this section, we discuss how to estimate region-wise return levels under homogeneity assumptions (Section 5.1). We also propose two possible extensions of the pooling approach from the previous sections to other hypotheses (Section 5.2) or other underlying model assumptions (Section 5.3).

5.1 Estimation of regional return levels and return periods

As pointed out in Kreienkamp *et al.* (2021); Tradowsky *et al.* (2022) among others, an estimated return period (RP) of T years for a given event and in a fixed reference climate (e.g., the preindustrial climate), obtained based on a fit of the GEV distribution to a pooled homogeneous sample, has the following interpretation: for each fixed location/tile within the region, one can expect one event of the same or larger magnitude within T (imaginary) years of observing the reference climate. We refer to this quantity as the local return period. Obviously, one would expect more than one event of similar magnitude happening at *at least one of the locations* of the pooling region. Likewise, for a given T , one would expect a higher T -year return level for the whole region. The latter corresponds to the value that is expected to be exceeded only once in T years at *at least one of the locations*.

Mathematically, using the notation from Section 2.1, the exceedance probability of value r at at least one among $D \geq 2$ locations in the reference climate corresponding to year t is given by

$$p_t(r) = P\left(\exists j \in \{1, \dots, D\} : M_j^{(t)} \geq r\right) = P\left(\max_{d=1, \dots, D} M_d^{(t)} \geq r\right),$$

such that the return period for event r of the region is $\text{RP}_t^{\text{reg}}(r) = \frac{1}{p_t(r)}$. Further, the T -year return level of the region in the climate corresponding to year t is the minimal value $\text{RL}_t^{\text{reg}}(T)$ for which

$$P\left(\max_{d=1, \dots, D} M_d^{(t)} \geq \text{RL}_t^{\text{reg}}(T)\right) \leq \frac{1}{T}$$

holds. Both quantities could be computed (exactly) if one had access to the distribution of $\max_{d=1, \dots, D} M_d^{(t)}$. For example, if the random variables $M_d^{(t)}$, $d = 1, \dots, D$ were independent, $p_t(r)$ could be further simplified to

$$p_t(r) = 1 - P\left(\max_{d=1, \dots, D} M_d^{(t)} \leq r\right) = 1 - (G_{(\mu(t), \sigma(t), \gamma)}(r))^D,$$

where G is the distribution function of the GEV distribution and where $\mu(t), \sigma(t)$ and γ denote the parameters at reference climate of year t from Equation (2) under the homogeneity assumption from Equation (3).

The locations are, however, usually not independent in applications. In the following, we propose a simulation-based estimation method that involves max-stable process models to account for the spatial dependence. As before, the R package `SpatialExtremes` (Ribatet, 2022) allows for fitting and simulating max-stable process models.

Algorithm 3. (Simulation-based estimation of the regionwise RL and RP)

- (1) Fit the scale-GEV parameters to the pooled homogeneous sample, resulting in the parameter vector $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\sigma}, \hat{\gamma}, \hat{\alpha})^\top$.
- (2) Transform the margins of the pooled data to approximately unit Fréchet by applying transformation from Equation (9) with the parameter estimate from Step 1. Then fit several max-stable process models to the obtained data and choose the best fit according to the information criterion CLIC.
- (3) Replicate for $b = 1, \dots, B$ the following steps:
 - (i) Generate one random observation $(y_{1,b}^{(t),*}, \dots, y_{D,b}^{(t),*})$ from the chosen max-stable process model.
 - (ii) Transform the margins to GEV margins, by applying the transformation in (10) with parameters as estimated in Step 1, resulting in the observation $(m_{1,b}^{(t),*}, \dots, m_{D,b}^{(t),*})$.
 - (iii) Compute the maximum $m_{\max,b}^{(t),*} = \max_{d=1, \dots, D} m_{d,b}^{(t),*}$.
- (4) The regionwise T -year return level $\text{RL}_{t,\text{reg}}(T)$ and the return period $\text{RP}_{t,\text{reg}}(r)$ of an event with value r can now be estimated from the empirical cumulative distribution function \hat{F}_t^* of the sample $(m_{\max,b}^{(t),*})_{b=1, \dots, B}$ through

$$\widehat{\text{RL}}_t^{\text{reg}}(T) = (\hat{F}_t^*)^{-1}(1 - 1/T), \quad \widehat{\text{RP}}_t^{\text{reg}}(r) = \frac{1}{1 - \hat{F}_t^*(r)}.$$

Especially, when we have estimated the local 100-year RL, we can get an estimate of the return time this event has for the whole region. Likewise, when we have an estimate of the local return period of an event with value r , we can estimate what the event value for that return period would be for the whole region.

We illustrate the estimators for the pooled data sets from Section 4. The estimates are based on $B = 100\,000$ simulation replications and are shown in Table 3. We see that the local 100-year return levels have substantially shorter region-wise return periods. In the region with 15 tiles (only cell 11 excluded), the estimated local 100-year RL at reference climate of 2021 can be expected to be exceeded once in approximately 19 years in at least one of the 15 tiles. We find a similar region-wise return period for the pooling region consisting of 14 tiles. In the pooling region consisting of 11 tiles, the local 100-year return level becomes a region-wise 33-year event. This comparably larger value arises from the smaller region that is considered: the smaller the region, the less likely it is that one of the locations exceeds a high threshold. Further, as expected, we find that the region-wise 100-year return levels at reference climate of 2021 are larger than their local counterparts. For the regions consisting of 15 and 14 tiles, this increase is approximately 26%, while it is approximately 17.3% for the region consisting of 11 tiles.

cells excluded	RL ₂₀₂₁ (100)	RP ₂₀₂₁ ^{reg} (RL ₂₀₂₁ (100))	RL ₂₀₂₁ ^{reg} (100)
11	52.74	18.90	66.40
11, 17	52.82	18.32	67.08
10, 11, 14, 15, 16	51.93	32.76	60.93

Table 3: Estimated local (second column) and regional (fourth column) 100-year RLs for reference climate 2021, for three possible choices of pooling regions as indicated by the first column. Column 3 shows the regional return periods of the local 100-year events.

5.2 A homogeneous scaling model with location-wise scaling factor

In this section, we maintain the temporal dynamics from the scale-GEV model from Equation (2). However, instead of testing for the homogeneity assumption from Equation (3), we additionally allow for a location-wise scaling factor under the null hypothesis. Such an approach can be useful when it is known that observations from different locations occur on different scales, but, apart from that, show a common probabilistic behaviour. In fact, a stationary version of the following model is commonly used in hydrology, where it is known as the Index Flood approach (Dalrymple, 1960).

More precisely, suppose that

$$M_{t,d} \sim c_d \exp\left(\frac{\alpha \text{GMST}'(t)}{\mu}\right) \text{GEV}(\mu, \sigma, \gamma) \quad \forall t, d, \quad (15)$$

where $c_d > 0$ is a location-specific scaling factor that we may fix to 1 at the location of primary interest (say $d = 1$, i.e., $c_1 = 1$). Writing $\mu_d = c_d\mu$, $\sigma_d = c_d\sigma$, $\alpha_d = c_d\alpha$, the model in Equation (15) can be rewritten as

$$M_{t,d} \sim \text{GEV}(\mu_d(t), \sigma_d(t), \gamma) \quad \forall t, d,$$

where

$$\mu_d(t) = \mu_d \exp\left(\frac{\alpha_d \text{GMST}'(t)}{\mu_d}\right), \quad \sigma_d(t) = \sigma_d \exp\left(\frac{\alpha_d \text{GMST}'(t)}{\mu_d}\right). \quad (16)$$

Note that the parameters $\mu_1, \dots, \mu_D, \sigma_1, \dots, \sigma_D, \alpha_1, \dots, \alpha_D$ satisfy the relationships

$$\frac{\mu_d}{\sigma_d} \equiv \delta, \quad \frac{\alpha_d}{\mu_d} \equiv \eta, \quad \frac{\alpha_d}{\sigma_d} \equiv \kappa$$

for certain parameters δ, η, κ ; in particular, $\mu_1, \dots, \mu_D, \sigma_1, \dots, \sigma_D, \alpha_1, \dots, \alpha_D$ can be retrieved from $\mu_1, \dots, \mu_D, \delta, \eta$ (note that the constraint on α_d/σ_d is not needed, but comes as a consequence of the first two relations). Fitting this model instead of fitting the scale-GEV distribution to each location separately has the advantage of reducing the number of parameters that need to be estimated substantially ($4 + (D - 1) = D + 3$ instead of $4D$ parameters). Once the local scaling factors are identified, we can bring all observations to the same scale by dividing each location by its location-specific scaling factor.

Now one can test whether such a local scaling model holds on a subset $A = \{d_1, \dots, d_k\} \subset \{1, \dots, D\}$ with cardinality $k = |A| \geq 2$, by testing the hypothesis

$$H_0^{\text{LS}}(A) : \exists \delta_A, \eta_A, \gamma_A \forall d \in A : \frac{\mu_d}{\sigma_d} = \delta_A, \quad \frac{\alpha_d}{\mu_d} = \eta_A, \quad \gamma_d = \gamma_A, \quad (17)$$

with a Wald-type test statistic. In this case, the latter is defined as

$$T_n^{\text{LS}}(A) = n(g_A(\hat{\boldsymbol{\theta}}))^\top \left(\mathbf{G}_A(\hat{\boldsymbol{\theta}}) \hat{\boldsymbol{\Sigma}}_n \mathbf{G}_A(\hat{\boldsymbol{\theta}})^\top \right)^{-1} g_A(\hat{\boldsymbol{\theta}}), \quad (18)$$

where $g_A : \mathbb{R}^{4D} \rightarrow \mathbb{R}^{3(k-1)}$ is given by

$$g_A(\boldsymbol{\theta}) = \left(\frac{\mu_{d_1}}{\sigma_{d_1}} - \frac{\mu_{d_2}}{\sigma_{d_2}}, \gamma_{d_1} - \gamma_{d_2}, \frac{\alpha_{d_1}}{\mu_{d_1}} - \frac{\alpha_{d_2}}{\mu_{d_2}}, \dots, \gamma_{d_{k-1}} - \gamma_{d_k}, \frac{\alpha_{d_{k-1}}}{\mu_{d_{k-1}}} - \frac{\alpha_{d_k}}{\mu_{d_k}} \right)^\top,$$

with Jacobian matrix $G_A(\boldsymbol{\theta}) \in \mathbb{R}^{3(k-1) \times 4D}$, since the hypothesis in Equation (17) may be rewritten as $H_0^{\text{LS}}(A) : g_A(\boldsymbol{\theta}) = 0$.

When considering this kind of modification, the bootstrap algorithms from Section 2.4, steps (5)-(7), must be adapted accordingly. In step (5), one has to estimate the parameter under the constraint of the considered null hypothesis by adapting the log-likelihood accordingly. The estimated parameters are then used during the transformation step (6). Further, the test statistic in steps (6) and (7) is replaced by $T_n^{\text{LS}}(A)$ from (18). Further details are omitted for the sake of brevity.

5.3 General homogeneous models with smooth parametrization

In this section, we consider general GEV models in which the location, scale and shape parameters are allowed to depend in a (fixed) differentiable way on some parameter vector $\boldsymbol{\vartheta} \in \mathbb{R}^q$ and some temporal covariate $c^{(t)} \in \mathbb{R}^p$ with $p, q \in \mathbb{N}$. More precisely, suppose that f_μ, f_σ and f_γ are (known) real-valued functions of $\boldsymbol{\vartheta}$ and c that are differentiable with respect to their first argument $\boldsymbol{\vartheta}$. We assume that, for each $d = 1, \dots, D$, there exists an unknown parameter $\boldsymbol{\vartheta}_d$ such that $M_d^{(t)} \sim \text{GEV}(\mu_d(t), \sigma_d(t), \gamma_d(t))$ with

$$\mu_d(t) = f_\mu(\boldsymbol{\vartheta}_d; c^{(t)}), \quad \sigma_d(t) = f_\sigma(\boldsymbol{\vartheta}_d; c^{(t)}), \quad \gamma_d(t) = f_\gamma(\boldsymbol{\vartheta}_d; c^{(t)}).$$

The global null hypothesis of interest within this model is assumed to be expressible as $h(\boldsymbol{\vartheta}_1, \dots, \boldsymbol{\vartheta}_D) = 0$ for a differentiable function $h : \mathbb{R}^{qD} \rightarrow \mathbb{R}^s$ with $s \in \mathbb{N}$.

An example is given by the linear shift model that is frequently considered when modelling temperature extremes in Extreme Event Attribution studies (see Philip *et al.*, 2020), where

$$\mu_d(t) = \mu_d + \alpha_d \text{GMST}'(t), \quad \sigma_d(t) \equiv \sigma_d, \quad \gamma_d(t) \equiv \gamma_d.$$

A possible global null hypothesis of interest could be

$$H_0 : \exists \boldsymbol{\vartheta} \in \mathbb{R} \times (0, \infty) \times \mathbb{R}^2 \forall d \in \{1, \dots, D\} : \boldsymbol{\vartheta}_d = \boldsymbol{\vartheta},$$

where $\boldsymbol{\vartheta} = (\mu, \sigma, \gamma, \alpha)^\top$ and $\boldsymbol{\vartheta}_d = (\mu_d, \sigma_d, \gamma_d, \alpha_d)^\top$.

When considering this kind of extension, one has to adapt the maximum likelihood estimator as well as the estimator of its covariance matrix, hence steps (1)-(2) and (5)-(7) in the bootstrap algorithms are affected. Further details are omitted for the sake of brevity.

6 Conclusion

Extreme event attribution studies can build upon a GEV scaling model. Depending on the analysed variable, it may be useful to apply spatial pooling and fit the GEV distribution to a pooled sample of observations collected at sufficiently homogeneous spatial locations as it has been done in Kreienkamp *et al.* (2021); Tradowsky *et al.* (2022); Vautard *et al.* (2015), among others. Here, we propose several statistical methods that enable the selection of a homogeneous pooling region from a larger initial region. The BH approach was found to be quite conservative, hence some heterogeneous locations are likely to be declared homogeneous. The IM approach is a more liberal alternative with a higher proportion of rejected locations that may contain some homogeneous ones. In subsequent analyses, the selected pooling region typically results in a classical bias-variance trade-off: the larger the pooling region, the smaller the variance. At the same time, the bias may be larger, given that some heterogeneous regions may have been declared homogeneous. In practice, the tests should always be complemented by expert knowledge on the driving meteorological/climatological background processes.

To make the statistical approach to select homogeneous pooling regions for attribution studies as described here usable for the extreme event attribution community, we have developed a software package that can be freely downloaded and used by applied researchers (Zanger, 2022). The selection of spatial pooling regions for attribution studies may hence be based on a combination of expert knowledge and thorough statistical tests. The experts applying the methods can thereby decide between a conservative approach, which tends to reject more locations and a liberal approach which tends to accept more locations as being homogeneous. This decision depends on the a priori knowledge about the meteorology of the analysed area and the specific requirements of the study.

If the ultimate interest is estimation of, for example, return levels, one may, as an alternative to the classical approach based on pooling selected locations, consider penalized maximum likelihood estimators with a penalty on large heterogeneities (Bücher *et al.*, 2021). A detailed investigation of the resulting bias-variance trade-off would be a worthwhile topic for future research.

Declaration

Funding This work has been supported by the integrated project “Climate Change and Extreme Events - ClimXtreme Module B - Statistics (subprojects B1.2, grant number: 01LP1902B and B3.3, grant number: 01LP1902L)” funded by the German Federal Ministry of Education and Research (BMBF).

Data and Code Availability All methods are implemented in the R package `findpoolreg` (Zanger, 2022) that is publicly available at <https://github.com/leandrazan/findpoolreg>. The data used throughout the case study is derived from the E-OBS gridded data set, publicly available at <https://www.ecad.eu/download/ensembles/download.php>.

Acknowledgements Computational infrastructure and support were provided by the Centre for Information and Media Technology at Heinrich Heine University Düsseldorf, which is gratefully acknowledged.

References

- Akaike H (1973). “Information theory and an extension of the maximum likelihood principle.” In “Second International Symposium on Information Theory (Tsahkadsor, 1971),” pp. 267–281. Akadémiai Kiadó, Budapest.
- Benjamini Y, Hochberg Y (1995). “Controlling the false discovery rate: a practical and powerful approach to multiple testing.” *J. Roy. Statist. Soc. Ser. B*, **57**(1), 289–300. ISSN 0035-9246. URL [http://links.jstor.org/sici?sici=0035-9246\(1995\)57:1<289:CTFDRA>2.0.CO;2-E&origin=MSN](http://links.jstor.org/sici?sici=0035-9246(1995)57:1<289:CTFDRA>2.0.CO;2-E&origin=MSN).

- Benjamini Y, Yekutieli D (2001). “The control of the false discovery rate in multiple testing under dependency.” *Ann. Statist.*, **29**(4), 1165–1188. ISSN 0090-5364. doi: 10.1214/aos/1013699998. URL <https://doi.org/10.1214/aos/1013699998>.
- Bücher A, Lilienthal J, Kinsvater P, Fried R (2021). “Penalized quasi-maximum likelihood estimation for extreme value models with application to flood frequency analysis.” *Extremes*, **24**(2), 325–348. ISSN 1386-1999. doi:10.1007/s10687-020-00379-y. URL <https://doi.org/10.1007/s10687-020-00379-y>.
- Bücher A, Segers J (2017). “On the maximum likelihood estimator for the generalized extreme-value distribution.” *Extremes*, **20**(4), 839–872. ISSN 1386-1999. doi:10.1007/s10687-017-0292-6. URL <https://doi.org/10.1007/s10687-017-0292-6>.
- Coles S (2001). *An introduction to statistical modeling of extreme values*. Springer Series in Statistics. Springer-Verlag London, Ltd., London. ISBN 1-85233-459-2. doi:10.1007/978-1-4471-3675-0. URL <https://doi.org/10.1007/978-1-4471-3675-0>.
- Cornes RC, van der Schrier G, van den Besselaar EJ, Jones PD (2018). “An ensemble version of the E-OBS temperature and precipitation data sets.” *Journal of Geophysical Research: Atmospheres*, **123**(17), 9391–9409. URL <https://doi.org/10.1029/2017JD028200>.
- Dalrymple T (1960). “Flood-frequency analyses, manual of hydrology: Part 3.” *Technical report*, USGPO,.
- Davison AC, Gholamrezaee MM (2012). “Geostatistics of extremes.” *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **468**(2138), 581–608. URL <https://doi.org/10.1098/rspa.2011.0412>.
- Davison AC, Padoan SA, Ribatet M (2012). “Statistical Modeling of Spatial Extremes.” *Statistical Science*, **27**(2), 161 – 186. doi:10.1214/11-STS376. URL <https://doi.org/10.1214/11-STS376>.
- Holm S (1979). “A simple sequentially rejective multiple test procedure.” *Scand. J. Statist.*, **6**(2), 65–70. ISSN 0303-6898. URL <https://www.jstor.org/stable/4615733>.
- Kreienkamp F, Philip SY, Tradowsky JS, Kew SF, Lorenz P, Arrighi J, Belleflamme A, Bettmann T, Caluwaerts S, Chan SC, *et al.* (2021). “Rapid attribution of heavy rainfall events leading to the severe flooding in Western Europe during July 2021.” URL <https://www.worldweatherattribution.org/heavy-rainfall-which-led-to-severe-flooding-in-western-europe-made-more-likely-by-climate-change/>.

- Lehmann EL, Romano JP (2021). *Testing statistical hypotheses*. Springer Texts in Statistics. Springer, Cham, fourth edition. ISBN 978-3-030-70577-0; 978-3-030-70578-7.
- Lilienthal J, Zanger L, Bücher A, Fried R (2022). “A note on statistical tests for homogeneities in multivariate extreme value models for block maxima.” *Environmetrics*, **33**(7), e2746. doi:<https://doi.org/10.1002/env.2746>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/env.2746>.
- Otto FEL, van der Wiel K, van Oldenborgh GJ, Philip S, Kew SF, Uhe P, Cullen H (2018). “Climate change increases the probability of heavy rains in Northern England/Southern Scotland like those of storm Desmond—a real-time event attribution revisited.” *Environmental Research Letters*, **13**(2), 024006. doi:10.1088/1748-9326/aa9663. URL <https://dx.doi.org/10.1088/1748-9326/aa9663>.
- Philip S, Kew S, van Oldenborgh GJ, Otto F, Vautard R, van der Wiel K, King A, Lott F, Arrighi J, Singh R, *et al.* (2020). “A protocol for probabilistic extreme event attribution analyses.” *Advances in Statistical Climatology, Meteorology and Oceanography*, **6**(2), 177–203. doi:10.5194/ascmo-6-177-2020. URL <https://ascmo.copernicus.org/articles/6/177/2020/>.
- R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Ribatet M (2022). *SpatialExtremes: Modelling Spatial Extremes*. R package version 2.1-0, URL <https://CRAN.R-project.org/package=SpatialExtremes>.
- Stephenson AG (2002). “evd: Extreme Value Distributions.” *R News*, **2**(2), 0. URL <https://CRAN.R-project.org/doc/Rnews/>.
- Tradowsky JS, Philip SY, Kreienkamp F, Kew SF, Lorenz P, Arrighi J, Belleflamme A, Bettmann T, Caluwaerts S, Chan SC, Ciavarella A, De Cruz L, de Vries H, Demuth N, Ferrone A, Fischer EM, Fowler HJ, Goergen K, Heinrich D, Henrichs Y, Lenderink G, Kaspar F, Nilson E, Otto FEL, Ragone F, Seneviratne SI, Singh RK, Skålevåg A, Termonia P, Thalheimer L, van Aalst M, Van den Bergh J, Van de Vyver H, Vannitsem S, van Oldenborgh GJ, Van Schaeybroeck B, Vautard R, Vonk D, Wanders N (2022). “Attribution of heavy rainfall events leading to the severe flooding in Western Europe during July 2021.” *Climatic Change*. In revision.
- van der Vaart AW (1998). *Asymptotic Statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge. ISBN 0-521-49603-9; 0-521-78450-6.
- van der Wiel K, Kapnick SB, van Oldenborgh GJ, Whan K, Philip S, Vecchi GA, Singh RK, Arrighi J, Cullen H (2017). “Rapid attribution of the August 2016

- flood-inducing extreme precipitation in south Louisiana to climate change.” *Hydrology and Earth System Sciences*, **21**(2), 897–921. doi:10.5194/hess-21-897-2017. URL <https://hess.copernicus.org/articles/21/897/2017/>.
- van Oldenborgh GJ, van der Wiel K, Sebastian A, Singh R, Arrighi J, Otto F, Haustein K, Li S, Vecchi G, Cullen H (2017). “Attribution of extreme rainfall from Hurricane Harvey, August 2017.” *Environmental Research Letters*, **12**(12), 124009. doi:10.1088/1748-9326/aa9ef2. URL <https://dx.doi.org/10.1088/1748-9326/aa9ef2>.
- Vautard R, Yiou P, van Oldenborgh GJ, Lenderink G, Thao S, Ribes A, Planton S, Dubuisson B, Soubeyroux JM (2015). “Extreme Fall 2014 Precipitation in the Cévennes Mountains.” *Bulletin of the American Meteorological Society*, **96**, S56–S60. doi:10.1175/BAMS-EEE_2014_ch12.1.
- Zanger L (2022). *findpoolreg: Find a possible pooling region for precipitation extremes*. R package version 0.1.0, URL <https://github.com/leandrazan/findpoolreg>.

Supplement to the Paper:
Regional Pooling in Extreme Event Attribution Studies:
an Approach Based on Multiple Statistical Testing

Leandra Zanger¹, Axel Bücher^{*1}, Frank Kreienkamp², Philip Lorenz², and
Jordis Tradowsky²

¹Heinrich-Heine Universität Düsseldorf, Mathematisches Institut,
Düsseldorf, Germany

²Deutscher Wetterdienst, Regionales Klimabüro Potsdam, Potsdam,
Germany

January 5, 2023

Abstract

This supplement contains mathematical details on the maximum likelihood estimator and the estimation of its covariance matrix, as well as additional simulation results and further details on the case study. References like (1) refer to equations from the main paper, while references like (A.1) or Figure B.2 refer to equations or Figures within this appendix.

A Mathematical Details

A.1 Maximum Likelihood estimation

Throughout this section, we provide mathematical details on the coordinate-wise maximum likelihood estimator from Equation (6). In particular, we motivate the approximate normality of $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\vartheta}}_1^\top, \dots, \hat{\boldsymbol{\vartheta}}_D^\top)^\top \in \Theta^D$ with mean $\boldsymbol{\theta} = (\boldsymbol{\vartheta}_1^\top, \dots, \boldsymbol{\vartheta}_D^\top)^\top$ and covariance $n^{-1}\boldsymbol{\Sigma}_n$ with $\boldsymbol{\Sigma}_n = (\boldsymbol{\Sigma}_{n;j,k})_{j,k=1}^D \in \mathbb{R}^{4D \times 4D}$ as defined in Equation (7). As in the stationary GEV-model, the derivations require $\gamma > -1/2$, see Bücher and Segers (2017).

We start by some explicit formulas for the functions appearing in Equations (6) and (7). For that purpose, let $l_{(\mu,\sigma,\gamma)}(x)$ denote the log density of the plain $\text{GEV}(\mu, \sigma, \gamma)$ distribution (see Appendix B in Bücher and Segers, 2017), i.e.,

$$l_{(\mu,\sigma,\gamma)}(x) = -\log(\sigma) - u_\gamma\left(\frac{x-\mu}{\sigma}\right) + (\gamma+1)\log\left(u_\gamma\left(\frac{x-\mu}{\sigma}\right)\right) \quad (\text{A.1})$$

^{*}Corresponding author: axel.buecher@hhu.de

for x such that $1 + \gamma \frac{x-\mu}{\sigma} > 0$; here

$$u_\gamma(z) = \begin{cases} (1 + \gamma z)^{-\frac{1}{\gamma}}, & \gamma \neq 0, \\ \exp(-z), & \gamma = 0. \end{cases}$$

Then, writing $\boldsymbol{\vartheta} = (\mu, \sigma, \gamma, \alpha)^\top$, the log-density $\ell_{\boldsymbol{\vartheta}}(x, c)$ from Equation (5) may be written as

$$\ell_{\boldsymbol{\vartheta}}(x, c) = l_{(\mu(c), \sigma(c), \gamma)}(x), \quad (\text{A.2})$$

where

$$\mu(c) = \mu(c \mid \mu, \alpha) = \mu \exp(\alpha c / \mu), \quad \sigma(c) = \sigma(c \mid \mu, \sigma, \alpha) = \sigma \exp(\alpha c / \mu).$$

We next derive formulas for the gradient and the Hessian of $\boldsymbol{\vartheta} \mapsto \ell_{\boldsymbol{\vartheta}}(x, c)$. For that purpose, let $\dot{l}_{(\mu, \sigma, \gamma)}(x)$ and $\ddot{l}_{(\mu, \sigma, \gamma)}(x)$ denote the respective gradient and Hessian of the standard GEV log density (see Appendix B in Bücher and Segers, 2017) for precise formulas). Note that, in view of the fact that the GEV distribution is a location scale family,

$$\dot{l}_{(\mu, \sigma, \gamma)}(x) = T_\sigma^{-1} \dot{l}_{(0, 1, \gamma)}\left(\frac{x - \mu}{\sigma}\right), \quad T_\sigma = \text{diag}(\sigma, \sigma, 1) \in \mathbb{R}^{3 \times 3}, \quad (\text{A.3})$$

$$\ddot{l}_{(\mu, \sigma, \gamma)}(x) = T_\sigma^{-1} \ddot{l}_{(0, 1, \gamma)}\left(\frac{x - \mu}{\sigma}\right) T_\sigma^{-1}. \quad (\text{A.4})$$

Next, consider the function $p_c : \Theta \rightarrow (0, \infty)^2 \times \mathbb{R}$ defined by $p_c(\mu, \sigma, \gamma, \alpha) = (\mu \exp(\alpha c / \mu), \sigma \exp(\alpha c / \mu), \gamma)^\top$, whose Jacobian is given by $B_c(\mu, \sigma, \alpha)^\top$, where

$$B_c(\mu, \sigma, \alpha) = \begin{pmatrix} \left(1 - \frac{\alpha c}{\mu}\right) \exp\left(\frac{\alpha}{\mu} c\right) & -\frac{\sigma \alpha c}{\mu^2} \exp\left(\frac{\alpha}{\mu} c\right) & 0 \\ 0 & \exp\left(\frac{\alpha}{\mu} c\right) & 0 \\ 0 & 0 & 1 \\ c \exp\left(\frac{\alpha}{\mu} c\right) & \frac{\sigma c}{\mu} \exp\left(\frac{\alpha}{\mu} c\right) & 0 \end{pmatrix}.$$

Then, in view of Equations (A.2) and (A.3), the multivariate chain rule yields

$$\begin{aligned} \dot{\ell}_{\boldsymbol{\vartheta}}(x, c) &= B_c(\mu, \sigma, \alpha) \cdot \dot{l}_{(\mu(c), \sigma(c), \gamma)}(x) \\ &= B_c(\mu, \sigma, \alpha) \cdot T_{\sigma(c)}^{-1} \cdot \dot{l}_{(0, 1, \gamma)}\left(\frac{x - \mu(c)}{\sigma(c)}\right). \end{aligned} \quad (\text{A.5})$$

In view of the multivariate product rule, this equation further implies

$$\begin{aligned}
\ddot{\ell}_{\boldsymbol{\vartheta}}(x, c) &= \begin{pmatrix} \dot{\ell}_{(\mu(c), \sigma(c), \gamma)}(x)^\top \dot{B}_{c,1}(\mu, \sigma, \alpha) \\ \dot{\ell}_{(\mu(c), \sigma(c), \gamma)}(x)^\top \dot{B}_{c,2}(\mu, \sigma, \alpha) \\ \dot{\ell}_{(\mu(c), \sigma(c), \gamma)}(x)^\top \dot{B}_{c,3}(\mu, \sigma, \alpha) \\ \dot{\ell}_{(\mu(c), \sigma(c), \gamma)}(x)^\top \dot{B}_{c,4}(\mu, \sigma, \alpha) \end{pmatrix} + B_c(\mu, \sigma, \alpha) \ddot{\ell}_{(\mu(c), \sigma(c), \gamma)}(x) B_c(\mu, \sigma, \alpha)^\top \\
&= \begin{pmatrix} \dot{\ell}_{(0,1,\gamma)}\left(\frac{x-\mu(c)}{\sigma(c)}\right)^\top T_{\sigma(c)}^{-1} \dot{B}_{c,1}(\mu, \sigma, \alpha) \\ \dot{\ell}_{(0,1,\gamma)}\left(\frac{x-\mu(c)}{\sigma(c)}\right)^\top T_{\sigma(c)}^{-1} \dot{B}_{c,2}(\mu, \sigma, \alpha) \\ \dot{\ell}_{(0,1,\gamma)}\left(\frac{x-\mu(c)}{\sigma(c)}\right)^\top T_{\sigma(c)}^{-1} \dot{B}_{c,3}(\mu, \sigma, \alpha) \\ \dot{\ell}_{(0,1,\gamma)}\left(\frac{x-\mu(c)}{\sigma(c)}\right)^\top T_{\sigma(c)}^{-1} \dot{B}_{c,4}(\mu, \sigma, \alpha) \end{pmatrix} \\
&\quad + B_c(\mu, \sigma, \alpha) T_{\sigma(c)}^{-1} \ddot{\ell}_{(0,1,\gamma)}\left(\frac{x-\mu(c)}{\sigma(c)}\right) T_{\sigma(c)}^{-1} (B_c(\mu, \sigma, \alpha))^\top, \quad (\text{A.6})
\end{aligned}$$

where we used Equations (A.3) and (A.4) for the second equality and where $\dot{B}_{c,j}(\mu, \sigma, \alpha)$ denotes the Jacobian in $\mathbb{R}^{3 \times 4}$ of the j th row of $B_c(\mu, \sigma, \alpha)$ (derivative with respect to $(\mu, \sigma, \gamma, \alpha)$). The latter can be derived explicitly by a tedious but straightforward calculation; we omit precise formulas for the sake of brevity.

We next motivate the claimed normal approximation. First of all, in view of the differentiability of $\boldsymbol{\vartheta} \mapsto \ell_{\boldsymbol{\vartheta}}$, the vector of maximum likelihood estimators is necessarily a zero of the gradient of the log-likelihood function, i.e., we have

$$0 = \frac{1}{n} \sum_{t=1}^n \begin{pmatrix} \dot{\ell}_{\hat{\boldsymbol{\vartheta}}_1}(M_1^{(t)}, c^{(t)}) \\ \vdots \\ \dot{\ell}_{\hat{\boldsymbol{\vartheta}}_D}(M_D^{(t)}, c^{(t)}) \end{pmatrix}.$$

Denoting the true parameter vector by $\boldsymbol{\theta}$, a Taylor expansion implies that

$$\begin{aligned}
0 &= \frac{1}{n} \sum_{t=1}^n \begin{pmatrix} \dot{\ell}_{\boldsymbol{\vartheta}_1}(M_1^{(t)}, c^{(t)}) \\ \vdots \\ \dot{\ell}_{\boldsymbol{\vartheta}_D}(M_D^{(t)}, c^{(t)}) \end{pmatrix} \\
&\quad + \left\{ \frac{1}{n} \sum_{t=1}^n \begin{pmatrix} \ddot{\ell}_{\boldsymbol{\vartheta}_1}(M_1^{(t)}, c^{(t)}) & 0 & \dots & 0 \\ 0 & \ddot{\ell}_{\boldsymbol{\vartheta}_2}(M_2^{(t)}, c^{(t)}) & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \ddot{\ell}_{\boldsymbol{\vartheta}_D}(M_D^{(t)}, c^{(t)}) \end{pmatrix} \right\} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + R_n \\
&\equiv L_{n,\boldsymbol{\theta}} + I_{n,\boldsymbol{\theta}} \cdot (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + R_n,
\end{aligned}$$

where R_n denotes higher order terms which are negligible. Solving for $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$, we obtain that

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \approx -I_{n,\boldsymbol{\theta}}^{-1} \cdot \sqrt{n}L_{n,\boldsymbol{\theta}}.$$

By Equation (A.6), each (4×4) block matrix $I_{n,d,\boldsymbol{\vartheta}_d} = \frac{1}{n} \sum_{t=1}^n \ddot{\ell}_{\boldsymbol{\vartheta}_d}(M_d^{(t)}, c^{(t)})$ on the block-diagonal of $I_{n,\boldsymbol{\theta}}$ is of the form

$$\frac{1}{n} \sum_{t=1}^n f(c^{(t)}) g(Z_d^{(t)}) \quad (\text{A.7})$$

for suitable functions f and g , where

$$Z_d^{(t)} = \{M_d^{(t)} - \mu_d(c^{(t)})\} / \sigma_d(c^{(t)}) \quad (\text{A.8})$$

with $\mu_d(c^{(t)}) = \mu_d \exp(\alpha_d c^{(t)} / \mu_d)$ and $\sigma(c^{(t)}) = \sigma_d \exp(\alpha_d c^{(t)} / \mu_d)$ is $\text{GEV}(0, 1, \gamma_d)$ -distributed and independent over t . Under suitable assumptions on $t \mapsto f(c^{(t)})$ (and hence on $c^{(t)}$), we obtain that the variance of $I_{n,d,\boldsymbol{\vartheta}_d}$ is of the order $1/n$. As a consequence, $I_{n,d,\boldsymbol{\vartheta}_d}$ is close to its expectation, that is, $I_{n,d,\boldsymbol{\vartheta}_d} = J_{n,d,\boldsymbol{\vartheta}_d} + o(1)$ with $J_{n,d,\boldsymbol{\vartheta}_d}$ defined just after Equation (7). More precisely, in an asymptotic framework where one assumes that $c^{(t)} = h(t/n)$ for some continuous function $h : [0, 1] \rightarrow \mathbb{R}$, expressions as in Equation (A.7) converge to $\int_0^1 f(h(t)) dt \times \mathbb{E}[g(Z_d)]$ with $Z_d \sim \text{GEV}(0, 1, \gamma_d)$ (note that both the integral and the expectation exist).

Next, consider $\sqrt{n}L_{n,\boldsymbol{\theta}}$. It suffices to argue that we may apply a suitable version of the Central Limit Theorem, under suitable assumptions on $t \mapsto c^{(t)}$. Similar as for $I_{n,\boldsymbol{\theta}}$, by Equation (A.5), each entry of $\sqrt{n}L_{n,\boldsymbol{\theta}}$ is of the form

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n f(c^{(t)}) g(Z_d^{(t)})$$

for certain functions f and g and for some $d \in \{1, \dots, D\}$. In view of the independence over t and the fact that $Z_d^{(t)} \sim \text{GEV}(0, 1, \gamma_d)$, we may for instance apply the Ljapunov CLT for independent triangular arrays, see Theorem 27.3 in Billingsley (1995). Since $\mathbb{E}[g(Z_d^{(t)})^p] < \infty$ for sufficiently small $p > 2$ and for the functions g of interest, a sufficient condition for its applicability is

$$\lim_{n \rightarrow \infty} \frac{\sum_{t=1}^n \{f(c^{(t)})\}^p}{[\sum_{t=1}^n \{f(c^{(t)})\}^2]^{p/2}} = 0,$$

which readily follows for instance if one assumes that $c^{(t)} = h(t/n)$ for some continuous function $h : [0, 1] \rightarrow \mathbb{R}$.

A.2 Covariance Estimation

Throughout this section, we provide an estimator for $\boldsymbol{\Sigma}_n = (\boldsymbol{\Sigma}_{n;j,k})_{j,k=1}^D$ defined in Equation (7). First of all, we denote by $\hat{J}_{n,d,\boldsymbol{\vartheta}_d}$ an (approximate) Hessian of the function

$$\boldsymbol{\vartheta}_d \mapsto \frac{1}{n} \sum_{t=1}^n \ell_{\boldsymbol{\vartheta}_d}(M_d^{(t)}, c^{(t)})$$

evaluated at its maximizer $\hat{\boldsymbol{\vartheta}}_d$, possibly obtained by numerical differentiation. Note that this matrix is routinely returned by standard implementations for maximization; for instance, the `optim`-function in R returns an output value `hessian`.

It remains to estimate the matrix

$$\mathbf{C}_{n,j,k} := \frac{1}{n} \sum_{t=1}^n \text{Cov} [\dot{\ell}_{\boldsymbol{\vartheta}_j}(M_j^{(t)}, c^{(t)}), \dot{\ell}_{\boldsymbol{\vartheta}_k}(M_k^{(t)}, c^{(t)})] \in \mathbb{R}^{4 \times 4}$$

for all $1 \leq j < k \leq D$. By Equation (A.5), we may write

$$\begin{aligned} \mathbf{C}_{n,j,k} &= \frac{1}{n} \sum_{t=1}^n B_{c^{(t)}}(\mu_j, \sigma_j, \alpha_j) T_{\sigma_j(c^{(t)})}^{-1} \\ &\quad \times \text{Cov} \left(\dot{l}_{(0,1,\gamma_j)}(Z_j^{(t)}), \dot{l}_{(0,1,\gamma_k)}(Z_k^{(t)}) \right) T_{\sigma_k(c^{(t)})}^{-1} (B_{c^{(t)}}(\mu_k, \sigma_k, \alpha_k))^\top. \end{aligned}$$

with $\dot{l}_{(0,1,\gamma)}$ the gradient of $l_{(0,1,\gamma)}$ from Equation (A.1) and with $Z_j^{(t)}$ as defined in Equation (A.8), which is $\text{GEV}(0, 1, \gamma_j)$ -distributed for any $j = 1, \dots, D$. Note that the cross covariance $\mathbf{\Gamma}_{j,k} = \text{Cov}(\dot{l}_{(0,1,\gamma_j)}(Z_j^{(t)}), \dot{l}_{(0,1,\gamma_k)}(Z_k^{(t)}))$ does not depend on t , and may hence be estimated empirically after replacing the true parameters by their estimators. More precisely, we obtain the estimator

$$\hat{\mathbf{C}}_{n,j,k} = \frac{1}{n} \sum_{t=1}^n B_{c^{(t)}}(\hat{\mu}_j, \hat{\sigma}_j, \hat{\alpha}_j) T_{\hat{\sigma}_j(c^{(t)})}^{-1} \hat{\mathbf{\Gamma}}_{n,j,k} T_{\hat{\sigma}_k(c^{(t)})}^{-1} B_{c^{(t)}}(\hat{\mu}_k, \hat{\sigma}_k, \hat{\alpha}_k)$$

where $\hat{\sigma}_j(c) = \hat{\sigma}_j \exp(\hat{\alpha}_j c / \hat{\mu}_j)$ and where $\hat{\mathbf{\Gamma}}_{n,j,k}$ denotes the empirical cross covariance matrix of the two samples $(\dot{l}_{(0,1,\hat{\gamma}_j)}(\hat{Z}_j^{(t)}))_{t=1}^n$ and $(\dot{l}_{(0,1,\hat{\gamma}_k)}(\hat{Z}_k^{(t)}))_{t=1}^n$ with

$$\hat{Z}_j^{(t)} = \frac{M_j^{(t)} - \hat{\mu}_j(c^{(t)})}{\hat{\sigma}_j(c^{(t)})}$$

and $\hat{\mu}_j(c) = \hat{\mu}_j \exp(\hat{\alpha}_j c / \hat{\mu}_j)$. The final estimator for $\mathbf{\Sigma}_n$ is $\hat{\mathbf{\Sigma}}_n = (\hat{\mathbf{\Sigma}}_{n;j,k})_{j,k=1}^D$ with

$$\hat{\mathbf{\Sigma}}_{n;j,k} = \hat{J}_{n,j,\vartheta_d}^{-1} \hat{\mathbf{C}}_{n,j,k} \hat{J}_{n,k,\vartheta_k}^{-1}.$$

B Additional results of the simulation study

B.1 Additional results for record length $n = 75$

We report the power properties obtained with the BH and IM method for procedure (B3) in Figures B.1 and B.2, respectively. Power properties obtained with the Holm method are shown in Figures B.3 (for (B2)) and B.4 (for (B3)).

B.2 Additional results for record length $n = 100$

Since the bootstrap procedures implicitly depend on the asymptotic distribution of the test statistic, we repeated the simulation study with a larger sample size, in order to investigate the sample size's impact on the performance of the bootstrap procedure. The location-wise sample size is increased to $n = 100$. Again, the FDR and FWER are reported (Table B.1), as well as the power plots for BH and (B2) in Figure B.5, and for IM and (B2) in Figure B.6. As expected, the error rates are again sufficiently controlled by those methods that claim to do so, while the power has substantially increased (on average by 50% for the BH method and by 18% for the IM method). The results for the other methods and (B3) were again similar.

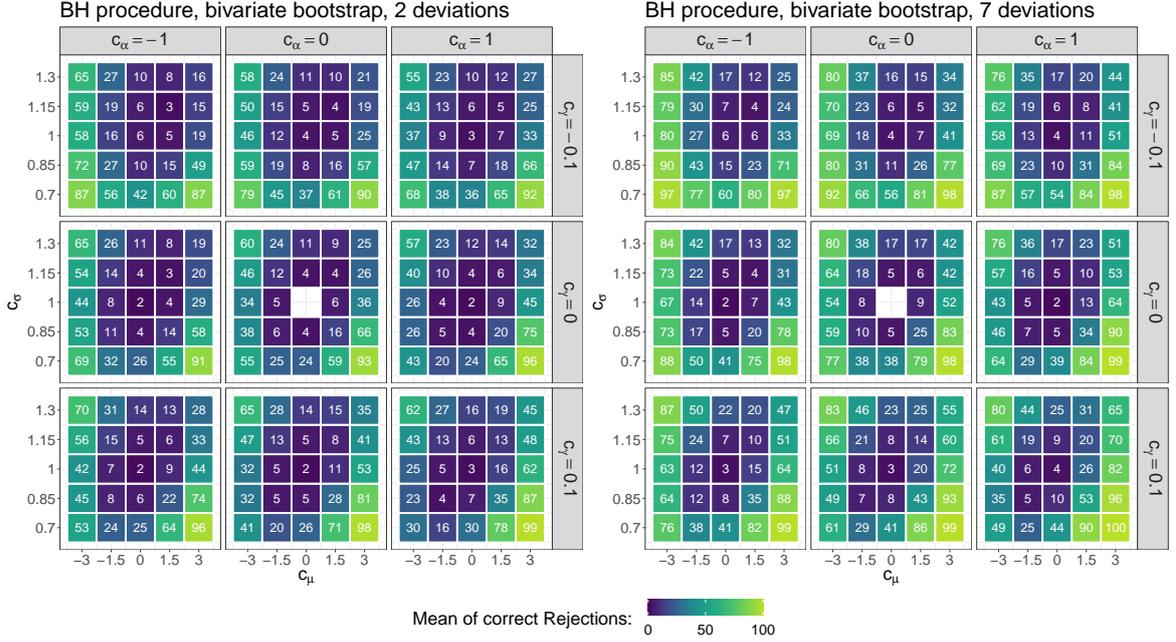


Figure B.1: Proportion of correct rejections in % obtained with the Benjamini Hochberg procedure at a level of 0.1, in the setting where two stations deviate from the rest (left column) or 7 stations deviate from the rest (right column), with the bootstrap procedure based on bivariate extreme value distributions. The axis and facets are as described in Figure 2.

C Additional results for the case study

The complete results of the bootstrap procedures applied to the 4×4 can be found in Table C.1. For the 6×6 grid, the complete results of the bootstrap based on bivariate extreme value distributions can be found in Table C.2, and the results for the bootstrap procedure based on max-stable processes in Table C.3.

The time series used throughout the case study are shown in Figure C.1. Along with the Apr-Sep maxima of 1950-2021, we plot values of

$$\hat{\mu}(t) = \hat{\mu} \exp\left(\frac{\hat{\alpha} \text{GMST}(t)}{\hat{\mu}}\right),$$

where $\hat{\mu}$ and $\hat{\alpha}$ are estimated on the data of location 21 only (blue line), the respective location d (red line) or the pooled data of the pair $(21, d)$ (green line), for $d \in \{1, \dots, 36\} \setminus \{21\}$. Note that these three lines should not differ much when the homogeneity assumption holds. On the other hand, perfectly matching lines do not imply that the homogeneity assumption is true, since they do not give any information about the scale and shape parameter of the distributions.

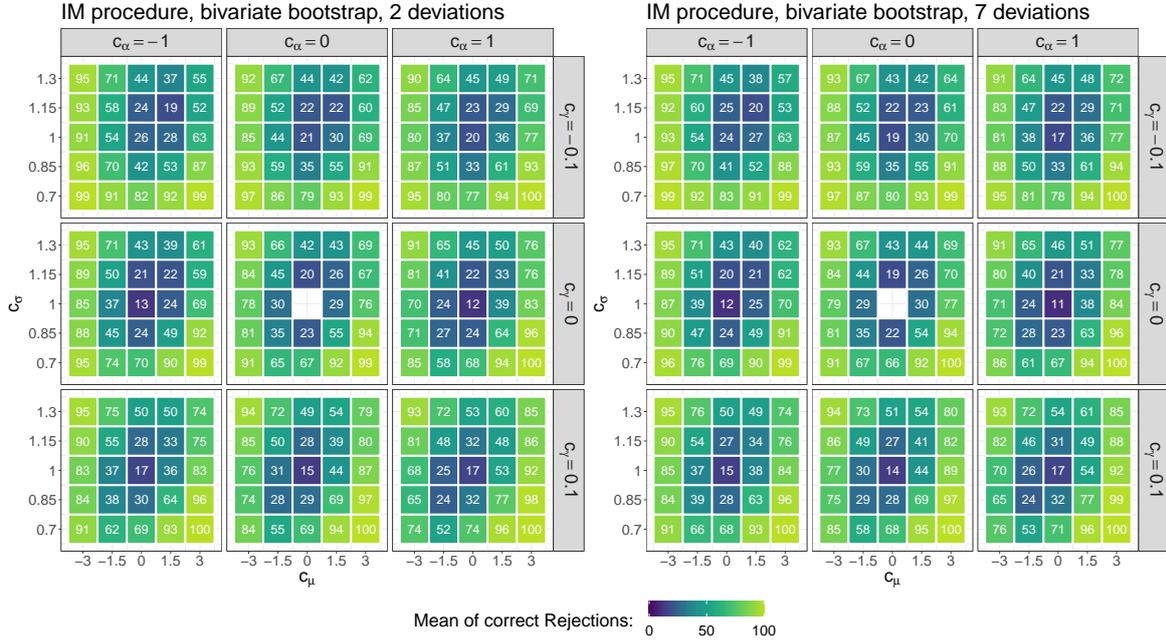


Figure B.2: Proportion of correct rejections in % obtained with the Benjamini Hochberg procedure at a level of 0.1, in the setting where two stations deviate from the rest (left column) or 7 stations deviate from the rest (right column), with the bootstrap procedure based on bivariate extreme value distributions. The axis and facets are as described in Figure 2.

Method	min FDR		max FDR		mean FDR		min FWER		max FWER		mean FWER	
	(B2)	(B3)	(B2)	(B3)	(B2)	(B3)	(B2)	(B3)	(B2)	(B3)	(B2)	(B3)
<i>Scenario 1: $A_{\text{dev}} = 2$</i>												
BH	6.9	5.2	12.4	11.4	9.4	7.8	8.7	6.8	23.9	20.4	15.8	13.4
Holm	2.7	1.4	10.3	8.0	6.4	4.6	6.8	3.9	12.9	9.4	9.5	6.7
IM	26.4	25.7	60.3	59.9	35.2	34.6	53.4	52.7	64.4	63.8	59.2	58.3
<i>Scenario 2: $A_{\text{dev}} = 7$</i>												
BH	4.0	2.7	10.7	8.3	5.6	5.0	6.3	5.1	32.4	32.6	21.6	19.9
Holm	0.8	0.5	10.3	6.5	2.9	2.1	4.2	3.0	11.2	8.0	7.1	5.2
IM	8.2	7.8	60.1	59.7	14.2	13.8	41.8	40.9	60.1	59.7	47.9	46.6

Table B.1: False Discovery Rate (FDR) and Familywise Error Rate (FWER) for the three p-value combination methods from Section 2.5 and the two bootstrap methods (B2) and (B3), obtained in the simulations with record length $n = 100$. The stated values are the minimum, maximum and mean across the 224 models from each scenario.

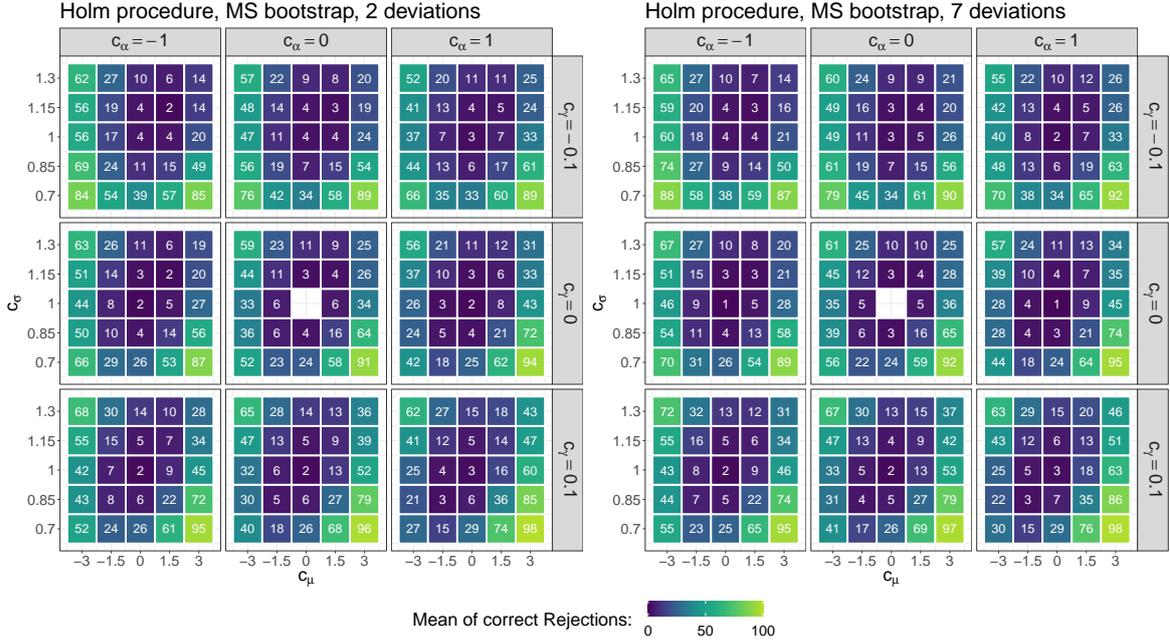


Figure B.3: Proportion of correct rejections in % obtained with the Holm procedure at a level of 0.1, in the setting where two stations deviate from the rest (left column) or 7 stations deviate from the rest (right column), with the bootstrap procedure based on max-stable processes. The axis and facets are as described in Figure 2.

		MS bootstrap (B2)			bivariate bootstrap (B3)		
Pair	t_n	p_{raw}	p_{BH}	p_{Holm}	p_{raw}	p_{BH}	p_{Holm}
(21, 11)	78.8	0.00	0.00	0.00	0.05	0.75	0.75
(21, 16)	15.6	1.60	10.64	22.39	1.45	9.90	20.29
(21, 15)	15.2	2.50	10.64	32.48	2.50	9.90	32.48
(21, 10)	13.6	3.40	10.64	40.78	3.30	9.90	36.58
(21, 14)	13.7	3.55	10.64	40.78	3.05	9.90	36.58
(21, 23)	12.1	5.30	13.24	52.97	6.30	15.74	62.97
(21, 20)	10.4	7.15	15.09	64.32	9.85	16.36	78.76
(21, 27)	10.6	8.05	15.09	64.37	8.10	16.36	72.86
(21, 9)	9.8	10.00	15.59	69.97	10.44	16.36	78.76
(21, 22)	9.4	10.39	15.59	69.97	11.69	16.36	78.76
(21, 8)	9.0	13.04	17.79	69.97	11.99	16.36	78.76
(21, 26)	7.2	20.34	25.42	81.36	22.04	27.55	88.16
(21, 17)	4.4	46.08	53.17	100.00	45.73	52.76	100.00
(21, 28)	3.9	52.17	55.90	100.00	50.92	54.56	100.00
(21, 29)	2.8	66.92	66.92	100.00	65.77	65.77	100.00

Table C.1: (Adjusted) p-values for all three methods from Section 2.5 obtained with both bootstrap methods applied to the 4×4 grid. Values that are significant at the 10%-level are in boldface. Results are based on $B = 2000$ bootstrap replications.

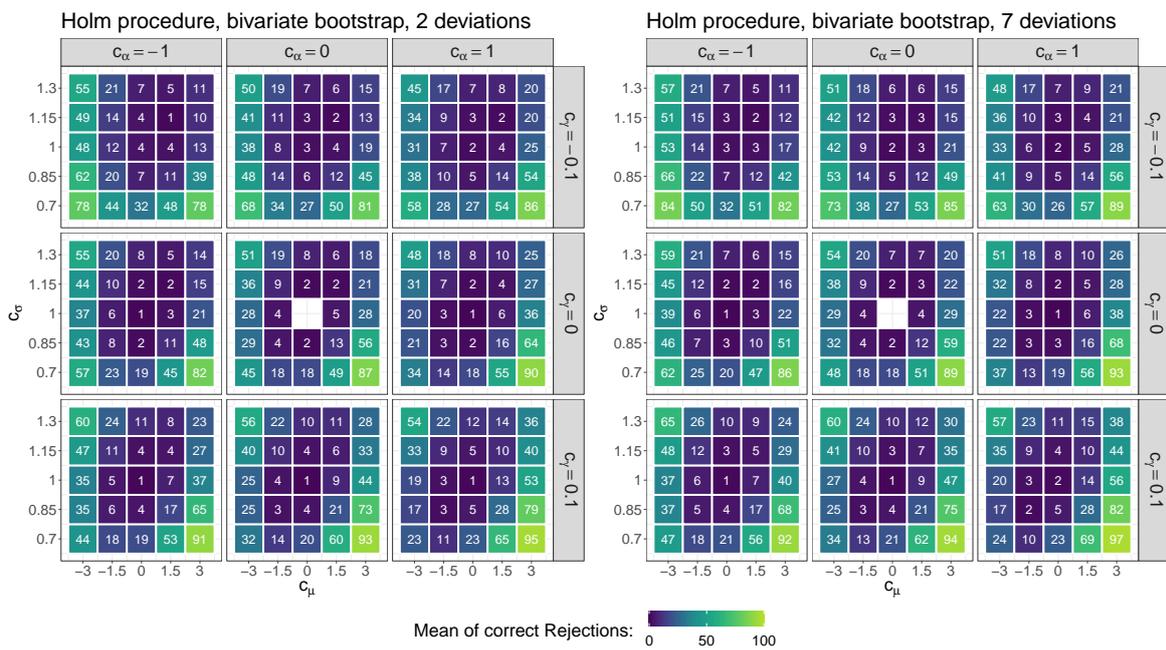


Figure B.4: Proportion of correct rejections in % obtained with the Holm procedure at a level of 0.1, in the setting where two stations deviate from the rest (left column) or 7 stations deviate from the rest (right column), with the bootstrap procedure based on bivariate extreme value distributions. The axis and facets are as described in Figure 2.

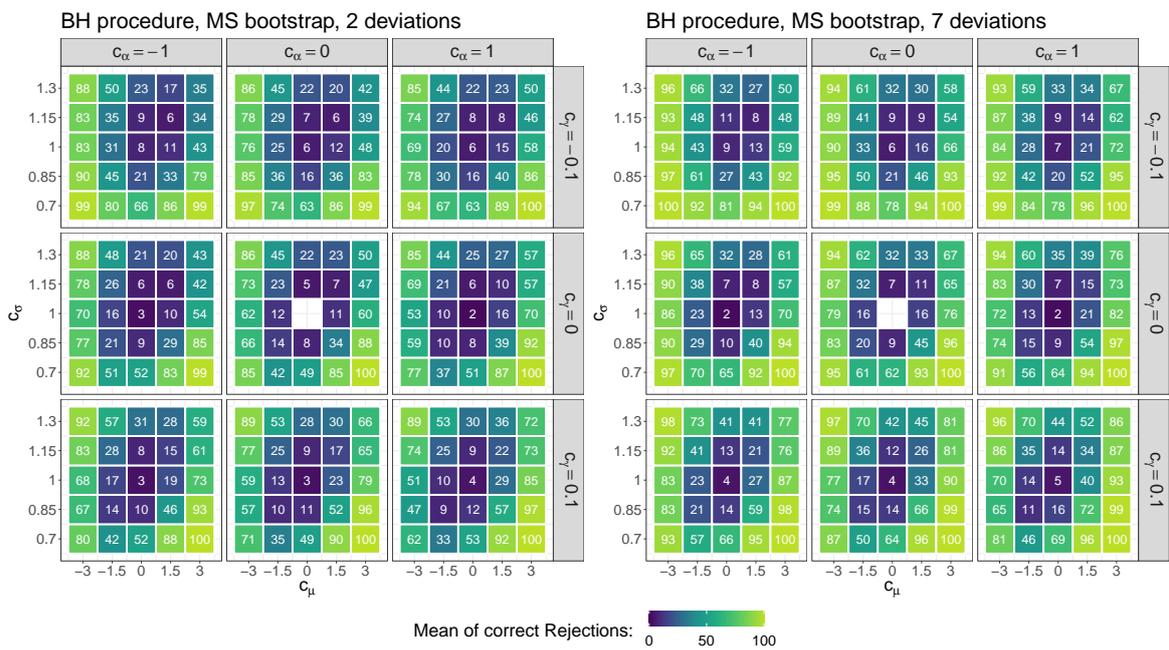


Figure B.5: Proportion of correct rejections in % obtained with the Benjamini Hochberg procedure at a level of 0.1, in the setting where two stations deviate from the rest (left column) or 7 stations deviate from the rest (right column), with the bootstrap procedure based on max-stable processes and record length $n = 100$. The axis and facets are as described in Figure 2.

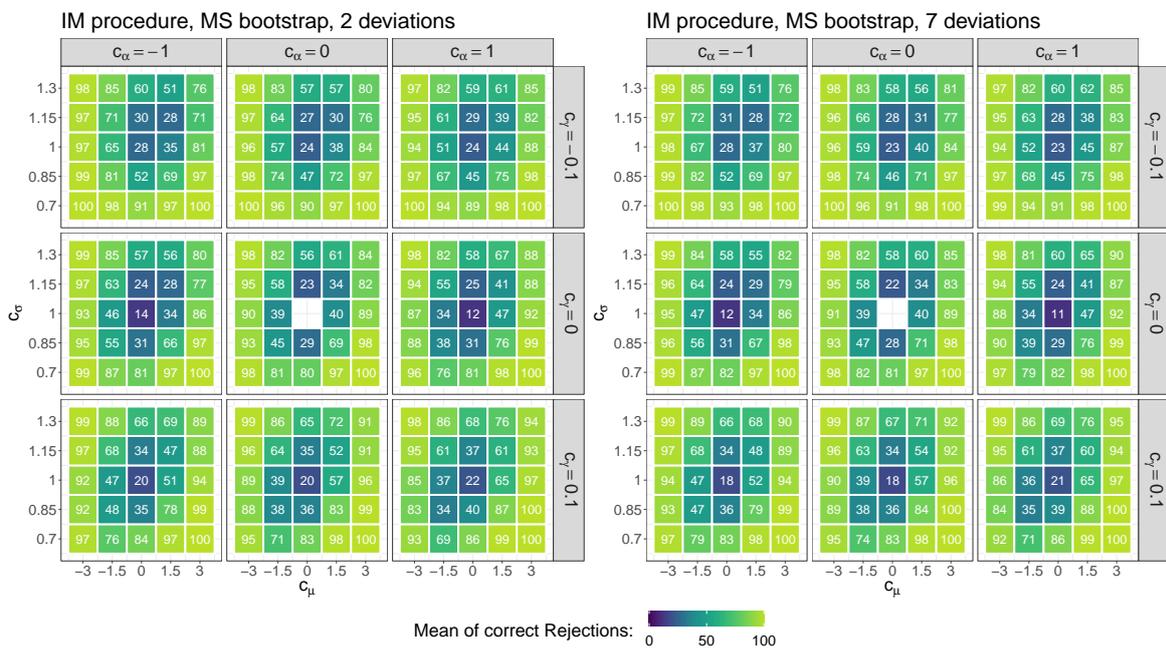


Figure B.6: Proportion of correct rejections in % obtained when ignoring the multiple testing problem, at a level of 0.1, in the setting where two stations deviate from the rest (left column) or 7 stations deviate from the rest (right column), with the bootstrap procedure based on max-stable processes and record length $n = 100$. The axis and facets are as described in Figure 2.

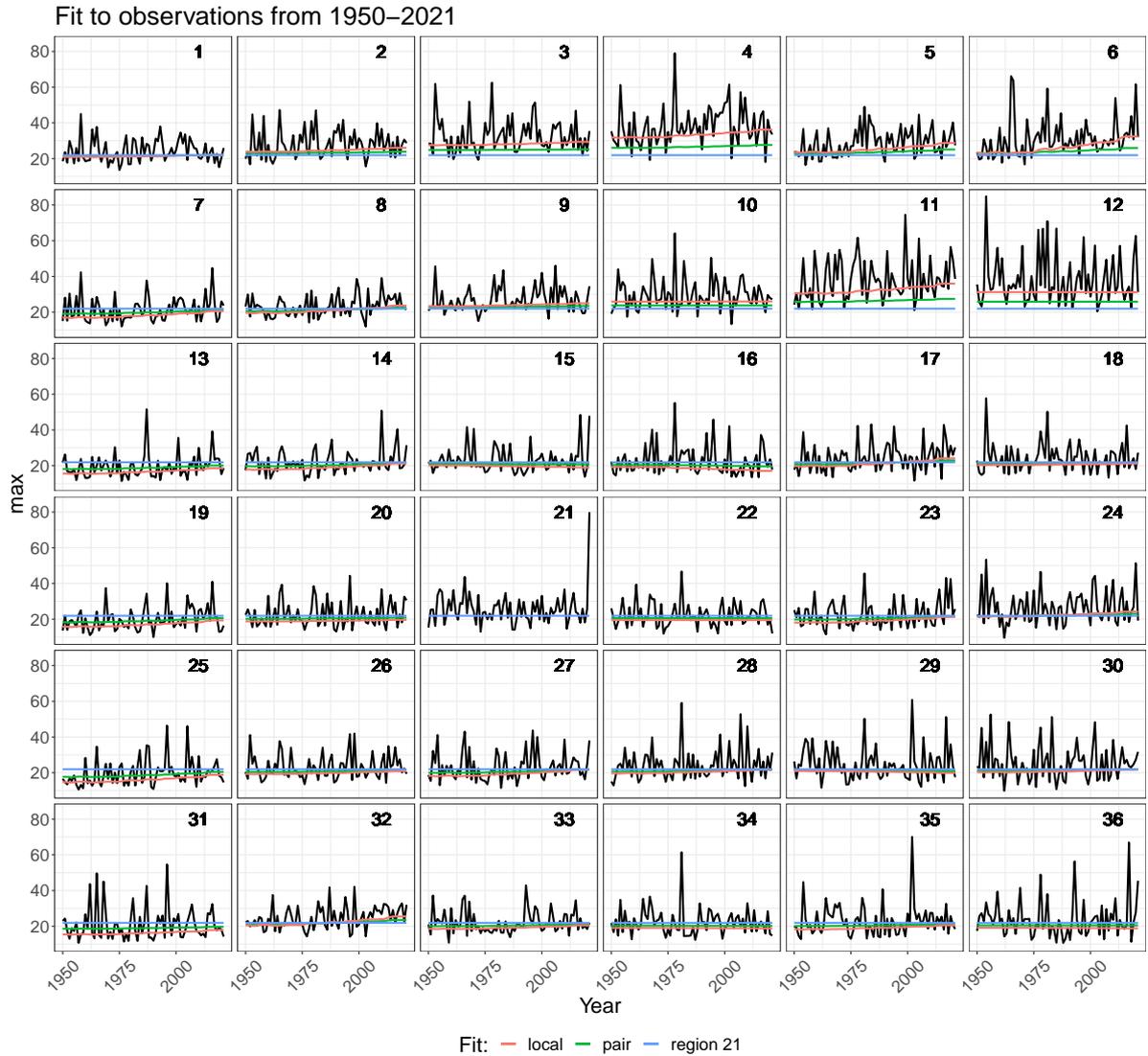


Figure C.1: Observations and fitted trend as estimated for each location d , $d = 1, \dots, 36$, (red line) as well as the trend obtained from the GEV fit for location 21 (blue line) and the GEV-fit obtained from the pooled sample consisting of the pair $(21, d)$, $d = 1, \dots, 36$ (green line). Location labels are given in the top right corner.

Pair	t_n	p_{raw}	p_{BH}	p_{Holm}
(21, 3)	41.5	0.10	0.50	3.50
(21, 4)	87.3	0.10	0.50	3.50
(21, 6)	46.0	0.10	0.50	3.50
(21, 11)	78.8	0.10	0.50	3.50
(21, 12)	75.2	0.10	0.50	3.50
(21, 13)	33.7	0.10	0.50	3.50
(21, 25)	38.5	0.10	0.50	3.50
(21, 19)	28.6	0.20	0.87	5.59
(21, 31)	33.3	0.30	1.17	8.09
(21, 7)	16.1	1.50	5.24	38.96
(21, 5)	15.7	1.70	5.40	42.46
(21, 16)	15.6	2.00	5.83	47.95
(21, 10)	13.6	2.70	6.99	62.04
(21, 15)	15.2	2.80	6.99	62.04
(21, 14)	13.8	4.10	9.56	86.01
(21, 33)	12.5	4.70	10.27	93.91
(21, 23)	12.1	6.89	14.19	100.00
(21, 20)	10.4	8.39	16.19	100.00
(21, 27)	10.6	8.79	16.19	100.00
(21, 36)	10.5	9.89	16.98	100.00
(21, 34)	9.9	10.19	16.98	100.00
(21, 2)	9.6	11.19	17.03	100.00
(21, 9)	9.8	11.19	17.03	100.00
(21, 22)	9.4	13.69	19.44	100.00
(21, 8)	9.0	13.89	19.44	100.00
(21, 35)	8.2	15.88	21.38	100.00
(21, 26)	7.2	19.78	25.64	100.00
(21, 30)	6.1	29.17	36.46	100.00
(21, 32)	5.8	33.07	39.91	100.00
(21, 17)	4.4	46.75	54.55	100.00
(21, 28)	3.9	52.75	59.55	100.00
(21, 29)	2.8	66.13	72.33	100.00
(21, 24)	2.7	70.23	74.49	100.00
(21, 1)	2.4	73.13	75.28	100.00
(21, 18)	1.8	83.82	83.82	100.00

Table C.2: (Adjusted) p-values for all three methods from Section 2.5 obtained with the bootstrap based on bivariate extreme value distributions. Values that are significant at the 10%-level are in boldface. Results are based on $B = 2000$ bootstrap replications.

Pair	t_n	p_{raw}	p_{BH}	p_{Holm}
(21, 3)	41.5	0.00	0.00	0.00
(21, 4)	87.2	0.00	0.00	0.00
(21, 11)	78.8	0.00	0.00	0.00
(21, 12)	75.2	0.00	0.00	0.00
(21, 25)	38.5	0.00	0.00	0.00
(21, 6)	46.0	0.10	0.50	3.00
(21, 13)	33.4	0.10	0.50	3.00
(21, 31)	33.3	0.20	0.87	5.59
(21, 19)	28.7	0.30	1.17	8.09
(21, 7)	16.0	1.80	6.29	46.75
(21, 5)	15.7	2.60	7.53	64.94
(21, 15)	15.2	2.70	7.53	64.94
(21, 16)	15.6	2.80	7.53	64.94
(21, 14)	13.7	3.70	9.24	81.32
(21, 10)	13.6	4.40	10.26	92.31
(21, 23)	12.1	5.09	11.15	100.00
(21, 33)	12.5	5.89	12.13	100.00
(21, 27)	10.6	8.59	15.96	100.00
(21, 20)	10.4	8.89	15.96	100.00
(21, 36)	10.5	9.19	15.96	100.00
(21, 9)	9.8	9.59	15.96	100.00
(21, 2)	9.6	10.49	15.96	100.00
(21, 34)	9.9	10.49	15.96	100.00
(21, 22)	9.4	10.99	16.03	100.00
(21, 8)	9.0	12.39	17.34	100.00
(21, 35)	8.2	16.38	22.05	100.00
(21, 26)	7.2	21.88	28.36	100.00
(21, 30)	6.1	28.77	35.96	100.00
(21, 32)	5.8	31.87	38.46	100.00
(21, 17)	4.4	46.25	53.96	100.00
(21, 28)	3.9	50.95	57.52	100.00
(21, 29)	2.8	66.33	70.99	100.00
(21, 24)	2.7	66.93	70.99	100.00
(21, 1)	2.4	71.03	73.12	100.00
(21, 18)	1.8	81.92	81.92	100.00

Table C.3: (Adjusted) p-values for all three methods from Section 2.5 obtained with the bootstrap based on max-stable processes. Values that are significant at the 10%-level are in boldface. Results are based on $B = 2000$ bootstrap replications.

References

- Billingsley P (1995). *Probability and measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, third edition. ISBN 0-471-00710-2. A Wiley-Interscience Publication.
- Bücher A, Segers J (2017). “On the maximum likelihood estimator for the generalized extreme-value distribution.” *Extremes*, **20**(4), 839–872. ISSN 1386-1999. doi:10.1007/s10687-017-0292-6. URL <https://doi.org/10.1007/s10687-017-0292-6>.

3 Outlook

The findings presented throughout Chapter 2 of this thesis offer several starting points for further research activities, some of which are briefly presented below.

A first appealing research topic would be the investigation of a maximum likelihood estimator $\hat{\vartheta}_{\text{ML}}$ for the GEV parameter vector ϑ based on sliding blocks of univariate (piecewise) stationary time series. So far, ML estimation based on sliding blocks has only been considered in the case of heavy tails, i.e., when the limiting distributions are of the Fréchet-type. Compared to estimation based on disjoint blocks, a substantial reduction of asymptotic estimation variance has been found, with about 19% for the shape and 14% for the scale parameter (Bücher and Segers (2018a)). It seems reasonable to assume that a similar result applies when not restricting to heavy tails, but considering the more general GEV distribution as the limit distribution instead. For deriving asymptotic results for $\hat{\vartheta}_{\text{ML}}$ in this scenario, the fact that the support of a GEV distribution depends on its parameters poses a difficulty, since standard theory does not apply. However, it is possible to identify the candidate of the asymptotic distribution of $\hat{\vartheta}_{\text{ML}}$ by applying a Taylor expansion to the gradient of the (pseudo) log-likelihood, as it has been done in some unpublished work of the author of this thesis. Further, based on this candidate, an estimator for the asymptotic covariance matrix can be proposed. It would thus be interesting to elaborate the details of the asymptotics and to investigate the properties of the proposed estimator.

The next research question that arises concerns the combination of sliding block maxima with non-stationary time series, which has not yet been considered in the literature. For this purpose, it seems convenient to write the underlying observations of the t -th block of length r as $X_1^{(t)}, X_2^{(t)}, \dots, X_r^{(t)}$, for $t = 1, \dots, T$, where T is the number of disjoint blocks that are observed. Further, when letting $I_t = \{1, \dots, r\}$ for $t \leq T - 1$ and $I_t = \{1\}$ for $t = T$, the sliding block maximum that is computed over a block that starts in year t is given by

$$M_{r,i}^{(t,\text{sb})} = \max \left\{ X_i^{(t)}, \dots, X_r^{(t)}, X_1^{(t+1)}, \dots, X_{i-1}^{(t+1)} \right\}$$

for $i \in I_t$. Again, the assumption is that $M_{r,i}^{(t,\text{sb})}$ is GEV-distributed with parameters $\mu(t), \sigma(t)$ and $\gamma(t)$ depending on some covariate $c^{(t)}$. Since the most common indicator for climate

change, the GMST anomaly, is only available at a minimal resolution of monthly values (see, e.g., Lenssen et al. (2019)), it might be ambiguous what covariate value $c^{(t)}$ to choose for $M_{r,i}^{(t, \text{sb})}$ in data applications, as one could e.g. choose the value that corresponds to the month in which the block started, or a weighted mean of the two months whose observations contribute to the sliding block maximum. However, when working out the theory and assuming a continuous covariate function, this should not pose a problem. Candidates for the limiting distributions of (pseudo) ML estimators based on sliding blocks in the shift and scale models from Equations (7) and (8) have been identified by the author of this thesis, and corresponding estimators for the covariance matrix have been proposed. An elaboration of the theoretical details and an examination of the suggested covariance matrix estimator would be of interest.

In both the stationary and the non-stationary setting, an assessment of uncertainties for parameter estimates based on their asymptotic distributions might be inadequate, in particular when block sizes or record lengths are short. Therefore, it would be useful to develop a bootstrap procedure to estimate (co-)variances for parameter estimators, or quantities derived from them, when working with sliding blocks. Here, the bootstrap samples should reflect the dependence structure of the original sliding block maxima sample, which could e.g. be achieved by an appropriate block bootstrap procedure.

Last but not least, the results on the multivariate models considered in Chapters 2.2 and 2.3 could be extended in several possible ways. First of all, one could consider to relax the assumption of exactly GEV-distributed margins, and instead assume a domain of attraction condition. Further, an attempt could be made to use sliding blocks in this framework and prove respective results. Finally, the case of more general non-stationary GEV models as introduced in Section 5.3 of the article presented in Chapter 2.3 could be amplified, e.g., it would be of interest to obtain a general formula for the covariance matrix estimator.

Bibliography

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory (Tsahkadsor, 1971)*, pages 267–281. Akadémiai Kiadó, Budapest.
- Alexander, L. V., Zhang, X., Peterson, T. C., Caesar, J., Gleason, B., Klein Tank, A. M. G., Haylock, M., Collins, D., Trewin, B., Rahimzadeh, F., Tagipour, A., Rupa Kumar, K., Revadekar, J., Griffiths, G., Vincent, L., Stephenson, D. B., Burn, J., Aguilar, E., Brunet, M., Taylor, M., New, M., Zhai, P., Rusticucci, M., and Vazquez-Aguirre, J. L. (2006). Global observed changes in daily climate extremes of temperature and precipitation. *Journal of Geophysical Research: Atmospheres*, 111(D5):1–22.
- Balkema, A. A. and de Haan, L. (1974). Residual life time at great age. *The Annals of Probability*, 2(5):792–804.
- Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester.
- Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, 57(1):289–300.
- Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, 29(4):1165–1188.
- Berbee, H. C. P. (1979). *Random walks with stationary increments and renewal theory*, volume 112 of *Mathematical Centre Tracts*. Mathematisch Centrum, Amsterdam.
- Berghaus, B. and Bücher, A. (2018). Weak convergence of a pseudo maximum likelihood estimator for the extremal index. *The Annals of Statistics*, 46(5):2307–2335.
- Billingsley, P. (1995). *Probability and measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, third edition.

- Blöschl, G., Hall, J., Parajka, J., Perdigão, R. A. P., Merz, B., Arheimer, B., Aronica, G. T., Bilibashi, A., Bonacci, O., Borga, M., Čanjevac, I., Castellarin, A., Chirico, G. B., Claps, P., Fiala, K., Frolova, N., Gorbachova, L., Gül, A., Hannaford, J., Harrigan, S., Kireeva, M., Kiss, A., Kjeldsen, T. R., Kohnová, S., Koskela, J. J., Ledvinka, O., Macdonald, N., Mavrova-Guirguinova, M., Mediero, L., Merz, R., Molnar, P., Montanari, A., Murphy, C., Osuch, M., Ovcharuk, V., Radevski, I., Rogger, M., Salinas, J. L., Sauquet, E., Šraj, M., Szolgay, J., Viglione, A., Volpi, E., Wilson, D., Zaimi, K., and Živković, N. (2017). Changing climate shifts timing of European floods. *Science*, 357(6351):588–590.
- Boukhelifa, M., Meddi, M., and Gaume, E. (2018). Integrated Bayesian Estimation of Intensity-Duration-Frequency Curves: Consolidation and Extensive Testing of a Method. *Water Resources Research*, 54:7459–7477.
- Bradley, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys*, 2:107–144. Update of, and a supplement to, the 1986 original.
- Brockwell, P. J. and Davis, R. A. (1987). *Time series: theory and methods*. Springer Series in Statistics. Springer, New York.
- Bücher, A. and Jennessen, T. (2020). Method of moments estimators for the extremal index of a stationary time series. *Electronic Journal of Statistics*, 14(2):3103–3156.
- Bücher, A., Lilienthal, J., Kinsvater, P., and Fried, R. (2021). Penalized quasi-maximum likelihood estimation for extreme value models with application to flood frequency analysis. *Extremes*, 24(2):325–348.
- Bücher, A. and Segers, J. (2014). Extreme value copula estimation based on block maxima of a multivariate stationary time series. *Extremes*, 17(3):495–528.
- Bücher, A. and Segers, J. (2017). On the maximum likelihood estimator for the generalized extreme-value distribution. *Extremes*, 20(4):839–872.
- Bücher, A. and Segers, J. (2018a). Inference for heavy tailed stationary time series based on sliding blocks. *Electronic Journal of Statistics*, 12(1):1098–1125.
- Bücher, A. and Segers, J. (2018b). Maximum likelihood estimation for the Fréchet distribution based on block maxima extracted from a time series. *Bernoulli*, 24(2):1427–1462.
- Bücher, A. and Zhou, C. (2021). A Horse Race between the Block Maxima Method and the Peak-over-Threshold Approach. *Statistical Science*, 36(3):360–378.
- Castellarin, A., Burn, D., and Brath, A. (2008). Homogeneity testing: How homogeneous do heterogeneous cross-correlated regions seem? *Journal of Hydrology*, 360(1-4):67–76.

- Chebana, F. and Ouarda, T. B. (2009). Index flood-based multivariate regional frequency analysis. *Water Resources Research*, 45(10):W10435.
- Chebana, F. and Ouarda, T. B. M. J. (2007). Multivariate L-moment homogeneity test. *Water Resources Research*, 43(8):W08406.
- Cheng, L., AghaKouchak, A., Gilleland, E., and Katz, R. W. (2014). Non-stationary extreme value analysis in a changing climate. *Climatic change*, 127(2):353–369.
- Cissokho, Y. and Kulik, R. (2021). Estimation of cluster functionals for regularly varying time series: sliding blocks estimators. *Electronic Journal of Statistics*, 15(1):2777–2831.
- Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics. Springer London, Ltd., London.
- Cooley, D. (2012). Return periods and return levels under climate change. In *Extremes in a changing climate: Detection, analysis and uncertainty*, pages 97–114. Springer, Dordrecht.
- Cornes, R. C., van der Schrier, G., van den Besselaar, E. J., and Jones, P. D. (2018). An ensemble version of the E-OBS temperature and precipitation data sets. *Journal of Geophysical Research: Atmospheres*, 123(17):9391–9409.
- Dalrymple, T. (1960). Flood-frequency analyses, manual of hydrology: Part 3. Technical report, USGPO.
- Davison, A. C. and Gholamrezaee, M. M. (2012). Geostatistics of extremes. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 468(2138):581–608.
- Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and their Application*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge.
- Davison, A. C., Padoan, S. A., and Ribatet, M. (2012). Statistical Modeling of Spatial Extremes. *Statistical Science*, 27(2):161–186.
- de Haan, L. (1984). A Spectral Representation for Max-stable Processes. *The Annals of Probability*, 12(4):1194–1204.
- de Haan, L. and Ferreira, A. (2006). *Extreme value theory: an introduction*. Springer, New York.
- Dehling, H. and Philipp, W. (2002). Empirical process techniques for dependent data. In *Empirical process techniques for dependent data*, pages 3–113. Birkhäuser Boston, Boston, MA.

- Dombry, C. (2015). Existence and consistency of the maximum likelihood estimators for the extreme value index within the block maxima framework. *Bernoulli*, 21(1):420–436.
- Dombry, C. and Ferreira, A. (2019). Maximum likelihood estimators based on the block maxima method. *Bernoulli*, 25(3):1690–1723.
- Drees, H. and Neblung, S. (2021). Asymptotics for sliding blocks estimators of rare events. *Bernoulli*, 27(2):1239–1269.
- Eastoe, E. F. and Tawn, J. A. (2009). Modelling non-stationary extremes with application to surface level ozone. *Journal of the Royal Statistical Society: Series C*, 58(1):25–45.
- Eden, J. M., Wolter, K., Otto, F. E., and van Oldenborgh, G. J. (2016). Multi-method attribution analysis of extreme precipitation in Boulder, Colorado. *Environmental Research Letters*, 11(12):124009.
- El Adlouni, S., Ouarda, T. B., Zhang, X., Roy, R., and Bobée, B. (2007). Generalized maximum likelihood estimators for the nonstationary generalized extreme value model. *Water Resources Research*, 43(3):W03410.
- Elamir, E. A. and Seheult, A. H. (2003). Trimmed L-moments. *Computational Statistics & Data Analysis*, 43(3):299–314.
- Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling extremal events: for insurance and finance*. Springer, Berlin.
- Ferreira, A. and de Haan, L. (2015). On the block maxima method in extreme value theory: PWM estimators. *The Annals of Statistics*, 43(1):276–298.
- Fill, H. D. and Stedinger, J. R. (1995). Homogeneity tests based upon Gumbel distribution and a critical appraisal of Dalrymple’s test. *Journal of Hydrology*, 166(1-2):81–105.
- Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(2):180–190.
- Gaume, E., Gaál, L., Viglione, A., Szolgay, J., Kohnová, S., and Blöschl, G. (2010). Bayesian MCMC approach to regional flood frequency analyses involving extraordinary flood events at ungauged sites. *Journal of Hydrology*, 394(1-2):101–117.
- Genest, C. and Segers, J. (2009). Rank-based inference for bivariate extreme-value copulas. *The Annals of Statistics*, 37(5B):2990–3022.
- Greenwood, J. A., Landwehr, J. M., Matalas, N. C., and Wallis, J. R. (1979). Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. *Water Resources Research*, 15(5):1049–1054.

- Gumbel, E. J. (1958). *Statistics of extremes*. Columbia University Press, New York.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics. Theory and Applications*, 6(2):65–70.
- Hosking, J. (2007). Some theory and practical uses of trimmed L-moments. *Journal of Statistical Planning and Inference*, 137(9):3024–3039.
- Hosking, J. and Wallis, J. (1993). Some statistics useful in regional frequency analysis. *Water Resources Research*, 29(2):271–281.
- Hosking, J. R. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society: Series B*, 52(1):105–124.
- Hosking, J. R. M. and Wallis, J. R. (1997). *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge University Press, Cambridge.
- Hosking, J. R. M., Wallis, J. R., and Wood, E. F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, 27(3):251–261.
- IPCC (2021). *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, volume In Press. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Jurado, O. E., Ulrich, J., Scheibel, M., and Rust, H. W. (2020). Evaluating the Performance of a Max-Stable Process for Estimating Intensity-Duration-Frequency Curves. *Water*, 12(12):3314.
- Katz, R. W., Parlange, M. B., and Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in Water Resources*, 25(8):1287–1304.
- Kojadinovic, I. and Naveau, P. (2017). Detecting distributional changes in samples of independent block maxima using probability weighted moments. *Extremes*, 20(2):417–450.
- Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference*. Springer Series in Statistics. Springer, New York.
- Koutsoyiannis, D., Kozonis, D., and Manetas, A. (1998). A mathematical framework for studying rainfall intensity-duration-frequency relationships. *Journal of Hydrology*, 206(1-2):118–135.

- Kreienkamp, F., Philip, S. Y., Tradowosky, J. S., Kew, S. F., Lorenz, P., Arrighi, J., Belleflamme, A., Bettmann, T., Caluwaerts, S., Chan, S. C., et al. (2021). Rapid attribution of heavy rainfall events leading to the severe flooding in Western Europe during July 2021. *World Weather Attribution* (www.worldweatherattribution.org).
- Landwehr, J. M., Matalas, N. C., and Wallis, J. R. (1979). Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles. *Water Resources Research*, 15(5):1055–1064.
- Leadbetter, M. R. (1983). Extremes and local dependence in stationary sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 65(2):291–306.
- Lehmann, E. L. and Romano, J. P. (2021). *Testing statistical hypotheses*. Springer Texts in Statistics. Springer, Cham, fourth edition.
- Lenssen, N., Schmidt, G., Hansen, J., Menne, M., Persin, A., Ruedy, R., and Zyss, D. (2019). Improvements in the GISTEMP uncertainty model. *Journal of Geophysical Research: Atmospheres.*, 124(12):6307–6326.
- Lettenmaier, D. P., Wallis, J. R., and Wood, E. F. (1987). Effect of regional heterogeneity on flood frequency estimation. *Water Resources Research*, 23(2):313–323.
- Lilienthal, J. (2019a). *Group-based regionalization in flood frequency analysis considering heterogeneity*. PhD thesis, Universitätsbibliothek Dortmund.
- Lilienthal, J. (2019b). *TLMoments: Calculate TL-Moments and Convert Them to Distribution Parameters*. R package version 0.7.5.
- Lilienthal, J., Fried, R., and Schumann, A. (2018). Homogeneity testing for skewed and cross-correlated data in regional flood frequency analysis. *Journal of Hydrology*, 556:557–571.
- Lilienthal, J., Kinsvater, P., and Fried, R. (2016). On the method of probability weighted moments in regional frequency analysis. *SFB 823 Discussion Paper*, 63/2016.
- Lilienthal, J., Zanger, L., Bücher, A., and Fried, R. (2022). A note on statistical tests for homogeneities in multivariate extreme value models for block maxima. *Environmetrics*, 33(7):e2746.
- Lo, A. (2017). Functional generalizations of Hoeffding’s covariance lemma and a formula for Kendall’s tau. *Statistics & Probability Letters*, 122:218–226.
- Lu, L. H. and Stedinger, J. R. (1992). Variance of two- and three-parameter GEV/PWM quantile estimators: formulae, confidence intervals, and a comparison. *Journal of Hydrology*, 138(1–2):247–267.

- Masselot, P., Chebana, F., and Ouarda, T. B. (2017). Fast and direct nonparametric procedures in the L-moment homogeneity test. *Stochastic Environmental Research and Risk Assessment*, 31(2):509–522.
- Naveau, P., Nogaj, M., Ammann, C., Yiou, P., Cooley, D., and Jomelli, V. (2005). Statistical methods for the analysis of climate extremes. *Comptes Rendus Geoscience*, 337(10-11):1013–1022.
- Nguyen, C., Gaume, E., and Payrastre, O. (2014). Regional flood frequency analyses involving extraordinary flood events at ungauged sites: further developments and validations. *Journal of Hydrology*, 508:385–396.
- Oorschot, J. and Zhou, C. (2020). All Block Maxima method for estimating the extreme value index. *ArXiv preprints: 2010.15950*.
- Otto, F. E. L., van der Wiel, K., van Oldenborgh, G. J., Philip, S., Kew, S. F., Uhe, P., and Cullen, H. (2018). Climate change increases the probability of heavy rains in Northern England/Southern Scotland like those of storm Desmond—a real-time event attribution revisited. *Environmental Research Letters*, 13(2):024006.
- Padoan, S. A., Ribatet, M., and Sisson, S. A. (2010). Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association*, 105(489):263–277.
- Philip, S., Kew, S., van Oldenborgh, G. J., Otto, F., Vautard, R., van der Wiel, K., King, A., Lott, F., Arrighi, J., Singh, R., et al. (2020). A protocol for probabilistic extreme event attribution analyses. *Advances in Statistical Climatology, Meteorology and Oceanography*, 6(2):177–203.
- Pickands, J. (1975). Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics*, 3(1):119–131.
- Prescott, P. and Walden, A. T. (1980). Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. *Biometrika*, 67(3):723–724.
- R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Renard, B., Garreta, V., and Lang, M. (2006). An application of Bayesian analysis and Markov chain Monte Carlo methods to the estimation of a regional trend in annual maxima. *Water Resources Research*, 42(12):W12422.
- Ribatet, M. (2022). *SpatialExtremes: Modelling Spatial Extremes*. R package version 2.1-0.

- Ribereau, P., Guillou, A., and Naveau, P. (2008). Estimating return levels from maxima of non-stationary random sequences using the Generalized PWM method. *Nonlinear Processes in Geophysics*, 15(6):1033–1039.
- Robert, C. Y., Segers, J., and Ferro, C. A. (2009). A sliding blocks estimator for the extremal index. *Electronic Journal of Statistics*, 3:993–1020.
- Rootzén, H. and Katz, R. W. (2013). Design life level: quantifying risk in a changing climate. *Water Resources Research*, 49(9):5964–5972.
- Schlather, M. (2002). Models for stationary max-stable random fields. *Extremes*, 5(1):33–44.
- Sen, P. K. (1968). Estimates of the regression coefficient based on Kendall’s tau. *Journal of the American Statistical Association*, 63(324):1379–1389.
- Seneviratne, S. I., Nicholls, N., Easterling, D., Goodess, C. M., Kanae, S., Kossin, J., Luo, Y., Marengo, J., McInnes, K., Rahimi, M., et al. (2012). Changes in climate extremes and their impacts on the natural physical environment. In *Managing the risks of extreme events and disasters to advance climate change adaptation: Special Report of the Intergovernmental Panel on Climate Change*, pages 109–230. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Šimková, T. (2017). Homogeneity testing for spatially correlated data in multivariate regional frequency analysis. *Water Resources Research*, 53(8):7012–7028.
- Šimková, T. (2018). L-moment homogeneity test in trivariate regional frequency analysis of extreme precipitation events. *Meteorological Applications*, 25(1):11–22.
- Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1):67–90.
- Smith, R. L. (1990). Max-stable processes and spatial extremes. *Unpublished manuscript*, 205:1–32.
- Stein, M. (2017). Should annual maximum temperatures follow a generalized extreme value distribution? *Biometrika*, 104(1):1–16.
- Stephenson, A. G. (2002). evd: Extreme Value Distributions. *R News*, 2(2):0.
- Stocker, T., Qin, D., Plattner, G.-K., Alexander, L., Allen, S., Bindoff, N., Bréon, F.-M., Church, J., Cubasch, U., Emori, S., Forster, P., Friedlingstein, P., Gillett, N., Gregory, J., Hartmann, D., Jansen, E., Kirtman, B., Knutti, R., Krishna Kumar, K., Lemke, P., Marotzke, J., Masson-Delmotte, V., Meehl, G., Mokhov, I., Piao, S., Ramaswamy, V., Randall, D., Rhein, M., Rojas, M., Sabine, C., Shindell, D., Talley, L., Vaughan, D., and Xie, S.-P. (2013). Technical summary. In Stocker, T., Qin, D., Plattner, G.-K.,

- Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P., editors, *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, pages 33–115. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Stott, P. A., Christidis, N., Otto, F. E. L., Sun, Y., Vanderlinden, J.-P., van Oldenborgh, G. J., Vautard, R., von Storch, H., Walton, P., Yiou, P., and Zwiers, F. W. (2016). Attribution of extreme weather and climate-related events. *WIREs Climate Change*, 7(1):23–41.
- Tradowsky, J. S., Philip, S. Y., Kreienkamp, F., Kew, S. F., Lorenz, P., Arrighi, J., Belleflamme, A., Bettmann, T., Caluwaerts, S., Chan, S. C., Ciavarella, A., De Cruz, L., de Vries, H., Demuth, N., Ferrone, A., Fischer, E. M., Fowler, H. J., Goergen, K., Heinrich, D., Henrichs, Y., Lenderink, G., Kaspar, F., Nilson, E., Otto, F. E. L., Ragone, F., Seneviratne, S. I., Singh, R. K., Skålevåg, A., Termonia, P., Thalheimer, L., van Aalst, M., Van den Bergh, J., van de Vyver, H., Vannitsem, S., van Oldenborgh, G. J., van Schaeybroeck, B., Vautard, R., Vonk, D., and Wanders, N. (2022). Attribution of heavy rainfall events leading to the severe flooding in Western Europe during July 2021. *Climatic Change*. In revision.
- van der Vaart, A. W. (1998). *Asymptotic Statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.
- van der Wiel, K., Kapnick, S. B., van Oldenborgh, G. J., Whan, K., Philip, S., Vecchi, G. A., Singh, R. K., Arrighi, J., and Cullen, H. (2017). Rapid attribution of the August 2016 flood-inducing extreme precipitation in south Louisiana to climate change. *Hydrology and Earth System Sciences*, 21(2):897–921.
- van Oldenborgh, G. J., Philip, S., Kew, S., van Weele, M., Uhe, P., Otto, F., Singh, R., Pai, I., Cullen, H., and AchutaRao, K. (2018). Extreme heat in India and anthropogenic climate change. *Natural Hazards and Earth System Sciences*, 18(1):365–381.
- van Oldenborgh, G. J., van der Wiel, K., Sebastian, A., Singh, R., Arrighi, J., Otto, F., Hausteijn, K., Li, S., Vecchi, G., and Cullen, H. (2017). Attribution of extreme rainfall from Hurricane Harvey, August 2017. *Environmental Research Letters*, 12(12):124009.
- Vautard, R., Yiou, P., van Oldenborgh, G. J., Lenderink, G., Thao, S., Ribes, A., Planton, S., Dubuisson, B., and Soubeyroux, J.-M. (2015). Extreme Fall 2014 Precipitation in the Cévennes Mountains. *Bulletin of the American Meteorological Society*, 96:56–60.
- Vigilione, A. (2012). *homtest: Homogeneity tests for Regional Frequency Analysis*. R package version 1.0-5.

- Viglione, A., Laio, F., and Claps, P. (2007). A comparison of homogeneity tests for regional frequency analysis. *Water Resources Research*, 43(3):W03428.
- Viglione, A., Merz, R., Salinas, J. L., and Blöschl, G. (2013). Flood frequency hydrology: 3. A Bayesian analysis. *Water Resources Research*, 49(2):675–692.
- Villani, C. (2008). *Optimal transport: old and new*, volume 338. Springer, Berlin.
- Westra, S. and Sisson, S. A. (2011). Detection of non-stationarity in precipitation extremes using a max-stable process model. *Journal of Hydrology*, 406(1-2):119–128.
- Zanger, L. (2022). *findpoolreg: Find a possible pooling region for precipitation extremes*. R package version 0.1.0.
- Zou, N., Volgushev, S., and Bücher, A. (2021). Multiple block sizes and overlapping blocks for multivariate time series extremes. *The Annals of Statistics*, 49(1):295–320.

Author Contribution Statement

The following list provides details on the contribution of the authors, in particular of the author of this thesis, to the respective research articles.

- 2.1 Bücher, A. and Zanger, L. (2023). On the Disjoint and Sliding Block Maxima method for piecewise stationary time series. Accepted for publication in *The Annals of Statistics*. Previous version publicly available at arXiv:2110.15576v1.

The idea of investigating the PWM estimator based on sliding block maxima can be attributed to the first author of the article. He also wrote the abstract, introduction and conclusion and made several changes to the structure and wording of the first version of the manuscript, which was written by the second author. The development of the framework for piecewise stationary time series reflecting the structure of the underlying observations in environmental studies was developed jointly by the two authors. The proofs were elaborated by the second author, accompanied by several comments, literature recommendations and support from the first author. All implementations used within the simulation and case studies were made by the second author, who also conducted these studies.

- 2.2 Lilienthal, J., Zanger, L., Bücher, A. and Fried, R. (2022). A note on statistical tests for homogeneities in multivariate extreme value models for block maxima. *Environmetrics*, 33(7):e2746 (DOI: 10.1002/env.2746).

The initial idea of developing a homogeneity test that is based on asymptotic theory for parameter estimators obtained from (trimmed) L-moments was developed by the first and the last author, who then also developed the corresponding theory involving the TLM-based estimators. The respective theoretical parts in Sections 2 and 3 of the research article were written by the first author, and so was the code for parameter estimation based on TLMs. The third author suggested to compare the TLM-based estimators to a (pseudo) ML estimator. The respective theory was elaborated by the second author, who wrote the corresponding theoretical parts in Sections 2 and 3 of the paper. The bootstrap procedure based on max-stable processes was developed and implemented by the second author. She also wrote Sections 4, 5 and 6 and carried out all simulations as well as the case study. The introduction and conclusion were contributed

by the third author, who also suggested several improvements on the original draft of the manuscript.

- 2.3 Zanger, L., Bücher, A., Kreienkamp, F., Lorenz, P. and Tradowsky, J. (2023). Regional Pooling in Extreme Event Attribution Studies: an Approach Based on Multiple Statistical Testing. Submitted to *Extremes* and publicly available at arXiv:2301.06295v1.

The last three authors drew attention to the problem of validating homogeneity assumptions when pooling data in extreme event attribution studies. The first author transferred this problem to the statistical setting, developed the statistical tests and bootstrap procedures. Except for the introduction and the conclusion, which were written by the second author, the initial version of the manuscript was drafted by the first author. The remaining authors contributed with several minor adaptations to an improved version of the manuscript. All implementations as well as the simulation and the case study were conducted by the first author.

Eidesstattliche Versicherung

Ich versichere an Eides Statt, dass die Dissertation von mir selbständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ erstellt worden ist. Die Dissertation wurde in der vorgelegten oder ähnlicher Form noch bei keiner anderen Institution eingereicht. Ich habe bisher keine erfolglosen Promotionsversuche unternommen.

Düsseldorf, den 01. März 2023

Leandra Zanger