Project no. GOCE-CT-2003-505539

Project acronym: ENSEMBLES

Project title: ENSEMBLE-based Predictions of Climate Changes and their Impacts

Instrument: Integrated Project

Thematic Priority: Global Change and Ecosystems

D5.22
A non-stationary index-flood model for precipitation extremes in transient RCM runs

Due date of deliverable: Month 50 (October 2008)
Actual submission date: Month 55 (March 2009)

Start date of project: 1 September 2004 Duration: 60 Months

Organisation name of lead contractor for this deliverable: KNMI

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the Consortium (including the Commission Services)	

A non-stationary index-flood model for precipitation extremes in transient RCM runs

Martin Hanel,[1] T. Adri Buishand,[1] and Christopher A. T. Ferro[2]

[1] Royal Netherlands Meteorological Institute (KNMI), De Bilt, Netherlands
[2] School of Engineering, Computing and Mathematics, University of Exeter, Exeter, UK

Submitted to Journal of Geophysical Research - Atmospheres

Abstract

The Generalized Extreme Value (GEV) distribution has often been used to describe the distribution of daily maximum precipitation in observed and climate model data. The model developed in this paper allows the GEV location parameter to vary over the region, while the dispersion coefficient (the ratio of the GEV scale and location parameters) and the GEV shape parameter are assumed to be constant over the region. This corresponds with the index-flood assumption in hydrology. It is further assumed that all three GEV parameters vary with time such that the relative change in a quantile of the distribution is constant over the region. The parameters of the non-stationary GEV model are estimated by the maximum likelihood method. Goodness of fit is tested with the Anderson-Darling statistic. A parametric bootstrap procedure is used to determine the local and field significance. The non-stationary GEV model is applied to the 1-day summer and 5-day winter precipitation extremes in the river Rhine basin in a simulation of the RACMO regional climate model for the period 1950–2100. The results are compared with gridded observations. Except for an underestimation of the dispersion coefficient of the 5-day winter maxima by about 35% the GEV parameters are reasonably well reproduced by RACMO. A positive trend in the dispersion coefficient is found in the summer season, which implies that the relative increase of a quantile increases with increasing return period. In the winter season there is a positive trend in the location parameter and a negative trend in the shape parameter. For large quantiles the latter counterbalances the effect of the increase of the location parameter. It is shown that the standard errors of the parameter estimates are significantly reduced in the regional approach compared to those of the estimated parameters from individual grid box values, especially for the summer maxima.

1. Introduction

Regional climate models (RCMs) nested into a global climate model provide useful information about potential local climate change. Precipitation extremes in RCM simulations have been analyzed in different ways. One method is to consider the change in a large empirical quantile of the daily precipitation amounts (e.g., the 99th) or the properties of the exceedances of such a quantile [e.g., *Durman et al*., 2001; *Christensen and Christensen*, 2004]. An alternative is to fit an extreme-value distribution to the largest daily precipitation amount in a season [e.g., *Frei et al*., 2006; *Goubanova and Li*, 2007] or year [e.g., *Huntingford et al*., 2003; *Fowler et al*., 2005; *Ekström et al*., 2005]. A number of these studies also deal with the maxima of multi-day precipitation amounts.

A problem with extreme precipitation is that the likelihood of detecting a systematic change at a grid box is generally small due to the large year-to-year variability. *Frei and Schär* [2001] mention, for instance, that a frequency change by a factor of 1.5 in events with an average return period of 100 days can be detected with a probability of only 0.2 in a 100-year record. The decrease of this probability with increasing event magnitude limits the detection of systematic changes in extreme events at a single grid box.

Spatial pooling has been considered to obtain meaningful changes in extremes. *Frei et al*. [2006] and *Goubanova and Li* [2007] averaged an estimated quantile of the extreme-value distribution over large regions. *Kendon et al.* [2008] studied the effectivity of spatial pooling for the detection of changes in the 95th percentile of wet-day precipitation. An alternative is to assume that the most uncertain parameters of the extreme-value distribution are constant over some region. The estimates of these parameters based on the pooled data across the region are then generally more precise than those from the data of an individual grid box leading to a reduction of the standard errors of the estimated quantiles of the distribution. This approach has its origin in hydrology where it is known as regional frequency analysis. The most popular method is the index-flood method. *Fowler et al*. [2005] and *Ekström et al*. [2005] applied this method to the 1-, 2-, 5-, and 10-day annual maximum precipitation amounts across the UK in two RCM simulations. Apart from a change in the distribution parameters between the control and future climate, these parameters do not vary over time in their application.

The purpose of this paper is to introduce an index-flood model with time-varying parameters as a tool to summarize changes of extreme precipitation in transient RCM simulations. The model is applied to daily precipitation in the river Rhine basin in the RACMO-ECHAM5 simulation. In this part of Europe, short-period convective storms may cause local flooding in summer, whereas in winter multi-day episodes may have adverse impacts over large areas. Similarly as in *Frei et al.* [2006], we analyze the 1-day precipitation maxima in summer and the 5-day precipitation maxima in winter.

The index-flood model is described in section 2. Section 3 provides some information about the river Rhine basin, the RACMO-ECHAM5 simulation, and the observational data sets that were used for validation. The results for the summer maxima are presented in section 4 and those for the winter maxima in section 5. Section 6 presents the conclusions.

2. Regional modeling of non-stationary precipitation extremes

2.1. Index-flood model

The idea behind the index-flood method is that the variables within a homogeneous region are identically distributed after scaling with a site-specific factor, the index flood. Then the *T*-year quantile $Q_T(s)$ of the distribution of the variable $X(s)$ at any given site s, i.e., the value that is exceeded with probability $1/T$, can be written as

$$Q_T(s) = \mu(s) q_T , \tag{1}$$

where $\mu(s)$ is the index flood and q_T is a regional dimensionless quantile function, in this context often called the growth curve. The mean or median of the distribution is usually chosen as the index flood.

The index-flood method has been used with different probability distributions. For seasonal and annual precipitation maxima the generalized extreme value (GEV) distribution is quite popular. This distribution is a three-parameter distribution that combines the three possible types of extreme value distributions (i.e., Gumbel, Fréchet, and reverse Weibull distributions). Its distribution function is given by

$$\begin{aligned} F(x) &= \exp\left\{ -\left[1 + \kappa\left(\frac{x-\xi}{\alpha} \right) \right]^{-\frac{1}{\kappa}} \right\}, && \kappa \neq 0 , \\ F(x) &= \exp\left\{ -\exp\left[-\left(\frac{x-\xi}{\alpha} \right) \right] \right\}, && \kappa = 0 , \end{aligned} \tag{2}$$

with ξ, α, and κ the location, the scale, and the shape parameter, respectively. The shape parameter controls the behavior of the tails of the distribution – positive values imply a heavy upper tail (Fréchet distribution).

Apart from support from extreme value theory to select the GEV distribution, it has often been found that this distribution describes the distribution of observed or simulated precipitation maxima well. For annual precipitation maxima of various durations *Schaefer* [1990], *Alila* [1999], and *Kyselý and Picek* [2007], using L-moment ratio diagrams, observed that the GEV distribution is generally superior to other candidate distributions. In addition, *Alila* [1999] and *Kyselý and Picek* [2007] found that a goodness of fit test based on the L-kurtosis did not reject the GEV distribution. *Buonomo et al.* [2007] and *Goubanova and Li* [2007] used the Kolmogorov-Smirnov goodness of fit test and concluded that the GEV distribution is appropriate for modeling precipitation extremes in RCM projections for most parts of Europe as well. However, problems were met in dry areas where most of the seasonal maxima were zero.

For the development of our non-stationary model it is convenient to use the location parameter as the index flood, i.e., $\mu(s) = \xi(s)$, rather than the mean or the median. If the seasonal maximum $X(s)$ at site s follows a GEV distribution with parameters $\xi(s)$, $\alpha(s)$, and $\kappa(s)$, then the scaled seasonal maximum $X(s)/\xi(s)$ has a GEV distribution with location parameter 1, scale parameter $\gamma(s) = \alpha(s)/\xi(s)$, and shape parameter $\kappa(s)$,

which does not vary over the region if $\gamma(s)$ and $\kappa(s)$ are constant, i.e., when $\gamma(s) = \gamma$ and $\kappa(s) = \kappa$. The dispersion coefficient γ is comparable to the coefficient of variation.

The T-year quantile of the scaled seasonal maximum $X(s)/\xi(s)$ follows from equation (2) by setting $F(q_T) = 1–1/T$, $\xi = 1$ and $\alpha = \gamma$:

$$q_T = 1 - \frac{\gamma}{\kappa}\left\{1 - \left[-\log\left(1 - \frac{1}{T}\right)\right]^{-\kappa}\right\}, \qquad \kappa \neq 0,$$
$$q_T = 1 - \gamma \log\left[-\log\left(1 - \frac{1}{T}\right)\right], \qquad \kappa = 0. \tag{3}$$

Note that $q_T = 1$ for $T = 1/(1–1/e) = 1.58$ years being the return period corresponding to the location parameter. The growth curve is determined by γ and κ. This is also the case if $X(s)$ is scaled by the mean [*Sveinsson et al.*, 2001] or the median [*Northrop*, 2004]. However, the index flood then depends on γ and κ, which is inconvenient in the case of temporal trends in these common parameters.

2.2. Non-stationary index-flood model

A few studies in the hydrological literature deal with non-stationarity in regional frequency analysis. *Cunderlik and Burn* [2003] assume temporal and spatial variation in both the location and scale parameter of the distribution. Linear trends in these parameters were estimated with a distribution-free method due to *Sen* [1968]. In a subsequent paper [*Cunderlik and Ouarda*, 2006] the scale parameter was assumed to be constant over the region of interest but still time-varying. The regional scale parameter was estimated as a weighted average of the at-site scale parameters. *Renard et al.* [2006] used a regional non-stationary GEV model to describe trends in annual maximum discharges. In that model the shape parameter was constant but the scale and location parameters varied over the region and there was a common linear trend in the location parameter. Statistical inference was based on a Bayesian analysis using Markov chain Monte Carlo methods. Other authors have successfully used a GEV distribution with time-varying parameters, e.g., *Kharin and Zwiers* [2005], *Adlouni et al.* [2007], *García et al.* [2007], and *Brown et al.* [2008], although not in the framework of regional frequency analysis.

Let $X(s, t)$ be the seasonal maximum at site s in year t. Using the location parameter of the GEV distribution as the index flood, the T-year quantile $Q_T(s, t)$ can be represented as

$$Q_T(s,t) = \xi(s,t) q_T(t), \tag{4}$$

where $q_T(t)$ is given by equation (3) but with time-dependent dispersion coefficient $\gamma(t)$ and shape parameter $\kappa(t)$. The location parameter $\xi(s, t)$ varies both in time and space. As in the non-stationary GEV model of *Renard et al.* [2006], the temporal trend in this parameter is assumed to be constant over the region of interest. A motivation for this is that changes in extreme precipitation are mainly associated with large-scale changes in the atmospheric conditions (changes of the amount of precipitable water due to temperature change and changes of the atmospheric circulation).

We propose the following model for the GEV parameters:

$$\xi(s,t) = \xi_0(s)\exp[\xi_1 I(t)] \quad (5)$$

$$\gamma(t) = \exp[\gamma_0 + \gamma_1 I(t)] \quad (6)$$

$$\kappa(t) = \kappa_0 + \kappa_1 I(t) \quad (7)$$

where $I(t)$ is the time indicator (time-dependent covariate). The choice of $I(t)$ is discussed in section 3.

In line with GEV models with a time-dependent scale parameter, the exponential expression is used in equation (6) to avoid negative values of the dispersion coefficient. The purpose of the exponential function in equation (5) is to achieve that the relative changes in the quantiles are constant over the region of interest. From equations (4) and (5) it follows that the relative change of the T-year quantile between years t_1 and t_2 at site s can be written as

$$\frac{Q_T(s,t_2)}{Q_T(s,t_1)} = \frac{\xi(s,t_2)}{\xi(s,t_1)}\frac{q_T(t_2)}{q_T(t_1)} = \exp\{\xi_1[I(t_2)-I(t_1)]\}\frac{q_T(t_2)}{q_T(t_1)}, \quad t_2 \geq t_1 \geq t_0, \quad (8)$$

which does not depend on s. Apart from the common usage of percentages in the case of changes in extreme precipitation, a reason to assume constant relative changes rather than absolute changes is that specific humidity and hence atmospheric moisture would increase roughly exponentially with temperature (about 6.5% per degree) according to the Clausius-Clapeyron relation [e.g., *Pall et al.*, 2007].

The parameters $\xi_0(s)$, ξ_1, γ_0, γ_1, κ_0, and κ_1 of the model were estimated by maximizing the log-likelihood

$$L = \sum_{s=1}^{S} L_s\left(\xi_0(s), \xi_1, \gamma_0, \gamma_1, \kappa_0, \kappa_1\right) \quad (9)$$

where $L_s(\xi_0(s), \xi_1, \gamma_0, \gamma_1, \kappa_0, \kappa_1)$ is the log-likelihood for the seasonal maxima at grid box s, and S is the number of grid boxes in the region. The number of parameters that has to be determined is thus S+5. Dealing usually with more than 50 grid boxes in one region it was difficult to estimate all parameters simultaneously. Therefore, a two-step procedure was applied [*Arnell and Gabriele*, 1988; *Buishand*, 1991]. In the first step, all the site-specific location parameters $\xi_0(s)$ were estimated keeping the regional parameters fixed. In the second step, we fixed the values of $\xi_0(s)$ on their estimates from the previous step and estimated the regional parameters ξ_1, γ_0, γ_1, κ_0, and κ_1. These estimates were then used again in step one. The two steps were repeated until convergence.

2.3. Uncertainty and model checking

The maximum likelihood method also provides standard errors of the estimates if the data are independent. In the case of spatial dependence the bootstrap can be used to assess the uncertainty of the parameters and quantiles of the distribution. Rather

than bootstrapping the data of the grid boxes individually, the data for a certain year are bootstrapped simultaneously in order to preserve the spatial dependence [cf. *Faulkner and Jones*, 1999; *Kharin et al.*, 2007]. Prior to resampling, the trend has to be removed from the maxima $X(s, t)$. This is done by the transformation [*Coles*, 2001]

$$\widetilde{X}(s,t) = \frac{1}{\hat{\kappa}(t)} \log\left[1 + \frac{\hat{\kappa}(t)}{\hat{\gamma}(t)}\left(\frac{X(s,t)}{\hat{\xi}(s,t)} - 1\right)\right], \tag{10}$$

where $\widetilde{X}(s,t)$ are the detrended seasonal maxima and $\hat{\xi}(s,t)$, $\hat{\gamma}(t)$, and $\hat{\kappa}(t)$ are the maximum likelihood estimates of the GEV parameters (these are obtained by replacing $\xi_0(s)$, ξ_1, γ_0, γ_1, κ_0, and κ_1 in equations (5)–(7) by their maximum likelihood estimates $\hat{\xi}_0(s), \hat{\xi}_1, \hat{\gamma}_0, \hat{\gamma}_1, \hat{\kappa}_0$, and $\hat{\kappa}_1$). Then a sample $t_1, \ldots, t_u, \ldots, t_N$ is drawn with replacement from the years $1, \ldots, N$ (here the number N of years is 150). A bootstrap sample of detrended seasonal maxima is obtained by taking the vector $\left(\widetilde{X}(1,t_u), \ldots, \widetilde{X}(s,t_u), \ldots, \widetilde{X}(S,t_u)\right)$ for each resampled year t_u. Finally, the sample is transformed back to the original scale according to

$$X(s,u) = \hat{\xi}(s,u)\left\{1 + \hat{\gamma}(u)\frac{\exp[\hat{\kappa}(u)\widetilde{X}(s,t_u)] - 1}{\hat{\kappa}(u)}\right\} \tag{11}$$

and the parameters are re-estimated.

The transformed maxima $\widetilde{X}(s,t)$ should have a standard Gumbel distribution if the model is correct (we refer to them as standard Gumbel residuals further), which is tested in this study by calculating the Anderson-Darling statistic for each grid box. The Anderson-Darling statistic A^2 is defined as [*Anderson and Darling*, 1952]

$$A^2 = N\int_{-\infty}^{\infty} \frac{[F_N(x) - F(x)]^2}{F(x)[1 - F(x)]}\, dF(x), \tag{12}$$

where $F_N(x)$ is the empirical distribution of the $\widetilde{X}(s,t)$ for the grid box of interest and $F(x)$ is the standard Gumbel distribution function, $F(x) = \exp[-\exp(-x)]$. The A^2 statistic summarizes the mean square distance between the two distributions, putting more weight on the tails of the distribution through the function $1/\{F(x)[1-F(x)]\}$. For testing goodness of fit of extreme value distributions it has been shown [e.g., *Shimokawa and Liao*, 1999; *Laio*, 2004] that this statistic is more powerful than the Kolmogorov-Smirnov and Cramer-von Mises statistics and the probability plot correlation coefficient. Here A^2 also tests the adequacy of assumptions about the GEV parameters (index-flood assumption and constant trends over the region of interest).

The presence of decadal variability has been checked by exploring the temporal pattern of residuals. For this purpose, it is convenient to work with residuals that

have a symmetric distribution, in particular the normal distribution. Standard normal residuals $\widetilde{X}_{\text{norm}}(s,t)$ are obtained by the transformation

$$\widetilde{X}_{\text{norm}}(s,t) = \Phi^{-1}\left\{\exp\left[-\exp\left(-\widetilde{X}(s,t)\right)\right]\right\}, \tag{13}$$

with Φ^{-1} the quantile function of the standard normal distribution.

3. Rhine basin and data used

The river Rhine basin has an area of 185,000 km^2 and is situated in the territory of nine European countries (Figure 1a). The basin stretches from the Alps in the south with mountain peaks higher than 4000 m to a flat delta in the Netherlands in the north. Annual mean precipitation is quite variable – the wettest part is the Alpine region with more than 3000 mm of precipitation per year in some areas, the driest part is the area around Mainz in the center of the Rhine basin where annual mean precipitation is about 400 mm. The overall annual mean precipitation is 910 mm.

The precipitation maxima in the output of the KNMI regional climate model RACMO driven by the ECHAM5 global climate model under the SRES A1B emission scenario were studied. The RACMO-ECHAM5 run begins in 1950 and ends in 2100. The horizontal resolution of the RACMO model is ≈ 25 km on a rotated longitude-latitude grid. There are 316 grid boxes whose center lies within the Rhine basin (Figure 1b).

To use the index-flood model homogeneous regions have to be defined. Therefore the GEV parameters were estimated at individual grid boxes for the 1-day summer (JJA) and 5-day winter (DJF) maxima for two time slices (1950–1990 and 2060–2100) assuming no time dependence within these slices. The subdivision of the area was based on the spatial pattern of the dispersion coefficient, because the estimated shape parameter is less precise. Spatial heterogeneity of the dispersion coefficient turned out to be stronger for the summer maxima than for the winter maxima. We finally arrived at a subdivision of the Rhine basin into 5 regions (Figure 1b) each including 48 to 97 grid boxes. Region 1 roughly corresponds to the Swiss part of the basin and region 5 to the Dutch part.

Figure 2 shows the change of the seasonal and annual mean precipitation between the periods 1950–1990 and 2070–2100. In the model output annual mean precipitation increases by about 5% over the whole basin, winter mean precipitation increases by more than 20% over most of the basin and summer mean precipitation decreases by 10–20%.

The model for the parameters defined in equations (5)–(7) requires the choice of the time indicator $l(t)$. The most straightforward approach is to use $l(t) = t$. Since the enhanced greenhouse effect is small during the first decades of the RCM simulation, a more complicated function of the year t is needed to allow the GEV parameters to stay constant or almost constant in this period. Such a function usually contains one or more unknown parameters which generally leads to more uncertain trend estimates. An alternative time indicator which is representative of the enhanced greenhouse effect is the global mean temperature. In our application a seasonal global mean temperature anomaly from the driving ECHAM5 model is used. This

anomaly is calculated with respect to the overall 1950–2100 mean temperature to achieve that the parameters $\xi_0(s)$ are approximately orthogonal to the common parameters ξ_1, γ_0, γ_1, κ_0, and κ_1, which significantly speeds up the two-stage estimation procedure. Using temperature anomalies with respect to some historical period such as 1960–1990 (or temperature itself) leads to a significant correlation between $\hat{\xi}_0(s)$ and $\hat{\xi}_1$. For example, if the historical period 1960–1990 is considered, the average correlation between these parameters is –0.87. This correlation is only 0.14 if the anomalies are calculated with respect to the overall mean. The global mean temperature anomaly for the summer and winter season is given in Figure 3. The increase between the periods 1950–1990 and 2070–2100 is ≈ 3 °C in the summer and ≈ 3.5 °C in the winter season.
To compare the distribution of extremes in the RACMO-ECHAM5 run with that in observations, the gridded observed daily precipitation amounts produced within the EU funded ENSEMBLES project [*Haylock et al.*, 2008] were used. These data (further denoted as ENS) are available on different grids including a rotated longitude-latitude grid with a resolution of ≈ 25 km, which makes the comparison with the RACMO data straightforward. The data cover the period 1950–2005. The density of stations used for gridding in the Rhine basin was variable (e.g., Netherlands ≈ 1 station per 400 km^2, Switzerland ≈ 1 station per 1300 km^2, and Germany ≈ 1 station per 3400 km^2).

Another gridded data set of observed daily precipitation comprising part of the river Rhine basin is the Alpine data set [*Frei and Schär*, 1998]. This data set (from here on denoted as ALP) is available on a regular longitude-latitude grid with a resolution of ≈ 25 km. The gridding of the ALP data was based on a significantly larger number of stations (≈ 1 station per 100–200 km^2) than the ENS data and more high-elevation stations were included. The period that is covered is shorter, from 1971 to 1995.

4. Summer maxima

4.1. Results

Figure 4 shows boxplots of estimated parameters and their trends for the 1-day summer maximum precipitation. These boxplots were obtained from 3000 bootstrap samples. The upper panels (Figures 4a–c) refer to the GEV parameters for the period 1950–1990. The estimated values of $\overline{\xi}$ (average location parameter over the S grid boxes in the region), γ, and κ were derived from equations (5)–(7) using the 1950–1990 average global summer mean temperature anomaly for $I(t)$.

In the RACMO-ECHAM5 simulation the average location parameter is about 32 mm in the Alpine area and about 21 mm in the rest of the basin. This difference is caused by the high seasonal mean precipitation amounts in the Alps. The dispersion coefficient varies between 0.32 and 0.37 in the RACMO-ECHAM5 simulation. The high value of the dispersion coefficient in region 3 could be related to the low mean precipitation in this region. High values of the coefficient of variation of observed annual precipitation maxima have often been found in relatively dry areas [see *Brath et al.*, 2003]. We do not have any explanation for the high values of the dispersion coefficient in region 5. The shape parameter is positive (Fréchet distribution).

Figures 4a–c also give the estimated parameters from the ENS and ALP data sets based on the non-stationary GEV model using the average global summer mean temperature anomaly from the HadCRUT3 data set of gridded observed

temperatures [*Brohan et al.*, 2006] for $I(t)$ in equations (5)–(7). The location parameter in the RACMO-ECHAM5 simulation is on average 10% larger than the location parameter from the ENS data. Part of this difference is caused by the low number of stations used for gridding in certain countries. The variability of the grid box average precipitation is then too low which results in a reduction of the precipitation maxima. The average estimate of the location parameter from the ENS data in region 1 is significantly lower than that from the ALP data which are based on a substantially larger number of stations. Furthermore, there is little difference between the estimated location parameter from the RACMO-ECHAM5 and ENS data in region 5 where the gridding of the ENS data was based on a relatively large number of stations. The dispersion coefficient and the shape parameter show a reasonable agreement in the ENS and ALP data sets for region 1. These two parameters are in most regions somewhat larger in the RACMO-ECHAM5 simulation than in the ENS data.
Figures 4d–f refer to the estimated trends in the GEV parameters $\xi(t)$, $\gamma(t)$, and $\kappa(t)$. The change of $\xi(t)$ and $\gamma(t)$ is given as the ratio of the mean values of these parameters for the periods 2070–2100 and 1950–1990, the change of $\kappa(t)$ is the difference in the mean of $\kappa(t)$ for the same periods. There is a notable positive trend in the dispersion coefficient in all five regions, while the trends in the location and the shape parameters are less clear.

To asses the reduction of the variance of the parameter estimates due to spatial pooling, the non-stationary GEV model was fitted for each individual grid box (i.e., without spatial pooling) and boxplots of the estimated parameters were calculated using the bootstrap. Then, for each region and each parameter the average boxplot was produced by averaging the boxplots from all grid boxes in the region (i.e., the lower whisker of the average boxplot was the average of the lower whiskers of the boxplots from all grid boxes in the region and so on). These average boxplots were compared with those in Figure 4. Table 1 gives the reduction of the interquartile range (range between the 25th and 75th percentile) for the summer season for the RACMO-ECHAM5 data. This reduction is substantial (30–80%). Spatial pooling has the largest influence on the uncertainty of the shape parameter. The reduction is larger for parameters describing trends.

The relative changes of quantiles (ratios of the mean values of quantiles in the periods 2070–2100 and 1950–1990) are shown in Figure 5. Despite the decrease of summer mean precipitation, the quantiles of the extremes increase. The change of the 2-year quantile is largely determined by the change of the location parameter. Therefore, there is only a small increase (up to 10%) of the 2-year quantile except for region 2 where a relatively large increase of the location parameter leads to an increase of this quantile of almost 30%. The relative increase of the 50-year quantile is larger in all regions except for region 2 because of the positive trend in the dispersion coefficient. The 50-year quantile increases by 10–30% in regions 1 and 3 and even by 50% in regions 4 and 5 where the positive trend in the dispersion coefficient is enforced by the positive trend in the shape parameter. The relatively small increase of the 50-year quantile in region 2 is caused by the decrease of the shape parameter. The uncertainty of the change of a given quantile is large, in general comparable with its magnitude.

A possibility to reduce the uncertainty of changes of quantiles is to join regions or to assume that certain regions have common parameters. To test for differences between regions the following statistic was used:

$$R = \sum_{i=1}^{n} (\hat{\theta}_i - \overline{\theta})^2 \quad (14)$$

with n the number of regions, $\hat{\theta}_i$ the estimate of the parameter of interest for region i and $\overline{\theta} = \sum_{i=1}^{n} \hat{\theta}_i / n$. The results of the test for the five regions in the Rhine basin are given in Table 2. The p-values were obtained using a bootstrap procedure as described in Appendix A. The differences between the regions are significant at the 0.1 level for all parameters except the trend parameter γ_1. The estimates of the trend parameters ξ_1 and κ_1 for region 2 in Figure 4 differ substantially from those for the other regions. Therefore, a restricted model with common trends of the GEV parameters in regions 1, 3, 4, and 5 was also fitted. Regardless different values of $\hat{\gamma}_0$ and $\hat{\kappa}_0$, the estimated changes of the quantiles for this restricted model are almost identical in these four regions (see Figure 6) and roughly correspond to the mean of the relative changes in these regions assuming no common parameters. The uncertainty is, however, significantly reduced. For the 50-year quantile in Figure 6 a 27% increase is found, i.e., 7.6% per degree, which is close to the value expected from the Clausius-Clapeyron relation. This is in agreement with *Lenderink and van Meijgaard* [2008] who found for the RACMO-ECHAM5 simulation that the average increase of large percentiles (99th and 99.99th) of the daily precipitation amounts over central Europe (between 46 °N–62 °N and 2 °W–22 °E) obeyed the Clausius-Clapeyron relation.

4.2. Model validation

The goodness of fit was tested using the A^2 statistic. For regions 1 and 3, Figure 7 gives the A^2 value for each grid box together with critical values for a test at the 0.1 significance level. These critical values were determined using a parametric bootstrap procedure (Appendices B and C). The local 0.1 critical values in Figure 7 apply to the goodness of fit test at an individual grid box. The likelihood that all A^2 values fall below these critical values is small. In the case of an adequate fit it is expected that 10% of the A^2 values exceed the local 0.1 critical value. This fraction is higher for regions 1 and 3 (≈ 20%). In order to evaluate the joint statistical significance the 0.1 global critical values in Figure 7 have to be considered. The chance that some A^2 value exceeds the line of these critical values is 0.1 if the data come from the assumed model. None of the A^2 values for region 1 is above this line, but in region 3 there are five grid boxes for which A^2 exceeds the global 0.1 critical value. Four of these grid boxes are situated near Mainz in the center of the region where the lowest precipitation in the Rhine basin is found. A separate model fit for these four grid boxes and three adjacent grid boxes with large A^2 values revealed a relatively high dispersion coefficient for this subregion. There was no evidence of lack of fit of the GEV distribution and the trend γ_1 in the dispersion coefficient did not deviate much from that for the rest of region 3. These seven grid boxes in this relatively dry area were excluded. In addition, four grid boxes in region 4 for which A^2 exceeds the global 0.1 critical value were excluded too. One of these grid boxes is located on the western border of the river Rhine basin, whereas the other three are situated in a relatively wet subregion, known as Sauerland, with grid box estimates of γ_0 lower than those for the rest of this region. The GEV model was then fitted again and the A^2 statistics and their critical values were recalculated. The results discussed in section 4.1 refer to the refitted model as well as Figures 4, 5, and 6.

Figure 8 shows the location of the excluded grid boxes and summarizes the results of the goodness of fit tests. In region 3 there remains one grid box for which A^2 exceeds the global 0.1 critical value.

Two additional checks were made to assess the presence of decadal variability: (1) the standard normal residuals were averaged over each of the five regions and smoothed using a locally weighted regression, "loess" [*Cleveland*, 1979], in order to find significant temporal patterns, (2) the average autocorrelation of the standard normal residuals was calculated for each of the five regions. Figures 9 and 10 show the results of these checks for region 1. Both pictures are representative of the other regions as well and both show no evidence of decadal variability.

5. Winter maxima

5.1. Results

Boxplots of the estimated GEV parameters for the 5-day winter maximum precipitation in the RACMO-ECHAM5 simulation for the period 1950–1990 are given in Figures 11a–c. As for the summer season the location parameter in the Alpine region is higher than in the rest of the basin. The dispersion coefficient shows a south north gradient. The shape parameter is almost zero in three of the five regions.

The RACMO-ECHAM5 simulation overestimates the location parameter by 10–30% and underestimates the dispersion coefficient by 35% with respect to the ENS data. For the 5-day winter maxima the reduction of variability in the ENS data due to the gridding of insufficient station data is smaller than for the 1-day summer maxima because of the stronger correlation between the 5-day winter maxima. The low number of stations used for gridding cannot explain the observed differences between the parameter estimates from the RACMO-ECHAM5 and ENS data. In contrast to the 1-day summer maxima the differences between the estimated location parameters from the ALP and ENS data are small for region 1. There is also a significant difference between the estimated location parameters from the RACMO-ECHAM5 and ENS data for the well-gauged region 5. The overestimation of the location parameter in the RACMO-ECHAM5 data is strongly related to the bias in winter mean precipitation (42%), which mainly results from the simulation of too many wet days. Part of this bias is caused by the systematic undercatch inherent to rain gauges for which the ENS data were not corrected. The bias in the mean is accompanied by an underestimation of the coefficient of variation (22%) of the daily mean precipitation amounts. The low relative variability of the daily values in the RACMO-ECHAM5 simulation partly accounts for the underestimation of the dispersion coefficient in the GEV model for the 5-day maxima across the basin.

The estimated trends of the GEV parameters are shown in Figures 11d–f. The location parameter increases and the shape parameter decreases significantly over the whole basin, while there is almost no change in the dispersion coefficient. The relative changes of the quantiles are given in Figure 12. Due to the increase of the location parameter the 2-year quantiles increase over the whole basin by 10–20%. The relative increase of these quantiles is, however, smaller than the relative increase of winter mean precipitation (Figure 2). For the 50-year quantiles the effect of the increase of the location parameter is counterbalanced by the decrease of the shape parameter resulting in only a slight and non-significant change of this quantile.

For the 5-day winter precipitation maxima the spatial pooling reduced the interquartile ranges of the parameter estimates from 17% for the trend in the location parameter up to 53% for the trend in the shape parameter. This reduction is lower than that for the 1-day summer maxima, which is due to the stronger spatial correlation between the 5-day winter precipitation maxima. In contrast to the summer maxima, the test for differences between regions indicates that for the 5-day winter precipitation maxima the trends in the GEV parameters can be assumed the same for the whole Rhine basin. However, the reduction of the uncertainty of the quantiles by fitting a model with common trend parameters ξ_1, γ_1, and κ_1 is not as large as that for the 1-day summer maxima. This is partly due to the larger correlation between the estimated parameters of different regions in winter and partly due to the fact that the uncertainty of the changes in quantiles is smaller in winter (compare the widths of the confidence bands in Figures 5 and 12).

5.2. Model validation

Figure 13 gives a summary of the goodness of fit testing for the winter season. As for the summer season the model was initially fitted to all grid boxes. Fifteen grid boxes with high values of A^2 were excluded. Most of these grid boxes are located on the border of region 1 or close to it, some of them at high altitude. Two excluded grid boxes are found on the border of region 4. After the exclusion of these grid boxes the model was refitted and the A^2 values were recalculated. The results discussed in section 5.1 refer to this refitted model. After refitting there remains one grid box with an A^2 value exceeding the global 0.1 critical value in region 2. The presence of decadal variability was checked in the same way as in the summer season. No significant signs of decadal variability were found by any of the approaches (not shown). However, there is a strong indication that the magnitude of the trend in the location parameter ξ_1 decreases with increasing altitude in the Swiss part of the Rhine basin (see Figure 14), showing that the assumption of a regionally constant ξ_1 could be too restrictive in regions with strong orography.

6. Conclusions

In the present study a non-stationary regional GEV model was introduced and applied to the 1-day summer and 5-day winter precipitation maxima in the transient RACMO-ECHAM5 run for the river Rhine basin in order to evaluate the changes in the properties of simulated precipitation extremes. The capability of the climate model to reproduce observed precipitation extremes was also assessed. The river Rhine basin was subdivided into 5 regions based on the grid box estimates of the dispersion coefficient for the RACMO-ECHAM5 data and the GEV model was applied to each of these regions. The model allows the location parameter to vary over the region of interest with common trend in time. The dispersion coefficient and the shape parameter are assumed constant over the region but varying with time.

The regional GEV model provides an informative summary of the differences between observed and simulated precipitation maxima as well as of the changes in the distribution of extremes. Looking at the parameters of the GEV distribution gives a better insight into the differences in distribution than looking at a single quantile only. In addition, the standard errors of the estimated common parameters are significantly reduced compared to the estimates based on the data of an individual grid box.

The values of estimated parameters in the period 1950–1990 for the 1-day summer precipitation extremes are reasonably well reproduced in the RACMO-ECHAM5 simulation. Part of the differences between the values from the ENS data can be ascribed to the low density of stations used for gridding. The distribution of the 5-day winter precipitation extremes is affected by strong positive biases in the number of wet days and winter mean precipitation. In particular, the dispersion coefficient of the GEV distribution is severely underestimated across the whole Rhine basin.

The changes of the distribution of the 1-day summer precipitation maxima are primarily related to the positive trend in the dispersion coefficient. Since there is almost no change in the location parameter, the changes in distribution are mainly found at large quantiles (e.g., the 50-year quantile) whereas there are only minor changes in quantiles close to median (i.e., the 2-year quantile). For the 5-day winter maxima the low quantiles (e.g., 2-year quantile) are increasing due to the increase of the location parameter. As the return period gets longer the effect of the positive trend in the location parameter is counterbalanced by the decrease of the shape parameter resulting in only minor positive or negative changes of large quantiles (e.g., the 50-year quantile).

The opposite direction of the changes in mean and 1-day maximum precipitation in summer is in agreement with earlier findings of *Christensen and Christensen* [2004] and *Frei et al.* [2006]. A relatively small change of the quantiles of extreme multi-day winter precipitation was also found by *Leander et al.* [2008] for the adjacent Meuse basin in a simulation of the RACMO model driven by the HadAM3H atmospheric model of the Hadley Centre. Despite a considerable increase in mean winter precipitation in this experiment there was little change in the distribution of the 10-day winter precipitation maxima and extreme river flows. The differences between changes in mean and extremes indicate that proportional adjustment of observed data can be very misleading.

Despite the reduction of standard errors due to spatial pooling of data, the changes in the quantiles of the extreme-value distributions are often not statistically significant. For the 2-year quantile of the 1-day summer maxima this can be attributed to the fact that the change in the location parameter is small. The estimates of the relative changes of the 50-year quantiles are strongly affected by the estimates of the dispersion coefficient and the shape parameter, which have large standard errors. For the summer season the uncertainty of the change in this quantile for regions 1, 3, 4, and 5 could be reduced considerably by assuming common trend parameters ξ_1, γ_1, and κ_1. The use of an ensemble of RACMO simulations driven by different simulations of the ECHAM5 global climate model is an option to improve the estimates of the changes in extreme value properties further.

The Anderson-Darling test shows that the model fits well for much of the Rhine basin. In the summer the model fails to fit in a relatively dry subregion with a relatively high dispersion coefficient and in a small relatively wet subregion. In the winter season the model did not fit well at a number of grid boxes on, or close to, the border of the Rhine basin, in particular in the Swiss part of the basin. As a consequence, a small number of grid boxes were excluded. A separate model fit using part of the excluded grid boxes suggests that formation of different, smaller regions could improve the goodness of fit, however, at the cost of increased uncertainty. Another possibility is the reformulation of the statistical model to allow

the dispersion coefficient to vary over the region of interest. In addition, for regions with strong orography it may be necessary to incorporate altitude-dependence of the trend in the location parameter.

Appendix A: Test for differences between regions

Let θ_i be one of the parameters ξ_1, γ_0, γ_1, κ_0, or κ_1 in the non-stationary GEV model for region *i* and let $\boldsymbol{\tau}$ be the vector of the other parameters. We want to test the hypothesis $\mathrm{H}_0 : \theta_1 = \theta_2 = \ldots = \theta_n$ using the statistic *R* in equation (14). The test consists of the following steps:

1. Calculate the value of the test statistic using equation (14) and denote this value *r*.
2. Calculate the standard Gumbel residuals using the $\hat{\theta}_i$ and the estimated values of the other parameters.
3. Re-estimate the other parameters $\hat{\boldsymbol{\tau}}_0$ given $\theta_1 = \theta_2 = \ldots = \theta_n = \overline{\theta}$.
4. Draw a bootstrap sample from the standard Gumbel residuals using resampling of years to preserve the spatial dependence structure (see section 2.3) and transform this sample back to the original scale using the parameter estimates $\overline{\theta}$ and $\hat{\boldsymbol{\tau}}_0$.
5. Re-estimate all parameters and re-calculate the test statistic as

$$r_b^* = \sum_{i=1}^{n} (\hat{\theta}_{b,i}^* - \overline{\theta}_b^*)^2 \,, \qquad \text{(A1)}$$

   with $\hat{\theta}_{b,i}^*$ the estimate of θ_i from bootstrap sample *b* and $\overline{\theta}_b^* = \sum_{i=1}^{n} \hat{\theta}_{b,i} / n$.
6. Repeat steps 4–5 until the desired number of bootstrap samples is obtained.

The *p*-value is the fraction of r_b^* values larger than *r*. The *p*-values in Table 2 are based on 500 bootstrap samples.

Appendix B: Determination of the critical values of the Anderson-Darling statistic

The critical values of the Anderson-Darling statistic A^2 in the literature usually refer to the situation of independent realizations from a distribution that is entirely specified under the null hypothesis. This does not apply to the standard Gumbel residuals $\widetilde{X}(s,t)$ at a given grid box, which are in fact dependent due to the use of estimated GEV parameters instead of their true but unknown values. It is well-known that parameter estimation has a substantial effect on the distribution of A^2 [e.g., *Laio*, 2004]. This appendix deals with the derivation of the local and global critical values of A^2 from bootstrap samples. The generation of these bootstrap samples is discussed in Appendix C. In our application *B* = 3000 bootstrap samples were generated.

Let *t*(*s*) be the value of A^2 from the climate model data at grid box *s* (*s* = 1, ..., *S*) and let $t_b^*(s)$ be the value of A^2 from bootstrap sample *b* (*b* = 1, ..., *B*) for this grid box.

For a chosen significance level α_{LOC}, the local critical values $c^{\alpha_{LOC}}(s)$ are obtained for each grid box as the *k*th largest value $t^*_{(k)}(s)$ of the $t^*_b(s)$, where $k = (1-\alpha_{LOC})(B+1)$.

The determination of the global critical values is based on an approach suggested by *Davison and Hinkley* [1997]. Let $c^{\alpha_{LOC}}_{-b}(s)$ be the local critical values that we get if we exclude bootstrap sample *b*. Then a bootstrap estimate of the global error rate α_{GLOB} is obtained as:

$$\alpha_{GLOB} = \frac{\#\{b : [t^*_b(s) \geq c^{\alpha_{LOC}}_{-b}(s), \quad \text{for any } s]\}}{B}, \tag{B1}$$

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

Appendix C: Comparison of two bootstrap procedures for goodness of fit testing

The determination of the critical values of the Anderson-Darling statistic A^2 requires simulation from the model under the null hypothesis. In particular, the preservation of spatial dependence is important. The bootstrap procedure outlined in section 2.3 to asses the uncertainty of the parameter estimates and quantiles is not appropriate for testing goodness of fit because the distribution of the $\tilde{X}(s,t)$ may deviate from the Gumbel distribution due to lack of fit of the GEV model and because of the occurrence of ties in the bootstrap samples. The latter influences the statistical properties of the empirical distribution function $F_N(x)$ in equation (12). In this appendix two alternatives are discussed:

- Replacement of resampled standard Gumbel residuals by samples from the standard Gumbel distribution, preserving the spatial structure of the ranks of the maxima as suggested by *Heffernan and Tawn* [2004]. This approach requires no assumptions about the underlying dependence structure of data.

- Sampling standard normal residuals from the multivariate normal distribution [*Hosking and Wallis*, 1997]. These residuals are assumed to be equicorrelated, i.e., the correlation $\rho_{i,j}$ between the residual at grid box *i* and the residual at grid box *j* equals $\rho_{i,j} = \rho$ for $i \neq j$ and $\rho_{i,j} = 1$ for $i = j$. In this case the multivariate normal dependence structure is introduced into the simulated samples.

In the following the procedures are referred to as "HT" and "MVN", respectively, and both are fully described below.

Bootstrap procedure based on the Heffernan and Tawn approach

1. Fit the statistical model to the original sample.
2. Calculate standard Gumbel residuals with the parameter estimates from step 1.
3. Bootstrap the residuals from step 2 (using resampling of years to preserve the spatial dependence as described in section 2.3).
4. Generate *S* independent samples of size *N* from the standard Gumbel distribution (*S* is the number of grid boxes and *N* the number of years).
5. Rearrange the values in the samples from step 4 such that the dependence structure of the ranks corresponds to that of the bootstrapped residuals from step 3.
6. Transform the rearranged standard Gumbel values from step 5 back to the original scale using the parameter estimates from step 1.
7. Fit the statistical model again.
8. Calculate standard Gumbel residuals with the parameter estimates from step 7 and calculate the A^2 statistics.
9. Repeat steps 3–8 until the desired number of bootstrap samples is obtained.

Parametric bootstrap procedure with sampling from the multivariate normal distribution

1. Fit the statistical model to the original sample.
2. Calculate standard normal residuals (see section 2.3) with the parameter estimates from step 1.
3. Calculate the average correlation $\hat{\rho}$ of the standard normal residuals.
4. Generate a sample of *S* equicorrelated standard normal variables with correlation $\hat{\rho}$.
5. Transform the sample from step 4 back to the original scale using the parameter estimates from step 1.
6. Fit the statistical model again.
7. Calculate standard Gumbel residuals with the parameter estimates from step 6 and calculate the A^2 statistics.
8. Repeat steps 4–7 until the desired number of bootstrap samples is obtained.

A simulation experiment was conducted to assess the validity of both approaches: 3000 samples of size 150 from an equicorrelated 30-dimensional normal distribution with known correlation were generated (think about 30 grid boxes in the RACMO-ECHAM5 simulation which has a length of 150 years). These samples (further denoted as control samples) were transformed according to the non-stationary GEV model

$$\xi(s,t) = \xi_0(s)\exp[\xi_1(t-40)_+] \quad \text{(C1)}$$

$$\gamma(t) = \exp[\gamma_0 + \gamma_1(t-40)_+] \quad \text{(C2)}$$

$$\kappa(t) = \kappa_0 + \kappa_1(t-40)_+ \quad \text{(C3)}$$

with $s = 1, \ldots, 30$; $t = 1, \ldots, 150$, and $(x)_+ = \max(x, 0)$. The values of the parameters were set to be representative of those obtained for the 1-day summer maximum precipitation in the Rhine basin, i.e., $\xi_0(s)$ ranged between 22 and 38, $\xi_1 = 0.00055$, $\exp(\gamma_0) = 0.37$, $\gamma_1 = 0.0013$, $\kappa_0 = 0.05$, and $\kappa_1 = 0.00015$.

For each sample the parameters of the GEV model were estimated and the values of the A^2 statistics were calculated. The 0.1 critical value from these simulations is denoted the "true" critical value. Further, for one of the control samples two sets of 3000 bootstrap samples were generated using the "HT" and "MVN" approaches, respectively, and the 0.1 local and global critical values of the A^2 statistic were calculated according to Appendix B.

Table C1 gives the local rejection rates of the null hypothesis as obtained from the control samples, i.e., the proportion of the A^2 values of these samples lying above the "HT" and "MVN" critical values. For the "MVN" critical values the rejection rate corresponds quite well with the nominal 0.1 significance level, but for the "HT" critical values the actual rejection rate is lower than 0.1 in the case of correlation and the difference grows with increasing correlation coefficient. Table C1 further shows that the "MVN" critical values resemble the "true" critical values and decrease with increasing correlation. By contrast the "HT" critical values do not depend on correlation. Though Table C1 refers to the local rejection rates and the local critical values, very similar results were obtained for the global test at the 0.1 significance level.

To understand why the critical values of the A^2 statistic are decreasing with increasing correlation, we have to examine how the estimates of the parameters are influenced by the data from a particular grid box. The estimate of $\xi_0(s)$ is largely determined by the maxima of the grid box of interest. If there is no or little correlation, the maxima of this grid box have little influence on the estimates of the other parameters γ_0, κ_0, ξ_1, γ_1, and κ_1. The influence of the maxima of the grid box of interest on the estimates of these parameters grows with increasing spatial correlation. As a result the fitted regional GEV model will describe the local maxima better and therefore the critical value of the A^2 statistic should be smaller than in the case of independence. The "MVN" and "true" critical values for $\rho = 0.99$ are close to the critical value for the case that all six parameters are estimated from the maxima at the grid box of interest only.

The reason of the failure of the "HT" approach in the case of goodness of fit testing is that the test statistic is insensitive to a permutation of the data, i.e., rearranging residuals at a grid box to preserve the spatial dependence of the ranks does not influence the value of the A^2 statistic. Unlike the "MVN" bootstrap samples, the values of the A^2 statistic do not exhibit any spatial correlation in the "HT" bootstrap samples. Although the "HT" approach is not suitable for goodness of fit testing, it can be used for the estimation of standard errors and the construction of confidence intervals, for which it was originally introduced by *Heffernan and Tawn* [2004].

It is not surprising that the "MVN" critical values do quite well because of the underlying multivariate normal dependence structure of the data. To study the robustness to the type of association at extreme levels, 3000 new samples were generated from our non-stationary GEV model but now with a dependence structure of a limiting extreme-value distribution. This was achieved by generating the standard Gumbel residuals from an equicorrelated multivariate Gumbel

distribution as described by *Stephenson* [2003]. The results (not shown) are very similar to those presented in Table C1 for a multivariate normal dependence structure from which it may be concluded that the "MVN" critical values are robust to the dependence structure.

Acknowledgments. We acknowledge the ENSEMBLES project, funded by the European Commission's 6th Framework Programme through contract GOCE-CT-2003-505539. The Alpine data set was kindly provided by MeteoSwiss.

References

Adlouni, S. El., T. B. M. J. Ouarda, X. Zhang, R. Roy, and B. Bobée (2007), Generalized maximum likelihood estimators for the nonstationary generalized extreme value model, *Water Resour. Res.*, *43*, W03410, doi:10.1029/2005WR004545.

Alila, Y. (1999), A hierarchical approach for the regionalization of precipitation annual maxima in Canada, *J. Geophys. Res.*, *104*(D24), 31,645–31,655.

Anderson, T. W., and D. A. Darling (1952), Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes, *Ann. Math. Stat.*, *23*(2), 193–212, doi:10.1214/aoms/1177729437.

Arnell, N. W., and S. Gabriele (1988), The performance of the two-component extreme value distribution in regional frequency analysis, *Water Resour. Res.*, *24*(6), 879–887.

Brath A., A. Castellarin, and A. Montanari (2003), Assessing the reliability of regional depth-duration-frequency equations for gaged and ungaged sites, *Water Resour. Res.*, *39*(12), 1367, doi:10.1029/2003WR002399.

Brohan, P., J. J. Kennedy, I. Harris, S. F. B. Tett, and P. D. Jones (2006), Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850, *J. Geophys. Res.*, *111*, D12106, doi:10.1029/2005JD006548.

Brown, S. J., J. Caesar, and C. A. T. Ferro (2008), Global changes in extreme daily temperature since 1950, *J. Geophys. Res.*, *113*, D05115, doi:10.1029/2006JD008091.

Buishand, T. A. (1991), Extreme rainfall estimation by combining data from several sites, *Hydrological Sciences Journal*, *36*(4), 345–365.

Buonomo, E., R. Jones, C. Huntingford, and J. Hannaford (2007), On the robustness of changes in extreme precipitation over Europe from two high resolution climate change simulations, *Q. J. R. Meteorol. Soc.*, *133*, 65–81, doi:10.1002/qj.13.

Christensen, O. B., and J. H. Christensen (2004), Intensification of extreme European summer precipitation in a warmer climate, *Global Planet. Change*, *44*, 107–117, doi: 10.1016/j.gloplacha.2004.06.013.

Cleveland, W. S. (1979), Robust locally weighted regression and smoothing scatterplots, *J. Am. Stat. Assoc.*, *74*, 829–836.

Coles, S. (2001), *An Introduction to Statistical Modeling of Extreme Values*, Springer-Verlag, New York.

Cunderlik, J. M., and D. H. Burn (2003), Non-stationary pooled flood frequency analysis, *J. Hydrol.*, *276*, 210–223.

Cunderlik, J. M., and T. B. M. J. Ouarda (2006), Regional flood-duration-frequency modeling in the changing environment, *J. Hydrol.*, *318*, 276–291.

Davison, A. C., and D. V. Hinkley (1997), *Bootstrap methods and their application*, Cambridge University Press.

Durman, C. F., J. M. Gregory, D. C. Hassell, R. G. Jones, and J. M. Murphy (2001), A comparison of extreme European daily precipitation simulated by a global and regional climate model for present and future climates, *Q. J. R. Meteorol. Soc.*, *127*, 1005–1015.

Ekström, M., H. J. Fowler, C. G. Kilsby, and P. D. Jones (2005), New estimates of future changes in extreme rainfall across the UK using regional climate model integrations. 2. Future estimates and use in impact studies, *J. Hydrol.*, *300*, 234–251, doi:10.1016/j.jhydrol.2004.06.019.

Faulkner, D. S., and D. A. Jones (1999), The FORGEX method of rainfall growth estimation III: Examples and confidence intervals. *Hydrol. Earth Syst. Sci.*, *3*, 205–212.

Fowler, H. J., M. Ekström, C. G. Kilsby, and P. D. Jones (2005), New estimates of future changes in extreme rainfall across the UK using regional climate model integrations, 1. Assessment of control climate, *J. Hydrol.*, *300*, 212–233, doi:10.1016/j.jhydrol.2004.06.017.

Frei C., and C. Schär (1998), A precipitation climatology of the Alps from high-resolution rain-gauge observations, *Int. J. Climatol.*, *18*(8), 873–900, doi: 10.1002/(SICI)1097-0088(19980630)18:8<873::AID-JOC255>3.0.CO;2-9.

Frei, C., and C. Schär (2001), Detection probability of trends in rare events: Theory and application to heavy precipitation in the Alpine region, *J. Clim.*, *14*, 1568–1584.

Frei, C., R. Schöll, S. Fukutome, J. Schmidli, and P. L. Vidale (2006), Future change of precipitation extremes in Europe: Intercomparison of scenarios from regional climate models, *J. Geophys. Res.*, *111*, D06105, doi:10.1029/2005JD005965.

García, J. A., M. C. Gallego, A. Serrano, and J. M. Vaquero (2007), Trends in block-seasonal extreme rainfall over the Iberian Peninsula in the second half of the twentieth century, *J. Clim.*, *20*, 113–130, doi:10.1175/JCLI3995.1.

Goubanova, K., and L. Li (2007), Extremes in temperature and precipitation around the Mediterranean basin in an ensemble of future climate scenario simulations, *Global Planet. Change*, *57*, 27–42, doi:10.1016/j.gloplacha.2004.06.010.

Haylock, M. R., N. Hofstra, A. M. G. Klein Tank, E. J. Klok, P. D. Jones, and M. New (2008), A European daily high-resolution gridded dataset of surface temperature and precipitation, *J. Geophys. Res.*, *113*, D20119, doi:10.1029/2008JD010201.

Heffernan, J. E., and J. A. Tawn (2004), A conditional approach for multivariate extreme values (with discussion), *J. R. Stat. Soc. B*, *66*(3), 497–546, doi:10.1111/j.1467-9868.2004.02050.x.

Hosking, J. R. M., and J. R. Wallis (1997), *Regional frequency analysis*, Cambridge University Press.

Huntingford, C., R. G. Jones, C. Prudhomme, R. Lamb, J. H. C. Gash, and D. A. Jones (2003), Regional climate model predictions of extreme rainfall for a changing climate, *Q. J. R. Meteorol. Soc.*, *129*, 1607–1621, doi:10.1256/qj.02.97.

Kendon, E. J., R. G. Jones, and E. Buonomo, (2008), Robustness of future changes in local precipitation extremes, *J. Clim.*, *21*, 4280–4297, doi:10.1175/2008JCLI2082.1.

Kharin, V. V., and F. W. Zwiers, (2005), Estimating extremes in transient climate change simulations, *J. Clim.*, *18*, 1156–1173, doi:10.1175/JCLI3320.1.

Kharin, V. V., F. W. Zwiers, X. Zhang, and G. C. Hegerl (2007), Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations. *J. Clim.*, *20*, 1419–1444, doi:10.1175/JCLI4066.1.

Kyselý, J., and J. Picek (2007), Regional growth curves and improved design value estimates of extreme precipitation events in the Czech Republic, *Clim. Res.*, *33*, 243–255.

Laio, F. (2004), Cramer-von Mises and Anderson-Darling goodness of fit tests for extreme value distributions with unknown parameters, *Water Resour. Res.*, *40*, W09308, doi:10.1029/2004WR003204.

Leander, R., T. A. Buishand, B. J. J. M. van den Hurk, and M. J. M. de Wit (2008), Estimated changes in flood quantiles of the river Meuse from resampling of regional climate model output, *J. Hydrol.*, *351*, 331–343, doi:10.1016/j.jhydrol.2007.12.020.

Lenderink, G., and E. van Meijgaard (2008), Increase in hourly precipitation extremes beyond expectations from temperature changes, *Nature Geoscience*, *1*, 511–514, doi:10.1038/ngeo262.

Northrop, P. J. (2004), Likelihood-based approaches to flood frequency estimation, *J. Hydrol.*, *292*, 96–113, doi:10.1016/j.jhydrol.2003.12.031.

Pall, P., M. R. Allen, and D. A. Stone (2006), Testing the Clausius-Clapeyron constraint on changes in extreme precipitation under CO_2 warming, *Clim. Dyn.*, *28*, 351–363, doi:10.1007/s00382-006-0180-2.

Renard, B., V. Garreta, and M. Lang (2006), An application of Bayesian analysis and Markov chain Monte Carlo methods to the estimation of a regional trend in annual maxima, *Water Resour. Res.*, *42*, W12422, doi:10.1029/2005WR004591.

Schaefer, M. G. (1990), Regional analyses of precipitation annual maxima in Washington State, *Water Resour. Res.*, *26*(1), 119–131.

Sen, P. K. (1968), Estimation of the regression coefficient based on Kendall's tau, *J. Am. Stat. Assoc.*, *63*, 1379–1389.

Shimokawa, T., and M. Liao (1999), Goodness of fit tests for type-1 extreme-value and 2-parameter Weibull distributions, *IEEE Trans. Reliab.*, *48*(1), 79–86.

Stephenson, A. (2003), Simulating multivariate extreme value distributions of logistic type, *Extremes*, *6*, 49–59.

Sveinsson, O. G. B., D. C. Boes, and J. D. Salas (2001), Population index flood method for regional frequency analysis, *Water Resour. Res.*, *37*(11), 2733–2748.

List of Figures

boxes in region 1 (black dots) are smoothed by locally weighted regression "loess" (black line).

Table 1. Reduction (%) of interquartile ranges of the parameter estimates due to spatial pooling for the summer (JJA) in the case of the RACMO-ECHAM5 data.

	region					
parameter	1	2	3	4	5	mean
ξ_1	37	32	34	31	39	35
γ_0	58	45	48	53	44	50
γ_1	67	60	60	61	58	61
κ_0	73	71	67	66	58	67
κ_1	80	75	75	72	66	74

Table 2. The *p*-values resulting from the test for differences between regions for the summer (JJA).

parameter	*p*-value
ξ_1	0.01
γ_0	0.00
γ_1	0.13
κ_0	0.00
κ_1	0.00

Table C1. Local rejection rates and critical values (nominal significance level of 0.1) for testing goodness of fit using the Anderson-Darling statistic. The "true" critical values are based on 3000 simulated samples from a non-stationary GEV model, the critical values "HT" and "MVN" are based on 3000 bootstrap samples from one of these simulations using respectively the Heffernan and Tawn approach and a multivariate normal distribution to preserve spatial dependence.

	rejection rate		critical value		
correlation	"HT"	"MVN"	"HT"	"MVN"	"true"
0.00	0.098	0.102	0.881	0.870	0.875
0.40	0.077	0.095	0.888	0.837	0.823
0.60	0.050	0.088	0.901	0.778	0.751
0.80	0.025	0.081	0.892	0.686	0.648
0.99	0.000	0.093	0.905	0.523	0.514

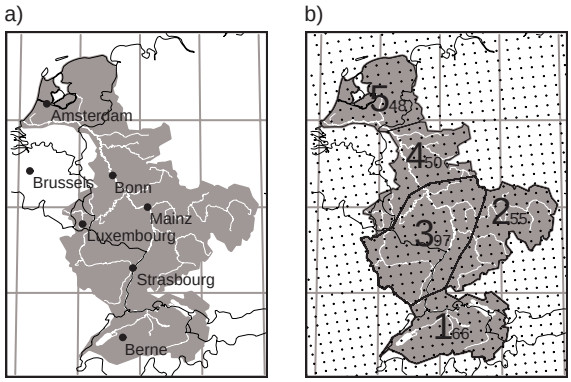

Figure 1. (a) The river Rhine basin. (b) Subdivision of the river Rhine basin into five regions. The numbers in subscript give the number of grid boxes included in the region. Dots represent the centers of the RACMO model grid boxes.

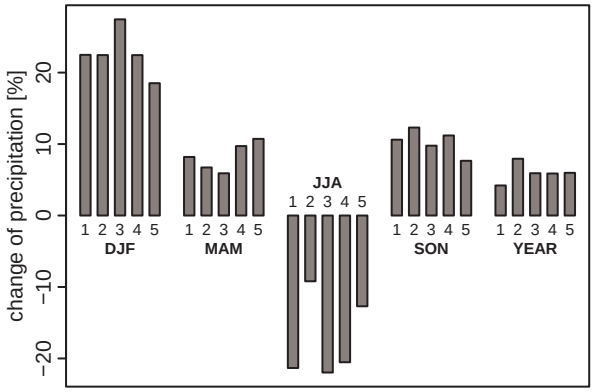

Figure 2. Relative change of the seasonal and annual mean precipitation between the periods 1950–1990 and 2070–2100 for all 5 regions of the Rhine basin.

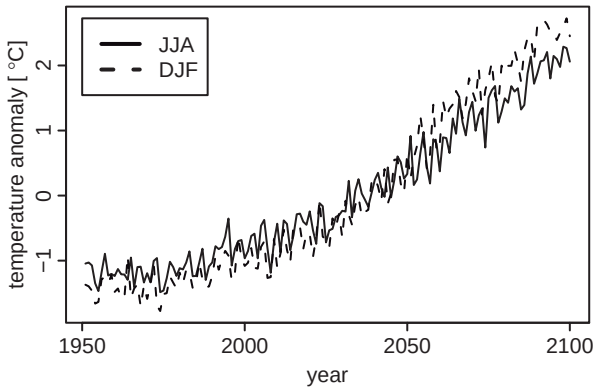

Figure 3. Summer (JJA) and winter (DJF) global mean temperature anomalies in the ECHAM5 simulation.

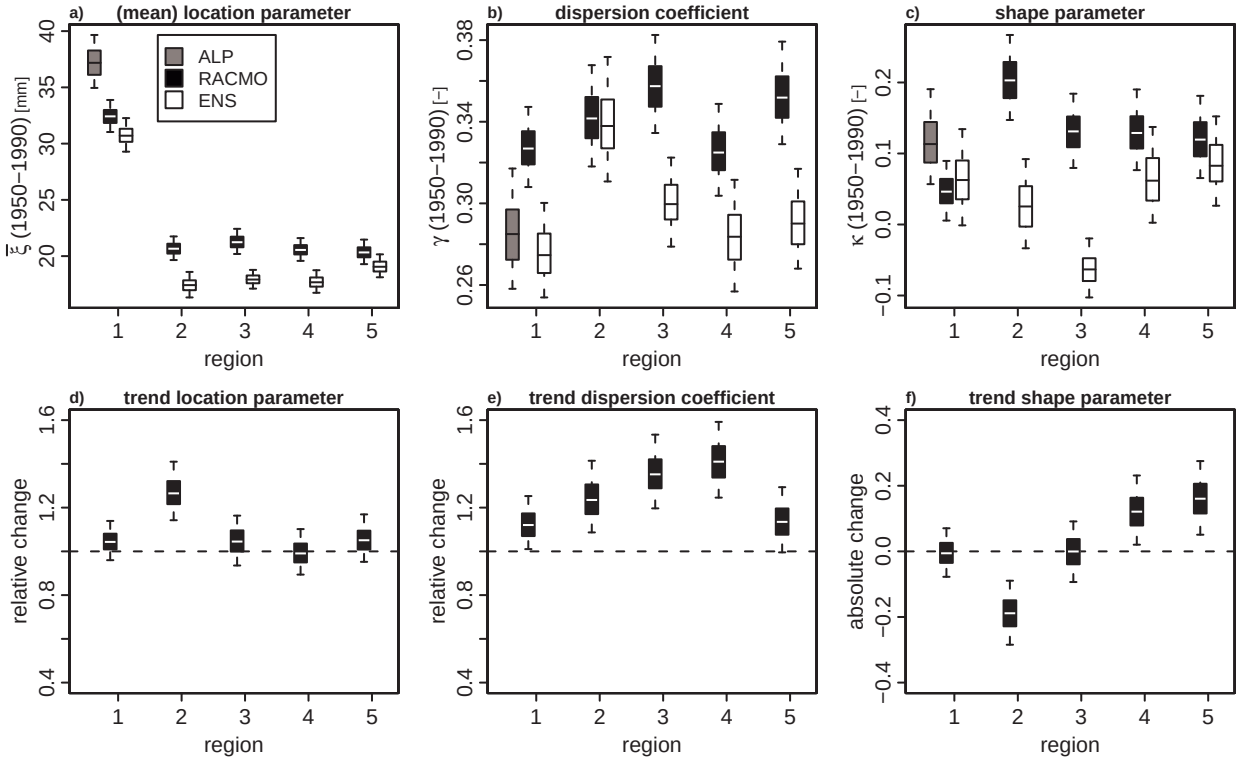

Figure 4. (a–c) Estimates of the GEV parameters for the 1-day summer (JJA) precipitation extremes for the period 1950–1990, averaged over the region in the case of the location parameter. (d–f) The changes of the GEV parameters for the 1-day summer (JJA) precipitation extremes between the periods 1950–1990 and 2070–2100. The boxplots were obtained from 3000 bootstrap samples. The boxes represent the interquartile range, the whiskers extend from the 5th to the 95th percentile of these bootstrap samples.

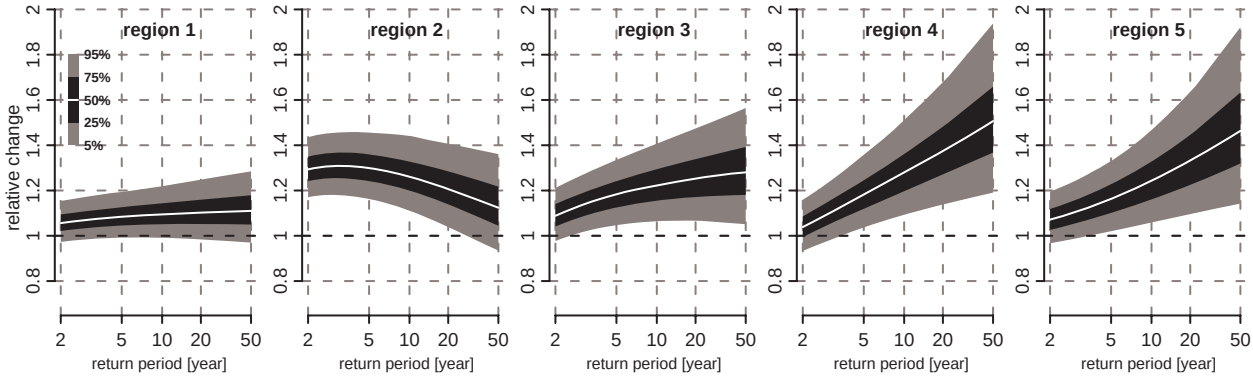

Figure 5. Relative changes of quantiles of the 1-day summer maximum precipitation between the periods 1950–1990 and 2070–2100 for all five regions. The confidence bands were obtained from 3000 bootstrap samples. The 5th, 25th, 50th, 75th and 90th percentile of these bootstrap samples are shown.

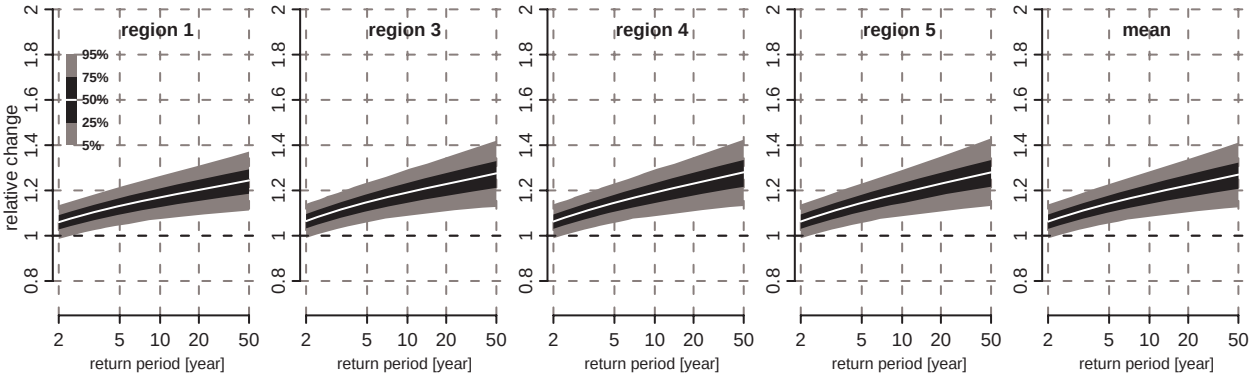

Figure 6. Same as Figure 5 but for the restricted model with common trends over regions 1, 3, 4 and 5. The panel on the right gives the average relative change of the four regions together with the average confidence band.

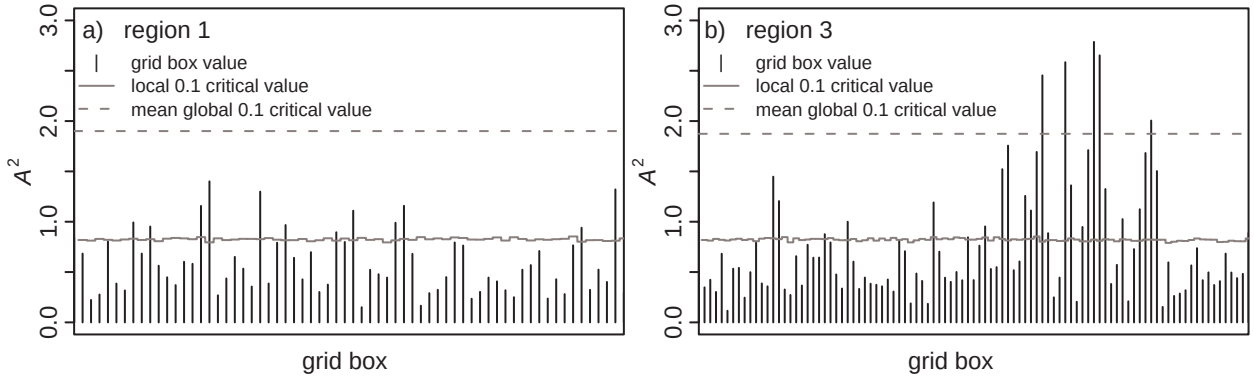

Figure 7. The values of the Anderson-Darling statistic for (a) region 1 and (b) region 3 for the 1-day summer (JJA) precipitation extremes.

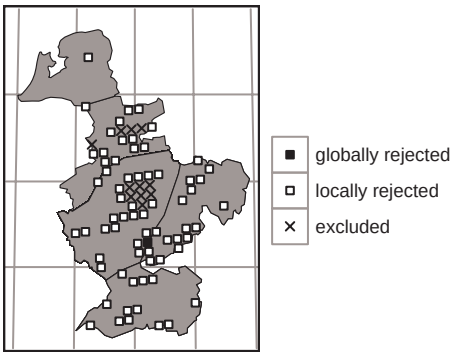

Figure 8. Summary of the goodness of fit testing of the non-stationary GEV model for the 1-day summer (JJA) precipitation extremes.

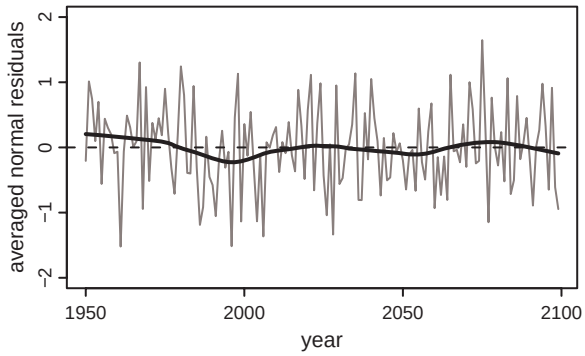

Figure 9. Averaged standard normal residuals (gray line) for the 1-day summer (JJA) precipitation extremes in region 1. The black line shows residuals smoothed by locally weighted regression "loess".

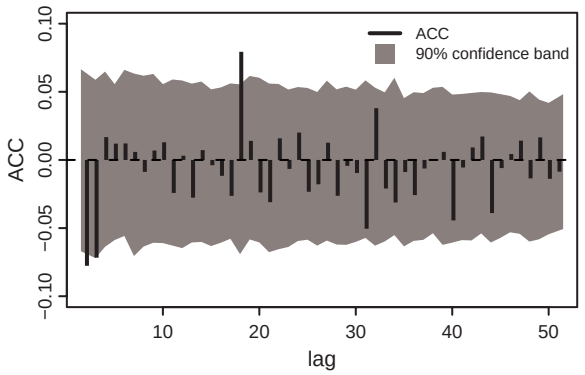

Figure 10. Average autocorrelation coefficients (ACC) of the standard normal residuals (vertical bars) for the 1-day summer (JJA) precipitation extremes in region 1. The 90% confidence band (shaded area) was obtained from 3000 bootstrap samples.

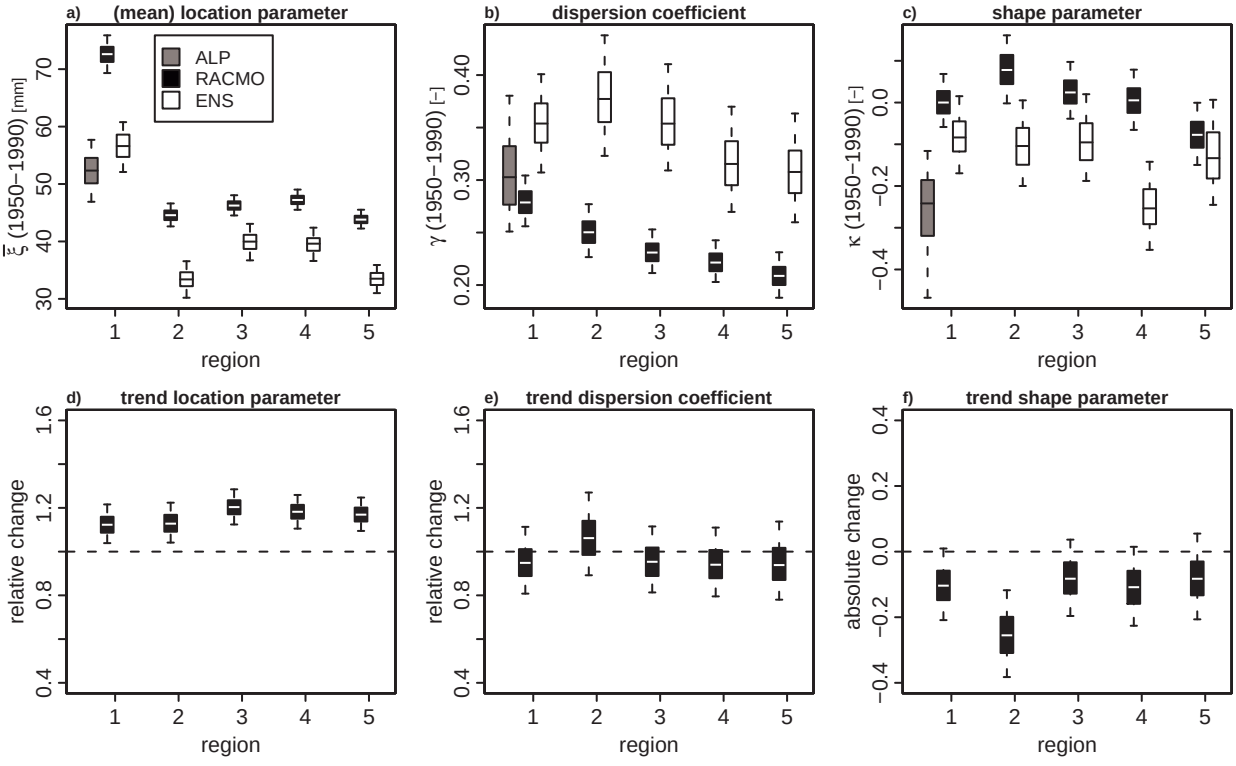

Figure 11. Same as Figure 4 but for the 5-day winter (DJF) precipitation extremes.

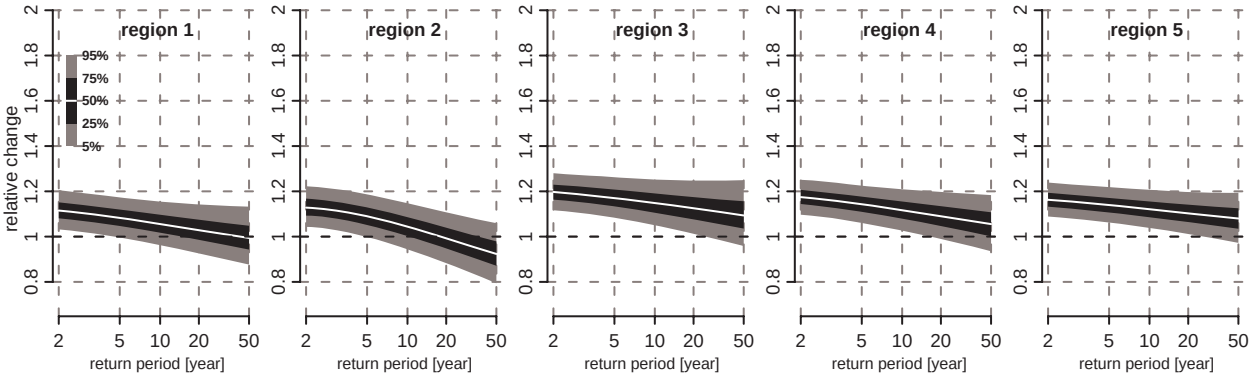

Figure 12. Same as Figure 5 but for the 5-day winter (DJF) precipitation extremes.

Figure 13. Same as Figure 8 but for the 5-day winter (DJF) precipitation extremes.

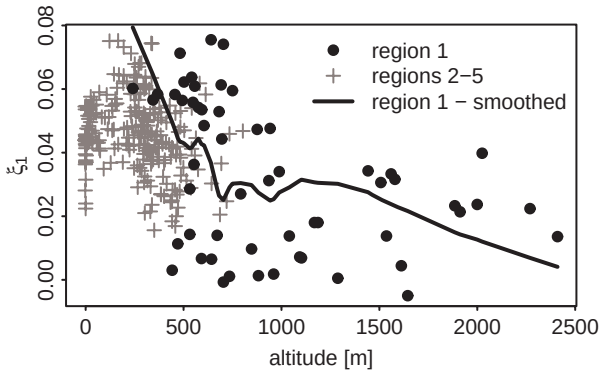

Figure 14. Grid box estimates of the trend in the location parameter as a function of altitude for the 5-day winter (DJF) precipitation maxima. The values for the grid boxes in region 1 (black dots) are smoothed by locally weighted regression "loess" (black line).