



Generalized extreme value shape parameter and its nature for extreme precipitation using long time series and the Bayesian approach

Galina Ragulina ^{a,b} and Trond Reitan ^{c,d}

^aNatural Hazard Department, Snow Avalanche and Rock Slides Division, Norwegian Geotechnical Institute (NGI), Oslo, Norway; ^bLandslides, Flood and River Management Department, Section of Hazard Mapping and Department of Hydrology, Section of Sediment and Erosion, Norwegian Water Resources and Energy Directorate (NVE), Oslo, Norway; ^cDepartment of Hydrology, Section of Hydroinformatics, Norwegian Water Resources and Energy Directorate (NVE), Oslo, Norway; ^dCEES, Department of Biosciences, University of Oslo, Oslo, Norway

ABSTRACT

Assessing the probability of extreme precipitation events is consequential in civil planning. This requires an understanding of how return values change with return periods, which is essentially described by the generalized extreme value (GEV) shape parameter. Some works in the field suggest a constant shape parameter, while our analysis indicates a non-universal value. We re-analysed an older precipitation dataset (169 stations) extended by Norwegian data (71 stations). We showed that while each set seems to have a constant shape parameter, it differs between the two datasets, indicating regional differences. For a more comprehensive analysis of spatial effects, we examined a global dataset (1495 stations). We provided shape parameter maps for two models and found clear evidence that the shape parameter depends on elevation, while the effect of latitude remains uncertain. Our results confirm an explanation in terms of dominating precipitation systems based on a proxy derived from the Köppen-Geiger climate classification.

ARTICLE HISTORY

Received 21 September 2015
Accepted 14 July 2016

EDITOR

D. Koutsoyiannis

ASSOCIATE EDITOR

not assigned

KEYWORDS

Extreme value theory; generalized extreme value (GEV); shape parameter; Bayesian approach; extreme precipitation; spatial modelling

1 Introduction

Extreme precipitation estimates are decisive for planning and design of important infrastructure, such as reservoir dams, avalanche mitigation measures, and power and transport lines. The accuracy of extreme precipitation estimates is therefore crucial for both economic and safety aspects (e.g. Coles and Tawn 1996, Blanchet *et al.* 2009, Eli *et al.* 2012, Dyrddal *et al.* 2014).

As mentioned by Serinaldi and Kilsby (2014), the history of extreme value theory (EVT) in its present formalization and its application to hydrological analyses is rooted in an extensive literature dating back to the 1940s. For a detailed historical survey on that subject, the reader is referred to Papalexiou and Koutsoyiannis (2013). The present paper focuses on the shape parameter, ξ , of the generalized extreme value (GEV) distribution. The GEV distribution encompasses three limiting distributions of extreme value depending on the value of the shape parameter (Coles 2001):

- $\xi > 0$ giving the heavy-tailed Fréchet case (EV2);
- $\xi = 0$ giving the light-tailed Gumbel case (EV1);
- $\xi < 0$ giving the short-tailed negative-Weibull case (EV3).

Koutsoyiannis (2004a, 2004b) has analysed the statistics of daily rainfall extremes and argued for the use of the EV2 distribution (with positive shape parameter) instead of the Gumbel distribution (EV1) when analysing rainfall data to avoid an underestimation of risk associated with extreme rainfall. L-moment estimation of the distribution's shape parameter, ξ , led Koutsoyiannis (2004a, 2004b) to conclude that $\xi \cong 0.15$ and that it is “constant for all examined geographical zones (Europe and North America)”. Recent work on the generalized Pareto (GP) distribution's shape parameter, $\xi_{\text{GEV}} \cong \xi_{\text{GP}}$ (Serinaldi and Kilsby 2014), supports $\xi > 0$ (Cavanaugh and Gershunov 2015, Cavanaugh *et al.* 2015). The conclusion made by Koutsoyiannis (2004a, 2004b) was drawn from a comparison of the spread of ξ estimates in the data with that of simulations where ξ is constant. However, a strict statistical test with the assumption of a constant ξ was not carried out. More recently, Papalexiou and Koutsoyiannis (2013) presented an extension of the work of Koutsoyiannis (2004b) to analyse daily rainfall from 15 137 precipitation stations worldwide with observation lengths of 40–63 years. They concluded, among other things, that:

- (a) “The record length strongly affects the estimate [L-moment method only was used] of the GEV shape parameter and long records are needed for reliable estimates;”
- (b) “The GEV shape parameter is expected to belong in a narrow range, approximately from 0 to 0.23 with confidence 99%,” and
- (c) “The geographical location of the globe may affect the value of the shape parameter.”

In addition to the work mentioned above, which primarily focuses on the proper choice of the GEV distribution, there is a significant body of work in which the shape parameter was evaluated as a by-product (e.g. Fowler and Kilsby 2003, Meddi and Toumi 2015). However, there are almost no studies dedicated to the nature of the GEV shape parameter itself. The value of this shape parameter directly influences estimated values of extreme precipitation, which in turn are crucial for dimensioning expensive engineering constructions.

The main aim of the present paper was to investigate the properties of the GEV shape parameter by using a statistical modelling approach to facilitate comparison of different hypotheses in a consistent fashion. We use Bayesian methods because some of our models are hierarchical and some prior knowledge exists. Bayesian analysis has been used previously for, amongst others, extreme precipitation modelling (e.g. Coles and Tawn 1996, Coles 2001, Coles *et al.* 2003, Smith 2005, Alston 2011, Eli *et al.* 2012, Sun *et al.* 2015).

Our analysis is conducted in three phases. In the first phase, to extend the analysis from Koutsoyiannis (2004b) consistently, we used the same time series. For a more comprehensive perspective, we extended these time series by additional data from 71 stations in Norway, each consisting of 99–131 years of observations. The objective of this phase of the study was to understand whether a statistical modelling and model comparison approach could broaden our knowledge of the shape parameter beyond the conclusions drawn by Koutsoyiannis (2004b). Three hypotheses about the shape parameter were tested in this part of the study, namely: a shape parameter that is (a) constant, (b) station specific, or (c) stochastic, but drawn from a common distribution. For this purpose, the same analysis was performed on a further extended (worldwide) dataset consisting of daily precipitation data from the Global Historical Climatology Network-Daily database (version 2.60, www.ncdc.noaa.gov/oa/climate/ghcn-daily) referred to as the Extended international dataset.

In the second phase of the study, we examined regional differences in the extended Koutsoyiannis dataset (with Norwegian data).

The third phase was a spatial analysis of the shape parameter using the Extended international dataset. Only those stations with a minimum of 99 complete years of measurements were chosen in order to meet the requirement (a) by Papalexiou and Koutsoyiannis (2013). Spatial patterns of the shape parameter’s distribution are discussed here.

The paper is organized as follows. Section 2 presents the data used in this study; Section 3 describes the statistical models for precipitation annual maxima; and Section 4 explains the tested hypotheses and the corresponding model structures. In Section 5 we discuss the choice of prior distributions to be used in Bayesian inference, and how Bayesian hypothesis testing is performed. Section 6 presents the results of the analyses, and in Section 7 we summarize our conclusions.

2 Data

In the first and the second phases of the study, we used the collection of 169 of the longest available rainfall records worldwide from Koutsoyiannis (2004b), each having 100–154 years of data, with annual maximum values (excluding the years with missing data). According to Koutsoyiannis (2004b), these time series were chosen after examination of some thousands of raingauge time series from Europe and the USA, namely data from the United States Historical Climatology Network (USHCN), Land Surface Observation Data of the UK Met Office, and data from the oldest stations of France, Italy and Greece (Fig. 1). Years with more than five missing daily values in two or more months were excluded.

Seventy-one additional time series, each having 99–131 years of data (excluding the years with missing data), were chosen among 3531 time series from precipitation stations in Norway (data from the web-service of the Norwegian Meteorological Institute (MET Norway), eKlima.no) (Fig. 1). Years with 36 or more missing daily values per year ($\geq 10\%$ of full year with data) were excluded.

These datasets were subdivided into groups and analysed in the following combinations:

- (1) All 169 time series from Koutsoyiannis (2004b) together (international data);
- (2) Norwegian data (71 time series);
- (3) USA vs Europe (141 vs 28);
- (4) Norwegian data vs International data (71 vs 169);
- (5) UK vs Norwegian data (24 vs 71);
- (6) South USA vs North USA (70 vs 71; 40°-parallel was chosen to be a separation line);
- (7) West USA vs East USA (17 vs 124; highly ridged vs relatively flat area); and

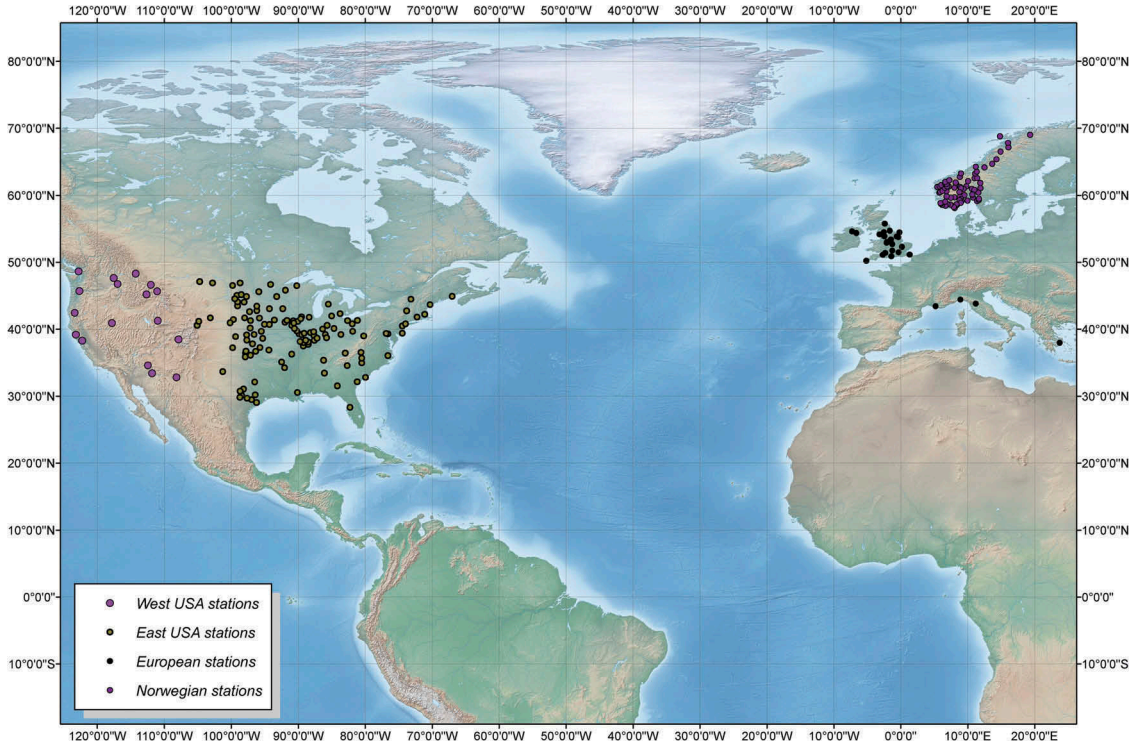


Figure 1. Map of positions of the stations used in the first phase of the analysis.

(8) Norwegian data vs West USA (71 vs 17).

The extended international dataset, used in both the first and third phases of the study, comprises 1495 daily precipitation time series (see Fig. 2 for geographical locations). These time series were selected from more than 15 000 precipitation records available in the Global Historical Climatology Network-Daily database (version 2.60, www.ncdc.noaa.gov/oa/climate/ghcn-daily) according to our requirements of 99-year minimum record length with less than 5% missing values. Raingauge geographical coordinates (longitude, latitude and elevation) were included in the datasets to perform spatial analysis.

3 Statistical models

For the analysis of precipitation time series, it is assumed that the distribution of yearly maximum precipitation for a given station follows the GEV distribution:

$$f(z|\mu, \sigma, \xi) = \frac{1}{\sigma} \left[1 + \xi \frac{(z - \mu)}{\sigma} \right]^{-1/\xi - 1} \exp \left\{ - \left[1 + \xi \frac{(z - \mu)}{\sigma} \right]^{-1/\xi} \right\} \quad (1)$$

where z is a yearly maximum precipitation, μ is a location parameter, σ is a scale parameter and ξ is the shape parameter. Here, the shape parameter is assumed not to be zero. If it is zero, the Gumbel distribution is used. The location and scale parameters can be assumed to differ from station to station according to how dry or wet the place is and how much variability there is in the yearly extremes. The analysis of Koutsoyiannis (2004a, 2004b) suggested that ξ was the same for all stations for precipitation data. The objective of the first phase of our study is to perform direct statistical tests of this hypothesis (ξ is universal for precipitation data) as well as to study whether the hypothesis holds upon inclusion of additional Norwegian data and possible reasons for regional differences of ξ values.

Let $i \in \{1, \dots, n\}$ denote the station number and let $\{z_{i,1}, \dots, z_{i,n_i}\}$ denote the n_i annual precipitation maxima for each station. Under the assumption of independent data, the combined likelihood is:

$$L(\{\mu_i, \sigma_i, \xi_i\}_{i \in \{1, \dots, m\}}) = \prod_{i=1}^m \prod_{j=1}^{n_i} \frac{1}{\sigma_i} \left[1 + \xi_i \frac{(z_{i,j} - \mu_i)}{\sigma_i} \right]^{-1/\xi_i - 1} \exp \left\{ - \left[1 + \xi_i \frac{(z_{i,j} - \mu_i)}{\sigma_i} \right]^{-1/\xi_i} \right\} \quad (2)$$

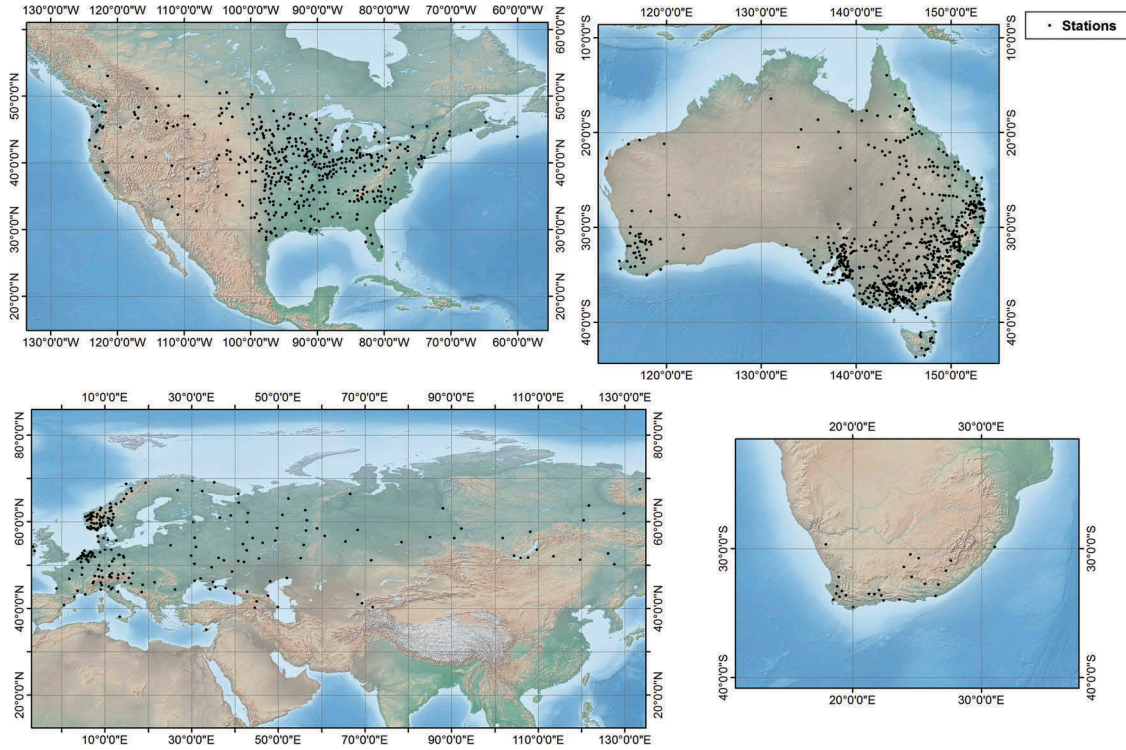


Figure 2. Map of positions of the stations from the Extended international dataset.

While this equation expresses the likelihood for all GEV parameters separately for each station, it can also accommodate models in which the shape parameter is the same for all stations, by setting

$$\xi_i = \xi \quad (3)$$

for all i .

4 Hypotheses

Three hypotheses were tested during the analysis of the Koutsoyiannis and Norwegian data (Phase I of the study):

- a.** The shape parameter differs from station to station (individual). If so, the expression for the likelihood in Equation (2) is used as it is.
- b.** The shape parameter has a parametric distribution determined by the data,

$$\xi_i \sim f(\xi | \underline{\lambda}) \quad (4)$$

where $\underline{\lambda}$ is the parameter set (as a vector) of the distribution of the shape parameter, ξ . That means that the shape parameter is regarded as a random effect. Since this requires a probabilistic treatment of the station-wise shape parameter, ξ , it becomes simpler to treat the entire parameter set in a Bayesian fashion than to split the analysis

into a Bayesian part for $\underline{\lambda}$ and a classic part for the other parameters (including those that determine the distribution of $\underline{\lambda}$).

- c.** The shape parameter is universal (common constant value). In this case $\xi_i = \xi$ for all stations, i (Equation (3)).

After the inference about the nature of ξ was made, a new block of hypotheses was created.

In the case that the results are in favour of hypothesis **b** or **c** for a given dataset, it is also possible to go further, since both these cases contain parameters that are “global” to that dataset. For hypothesis **b**, the parameters describing the distribution of the station-wise ξ are global. For hypothesis **c**, it is the shape parameter, ξ , itself that is global.

When data from two different regions are involved, it is of interest to test if the global parameters are the same or different in the two sets. This testing constitutes Phase II of the analysis. Thus, instead of a constant value of ξ for all stations (hypothesis **c**), there could be one shape parameter, ξ_1 , if a station belongs to dataset D_1 , and another, ξ_2 , if that station belongs to dataset D_2 . Hence, it is possible to test if the shape parameter is truly global or rather regional (belonging to a particular selection of stations). The regional model (hypothesis) thus takes the form:

- d. $\xi_i = \xi_{\text{region1}}$ if station i belongs to the first region
and $\xi_i = \xi_{\text{region2}}$ if it belongs to the second region.

The results from the regional analysis in Phase II encouraged us to try spatial analysis for the Extended international dataset. Considering the ξ as realizations of a random field, their values can be described by a multivariate distribution accounting for mutual (spatial) dependence. This can be seen as an extension of hypothesis **b**, as each ξ_i is again assigned a distribution, but this distribution also describes the correlation between the ξ at different places. We assigned a multivariate normal distribution, so that:

$$\underline{\xi} \sim N(\underline{\mu}_\xi, \Sigma_\xi) \quad (5)$$

where $\underline{\xi} = (\xi_1, \xi_2, \dots, \xi_m)$ is a vector of all ξ , $\underline{\mu}_\xi$ is a vector of expected values for the ξ (assumed the same for all stations in the simplest model) and Σ_ξ is a covariance matrix. Assuming the same variance for each station, each element can be described by

$$\Sigma_{\xi,i,j} = \sigma_\xi^2 \rho_{\xi,i,j} \quad (6)$$

where $\rho_{\xi,i,j}$ is the correlation between ξ_i and ξ_j and σ_ξ^2 is the variance of the ξ . Many sophisticated models for the correlation can be made, but without any indication of which would be the right one, we opted for what is arguably the simplest one:

$$\rho_{\xi,i,j} = e^{-r_{i,j}/R} \quad (7)$$

where $r_{i,j}$ is the distance between stations i and j , and R is a parameter that encodes the characteristic correlation length, i.e. R is the distance for which the correlation drops by a factor of e and, by analogy to the mean-reverting process, can be called the ‘‘characteristic distance’’ of the spatial field. Distances between stations were calculated using a spherical approximation, so that:

$$r_{i,j} = E \arccos\left(\sin(\varphi_i) \sin(\varphi_j) + \cos(\varphi_i) \cos(\varphi_j) \cos(\vartheta_i - \vartheta_j)\right) \quad (8)$$

where φ_i, φ_j is the latitude in radians, ϑ_i, ϑ_j is the longitude in radians of stations i and j , respectively, and E is the radius of the Earth.

Calculating the probability density of a huge array of values can be time consuming, making likelihood calculations computationally intense due to the need of finding the determinant and inverse of the covariance matrix or the corresponding correlation matrix. It was thus determined to discretize the parameter R (in 40

values ranging exponentially from 1 km to the circumference of the Earth) and calculate the determinant and inverse of the correlation matrix for each discretized value before doing the inference. This significantly speeded up the analysis.

Additional structure to a spatial model could be assigned by correcting for the height and latitude of each station in the expected value using linear models. We created four spatial model variants this way, which were the models examined in Phase III of our analysis:

- e. Global expected value:

$$\underline{\xi} \sim N(\mu_{\xi_0} \underline{1}, \Sigma_\xi) \quad (9)$$

where μ_{ξ_0} is the global expected value and $\underline{1}$ is a vector of ones.

- f. Expected value depends linearly on height:

$$\underline{\xi} \sim N(\mu_{\xi_0} \underline{1} + \beta_h \underline{h}, \Sigma_\xi) \quad (10)$$

where \underline{h} is a vector consisting of the heights of each station and β_h is a regression parameter that describes how the expected shape parameter changes with height. Hence, if $\beta_h = 0.0001$, for instance, the expected shape parameter changes from μ_0 at sea level to $\mu_0 + 0.1$ at a height of 1000 m.

- g. Expected value depends linearly on height and latitude:

$$\underline{\xi} \sim N\{\mu_{\xi_0} \underline{1} + \beta_h \underline{h} + \beta_{\phi_1} \underline{\phi}, \Sigma_\xi\} \quad (11)$$

where $\underline{\phi}$ is a vector of latitudes and β_{ϕ_1} is a regression parameters describing how much the expected value changes with changing latitude.

- h. Expected value depends linearly on height and linearly and quadratic on latitude:

$$\underline{\xi} \sim N\left(\mu_{\xi_0} \underline{1} + \beta_h \underline{h} + \beta_{\phi_1} \underline{\phi} + \beta_{\phi_2} \underline{\phi}^2, \Sigma_\xi\right) \quad (12)$$

where $\underline{\phi}^2$ is a vector of squared latitudes and β_{ϕ_2} is a regression parameter describing how much the expected value changes with squared latitude.

5 Priors and model selection

5.1 Priors

As this study contains Bayesian hierarchical models, Bayesian inference is necessary when it comes to both parameter estimation and model choice. The middle part of the hierarchy could conceivably be handled as random effects in a frequentist setting, but, since there is no ready-made methodology for our models, a Bayesian treatment offered fewer inferential and

numerical challenges. In addition, we are interested in the uncertainty of the shape parameter, which can be read directly from the posterior distribution of a Bayesian analysis, but which would require extra steps in a classic analysis.

In both cases, a prior distribution for the parameters is required. Prior distributions are used to represent a set of beliefs about the parameter of interest (Eli *et al.* 2012). In Bayesian statistics, this prior knowledge about the parameters is updated by the data using Bayes' theorem:

$$f(\underline{\theta}|\underline{D}) = f(\underline{D}|\underline{\theta})f(\underline{\theta})/f(\underline{D}) \quad (13)$$

where \underline{D} is the data, $\underline{\theta}$ is the parameter set, $f(\underline{\theta})$ is the prior distribution, $f(\underline{D}|\underline{\theta})$ is the likelihood and

$$f(\underline{D}) = \int f(\underline{D}|\underline{\theta})f(\underline{\theta})d\underline{\theta} \quad (14)$$

is a normalizing constant known as the marginal likelihood. One does not need to know the marginal likelihood in order to sample from the posterior distribution, but it is crucial for calculating model probabilities in Bayesian model selection (see Appendix for details).

The return period for a given outcome, x , in extreme value analysis is defined as:

$$T = 1/P(X > x) \quad (15)$$

i.e. the expected number of years before that value will be exceeded. The outcome x is then called the T -year return value. According to Coles and Tawn (1996), using the prior knowledge of an expert hydrologist, a Bayesian 95% interval estimate of the 100-year return level for daily rainfall was found to be approximately half of the width of the corresponding likelihood-based confidence interval.

For this study, we used a relatively wide (in terms of having a larger variance and credibility bands of predictions than would be expected from the data) and unstructured (in terms of assuming parametric independence rather than examining which parameter combinations yield reasonable predictions) prior, as well as a narrower and structured prior. The wide prior (assuming that there is little information available about the process apart from the data) was taken from Smith (2005), where wide marginal distributions for the GEV parameters were set directly and independently. Smith (2005) used the following prior distribution for each parameter:

$$\mu \sim N(0, 1000) \quad (16)$$

$$\sigma \sim \log N(0, 100) \quad (17)$$

$$\xi \sim N(0, 10) \quad (18)$$

The joint prior distribution is an independent combination of these normal single parameter distributions with large variances. The variances are chosen large enough to make the distributions almost flat, corresponding to prior ignorance. This represents a problem in two respects. Firstly, the prior can correspond to a distribution on the quantiles that is entirely unrealistic, in which case it can be described as being too wide. It is not easy to have intuition as to whether a particular combination of GEV parameter values is realistic or not. However, priors yielding a physically unrealistic 10-year return level (i.e. the 0.9 quantile of the GEV distribution) should be discarded. If this is the case, the prior cannot be said to represent our knowledge on the topic and is thus unrealistic. As it turns out, Smith's prior gives a 27% chance of a negative 10-year return value and similarly a 26% chance of it being larger than 10^{30} .

Secondly, while a too wide prior is seldom seen as a problem for parameter estimation when enough data are available, it can, nonetheless, pose serious problems for model selection. An observation named Bartlett's paradox (Bartlett 1957) states that the Bayes factor for a model compared to a simpler model (a zero hypothesis) can go to zero when the width of the prior distribution goes to infinity. In other words, a too wide prior can give an unfair advantage to a simpler model, and it might be difficult for the data to overcome this bias in the model selection. However, care must be taken when specifying a narrow prior, so that it does not penalize any parameter estimates deemed reasonable.

In addition, a more specific prior considerably reduces the uncertainty connected to value estimates. An example of the effect of a narrow band for the shape parameter, ξ (posterior band with Coles and Tawn (1996) prior described below), compared to one with the wide ξ prior (Smith's prior, Equation (18)) is given in Figure 3. Data from the Norwegian precipitation station no. 1650, Strømsfoss sluse (130 years with data), were used.

There is another prior for GEV models on precipitation data, elicited by Coles and Tawn (1996). This prior assigns a gamma distribution to the 10-year return value, T_{10} , the 100-year return value minus the 10-year return value ($T_{100} - T_{10}$), and the 1000-year return value minus the 100-year return value ($T_{1000} - T_{100}$). Thus, one gets a reasonable range of return values, given that reasonable (well-founded) hyper-parameters for the gamma distributions are specified. As the gamma distribution only has two parameters, this can be done by specifying (for instance) the 95% prior credibility band. For this study, we set a 95% credibility band for T_{10} , $T_{100} - T_{10}$ and $T_{1000} - T_{100}$ to the interval 3–600 mm. Such a prior

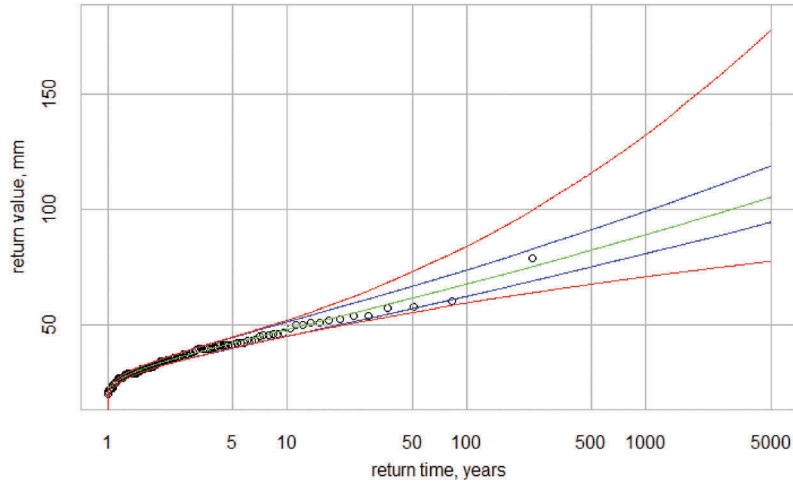


Figure 3. Effect of a narrow posterior band with the Coles & Tawn prior (Equations (21)–(23)) for the shape parameter, ξ , compared to one with the wide ξ prior (Smith's prior, Equations (16)–(18)). Green line: GEV distribution fitted to the data (Coles & Tawn prior); blue lines: 95% posterior credibility interval (Coles & Tawn prior); red lines: 95% posterior credibility interval (Smith's prior).

then implies a distribution for the parameter set also, which can be calculated using the transformation rule for probability distributions. Even though the distributions for T_{10} , $T_{100} - T_{10}$ and $T_{1000} - T_{100}$ are set independently, there is a dependency between the GEV parameters in the prior distribution (hereafter, this elicitation approach is referred to as the Coles & Tawn prior). If we define the vector:

$$\underline{\theta} \equiv (\mu, \sigma, \xi) \quad (19)$$

and the vector:

$$\underline{\varphi} \equiv (T_{10}, T_{100} - T_{10}, T_{1000} - T_{100}) \quad (20)$$

then the Coles & Tawn prior can be expressed as:

$$T_{10} \sim \gamma(\alpha = 0.91, \beta = 172) \quad (21)$$

$$T_{100} - T_{10} \sim \gamma(\alpha = 0.91, \beta = 172) \quad (22)$$

$$T_{1000} - T_{100} \sim \gamma(\alpha = 0.91, \beta = 172) \quad (23)$$

$$f(\underline{\theta}) = \left| \frac{\partial \underline{\varphi}}{\partial \underline{\theta}} \right| f_{\varphi}(\underline{\varphi}(\underline{\theta})) \quad (24)$$

where f_{φ} is the distribution of $\underline{\varphi}$ described in Equations (21)–(23). Thus, the prior distribution is readily available for the station-wise analysis. However, for models with common shape parameter or shape parameter from a common distribution, care must be taken. In the case of an underlying common distribution (hypothesis **b**) for the shape parameter, it is this distribution that determines the prior for the shape parameter, not the individual return periods. Similarly, for a constant ξ (hypothesis **c**), this cannot be specified by the return values of a single station, so that a separate

prior distribution is used instead. In this case, we used a normal prior distribution:

$$\xi \sim N(0, 0.5) \quad (25)$$

which has a 95% credibility band roughly from -1 to $+1$. Given the shape parameter, location and scale parameter, a distribution can be determined by the distribution of T_{10} and $T_{100} - T_{10}$, which again were assigned a gamma distribution with 95% prior credibility interval 3–600 mm. Thus, for hypothesis **c**:

$$f(\xi, \{\mu_i, \sigma_i\}_{i=1, \dots, m}) = f(\xi) \prod_{i=1}^m f(\{\mu_i, \sigma_i\} | \xi) \quad (26)$$

If ξ_i is considered to belong to one of two regions (hypothesis **d**), ξ_{region1} or ξ_{region2} , then similarly:

$$\begin{aligned} f(\xi_{\text{region1}}, \xi_{\text{region2}}, \{\mu_i, \sigma_i\}_{i=1, \dots, m}) \\ = f(\xi_{\text{region1}}) f(\xi_{\text{region2}}) \prod_{i=1}^m f(\{\mu_i, \sigma_i\} | \xi_{\text{region of } i}) \end{aligned} \quad (27)$$

For the distributional model for ξ , i.e. hypothesis **b**, we get

$$\begin{aligned} f(\mu_{\xi}, \sigma_{\xi}, \{\mu_i, \sigma_i, \xi_i\}_{i=1, \dots, m}) \\ = f(\mu_{\xi}) f(\sigma_{\xi}) \prod_{i=1}^m f(\xi_i | \mu_{\xi}, \sigma_{\xi}) f(\{\mu_i, \sigma_i\} | \xi_i) \end{aligned} \quad (28)$$

where μ_{ξ} and σ_{ξ} are the two parameters of the distribution for ξ . For a spatial model, where the distribution

of the shape parameters, $f\left(\frac{\xi}{\underline{\xi}}\right)$, is described in Equations (5)–(12) and (30)–(32), we similarly get:

$$f\left(\{\mu_i, \sigma_i, \xi_i\}_{i=1, \dots, m}\right) = f\left(\frac{\xi}{\underline{\xi}}\right) \prod_{i=1}^m f(\{\mu_i, \sigma_i\}|\xi_i) \quad (29)$$

Only normal and lognormal models for ξ were tried; both have just one location parameter and one scale parameter. The results based on the lognormal model are not shown, since the model gave a worse fit compared to the normal model.

For the spatial models, we used a uniform prior on the log scale for the discretized characteristic correlation length, R . Thus, this prior was simply constructed by specifying that all log ranges are equally probable and that the range is more than 1 km (most stations were placed more than that distance apart from each other) and not so large that two stations on the opposite side of Earth are strongly correlated. For the regression coefficient, we used a normal distribution so that:

$$\beta_h \in (-0.0001, 0.0001) \quad (30)$$

$$\beta_{\phi_1} \in (-0.0022, 0.0022) \quad (31)$$

$$\beta_{\phi_2} \in (-0.000025, 0.000025) \quad (32)$$

with 95% prior probability. This means that we do not expect ξ to change by more than 0.1 for every 1000 m, nor to change by more than 0.2 from the Equator to a pole.

5.2 Hypothesis testing

Bayesian hypothesis testing is based on the Bayesian model likelihood (BML, see Appendix). From this, one can calculate model probabilities (where one can compare multiple hypotheses), or the Bayes factor (where one compares two hypotheses). Using the latter approach, Bayesian hypothesis testing can be represented by an analysis of the Bayes factors (Jeffreys 1961), which compares the data prediction strength of one hypothesis with the data prediction strength of another. The Bayes factor interpretation scale from

Table 1. Jeffreys' interpretation scale for the Bayes factors.

Bayes factor	Strength of evidence
<1	Negative (supports the other model)
1–3	Barely worth mentioning
3–10	Substantial
10–30	Strong
30–100	Very strong
>100	Decisive

Jeffreys (1961) is given in Table 1, which provides an evaluation of how many times data are more probable under Model 1 compared to Model 2. However, the amount of time series (stations) analysed is of high importance. For example, if for every station the data are just 5% more probable under Model 1 than under Model 2, with data from 169 stations this gives a Bayes factor of 10^3 – 10^4 , which represents very significant evidence.

In other words, even a small difference per station will provide strong evidence if the number of stations is high enough; and the opposite will occur if the number of stations is low, so that a poor value for the Bayes factor can be expected with almost no regard to the strength of an effect. For instance, a low Bayes factor of 3 for 17 stations (as is shown in Table 2) means that the data under Model 1 are 6% more probable than under Model 2. This difference is 1% bigger than in the previous example. However, because only a few stations were analysed, the resulting Bayes factor (equal to 3) cannot be considered as strong evidence (see Jeffreys 1961 and Table 1). That is why it is essential to be careful when making an inference based on data from just a few stations.

Different priors may also lead to different parameter estimates. As shown in Table 2, the estimates of the shape parameter, ξ (medians), differ depending on the choice of prior. However, since Smith's prior has a rather unrealistic nature, whereas the Coles & Tawn prior incorporates reasonable assumptions (T_{10} , T_{100} – T_{10} and T_{1000} – T_{100}), the estimates based on the Coles & Tawn prior are supposed to be more trustworthy than those based on Smith's prior. Nevertheless, the posterior 95% credibility intervals from both priors overlap a lot.

While the shape parameter analysis itself was performed using Bayesian methodology, we performed

Table 2. International, Norwegian and Extended international datasets: Bayes factors for distributional ξ vs individual ξ and constant ξ vs distributional ξ models, and estimates of the shape parameter's median and posterior 95% credibility interval. Smith's prior is described in Equations (16)–(18), while the Coles & Tawn (C&T) prior is described in Equations (21)–(23).

Datasets	Prior	Distributional ξ vs individual ξ	Constant ξ vs distributional ξ	Median	Posterior 95% credibility interval
International (169 stations)	Smith	10^{319}	10^{167}	0.108	0.098–0.119
	C&T	10^{121}	4000	0.117	0.106–0.127
Norwegian (71 stations)	Smith	10^{129}	10^{70}	0.034	0.022–0.049
	C&T	10^{60}	20	0.044	0.028–0.062
Extended international (1495 stations)	C&T	10^{1114}	5×10^5	0.119	0.115–0.123

some simple classic tests on the estimated shape parameters from the spatial analysis, in order to see if climate classes or climate groups could explain some of the differences. This was done using p -values from ANOVA testing. Also, the BIC (Bayesian information criterion), where a classic estimate of the likelihood is penalized by the model complexity (using a Bayesian justification for the penalty term), was utilized for selecting among climate classes and climate groups.

6 Results and discussion

6.1 Phase I: global, distributional or constant shape parameters for the International, Norwegian and Extended international datasets

As described in Section 4, in the first phase of this study we tested hypotheses *a*, *b* and *c*.

The results of the Bayesian comparison between the models for the International, Norwegian and Extended international datasets, as well as the estimates of the shape parameter's median and posterior 95% credibility interval, are given in Table 2. In all cases, the model selection suggests a constant shape parameter (hypothesis *c*). According to Table 2, the evidence for the shape parameter, ξ , being global (common constant value; hypothesis *c*), is decisive for both datasets with both priors (Bayes factor: 4000–10³¹⁹), except for the Norwegian dataset with the Coles & Tawn prior, where the evidence is still strong (Bayes factor = 20).

6.2 Phase II: examining regional differences

The significant difference between the ξ estimates for the International and Norwegian datasets (0.117 and 0.044, respectively) in combination with the decisive evidence for the global constant value of this parameter (Table 2), suggested that further investigation was required to find out whether the shape parameter is globally constant or gradually changing, giving rise to regional differences. To this end, a new set of hypotheses was created to make regional inference on ξ (hypothesis *d*, described in Section 4).

The results of the Bayesian comparison between the models for the International and Norwegian datasets and their subsets, such as USA, Europe, UK, South USA, North USA, East USA and West USA (described in Section 2), are presented in Table 3. A dataset in each row of Table 3 represents the data from the combined set of stations for the two regions. When evidence is in favour of one constant shape parameter, rather than two regional ones, the Bayes factor for model *c* vs model *d* is given, rather than the Bayes factor for model *d* vs model *c*:

Table 3. Bayes factors for a regional model, *d*, vs constant shape parameter model, *c*, for various selections of two regions. Smith's prior is described in Equations (16)–(18), while the Coles & Tawn (C&T) prior is described in Equations (21)–(23).

Dataset	Prior	Common constant ξ vs separate constant ξ	Separate constant ξ vs common constant ξ
USA vs Europe	Smith	4000	–
	C&T	4	–
International vs Norway	C&T	–	2×10^{10}
Norway vs UK	C&T	–	800
South USA vs North USA	C&T	1700	–
West USA vs East USA	C&T	–	3
West USA vs Norway	C&T	20	–

$$B_{c,d} = 1/B_{d,c} \quad (33)$$

We found evidence of regional differences between the International dataset and Norway, between the UK and Norway, and between West and East USA. We did not find evidence for regional differences between the USA and the UK, between South and North USA, or between West USA and Norway. Estimated medians and posterior 95% credibility intervals of the shape parameter for the same datasets/subsets are shown in Table 4.

According to the results summarized in Table 4, ξ assumes values between 0.028 and 0.156 (95% credibility interval) and varies depending on the region. A positive ξ supports Koutsoyiannis' conclusion that the EV2 distribution ($\xi > 0$) should be used for inference on extreme precipitation instead of the Gumbel distribution ($\xi = 0$), since the latter underestimates the values (Koutsoyiannis 2004a, 2004b).

We found decisive evidence that ξ is different for Norway than for the UK (Table 3). This indicates that neither the location on the western coast of the European continent nor the proximity to an ocean has a significant influence on the value of ξ . Moreover, we found even stronger evidence for the South and North USA having a common shape parameter. This rejects the hypothesis that ξ depends strongly on latitude.

Nevertheless, ξ is undoubtedly different for the Norwegian stations compared to the International (Bayes factor = 2×10^{10} ; Table 3); in addition there is

Table 4. Estimates of the shape parameter, ξ , median and posterior 95% credibility interval using the Coles & Tawn prior (Equations (21)–(23)).

Datasets	Median	Posterior 95% credibility interval
Norway (71 stations)	0.044	0.028–0.062
USA (141 stations)	0.115	0.103–0.128
Europe (28 stations)	0.124	0.101–0.151
UK (24 stations)	0.124	0.097–0.156
West USA (17 stations)	0.069	0.038–0.100
East USA (124 stations)	0.120	0.106–0.135

evidence of a difference between the western and eastern parts of USA.

Blanchet *et al.* (2009) differentiated between extreme rainfall and extreme snowfall distributions, referring to Katz *et al.* (2002). In an attempt to explain our results, we investigated the hypothesis of the shape parameter being dependent on the form of precipitation (rain/snow) by analysing the Norwegian dataset from the following subgroups for selecting annual maxima:

- (1) Norway (summer, 71 stations): only observations from 1 May to 1 October were used;
- (2) Norway (winter, 71 stations): only observations from 1 October to 1 May of the following year were used;
- (3) Norway (summer rain, 61 stations): only observations specified as “rain” in periods from 1 May to 1 October were used;
- (4) Norway (the rest, 71 stations): all observations from the Norwegian dataset excluding Norway (summer rain) were used;
- (5) Norway (rain, 61 stations): only observations specified as “rain” were used; and
- (6) Norway (snow, 71 stations): only observations specified as “snow” and/or “sleet” were used.

Our results, presented in Table 5, showed no dependence of the shape parameter on the precipitation form in Norway, and the hypothesis was eventually rejected.

Strong evidence (Bayes factor = 20) was found for a common constant shape parameter for West USA and Norway. Both data regions belong to mountain areas; however, the stations within each dataset are situated at different elevations. Possibly it is the ruggedness of an area that has an influence on precipitation pattern and, hence, on the shape parameter, rather than the height above sea level alone. However, a spatial analysis was needed to investigate this question.

Table 5. Bayes factors (Norwegian data subsets; Coles & Tawn prior only, see Equations (21)–(23)): common constant ξ vs separate constant ξ .

Datasets	Common constant ξ vs separate constant ξ	Separate constant ξ vs common constant ξ
International vs Norway (rain)	–	4×10^7
International vs Norway (summer)	–	8×10^7
Norway (summer rain) vs UK	–	10
Norway (summer rain) vs Norway (the rest)	2.3	–
Norway (summer) vs Norway (winter)	40	–
Norway (rain) vs Norway (snow, sleet)	20	–

6.3 Phase III: spatial analysis of the Extended international dataset

The second part of the data analysis (Section 6.2) has given strong evidence for there being regional differences in the shape parameter. Since the Extended international dataset contained stations distributed globally, we considered spatial models rather than regional or global models to be more appropriate for this analysis. In this manner, we were able to extract the spatial patterns suggested by the data, as well as find some characteristics of these patterns in the form of characteristic correlation length R (described in Section 4) and regression coefficients.

From the estimated BML (see Appendix) values, we found evidence that height and possibly latitude were variables that worked well as linear predictors for the shape parameter, while squared latitude did not improve the results. In other words, hypotheses f and g were preferred over hypotheses e and h (Section 4). The Bayes factor, which discerns between presence or absence of a height dependency, was $B_{fvse} \approx 10^{25}$, while the Bayes factor for a linear latitude dependency vs a second-order polynomial dependency was $B_{gvsh} \approx 10^{50}$. However, due to computational limitations, we could not resolve which of hypotheses f and g had evidence on its side (as the BMLs were approximately equal for the two hypotheses).

Despite the numerical uncertainty connected to the estimation, the analysis suggested that $R = 250$ km, with a 95% credibility interval (150 km, 500 km). Thus, regions corresponding to small countries or administrative regions within larger countries can be assigned roughly the same shape parameter, but larger areas such as continents can be expected to be heterogeneous. This can also be seen in the maps of the estimated ξ values (medians; Figs. 4 and 5). Corresponding maps with standard deviations of the estimates are shown in Figures 6 and 7. Both maps present estimated ξ values for 0 m elevation (sea level). The model g (hypothesis g) considers latitude, as shown in Equation (11).

Some similarities can be seen in the spatial distribution patterns of ξ estimates simulated by the two models f and g . The eastern coast of the USA has lower values than its western coast; Australia has the opposite situation combined with stronger gradients. The whole of Scandinavia and the northwestern parts of Russia show low values. The southern part of Norway has the lowest estimated values of ξ . Eastern Siberia, the coastal area of the Caspian Sea and the northern coast of the Black Sea have relatively high values. The highest values were estimated for Australia by both models.

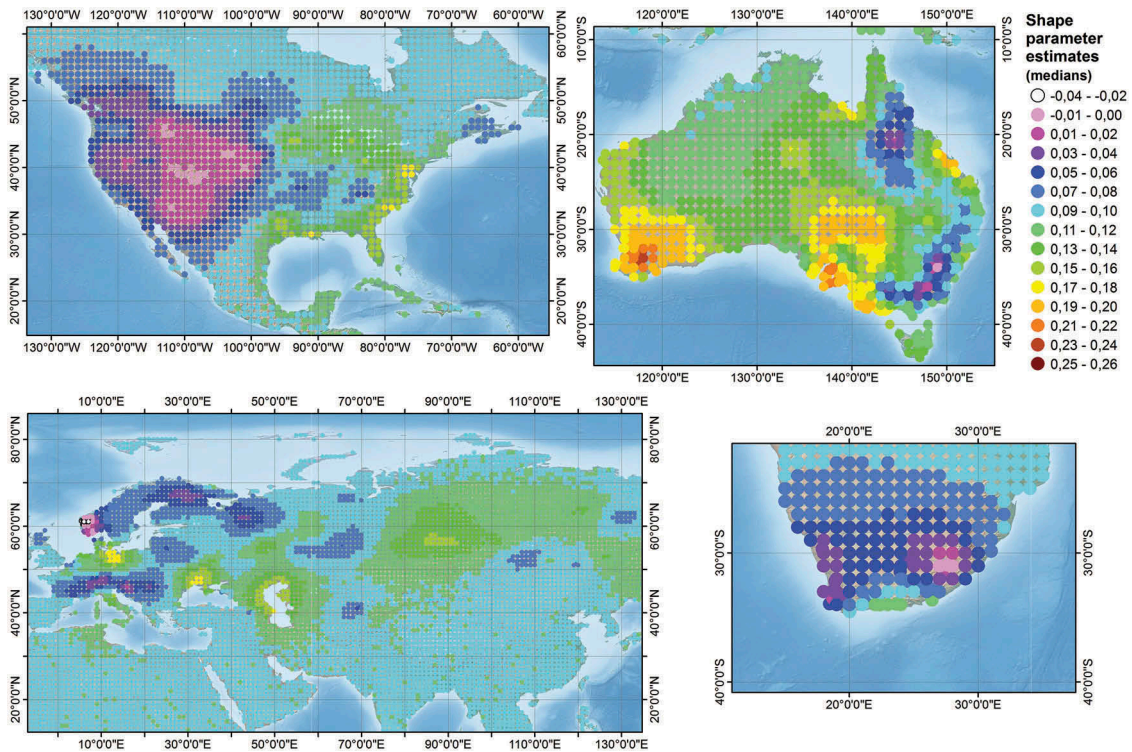


Figure 4. Map of the estimated ξ : model f (hypothesis f).

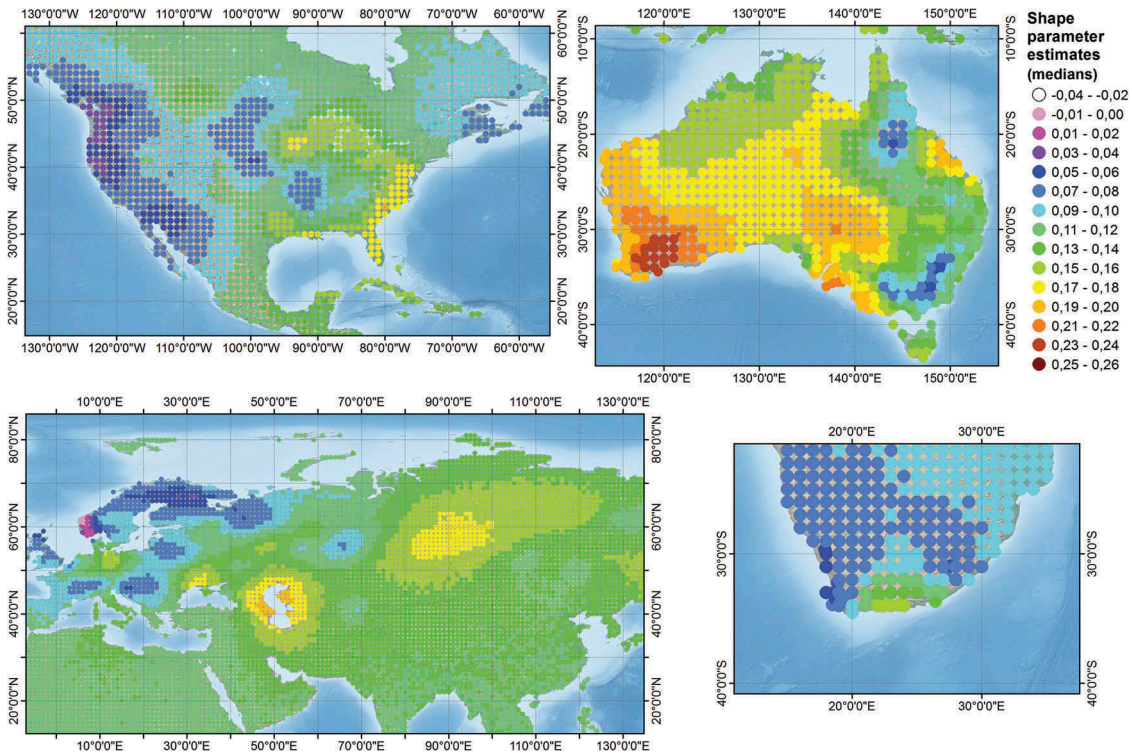


Figure 5. Map of the estimated ξ : model g (hypothesis g).

While it may seem obvious from Figures 6 and 7 that the estimated shape parameter generally has a lower value for model f than for model g , two notes of caution should be made:

- (1) This is a trend per area rather than per station. It turned out that the shape parameter estimates for model f were not different from the shape parameter estimates for model g when averaged

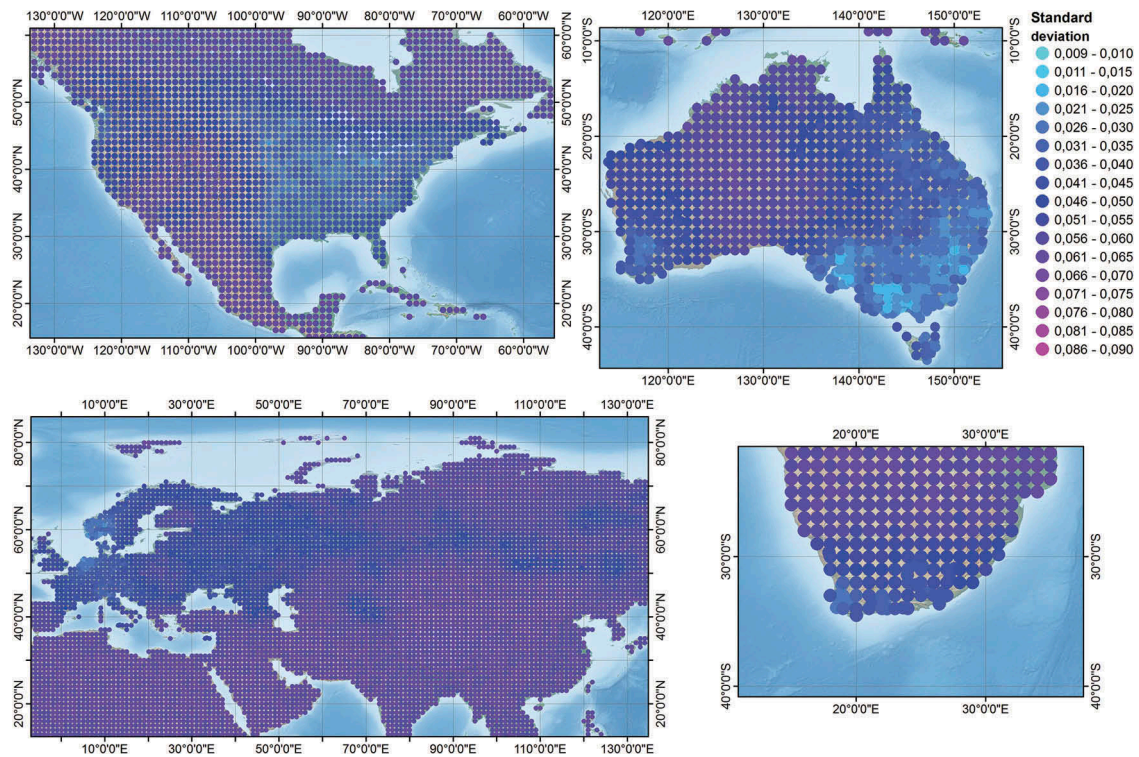


Figure 6. Map of the estimated standard deviations: model f (hypothesis f).

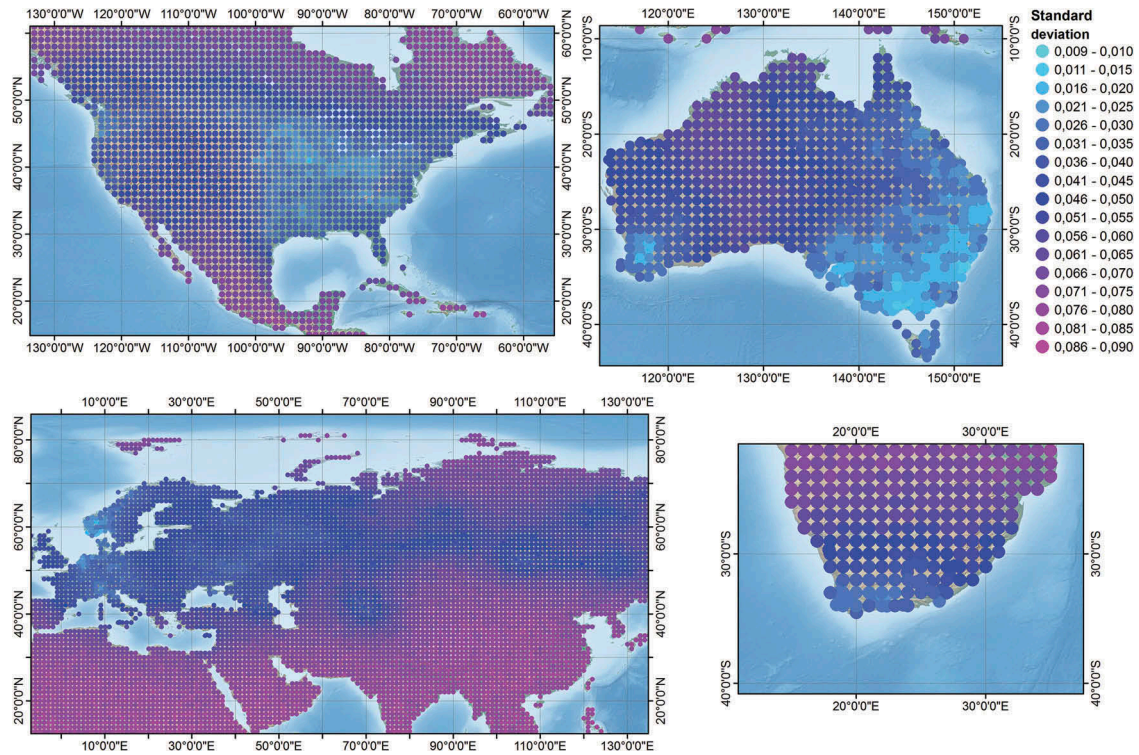


Figure 7. Map of the estimated standard deviations: model g (hypothesis g).

out over the stations. However, averaged out over an area, the shape parameter estimates can be different in the two models.

(2) These are estimates for the situation at sea level, thus not corrected for the real geographical height of each point. (Correction for height

would entail a much finer resolution of the spatial analysis as well as usage of a global map with similarly fine resolution for elevation.)

The estimate for the regression parameter for height, β_h , was approximately -0.072 per 1000 m for both models, so the shape parameter drops for high altitudes.

6.4 Discussion

Few attempts have been made in the literature to explain the origin of the spatial differences in the parameters for the distribution of extreme precipitation. One example of such an analysis was performed by Blanchet *et al.* (2009). However, they used only observations of extreme snowfall measured during 10 consequent years in the Swiss Alps (239 stations). As shown in Papalexioiu and Koutsoyiannis (2013), such short time series are not sufficient to make a good inference on the shape parameter and, hence, extreme precipitation distribution. Nevertheless, Blanchet *et al.* (2009) found an indication of nonlinear dependence of the shape parameter on altitude and mentioned a possible correlation with dominating precipitation systems.

Another investigation performed by Dyrddal *et al.* (2014) that proposed a statistical analysis directly on areal 24-h precipitation from a gridded dataset in Norway, suggested that the shape parameter varies spatially according to the dominating precipitation systems and, most probably, to the degree of orographic enhancement. Villarini and Smith (2010), in their study of flood peak distributions for the eastern United States, showed that tropical cyclones have a large impact on the GEV shape parameter values.

Unfortunately, it was not possible in the present work to test the hypothesis that the shape parameter depends on the dominating precipitation systems, since a worldwide spatial classification of dominating precipitation systems does not exist yet. However, we made an attempt to use the existing Köppen-Geiger climate classification (http://webmap.ornl.gov/wcsdown/wcsdown.jsp?dg_id=10012_1, Peel *et al.* 2007) for this purpose. Sernaldi and Kilsby (2014) pointed out some similarities between the Köppen-Geiger classification zones and the spatial behaviour of their estimates of the generalized Pareto distribution shape parameter. We re-grouped the 32 climate zones (Table 6) into 10 zones according to precipitation conditions (Table 7). Shape parameter estimates (medians, at 0 elevation) for the Extended international dataset stations (both *f* and *g* models) were then plotted against the climate zones

Table 6. Köppen-Geiger climate classification (Peel *et al.* 2007).

No.	Code	Type	Number of stations
1	Af	Tropical /Rainforest	5
2	Am	Tropical /Monsoon	4
3	Aw	Tropical /Savannah	10
4	BWh	Arid /Desert /Hot	39
5	BWk	Arid /Desert /Cold	32
6	BSh	Arid /Steppe /Hot	114
7	BSk	Arid /Steppe /Cold	181
8	Csa	Temperate /Dry summer /Hot summer	39
9	Csb	Temperate /Dry summer /Warm summer	52
10	Csc	Temperate /Dry summer /Cold summer	9
11	Cwa	Temperate /Dry winter /Hot summer	11
12	Cwb	Temperate /Dry winter /Warm summer	9
13	Cwc	Temperate /Dry winter /Cold summer	12
14	Cfa	Temperate /Without dry season /Hot summer	335
15	Cfb	Temperate /Without dry season /Warm summer	190
16	Cfc	Temperate /Without dry season /Cold summer	2
17	Dsa	Cold /Dry summer /Hot summer	2
18	Dsb	Cold /Dry summer /Warm summer	5
19	Dsc	Cold /Dry summer /Cold summer	2
20	Dsd	Cold /Dry summer /Very cold winter	3
21	Dwa	Cold /Dry winter /Hot summer	1
22	Dwb	Cold /Dry winter /Warm summer	5
23	Dwc	Cold /Dry winter /Cold summer	8
24	Dwd	Cold /Dry winter /Very cold winter	3
25	Dfa	Cold /Without dry season /Hot summer	176
26	Dfb	Cold /Without dry season /Warm summer	150
27	Dfc	Cold /Without dry season /Cold summer	54
28	Dfd	Cold /Without dry season /Very cold winter	3
29, 30	ET	Polar /Tundra	16, 5
31, 32	EF	Polar /Frost	3, 0

and the “climate-precipitation” groups to which the stations belong (according to the Köppen-Geiger climate classification). The resulting box plots are shown in Figures 8(a) and (b) and 9(a) and (b) (respectively for the full classification and for the re-grouped).

There are visible differences in the shape parameter estimates from different “climate-precipitation” zones. ANOVA tests suggest that there is strong evidence (very low p -values, smaller than about 10^{-15}) for the residuals having different means for different climate zones as well as for different “climate-precipitation” groups, for both model *f* and model *g* residuals. Climate zones were compared to “climate-precipitation” groups using BIC, which suggested that the latter were better for predicting the residuals. However, to investigate this dependency more research is needed. Spatial correlation might invalidate these results, as ANOVA analysis does not take this correlation into account.

Table 7. Re-grouped Köppen-Geiger climate classification with focus on precipitation.

No.	Type	Number of stations
1	Rainforest	5
2	Monsoon	4
3	Savannah	10
4	Desert	39
5	Steppe	295
6	Dry summer	112
7	Dry winter	49
8	Without dry season	910
9	ET (Polar /Tundra)	21
10	EF (Polar /Frost)	3

As an overall summary of our analysis, we obtained a global average (expected value) of the shape parameter, which equals 0.139 (model *f*, estimates for sea level), with a 95% credibility interval ranging from 0.127 to 0.150. Thus the estimate of 0.15 by Koutsoyiannis (2004b) obtained from a smaller dataset is still consistent with, but lies at the high end of, the range of values suggested by the present analysis of the extended dataset.

Confirming the statement of Dyrddal *et al.* (2014) regarding the shape parameter’s dependency on the degree of orographic enhancement, we found a height dependency, suggesting a decrease of the shape parameter of about 0.07 per 1000 m height, with:

$$\beta_h \times 1000 \text{ m} \in (-0.088, -0.056) \quad (34)$$

as a 95% credibility interval. This means that at a height of 2000 m, the expectation for the shape parameter will drop from about 0.14 to zero (the shape parameter of the Gumbel distribution). However, extrapolating the decrease of the shape parameter to negative values can be seen as physically inappropriate (Koutsoyiannis 2004a). This suggests that the dependency on elevation found may be nonlinear in reality.

The stationary standard variation of the spatial process was estimated to be 0.05, with (0.04, 0.06) as a 95% credibility interval for this parameter. Since the global expected value was approximately 0.14, this means

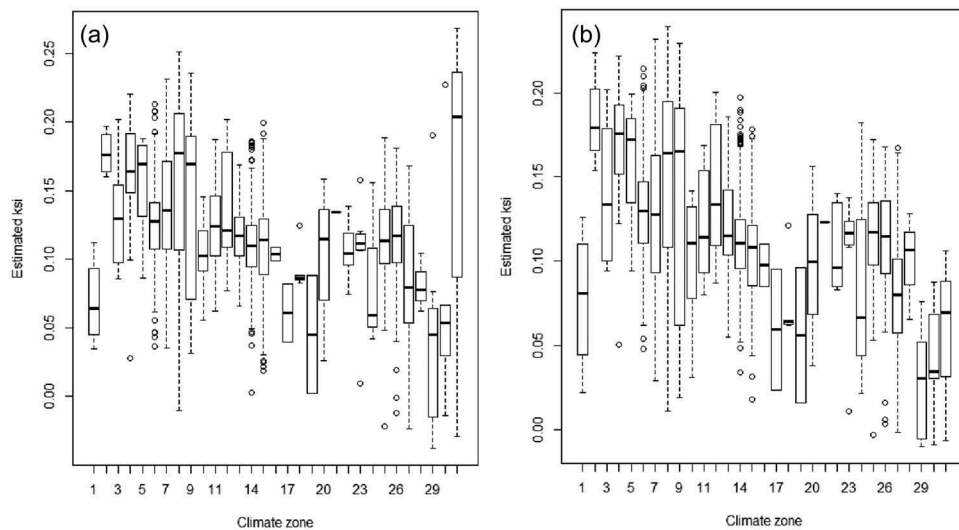


Figure 8. Box plots for the full climate classification: (a) model *f*; (b) model *g*.

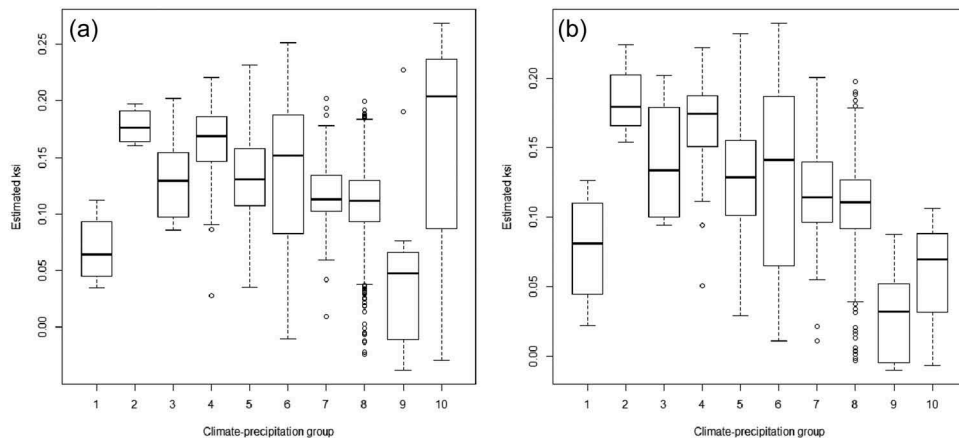


Figure 9. Box plots for the re-grouped “climate–precipitation” classification: (a) model *f*; (b) model *g*.

that, at sea level, 95% of all stations are expected to have a shape parameter of between 0.04 and 0.24. At 1000 m height, this changes to $\xi_i \in (-0.03, 0.17)$ for 95% of all stations.

7 Conclusions

The main conclusions of our study can be summarized as follows:

- The shape parameters of various stations are neither entirely separate nor independent if a distribution is assigned to them (hypotheses **a** and **b** were rejected). At the same time, the results presented here show that the shape parameter is not a universal, globally common value, but a regionally (large-scale) common value. Regions corresponding to small countries or administrative regions within larger countries can be assigned roughly the same shape parameter, but larger areas such as continents can be expected to be heterogeneous.
- The global average (expected value) for the shape parameter is equal to 0.139, with a 95% credibility interval ranging from 0.127 to 0.150.
- Shape parameter decreases with elevation: by 0.07 per 1000 m height, with $\beta_h \times 1000 \text{ m} \in (-0.088, -0.056)$ as a 95% credibility interval.
- There is no detectable dependency of the shape parameter on the precipitation form (rain/snow).
- It is very likely that the shape parameter varies according to dominating precipitation systems. However, more research is needed to define this dependency.
- It remains unclear whether the shape parameter changes with latitude in a systematic fashion.
- Maps of the shape parameter's global distribution were created (Figs. 4 and 5) and can be used for estimating extreme precipitation for engineering purposes. However, uncertainty of ξ (local) values cannot be neglected, and this uncertainty (see Figs. 6 and 7) is close to the variability of local (median) estimates across the globe.

Acknowledgments

The authors would like to thank Demetris Koutsoyiannis, Professor of Hydrology and Analysis of Hydrosystems, former Head of the Department of Water Resources and Environmental Engineering, and Dean of the School of Civil Engineering of the National Technical University of Athens, for sharing his own database with us, which allowed us to perform both the study and the comparison. We would also like to thank Professor Koutsoyiannis for the useful advice provided

throughout the study. We would like to thank CEES and the University of Oslo for access to the Abel computer cluster. Advice and support given by Dr Anita Verpe Dyrredal, researcher at the Norwegian Meteorological Institute (MET), and Dr Henning Löwe, scientist at WSL Institute for Snow and Avalanche Research SLF, Davos, were very helpful and useful – we are grateful to them. We would like to thank Emma Jane Wade for her great help throughout the language control of the manuscript. Finally, the authors thank the two reviewers, Francesco Serinaldi and Hristos Tyralis, for their careful and constructive suggestions that greatly improved the manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The authors would like to thank the NIFS-project (cooperation project between the Norwegian Water Resources and Energy Directorate (NVE), Norwegian railway authority (Jernbaneverket) and Norwegian public roads administration (Statens vegvesen)) for providing financial support for the first phase of the study. We express our deep gratitude to the Norwegian Geotechnical Institute (NGI) for the financial support to the second and the third phases of the study, which made it possible to complete the investigation and publish the results.

ORCID

Galina Ragulina  <http://orcid.org/0000-0002-8853-399X>
Trond Reitan  <http://orcid.org/0000-0003-0793-8731>

References

- Alston, S.C., 2011. *A Bayesian spatial analysis of extreme precipitation*. Dissertation submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the Degree of Doctor of Philosophy.
- Bartlett, M., 1957. A comment on D. V. Lindley's statistical paradox. *Biometrika*, 44, 533–534. doi:10.1093/biomet/44.3-4.533
- Blanchet, J., Marty, C., and Lehning, M., 2009. Extreme value statistics of snowfall in the Swiss Alpine region. *Water Resources Research*, 45, W05424. doi:10.1029/2009WR007916
- Cavanaugh, N.R. and Gershunov, A., 2015. Probabilistic tail dependence of intense precipitation on spatiotemporal scale in observations, reanalyses, and GCMs. *Climate Dynamics*, 45, 2965–2975. doi:10.1007/s00382-015-2517-1
- Cavanaugh, N.R., et al., 2015. The probability distribution of intense daily precipitation. *Geophysical Research Letters*, 42, 1560–1567. doi:10.1002/2015GL063238
- Coles, S.G., 2001. *An introduction to statistical modeling of extreme values*. Springer Series in Statistics. Berlin: Springer-Verlag.
- Coles, S.G., Pericchi, L.R., and Sisson, S., 2003. A fully probabilistic approach to extreme rainfall modeling.

- Journal of Hydrology*, 273, 35–50. doi:10.1016/S0022-1694(02)00353-0
- Coles, S.G. and Tawn, J.A., 1996. A Bayesian analysis of extreme rainfall data. *Applied Statistics*, 45, 463–478. doi:10.2307/2986068
- Dyrddal, A.V., et al., 2014. Estimating extreme areal precipitation in Norway from a gridded dataset. *Hydrological Sciences Journal*. doi:10.1080/02626667.2014.947289
- Eli, A., Shaffie, M., and Zin, W.Z.W., 2012. Preliminary study on Bayesian extreme rainfall analysis: A case study of Alor Setar, Kedah, Malaysia. *Sains Malaysiana*, 41 (11), 1403–1410.
- Fowler, H.J. and Kilsby, C.G., 2003. A regional frequency analysis of United Kingdom extreme rainfall from 1961 to 2000. *International Journal of Climatology*, 23 (11), 1313–1334. doi:10.1002/(ISSN)1097-0088
- Geyer, C.J., 1991. Markov chain Monte Carlo maximum likelihood. In: E.M. Keramides, ed. *Computing science and statistics (Proceedings of the 23rd symposium on the interface)*. Fairfax Station, Virginia: Interface Foundation, 156–163.
- Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57 (1), 97–109. doi:10.1093/biomet/57.1.97
- Jeffreys, H., 1961. *Theory of probability*. 3rd ed. Oxford Classic Texts in the Physical Sciences. Oxford: Oxford University Press.
- Katz, R.W., Parlange, M.B., and Naveau, P., 2002. Statistics of extremes in hydrology. *Advances in Water Resources*, 25 (8–12), 1287–1304. doi:10.1016/S0309-1708(02)00056-8
- Koutsoyiannis, D., 2004a. Statistics of extremes and estimation of extreme rainfall: I. Theoretical investigation. *Hydrological Sciences Journal*, 49 (4), 575–590. doi:10.1623/hysj.49.4.575.54430
- Koutsoyiannis, D., 2004b. Statistics of extremes and estimation of extreme rainfall: II. Empirical investigation of long rainfall records. *Hydrological Sciences Journal*, 49 (4), 591–610. doi:10.1623/hysj.49.4.591.54424
- Meddi, M. and Toumi, S., 2015. Spatial variability and cartography of maximum annual daily rainfall under different return periods in Northern Algeria. *Journal of Mountain Science*, 12 (6), 1403–1421. doi:10.1007/s11629-014-3084-3
- Metropolis, N., et al., 1953. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21 (6), 1087–1092. doi:10.1063/1.1699114
- Papalexiou, S.M. and Koutsoyiannis, D., 2013. Battle of extreme value distributions: A global survey on extreme daily rainfall. *Water Resources Research*, 49, 187–201. doi:10.1029/2012WR012557
- Peel, M.C., Finlayson, B.L., and McMahon, T.A., 2007. Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences*, 11 (5), 1633–1644. doi:10.5194/hess-11-1633-2007
- Reitan, T. and Aas, K., 2010/2011. A new robust importance sampling method for measuring VaR and ES allocations for credit portfolios. *Journal of Credit Risk*, 6 (4), 1–37.
- Reitan, T. and Petersen-Øverleir, A., 2009. Bayesian methods for estimating multi-segment discharge rating curves. *Stochastic Environmental Research and Risk Assessment*, 23, 627–642. doi:10.1007/s00477-008-0248-0
- Roberts, G.O., Gelman, A., and Gilks, W.R., 1997. Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability*, 7, 110–120. doi:10.1214/aoap/1034625254
- Serinaldi, F. and Kilsby, C.G., 2014. Rainfall extremes: toward reconciliation after the battle of distributions. *Water Resources Research*, 50, 336–352. doi:10.1002/wrcr.v50.1
- Smith, E., 2005. *Bayesian modelling of extreme rainfall data*. Thesis submitted for the degree of Doctor of Philosophy at the University of Newcastle upon Tyne.
- Sun, X., et al., 2015. A global analysis of the asymmetric effect of ENSO on extreme precipitation. *Journal of Hydrology*, 530, 51–65. doi:10.1016/j.jhydrol.2015.09.016
- Villarini, G. and Smith, J.A., 2010. Flood peak distributions for the eastern United States. *Water Resources Research*, 46, W06504. doi:10.1029/2009WR008395

Appendix

Here, we present the Markov chain Monte Carlo (MCMC) method used in model analysis and the estimation method for calculating the Bayesian model likelihood (BML).

MCMC method

Bayesian methodology seeks for each model in question to analyse the posterior distribution (see Equations (13)–(14)).

This represents knowledge concerning the model parameters after the data have been incorporated into the knowledge base. MCMC consists in constructing a special Markov chain, which samples from the posterior distribution even when the marginal likelihood (BML) is not known. In general, this is achieved by an iterative algorithm, in which each proposed change in parameter values is subject to rejection or acceptance. The proposal distribution can depend on the previous parameter values, so that:

$$\theta_{\text{proposed}} \sim h(\theta_{\text{proposed}} | \theta_{\text{previous}}) \quad (\text{A1})$$

This is accepted with probability:

$$p_{\text{accept}} = \min \left(1, \frac{f(\mathbf{D} | \theta_{\text{proposed}}) f(\theta_{\text{proposed}})}{f(\mathbf{D} | \theta_{\text{previous}}) f(\theta_{\text{previous}})} \right) \quad (\text{A2})$$

$$\left/ \frac{h(\theta_{\text{proposed}} | \theta_{\text{previous}})}{h(\theta_{\text{previous}} | \theta_{\text{proposed}})} \right)$$

This is known as the Metropolis-Hastings algorithm (Hastings 1970). This Markov chain will converge towards the posterior distribution.

Typically one uses a “burn-in” phase, in which the algorithm is run but the samples are not kept, in order to allow the Markov chain to converge to the target distribution. Also, since there is dependency between one sample and the next, it is usual to run a set of samples in between those that are kept so that each kept sample is approximately independent, which we will call spacing. In our case, a burn-in phase of 4000 samples gave stable results, as judged by repeated estimates and by later BML calculations. The spacing between samples was set to 40, which gave samples whose independence could not be rejected using a simple auto-regressive test. Then 100 samples were

returned, meaning that a total of 8000 MCMC iterations were performed. For the hierarchical spatial models, this was all we could afford, as the combination of spatial modelling, thousands of parameters (one μ , σ and ξ parameter for each of 1495 stations), a large dataset and 8000 iterations meant an execution time of nearly a month with our available computing resources at the time.

One can make proposal distributions for the whole parameter set. However, it is also possible to traverse the parameters one parameter at a time and when a particular parameter is handled the proposal is performed for the value of that parameter while the values of the other parameters are held constant. This is often easier to implement. In hierarchical models, such as some of those being examined in our study, such a parameter set traversal is much simpler to implement. Thus, we used parameter set traversal rather than seeking a single proposal for the whole parameter set.

The implementation and execution efficiency will depend on the proposal distribution of each parameter. As we have many different models and many parameters per model (at least for the hierarchical models), implementation efficiency was prioritized.

If the proposal distribution is symmetric along the previous values, so that:

$$h(\boldsymbol{\theta}_{\text{proposed}}|\boldsymbol{\theta}_{\text{previous}}) = h(\boldsymbol{\theta}_{\text{previous}}|\boldsymbol{\theta}_{\text{proposed}}) \quad (\text{A3})$$

one term from the acceptance probability falls away and one is left with the original Metropolis algorithm (Metropolis *et al.* 1953). This can be achieved, for instance, by letting:

$$\boldsymbol{\theta}_{\text{proposed}} = \boldsymbol{\theta}_{\text{previous}} + \sigma_{\text{prop}}\epsilon \quad (\text{A4})$$

where ϵ is standard normally distributed and is called the random walk MCMC method (RWMCMC). The RWMCMC method only depends in efficiency on the standard deviation of the proposal distribution, σ_{prop} . An optimal value for σ_{prop} can be obtained manually, but since we have many models and a great many parameters (for the hierarchical models), an automated method for optimizing σ_{prop} was sought. An acceptance rate of approximately 1/3 is deemed near optimal (Roberts *et al.* 1997). We thus added a step in the burn-in phase in which the σ_{prop} for each parameter was adjusted up if there was more than a 1/3 acceptance rate and down if there was less than a 1/3 acceptance rate.

We also included the use of the parallel tempering algorithm (Geyer 1991), in order to deal with the possibility of multimodality in the posterior distribution. Tempering consists in that a number, N , of MCMC chains are run in parallel, one for the target distribution (the posterior distribution), $f(\boldsymbol{\theta}|\mathbf{D})$, and $N - 1$ for a modified distribution proportional to $f(\boldsymbol{\theta}|\mathbf{D})^{1/T_i}$, for a set of “temperatures”, $\{T_i > T_{i-1}\}_{i=2,\dots,N}$. Thus, each higher distribution will be a smoother version of the distribution preceding it. Sometimes, the algorithm will propose switching the values of neighbouring chains. This makes it possible for the original chain to bridge the gap between two modes in the target distribution. For the spatial models, we needed $N = 12$ parallel chains, with T_i gradually increasing from 1 to 4, in order to get stable results. Even so, some runs failed, but we achieved three runs that gave consistent results.

BML estimation

The BML (Equation (14)) of a model can be used for calculating Bayesian model probabilities or the Bayes factors. From Bayes’ theorem, the posterior model probability is:

$$P(M|\mathbf{D}) = \frac{P(M)f(\mathbf{D}|M)}{\sum_{i=1}^m P(M_i)f(\mathbf{D}|M_i)} \quad (\text{A5})$$

where m is the total number of models considered, $P(M)$ is the prior model probability and $\{M_i\}_{i=1,\dots,m}$ is the set of models considered. If no model is preferred *a priori*, so that $P(M) = 1/m$, then the model with the highest BML will be the model with the highest posterior model probability. Thus, the BML or the log-BML can be used similarly to an information criterion, in that model selection consists in finding the model with the highest BML. The Bayes factor for model M_i versus model M_j is simply:

$$B_{i \text{ vs } j} = f(\mathbf{D}|M_i)/f(\mathbf{D}|M_j) \quad (\text{A6})$$

Thus, again, the BML is the only thing that needs calculation.

The BML will often not be analytically available. It can, however, be estimated using the Monte Carlo technique known as importance sampling. We used a particular importance sampler, developed by Reitan and Petersen-Øverleir (2009) and also explored by Reitan and Aas (2010/2011), to calculate the BML, which is briefly described here. (It was later used in other applications.) If one needs to estimate an integral:

$$\hat{I} = \int g(x)f(x)dx \quad (\text{A7})$$

where $f(x)$ is a distribution and $g(x)$ is a function, then a standard Monte Carlo approach would be to sample from $f(x)$ N times and estimate the integral as:

$$\hat{I} = \sum_{i=1}^N g(x_i) \quad (\text{A8})$$

where $x_i \sim f(x)$. An importance sampler replaces the distribution $f(x)$ with a different distribution $h(x)$ and estimates the integral as:

$$\hat{I} = \sum_{i=1}^N g(x_i)f(x_i)/h(x_i) \quad (\text{A9})$$

where now $x_i \sim h(x)$. If $h(x)$ is similar in form to $g(x)f(x)$, this alternative proposal distribution can be quite effective. Since the posterior distribution is proportional to $f(\mathbf{D}|\boldsymbol{\theta}_M, M)f(\boldsymbol{\theta}_M|M)$, a proposal distribution is obtained by extracting the average and variance from the MCMC samples. One can then make a multivariate normal proposal distribution which will have approximately the same first- and second-order moments as the posterior distribution. The multivariate normal distribution can be either independent for each parameter (Reitan and Aas 2010/2011), or also catch the linear dependencies between the parameters in the posterior distribution (Reitan and Petersen-Øverleir 2009). The latter means will be closer to the posterior distribution, but would mean calculating the covariance matrix of the MCMC samples, which is not practical when the parameter set is very large. Thus, in this application we opted for independence between parameters in the proposal distribution.