Meteorologische Zeitschrift, Vol. **25**, No. 5, 543–562 (published online August 9, 2016)

Impact of the horizontal resolution on the simulation of extremes in COSMO-CLM

Oliver Gutjahr[1*], Lukas Schefczyk[1], Philipp Reiter[2,3] and Günther Heinemann[1]

[1]Department of Environmental Meteorology, University of Trier, Germany
[2]Rhineland-Palatinate Centre of Excellence for Climate Change Impacts, Trippstadt, Germany
[3]Department of Physical Geography, University of Trier, Germany

(Manuscript received July 31, 2014; in revised form August 6, 2015; accepted March 11, 2016)

Abstract

The simulation of extremes using climate models is still a challenging task. Currently, the model grid horizontal resolution of state-of-the art regional climate models (RCMs) is about 11–25 km, which may still be too coarse to represent local extremes realistically. In this study we use dynamically downscaled ERA-40 reanalysis data of the RCM COSMO-CLM at 18 km resolution, downscale it dynamically further to 4.5 km and finally to 1.3 km to investigate the impact of the horizontal resolution on extremes. Extremes are estimated as return levels for the 2, 5 and 10-year return periods using 'peaks-over-threshold' (POT) models. Daily return levels are calculated for precipitation and maximum 2 m temperature in summer as well as precipitation and 2 m minimum temperature in winter. The results show that CCLM is able to capture the spatial and temporal structure of the observed extremes, except for summer precipitation extremes. Furthermore, the spatial variability of the return levels increases with resolution. This effect is more distinct in case of temperature extremes due to a higher correlation with the better resolved orography. This dependency increases with increasing horizontal resolution. In comparison to observations, the spatial variability of temperature extremes is better simulated at a resolution of 1.3 km, but the return levels are cold-biased in summer and warm-biased in winter. Regarding precipitation, the spatial variability improves as well, although the return levels were slightly overestimated in summer by all CCLM simulations. In summary, the results indicate that an increase of the horizontal resolution of CCLM does have a significant effect on the simulation of extremes and that impact models and assessment studies may benefit from such high-resolution model output.

Keywords: Extreme values, peaks over threshold, return level, GPD, COSMO-CLM, precipitation, temperature

1 Introduction

The severe impact which extremes may have on social, ecological or natural systems in a future climate (Ghil et al., 2011) was emphasized in the Special Report on Extreme Events (SREX) of the Intergovernmental Panel on Climate Change (IPCC, 2012). Because of these tremendous consequences, the study of extremes under climate change is an ongoing field of research.

The primary tool for such studies are climate models because they provide continuous spatial and temporal fields of e.g. temperature, wind and precipitation and are based on physical laws. With regional climate models (RCMs) it is possible to dynamically downscale output from global climate simulations or reanalysis products and they have proven to be important tools in climate sciences to add further details and complexity (Rummukainen, 2010; Prein et al., 2013). Current state-of-the-art RCMs are run with horizontal resolutions of about ~ 11–25 km as in the EURO-CORDEX project (Jacob et al., 2014), which is part of the global CORDEX framework (Giorgi et al., 2009), the newest successor of the PRUDENCE (50 km) and ENSEMBLES (25 km) projects. Over the last decade, large efforts have been made to improve the resolution of RCMs. With increasing computing power, higher resolutions became feasible (Prein et al., 2015), e.g. 7 km (Feldmann et al., 2010), 4.5 km (Gutjahr and Heinemann, 2013) and e.g. 2.8 km (Prein et al., 2013). According to Weisman et al. (1997), the threshold for the horizontal resolution classifying RCMs as convection-resolving is about 4 km. Individual studies at such high resolutions could be performed for small areas approaching the 1 km scale (e.g. Kendon et al., 2012; Knote et al., 2010). Yet, the resources are not sufficient to allow long-term simulations at such high resolutions. However, such resolutions are required by impact assessment models (Seo et al., 2009).

Increasing the horizontal resolution is expected to improve the simulation of near-surface fields, such as temperature, due to a better resolved orography, and precipitation, due to the better representation of physical processes (Giorgi, 2006; Hohenegger et al., 2008), such as explicit treatment of deep convection by turning off the convective parameterization scheme. Thereby

*Corresponding author: Oliver Gutjahr, Department of Environmental Meteorology, University of Trier, Behringstrasse 21, 54286 Trier, Germany, e-mail: gutjahr@uni-trier.de

the physical structure of organized convection in the warm season is expected to improve. It could be demonstrated that RCMs at the convection-permitting scale improve precipitation patterns (Boberg et al., 2010) and the intensity distribution (Prein et al., 2013) due to improved physical processes in summer. Summer temperature fields could be improved because of a better representation of the orography (Prein et al., 2013). Prein et al. (2013) investigated furthermore the effect of horizontal resolution of a RCM on the simulation of extremes. They found that convection-resolving RCM simulations with about 3 km resolution improve the hourly temperature extremes over the Alps. Knote et al. (2010) showed, for the same model domain as within this study, that temperature extremes are dependent on the height of orography. However, they focused only on the summer season. Tölle et al. (2014) already applied successfully the RCM used within this study at 1.3 km resolution based on our configuration.

In this study we further investigate whether extremes are better simulated by increasing the horizontal resolution systematically. Therefore we performed simulations in a multiple dynamical downscaling strategy from 18 km over 4.5 km down to a convection permitting resolution of 1.3 km. Compared to the previous study of Knote et al. (2010) in this geographic area, we extend the analysis of extremes to the winter season as well and compare the extreme values with observations.

This paper is organized as follows: Section 2 describes the configurations of the regional climate model and the model domain. The methodology of estimating extreme values of certain return periods is described in Section 3. The impacts of the horizontal resolution on these extreme values are shown in Section 4 and discussed in Section 5.

2 Model configuration, domain and data sets

2.1 The regional climate model COSMO-CLM

The study is performed with the regional climate model COSMO-CLM (CCLM; version 4.8_clm11), which is the climate version of the weather prediction model COSMO ("COnsortium of Small-scale MOdelling") coupled to the multi-layer soil model TERRA-ML (Schrodin and Heise, 2001; Heise et al., 2006; Doms et al., 2011). The CCLM evolved from the former "Lokal-Modell" (LM) of the German Weather Service (DWD) (Steppeler et al., 2003) and is applied and further developed by the Climate Limited-area Modelling (CLM) Community (Rockel et al., 2008). CCLM is a non-hydrostatic limited-area atmospheric prediction model developed for applications on the meso-β and meso-γ scale (Schättler et al., 2009) and relies on the primitive thermo-dynamical equations describing compressible flow in a moist atmosphere.

The global data were dynamically downscaled in a multi-nesting chain approach. The starting point of this nesting chain was the ERA-40 reanalysis data set (Uppala et al., 2005) and the first nest of CCLM was performed by Keuler et al. (2012) at a horizontal resolution of 18 km (CCLM18). The second nest was done for Mid-Europe (Figure 1a) with a resolution of 4.5 km (CCLM4.5) within this study and a similar downscaling of the climate projection was done by Gutjahr and Heinemann (2013). The final nest at 1.3 km (CCLM1.3) was done within this study for western Germany, parts of France and Belgium (Figure 1c).

In this paper, we performed 10-year long hindcasts of the present climate (1991–2000, ERA-40). In addition, the simulation period of CCLM4.5 was extended until 2010, so that 18 years were available. The overlapping time period of all models with the observational data sets is the period 1993–2000. Thus we restricted the analysis of return levels to these eight years, but compared the estimates from eight years and 18 years in case of CCLM4.5 to assess the influence of the time series length on the estimation of return levels.

Return levels were estimated from the three CCLM simulations and two observational data sets for daily precipitation (PR) and 2 m maximum (TX) in summer (June-July-August, JJA) and PR and 2 m minimum temperature (TN) in winter (December-January-February, DJF).

2.1.1 Study area

The investigation area was Rhineland-Palatinate (RLP) and Saarland in western Germany and its adjacent areas Lorraine (France) and Luxembourg (Saar-Lor-Lux) (Figure 1a, blue box and in detail in Figure 1b–d). This model domain is in particular interesting for this study since it is characterized by variations of the topography at short distances; with flat areas along the rivers Rhine and Moselle and higher orography in adjacent areas. The highest elevation in this domain is the Erbeskopf (Hunsrück) with 816 m a.s.l. The climatic conditions change as well within short distances. The Rhine and Moselle Valley are among the warmest and driest regions in Germany, whereas the climate of Hunsrück, Eifel and Westerwald is wet and rough (MUFV, 2007). The climatic mean 2 m temperature (1981–2010) of RLP is 9.2 °C with a mean annual precipitation of 825 mm/year (DWD). The mean temperature in summer (JJA) is 17.2 °C with an average precipitation amount of 207 mm and in winter (DJF) 1.4 °C and 216 mm, respectively. The dominant precipitation type is frontal or orographic rainfall in winter, whereas in summer convection becomes an important process for precipitation generation.

2.1.2 Configurations of the models

A Runge-Kutta scheme of order 3 (Wicker and Skamarock, 2002) combined with a 5th order advection scheme (Doms and Baldauf, 2015) is used to solve the

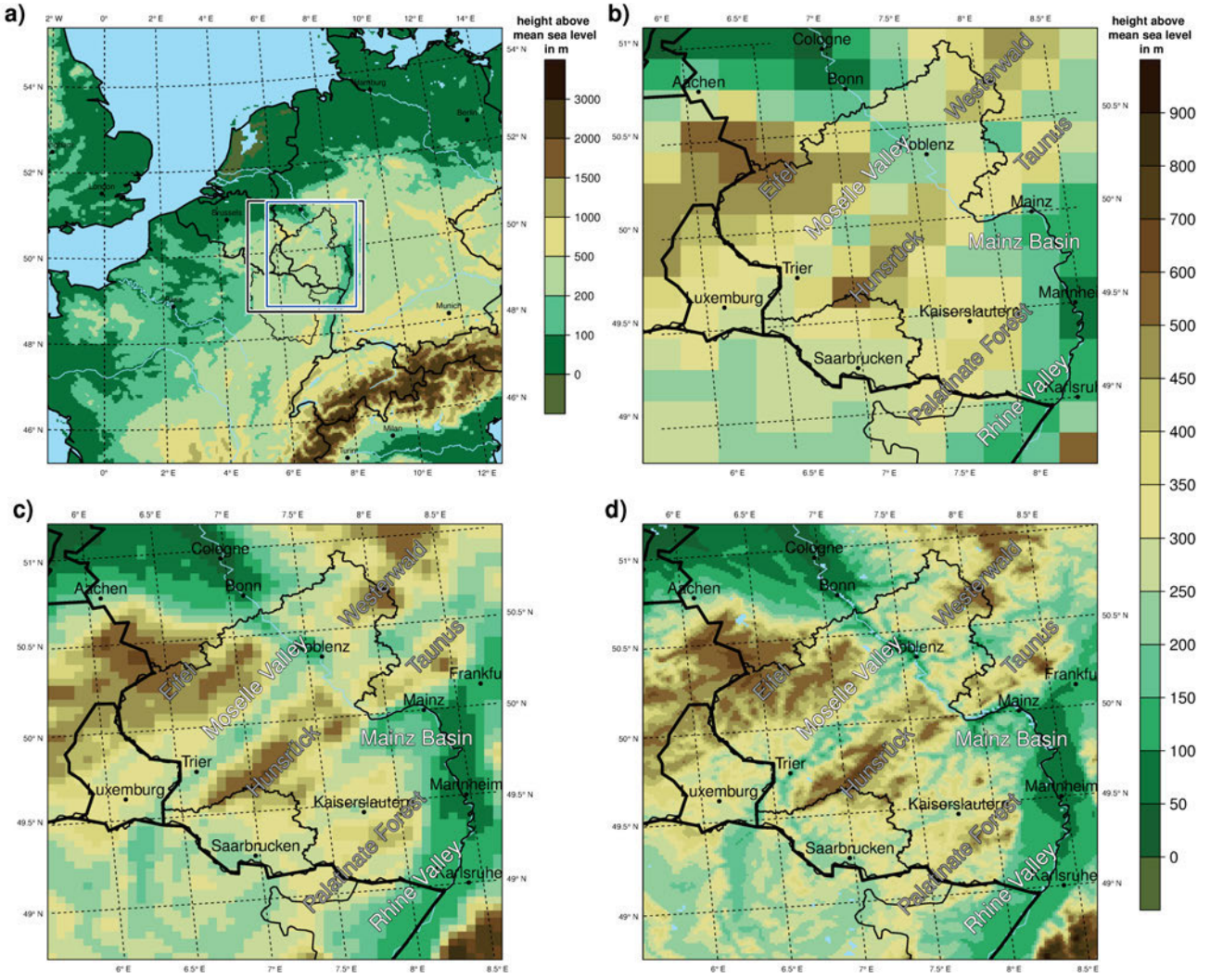

Figure 1: Model domain and height of orography of a) CCLM4.5 (254 × 254 grid boxes). The black box marks the domain of the CCLM1.3 model and the blue box is the investigation area. In b) the subdomain of Rhineland-Palatinate and the Saar-Lor-Lux region for CCLM18 (12 × 14 grid boxes) is shown, in c) for CCLM4.5 (51 × 62 grid boxes) and in d) for CCLM1.3 (172 × 208 grid boxes).

thermo-dynamic equations in addition with a diagnostic TKE-closure of level 2.0 (Mellor and Yamada, 1974). Thereby, three grid boxes were used for the lateral nudging. Prognostic precipitation is calculated from a bulk-formulation of a three-category-ice scheme, including cloud ice, snow and graupel based on Kessler (1969). For the radiation transfer the scheme of Ritter and Geleyn (1992) is called every hour in the simulation. The TERRA-ML soil model was used with 10 soil layers and a maximum depth of 14.58 m. In the coarser models (CCLM18, CCLM4.5) the Tiedtke-scheme is used (Tiedtke, 1989), whereas in CCLM1.3 only the shallow convection scheme of the Tiedtke parameterization was active. No spectral nudging was used in all three simulations.

CCLM4.5 was configured with 254 × 254 horizontal grid boxes and a time step of 45 s at 4.5 km resolution (Figure 1a) and CCLM1.3 with 220×220 horizontal grid boxes and a time step of 12 s at 1.3 km.

CCLM18 was originally defined with a slightly different rotated North Pole by Keuler et al. (2012), so that we performed a bilinear interpolation for precipitation and temperature onto a 18 km grid for the RLP & Saar-Lor-Lux subdomain (Figure 1b) in the rotated system of CCLM4.5 and CCLM1.3. From the CCLM18 of Keuler et al. (2012) we only use the data of the period 1991–2000.

CCLM uses a staggered Arakawa-C/Lorenz grid in a rotated geographical coordinate system with terrain-following vertical coordinates (sigma-levels). All three CCLM models were run with 40 vertical layers, whereof 14 are below 2 km height. Although recent RCMs are configured with > 40 layers, the decision for our configuration was a trade-off between the number of levels and the available computing time. The number of vertical levels in the lowest 2 km is considered to capture boundary layer processes sufficiently.

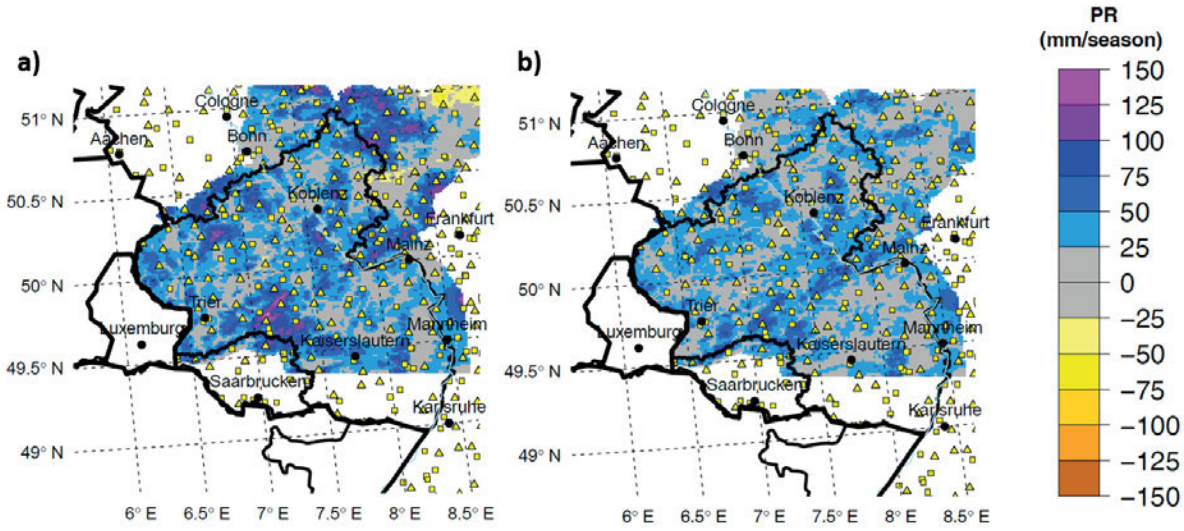

Figure 2: Differences in a) winter (DJF) and b) summer (JJA) precipitation of REGNIE (REG1.3) minus InterMet (INT1.3) at a resolution of 1.3 km for the period 1993–2000. Overlain are the stations that measure precipitation (yellow triangle) or precipitation and temperature (yellow squares).

2.2 Observations

On the horizontal scale of 1 km it is rarely given that observational products exist, thus it is fortunate that for the RLP area such a data set is available; in case of precipitation even two. As observations we use the REGNIE data set (Rauthe et al., 2013) of the German Weather Service (DWD) for precipitation and the InterMet data set (Dobler et al., 2004) of the "Landesamt für Umwelt, Wasserwirtschaft und Gewerbeaufsicht" (LUWG) for precipitation and temperature. Both are products of interpolated station data, but differ in their interpolation method as well as in the number and locations of the stations used.

In Figure 2 the differences in winter and summer precipitation of both data sets are shown as well as all observational stations for precipitation and 2 m temperature in this domain. However, there is no information available which stations have been used for REGNIE and InterMet. There are two distinct deviations between the two observational data sets: (i) REGNIE is wetter in almost all of the domain, except in valleys and (ii) in particular in areas with high orography REGNIE shows much higher precipitation values of up to more than 150 mm (Hunsrück) in winter (Figure 2a). Although the deviations are larger in winter, the bias seems to be systematic, since the spatial pattern of the differences in summer (Figure 2b) is the same as in winter. These differences can be related to three aspects: (i) the REGNIE data set is corrected for undercatchment following Richter (1995), whereas InterMet is not, (ii) the number of stations used in the spatial interpolation, and (iii) differences in the interpolation method, in particular the consideration of orography, which seems to be higher in REGNIE.

From Figure 2 it is obvious that the density of stations that measure temperature is sparser than the precipitation network. It can be expected that interpolated temperature fields of InterMet are too smooth, especially in high orographic areas where the station network is relatively sparse.

Since the InterMet data set was only available for the period 1993–2002, we restricted our analysis to the period 1993–2000. Furthermore, InterMet was only available for the state of Rhineland-Palatinate (excluding the southernmost part). For an adequate comparison, we therefore restricted the REGNIE data and the CCLM simulation output to the same spatial and temporal coverage as well. The native resolution of both observational data sets is about 1 km and both data sets have been bilinear interpolated to the 1.3 km grid of CCLM, thus we denote REGNIE as REG1.3 and InterMet as IMET1.3. An overview of all the data sets used is given in Table 1.

The REGNIE data set is quality checked so that we used it as reference. InterMet has an hourly temporal resolution, therefore we estimated the daily maximum/minimum temperature (TN/TX) based on hourly values. Prior to the use within this study, we made a quality check for the InterMet temperature data and filtered values outside the range of mean + standard deviation × 3.5 in case of maximum temperature and mean − standard deviation × 5 in case of minimum temperature. These thresholds were chosen to remove obvious outliers. The precipitation data was used without removing any potential outliers. The reason for this decision was that a simple approach, as for the temperature, is not suited for discrete data and would possibly remove the most extreme values which we cannot say are possibly true or false. Both data sets were not corrected for inhomogeneities.

Table 1: Overview of CCLM simulations and observational data sets.

Data set	Abbreviation	Resolution	Used time period	Source/Documentation
COSMO-CLM 18 km	CCLM18	18 km	1993–2000	Keuler et al. (2012)
COSMO-CLM 4.5 km	CCLM4.5	4.5 km	1993–2000	this study
COSMO-CLM 1.3 km	CCLM1.3	1.3 km	1993–2000	this study
InterMet	IMET1.3	1.3 km	1993–2000	Dobler et al. (2004)
REGNIE	REG1.3	1.3 km	1993–2000	DWD

2.3 Performance of the driving model CCLM18

We give a brief summary of the performance of the driving model CCLM18 (CON024 in Keuler et al. (2012)) for Mid-Europe (ME) in reference to the observational data set E-OBS (Haylock et al., 2008). The annual total precipitation (PR) of CCLM18 is about +60 mm/a (+8 %) too wet (see Table 5 in Keuler et al. (2012)). The seasonal bias of PR is about +30 to +45 mm in DJF and about −15 to −45 mm in JJA (see Figure 11 in Keuler et al. (2012)). The annual mean of TX is about −1.0 °C too cold (Table 6 in Keuler et al. (2012)), in winter this error is only −0.5 °C (Figure 12 in Keuler et al. (2012)). In contrast, TN is about +0.3 °C too warm in the annual mean (Table 7 in Keuler et al. (2012)), but about +1 °C too warm in DJF (Figure 13). The pressure at mean surface level is about −1.0 hPa too low in JJA and +0.5 to +1.0 hPa too high in DJF for Mid-Europe.

From these comparisons, it can be expected that CCLM4.5 and CCLM1.3 are too wet (dry) in winter (summer), with slighty too cold (warm) daily maximum (minimum) temperatures.

3 Application of the 'peaks-over-threshold' model

3.1 The General Pareto distribution

To assess extreme values, a 'peaks-over-threshold' (POT) model (Coles, 2001), which defines extremes as exceedances above a certain threshold u, was fitted to every grid box and return levels (RLs) of the 2, 5- and 10-year return periods were calculated. The second theorem or Pickands-Balkema-De Haan theorem (Pickands, 1975; De Haan and Ferreira, 2006) of the extreme value theory states that the intensities of independent and identically distributed extremes follow the General Pareto distribution (GPD) asymptotically, if scaled appropriately, and that their frequencies are a Poisson process (Ghil et al., 2011). The GPD is then defined for threshold exceedances $x = (X - u|X > u)$ of a vector X as:

$$GPD(x) = \begin{cases} 1 - \left(1 + \frac{\xi x}{\tilde{\sigma}}\right)^{-1/\xi}, & \text{if } \xi \neq 0, \\ 1 - \exp\left(-\frac{x}{\tilde{\sigma}}\right), & \text{if } \xi = 0. \end{cases} \quad (3.1)$$

with u the threshold, $\tilde{\sigma} = \sigma + \xi(u - \mu)$ the reparameterized scale parameter, σ the scale parameter, μ the location parameter and ξ the shape parameter.

The parameters of the GPD were estimated with the maximum-likelihood estimation (MLE). In case of precipitation, we implemented a censoring on the shape parameter ξ (Eq. (3.2)) to prevent unrealistic values caused by the small sample size or outliers. These are more likely in precipitation data than in temperature data because temperature series have a higher auto-correlation. The censoring is performed as in Martins and Stedinger (2000) by applying a shifted beta-prior to the log-likelihood during the MLE estimation:

$$P(\xi) = (0.5 - \xi)^{p-1}(0.5 + \xi)^{q-1}/B(p, q), \quad (3.2)$$

with P the probability of ξ, p=6, q=9 and $B(p, q) = \Gamma(p)\Gamma(q)/\Gamma(p + q)$ the Beta-function. The censoring ensures that the fitted ξ values are within the interval $[-0.5, 0.5]$, with the highest weight at the mean of the beta prior $\xi = 0.11$. Then the joint density of the prior and the GPD can be computed as:

$$GPD(\mu, \sigma, \xi) = L(\mu, \sigma, \xi|x) \cdot P(\xi) \quad (3.3)$$

with L the likelihood. Or it can be computed in terms of the log-likelihood with an additive part:

$$ln[GPD(\mu, \sigma, \xi)] = ln[L(\mu, \sigma, \xi|x)] + ln[P(\xi)] \quad (3.4)$$

The effect of the prior is thus to weight the MLE-estimates and hence to prevent unrealistic values.

3.2 Data preprocessing

To ensure temporal independence of extreme events for our analysis, we preprocessed the time series of every grid box.

Prior to the analysis, we applied a detrending to remove linear trends. Since it is likely that a threshold exceedance is accompanied by another (temporal clustering), the requirement of independence would be violated, in particular temperature often exhibits a high temporal correlation. To obtain independent exceedances, the 'run-length' declustering after Leadbetter et al. (1989) was applied to every grid box separately. This approach defines a window or cluster of length r_l enclosing an exceedance in which only the largest value is kept for the subsequent analysis. We estimated r_l as the largest time-lag of the auto-correlation function (ACF) that is significant on the 95 % level following Knote et al. (2010). Thereby, r_l was calculated for every year and then the mean was used. Typical values for r_l are 1–2 days for precipitation and about one week for temperature.

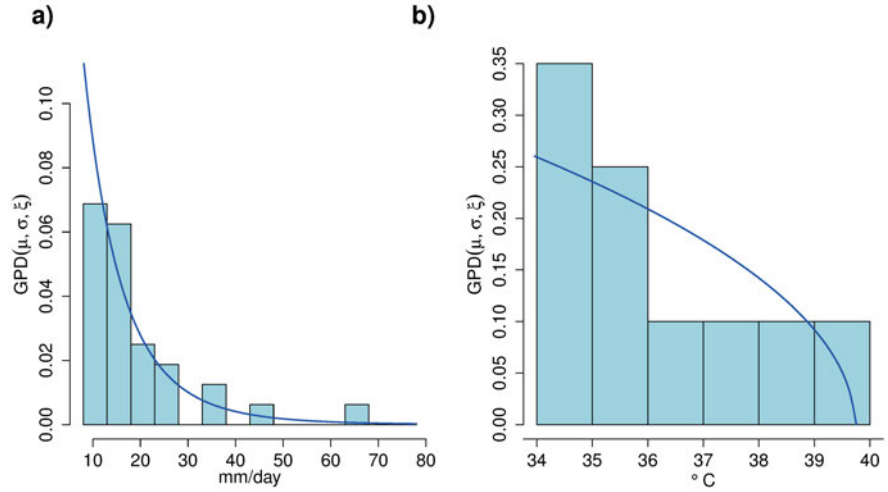

Figure 3: Threshold exceedances and fitted General Pareto distributions (GPD, blue lines) for a) summer precipitation extremes (PR) and b) summer 2 m maximum temperature extremes (TX), estimated from CCLM1.3 (1993–2000) at the location of the city of Trier.

Table 2: Estimated parameters for the GPD for daily precipitation (PR) and maximum 2 m temperature (TX) from CCLM1.3 (1993–2000, summer) at the location of the city of Trier. Thereby n denotes the number of threshold exceedances before the 'run length' declustering, r_l the 'run length' based on the ACF, n_{r_l} the exceedances after the declustering and u, σ and ξ are the estimated location, scale and shape parameter of the GPD, respectively.

Variable	n	r_l	n_{r_l}	u	σ	ξ
PR	37	1	32	8.11	8.90	0.14
TX	74	6	20	306.96	3.84	−0.67

The threshold u was set to the local 95th percentile value of the seasonal precipitation time series at every grid box and to the 90th percentile for the minimum and maximum 2 m temperature. This means that we assume 5 % of the precipitation data and 10 % of the temperature data at every grid box to be extreme values. Thus prior to the 'run-length' declustering the number of exceedendes is equal at every grid box. The threshold for temperature is chosen to be lower because the high autocorrelation of temperature causes a large r_l, which would in turn reduce the number of exceedances considerably in comparison to precipitation. With the threshold set to the 90th percentile, we expect to approximately balance the number of threshold exceedances of both variables after the declustering.

To exemplify this procedure, we fitted a GPD to the time series of PR and TX in summer from CCLM1.3 at the location of the city of Trier (49.7596 ° N, 6.6439 ° E). Table 2 shows the number of exceedances before (n) and after the declustering (n_{r_l}). For PR the 'run-length' (r_l) is only 1 day, whereas for TX it is 6 days. Thus, the number of threshold exceedances reduces only slightly for PR but considerably in case of TX. This demonstrates the reasoning of setting the threshold for TX initially to the 90 % percentile in order to obtain n_{r_l} of approximately the same order. Furthermore, this example shows typical estimates of ξ for PR and TX. In case of PR ξ is often slightly positive, at least in our domain, whereas in case of TX it is negative. A negative ξ means that there is an upper bound for estimated return levels, which is realistic from physical reasoning. Figure 3 shows the fitted GPDs to PR and TX and the histograms of the threshold exceedances. The shape of the GPDs is distinctly different for PR and TX due to the sign of the shape parameter ξ.

3.3 Return level calculation

After fitting the GPD, return levels (RLs), i.e. values which are exceeded statistically once in a defined period of time M, can be estimated by rearranging Eq. (3.1) (see Coles (2001) for details):

$$z_m = \begin{cases} u + \frac{\sigma}{\xi}\left[\left(M\, n_y \zeta_u\right)^{\xi} - 1\right], & \text{if } \xi \neq 0 \\ u + \sigma\, \log\left(M\, n_y \zeta_u\right), & \text{if } \xi = 0 \end{cases} \tag{3.5}$$

with u, σ, ξ the location, scale and shape parameter of the GPD, M the return period in years, n_y the number of observations per year and ζ_u the exceedance probability.

3.4 Confidence intervals and sampling uncertainty

Confidence intervals and sampling uncertainty were calculated by a parametric bootstrap approach (with n = 500 repetitions) following Zwiers and Kharin (1998), except that we used a block-bootstrap. Furthermore, a signal-to-noise ratio (SNR) analogue to Früh et al. (2010) has been calculated to identify a maximum return period that should be assessed. In the block-bootstrap a synthetic time series of the same length as the original one is constructed by resampling with replacement. Thereby blocks of sequential values are resampled, not

individual values. The block-length is estimated from the ACF and is identical to the 'run-length' r_l of Section 3.1, which was estimated from the original time series.

RLs are calculated, as presented in Section 3.3, for each of these 500 repetitions (note that now a different r_l may occur for the declustering) and the 90 % confidence interval is derived from the distribution of the RLs. In case of TX and TN, it may happen that the parameter estimation of the GPD fails for individual grid boxes. This is because we do not apply any prior for those variables but have rather short time periods so that unrealistic RLs may be returned by the POT models. Thus we remove grid boxes with such outliers prior to the analysis by building z-scores of the 5-year RLs and remove any grid box whose z-score > 5×SD of the bootstrapped 5-year RL field.

4 Results

4.1 Validation against observations

Previous to the analysis of estimated RLs, we first compared the direct model output fields of total precipitation (PR), 2 m maximum (TX) and minimum temperature (TN) to the observations REG1.3 (PR) and IMET1.3 (PR, TX, TN). We used quantile-quantile (QQ) plots, standard statistics such as the root-mean-squared error (RMSE), Pearson's correlation coefficient (r) and the standard deviation (SD) and combine them in a compact Taylor diagram. Furthermore, we compared the structure of PR fields by using the Structure-Amplitude-Location (SAL)-analysis (Wernli et al., 2008).

4.1.1 Comparison of daily fields

In Figure 4 the results are shown for domain-averaged PR and TX in summer (JJA) as well as PR and TN in winter (DJF). First of all, in case of precipitation (Figure 4ab) IMET1.3 differs from REG1.3, showing a lower SD and a slight offset in the RMSE. The QQ plots (Figure 5ab) reveal that REG1.3 is wetter compared to IMET1.3 and contains higher extreme values. However, these deviations are within the 95 % confidence interval. The reason for these deviations are: (i) differences in the interpolation methods, in particular the weight of orography is higher in REGNIE causing a larger spatial variability and thus a higher SD, and (ii) deviations in the number and location of stations used for the interpolation causing differences in the absolute values.

The precipitation of the CCLM simulations is shown in reference to REG1.3. In summer (Figure 4a), the temporal correlations between the CCLM precipitation fields and REG1.3 are low ($r \approx 0.4$–0.5), with a RMSE of about 1.1 mm/day and a SD close to 1 mm/day. The RMSE value can be explained by a dry bias for lower intensities of summer precipitation (Figure 5a) that is already present in the driving CCLM18 simulation (Section 2.3) and thus transferred to CCLM4.5 and CCLM1.3; although, it is only significant in case of CCLM1.3 as the deviation is outside the 95 % confidence interval. However, extreme percentiles are considerably overestimated, but in total this still results in the above mentioned dry bias. The QQ plots are based on spatial averages and therefore the upper quantiles are not directly related to extreme events, e.g. days with 20 mm/day at every grid box in the domain results in a high spatial average. In contrast a severe extreme event, which only occurs in a small part of the domain, might result in a lower spatial average, in particular in summer.

On August 7, 1995 a very high precipitation extreme event with more than 150 mm/day was simulated in all CCLM simulations, which is much weaker in the observational data set (80 mm/day). However, this event does only have a minor effect on the QQ plots, but we investigate the effect of the event on the extremes below.

The bad temporal correlation in summer may be explained by: (i) errors in the synoptic forcing from the driving model and (ii) the fact that summer precipitation consists to a large proportion of convective precipitation, which may be too small-scale to be registered at observational stations but is represented in CCLM. Thus it is likely that the temporal, but also the spatial correlations, are reduced. (iii) A further reduction might be caused by the stochastic nature of summer convection. Particularly during low forcing synoptic situations, when convection is primarily forced locally and not from the boundary conditions. As it cannot be expected that the model represents observed weather on a specific date or time in such situations, the temporal correlations and the SAL skill score (Section 4.1.2) may be reduced.

In winter (Figure 4b), the correlations ($r \approx 0.8$–0.85) are much higher than in summer. Furthermore, the RMSE and the SD improve to 0.6 mm/day and ≈ 1 mm/day, respectively. The QQ plots (Figure 5b) show an underestimation of extreme winter precipitation, with a higher bias for intensities between 20 to 35 mm/day, but these underestimations are still within the confidence interval.

The observations of summer TX (Figure 4c) and winter TN (Figure 4d) are well captured by the simulations ($r \approx 0.9$–0.95), with a slightly too low SD and a RMSE of about 0.4 °C. The QQ plot for TX in summer (Figure 5c) shows too cold temperatures for the higher quantiles (> 28 °C) for all CCLM simulations, when compared to IMET1.3. In contrast, the coldest TN in winter are too warm in the CCLM simulations. These bias were already found for the driving CCLM18 (see Section 2.3)

One issue that arises from the Taylor-plots (Figure 4) is the fact that increasing the horizontal resolution of the CCLM model does not have a major impact on the temporal correlation of domain-averaged precipitation and temperature. However, this analysis considers only domain averages, which seem not to be influenced significantly by the horizontal resolution. Although we did not investigate the reason for this behaviour, it might also

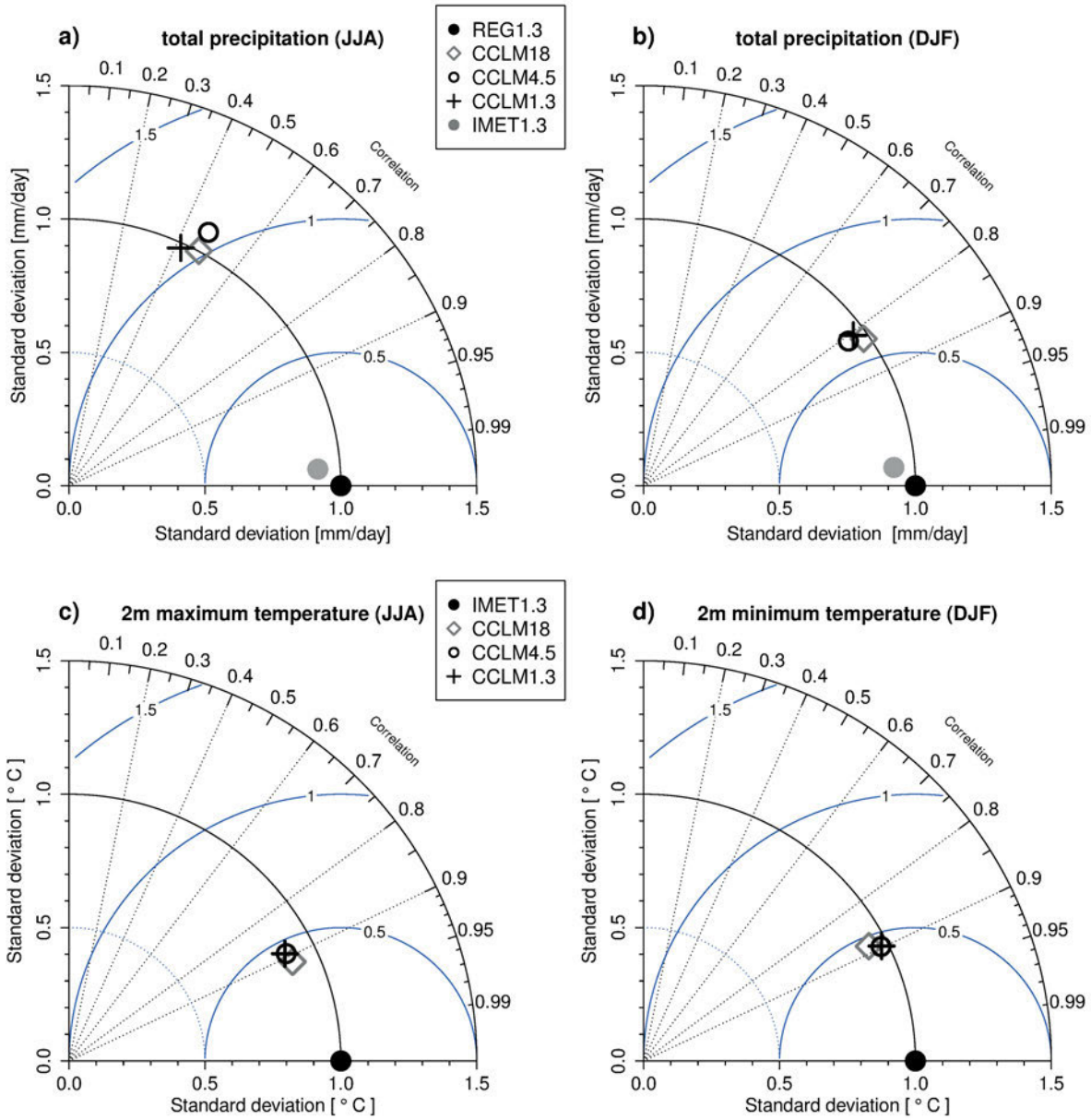

Figure 4: Taylor diagrams of domain-averaged a) seasonal precipitation (PR) of CCLM18, CCLM4.5, CCLM1.3 and IMET1.3 versus REG1.3 in summer and b) in winter. In c) for 2 m maximum temperature (TX) versus IMET1.3 in summer and in d) for 2 m minimum temperature (TN) in winter.

relate to errors in the synoptic fields of CCLM18, which may then cause errors in the correlation structure of the nested simulations.

4.1.2 SAL analysis for daily precipitation fields

The SAL analysis is a three-dimensional quality measure, which separately compares the structure (S), amplitude (A) and location (L) of a quantitative precipitation forecast within a domain of interest (Wernli et al., 2008) and is commonly used to analyse the structure of precipitation in time and space. We applied the SAL-analysis to daily observed and modelled precipitation fields and used a threshold of 10 mm/day to identify extreme precipitation objects, which roughly corresponds to the domain-averaged threshold of the GPD from Section 3.2. A precipitation object is defined as coherent grid boxes that exceed the threshold. Thereby we did the evaluation at the native horizontal resolution of the models, respectively.

The amplitude component (A) describes the normalized differences of the domain averaged precipitation and ranges from −2 to +2. Negative (positive) A values mean an underestimation (overestimation) of the precipitation amount and $A = 0$ denotes a perfect agreement. The location component (L) is calculated from two additive terms. The first compares the location of the centre of mass of the domain-wide precipitation and the second calculates the distances between the centre of mass of

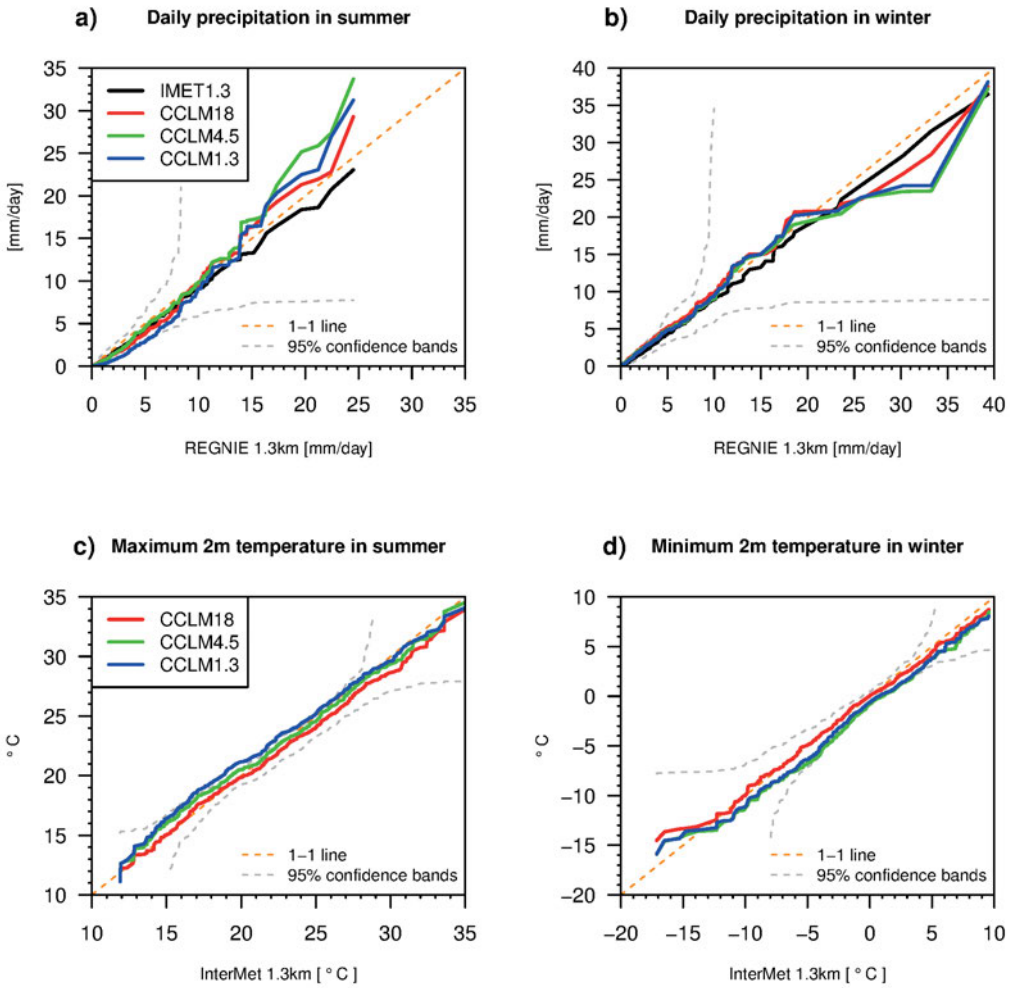

Figure 5: Quantile-quantile plots of domain-averaged a) daily precipitation (PR) in summer (JJA) and b) winter (DJF) as well as c) maximum 2 m temperature (TX) in JJA and d) minimum 2 m temperature (TN) in DJF. The REG1.3 data set is used as reference for precipitation and IMET1.3 for temperature. The 95 % confidence interval (dashed grey lines) is based on the Kolmogorov-Smirnov statistic (Doksum and Sievers, 1976).

the domain-wide precipitation and the centre of mass of individual precipitation objects. This is necessary since different precipitation fields can have the same domain wide centre of mass. The L component ranges from 0 to 2, where 0 means a perfect match. Finally, the structure component (S) compares the volume and the shape of precipitation objects and ranges from −2 to 2. Negative S values mean that simulated precipitation objects are too small or too peaked in comparison to the observations, whereas positive values mean that the simulated precipitation objects are too widespread, e.g. if convective cells are observed but coherent precipitation objects are simulated (Prein et al., 2013). See Wernli et al. (2008) for the details of this method.

In Figure 6 the results of the SAL analysis are shown for winter and summer. In general, the mean A-S value pairs are close to 0 for all models and seasons, with a slightly negative A component and a slightly positive S component. This means that the precipitation amount of the extremes are slightly underestimated in the mean and the precipitation objects are slightly too widespread in comparison to REGNIE. The L component is acceptably low with values up to 0.20. In winter, the L component is lower than in summer, which indicates that winter extreme precipitation objects are better captured by the models. In both seasons the scatter of the A component is much larger than the scatter of the S component, which is shown as the 25 % and 75 % quantiles of A and S in the figures (dashed red boxes).

Deviations from this pattern were found for CCLM1.3 in summer. Here the mean precipitation amount shows a larger negative amplitude value $A = -0.54$ and the structure component is negative with $S = -0.33$. The bias in the amount is caused by the general dry bias of CCLM1.3 in summer. A negative S means that the precipitation objects are more peaked than in the observations. This indicates that the increase of the horizontal resolution in CCLM causes smaller, more peaked precipitation objects or convective cells, which confirms the result of Prein et al. (2013).

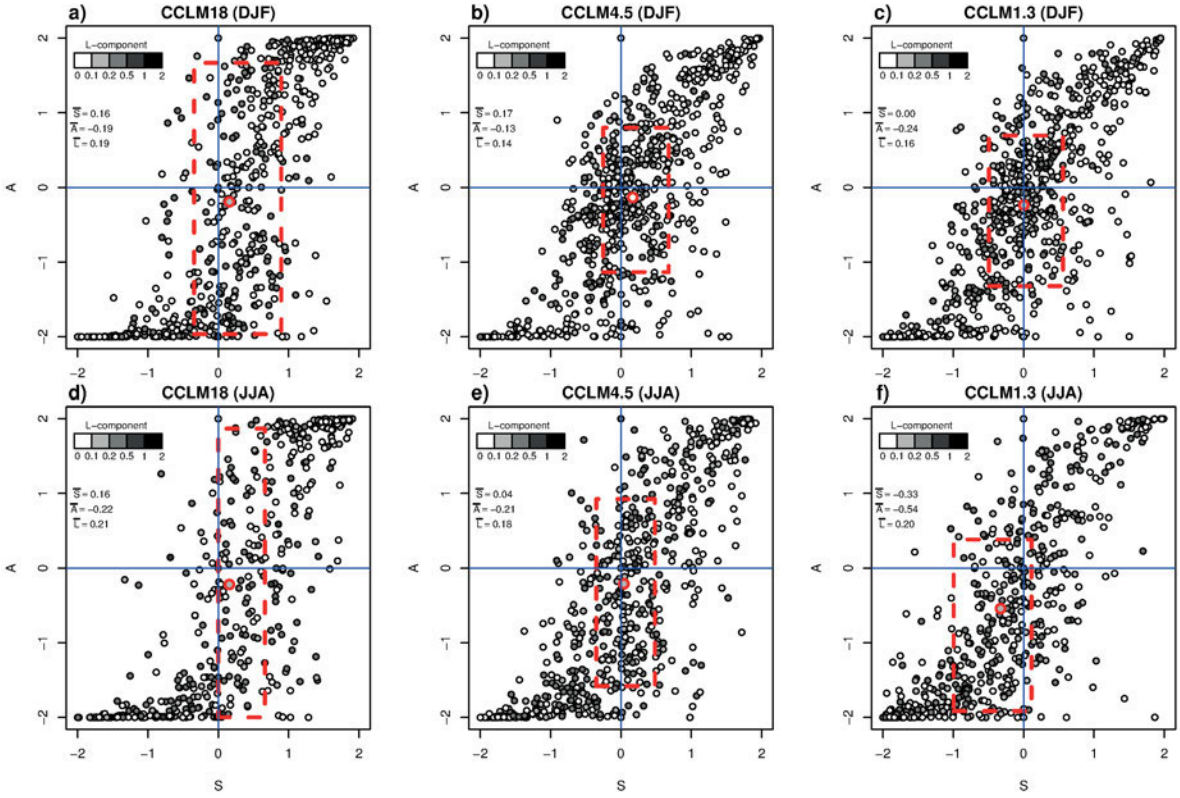

Figure 6: SAL diagrams for winter (DJF) and summer (JJA) precipitation extreme events of CCLM in comparison with REGNIE on a daily time scale. The red encircled points mark the mean of the A-S pairs and the red dashed lines mark the 25 % and 75 % of the S and A components. In addition, the mean values of the SAL components are shown within the figures.

4.2 Return level estimates for total precipitation

We restrict the analyses of RLs to return periods of up to 10 years due to the limited length of observations (8 years). However, we concentrate on the 2 and 5-year RLs because the 10-year RL is already an extrapolation. To assess the uncertainty of the return level estimation, the signal-to-noise ratio (SNR) is calculated for the block-bootstraped RLs of the observations and the CCLMs.

In Figure 7 the SNRs for different return periods are shown for summer and winter PR. The 10-year RL of REG1.3 in JJA corresponds approximately to a SNR of 2.3 and in DJF to a SNR of 2.1. This means that the RL value of a 10-year return period is about twice the uncertainty (or 90 % confidence interval). The SNR values from the CCLM simulations are systematically lower in JJA, with SNR between 1.75 and 2.0 for the 10-year return period, but higher in winter with a SNR between 2.3 and 2.4.

Since it is likely that the event on 7th August 1995 does not correspond to a return period of 8 years, we calculated the RLs in JJA for CCLM4.5 from the 8-year simulation (1993–2000) with and without the event and from the 18-year simulation (1993–2010) with this extreme event and compared the spatial patterns (see Section 4.2.2).

4.2.1 Spatial variability

The impact of the horizontal resolution on extremes is measured in terms of the mean and spatial variability of return levels. We define the spatial variability as the range of RLs within the domain. In Figure 8 the spatial variability of the 2, 5, and 10-year bootstrap-mean RLs are shown along with the domain-averaged 90 % confidence intervals. As point symbols the mean and the bootstrap-mean of the domain-averaged RLs are shown. Since they are very similar, a bootstrap sample size of 500 seems sufficient to capture the RL distribution.

The inter-quantile range (IQR, 25–75 % percentile) of the estimated 2-year RLs of daily PR in summer from REG1.3 (Figure 8a) ranges between 25 and 30 mm/day, the 5-year RLs between 31 and 37 mm/day and the 10-year RLs between 36 and 45 mm/day. However, much higher RLs occur for individual grid boxes. Regarding the uncertainty of the estimates, the confidence intervals become larger with increasing return period. Compared to REG1.3, IMET1.3 shows slightly lower (< 5 mm/day) RL estimates, with the most extreme RLs (upper whiskers) considerably lower for all return periods.

The RLs from CCLM are systematically higher, about +5 mm/day, over all return periods. The spatial variability (i.e. the whole range of the boxplots) shows a dependency on the horizontal resolution of the model. Although the spatial variability generally increases with

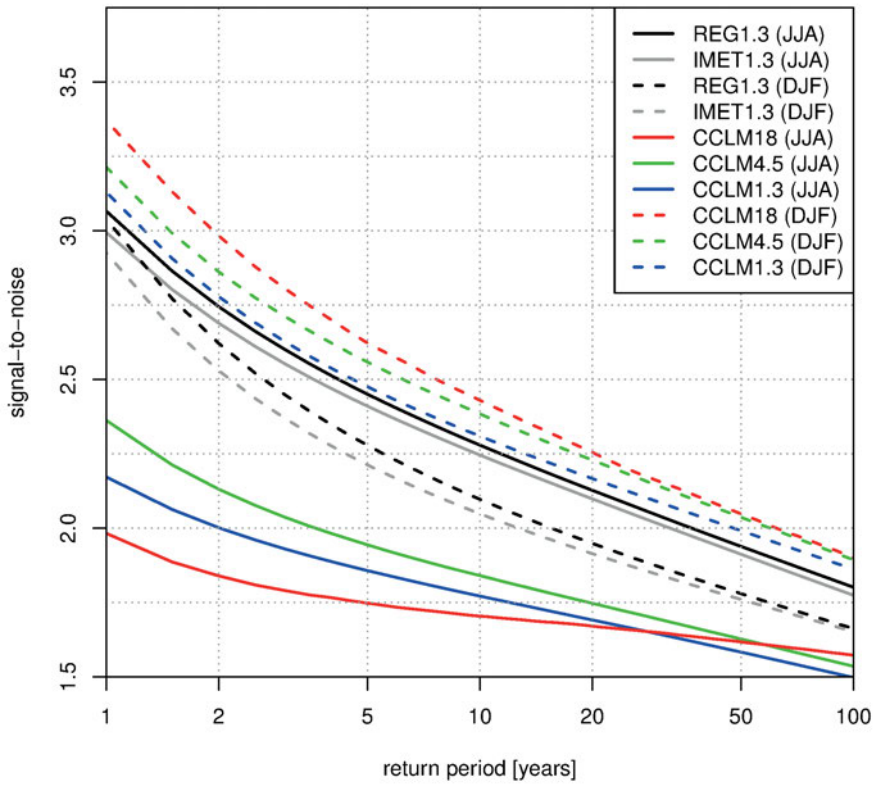

Figure 7: Signal-to-noise ratio (SNR) of domain-averaged bootstrap-mean return levels of precipitation (PR) in summer (JJA) and winter (DJF) estimated from the observational data sets InterMet (IMET1.3) and REGNIE (REG1.3) and from the CCLM simulations (only JJA) at different horizontal resolutions.

the return period (considering the whole range of the boxplots), the relevant result is that a higher resolution of 1.3 km and 4.5 km produces RLs, whose spatial variability resembles those of REG1.3 and is larger than at 18 km.

In winter, the spatial variability increases even more with the horizontal resolution. The IQR increases by approximately 5 mm/day from CCLM18 to CCLM4.5 and CCLM1.3 for all return periods. The range of the confidence intervals becomes again larger with increasing return periods. This shows a clear effect of the horizontal resolution on extreme precipitation. The differences between CCLM4.5 and CCLM1.3 are only minor, and the estimated RLs agree well with IMET1.3 and REG1.3.

4.2.2 Spatial patterns

In order to investigate whether the spatial patterns improve, we map the spatial pattern of the 5-year RLs of PR in summer (Figure 9). The 5-year RLs derived from IMET1.3 (Figure 9a) and REG1.3 (Figure 9b) show deviations, which are mapped as differences in Figure 9c. For most of the domain the differences are of the order of ±5 mm/day. However, in areas where the highest RLs were estimated from REG1.3 the differences are considerably larger, with up to +30 mm/day in REG1.3. The highest RLs from observations were estimated in the Westerwald, Hunsrück, Eifel, Palatinate Forest and close to the river Rhine to the west of the city Mainz. Moreover, spots of higher RLs (> 45 mm/day) are scattered loosely over the domain, which indicates a possible convective origin. In most of the domain the 5-year RLs are < 45 mm/day. The lowest RLs of < 25 mm/day were estimated for the northern Moselle Valley to the west of Koblenz and for the eastern parts of the Eifel. Overall, the spatial pattern is comparable between both observational data sets.

Comparing the spatial patterns of the observations with those of the CCLM simulations (Figure 9d–f) the patterns from CCLM show RLs in the range of 25–55 mm/day for most of the domain; in the northern part of the domain values > 55 mm/day were estimated (CCLM4.5 and CCLM1.3). The pattern of CCLM18 (Figure 9d) deviates in particular with regard to RLs > 45 mm/day. In contrast to the observations, the highest RLs occur in the centre of RLP. More similar to the patterns of the observations are the patterns of CCLM4.5 and CCLM1.3 (Figure 9e–f). The highest RLs occur in the mountainous areas, although they are systematically overestimated, as discussed before.

In order to analyse the effect of the extreme event on 7th August 1995, we additionally estimated the RLs in JJA from an 18 year long simulation (1993–2010) and repeated the estimation for the 8 year long simulation (1991–2000), but removed the event prior to the application of the POT model. The spatial patterns of both are shown in Figure 10. The obtained patterns are very similar and the RLs are systematically lower than Figure 9e,

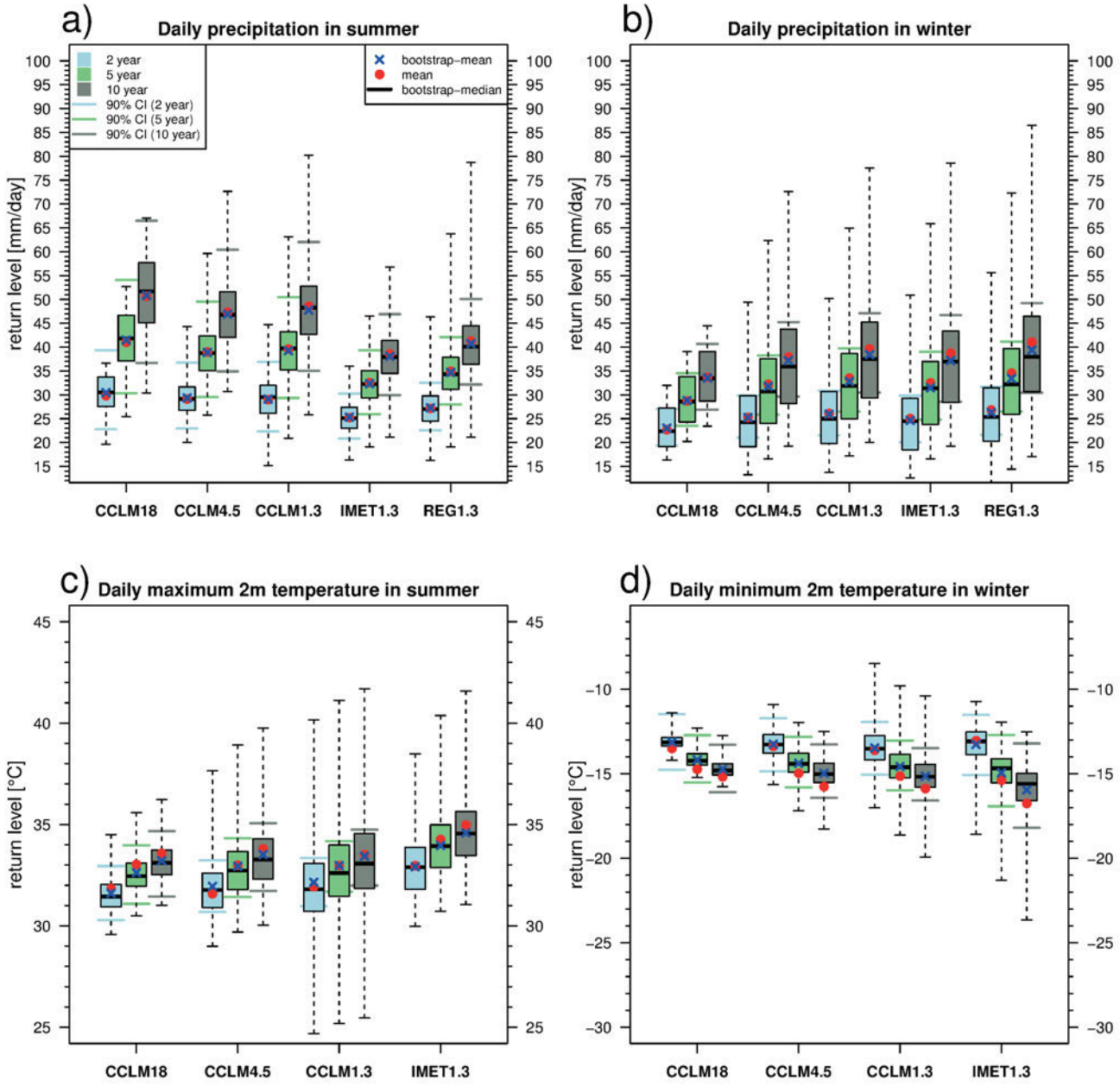

Figure 8: Spatial boxplots of bootstrap-mean 2, 5 and 10-year return levels from 'peaks-over-threshold' models estimated from the period 1993–2000 for precipitation (PR) in a) summer (JJA), b) winter (DJF) and in c) 2 m maximum temperature (TX) in JJA and in d) 2 m minimum temperature (TN) in DJF. The coloured horizontal lines denote the domain-averaged 90 % confidence interval from the block-bootstrap, which is a measure for the uncertainty of the domain-averaged mean values (red points).

but also slightly lower than the RLs estimated from the observations (Figure 9a–b). The spatial means ±1 SD are 29.5 ± 3.74 mm/day (8 years) and 29.4 ± 3.7 mm/day (18 years), respectively. From these results we can conclude that the event on 7th August 1995 has a return period of > 8 years, at least of 18 years, and causes the systematic shift in the RL estimation of summer PR. It also explains the higher RLs estimated from the other CCLM simulations.

In Table 3 the results of the spatial correlations are shown. The Pearson correlation coefficients between PR return levels in summer and the height of the orography show nearly no correlation (r between 0.07 and 0.21). This confirms that summer precipitation extremes may be mainly caused by convective cells which are generally not correlated to the height of the underlying orography.

In winter, the spatial patterns of the RLs from the observations (Figure 11a,b) are very similar. The highest RLs (> 45 mm/day) were estimated in the mountain areas of Hunsrück and Eifel in the South-west and Westerwald in the North of the domain. The lowest RLs (< 25 mm/day) were estimated along the Rhine river. The absolute differences between the RLs (Figure 11c) are of the same order as in summer and the largest deviations occur approximately in the same areas, with the

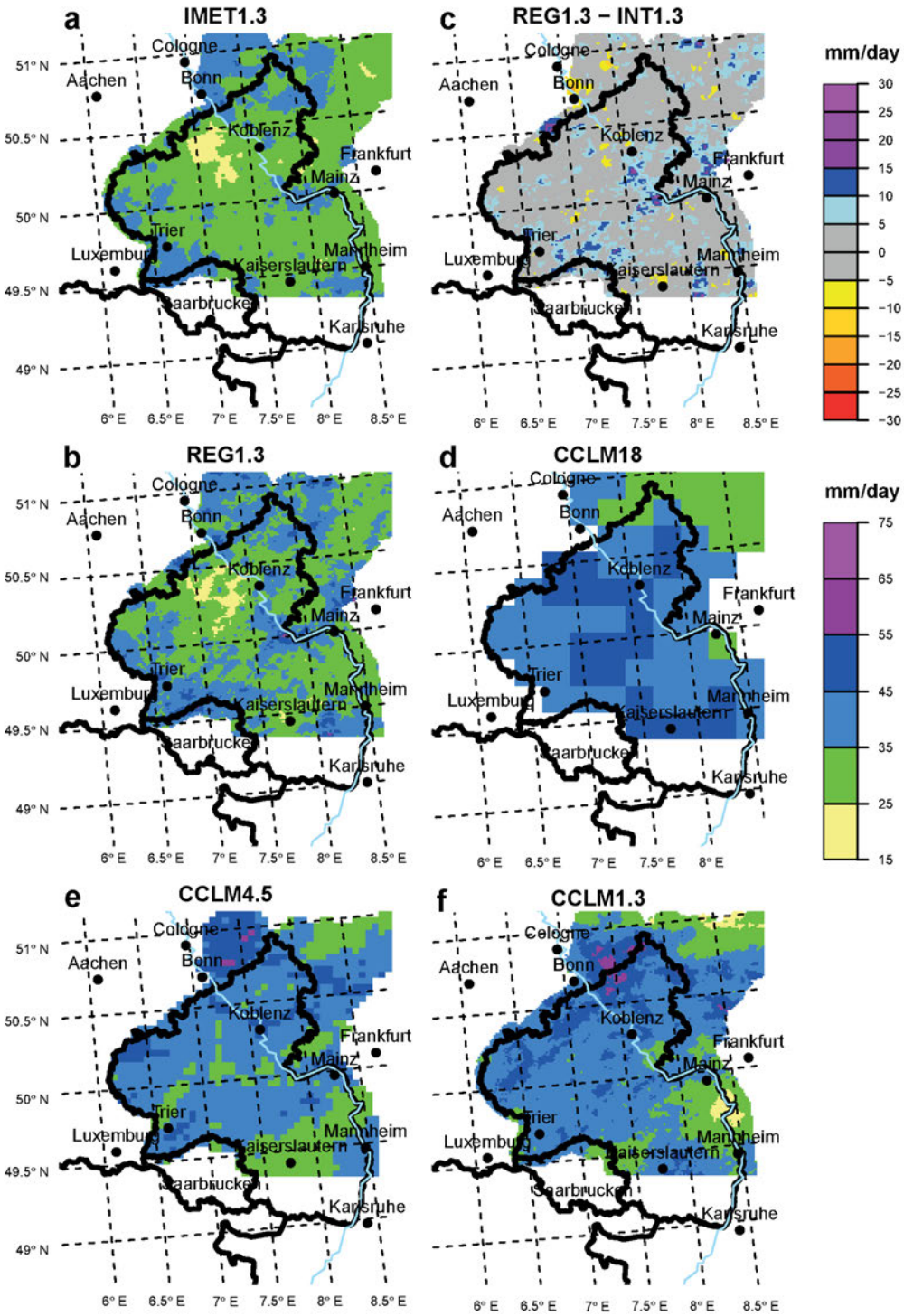

Figure 9: Estimated 5-year return level of total precipitation (PR) in summer (JJA) from the 'peaks-over-threshold' models for a) InterMet (IMET1.3), b) REGNIE (REG1.3), c) REGNIE minus InterMet, d) COSMO-CLM 18 km (CCLM18), e) COSMO-CLM 4.5 km (CCLM4.5), f) COSMO-CLM 1.3 km (CCLM1.3) from the period 1993–2000.

largest positive anomaly along the southwest-northeast striking mountain ridges of the Hunsrück.

The general spatial pattern is well captured by all CCLM simulations, although absolute values differ considerably between the simulations. The spatial pattern of the estimated RLs from CCLM18 (Figure 11d) resemble the large-scale pattern of the observations (Figure 11a), but RLs are considerably underestimated in high orographic areas. The 5-year RLs of REG1.3 are up to 75 mm/day in the Hunsrück and about 65 mm/day in the Eifel and Westerwald, whereas in CCLM18 the RLs are only about 45 mm/day at a maximum, which is an underestimation of about 20–30 mm/day.

At a horizontal resolution of 4.5 km (Figure 11e) the spatial pattern improves considerably. Now RLs of about 65 mm/day occur in the Hunsrück and Westerwald, only

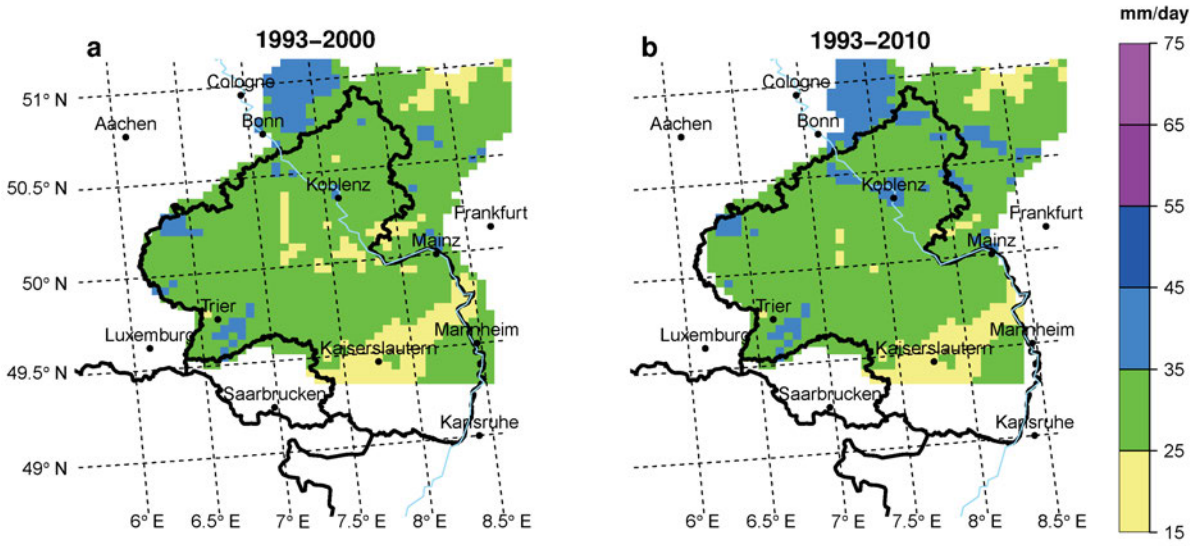

Figure 10: Estimated 5-year return level of total precipitation (PR) in summer (JJA) from the 'peaks-over-threshold' model for COSMO-CLM 4.5 km (CCLM4.5) from the period a) 1993–2000 without the event on 7th August 1995 and b) 1993–2010 with the event.

Table 3: Spatial correlation coefficients (Pearson's r) between estimated 5-year and 2-year return levels of precipitation (PR), 2 m minimum/maximum temperature (TN/TX) and orography. In addition, for the 2-year return levels the slope of fitted linear regressions are shown. The orography field was interpolated onto the same horizontal resolution of the corresponding data set.

Data set	PR (JJA)			PR (DJF)			TX (JJA)			TN (DJF)		
	5 yr	2 yr	mm/100 m	5 yr	2 yr	mm/100 m	5 yr	2 yr	°C/100 m	5 yr	2 yr	°C/100 m
CCLM18	0.07	0.19	0.66	0.67	0.59	2.51	−0.92	−0.80	−0.80	−0.52	−0.41	−0.21
CCLM4.5	−0.13	0.09	−0.28	0.54	0.56	3.50	−0.92	−0.91	−1.16	−0.53	−0.58	−0.40
CCLM1.3	0.21	0.27	0.95	0.60	0.62	3.55	−0.88	−0.90	−1.46	−0.53	−0.58	−0.44
IMET1.3	0.01	0.08	0.20	0.60	0.58	3.28	−0.67	−0.77	−0.80	−0.43	−0.42	−0.36
REG1.3	0.27	0.33	1.00	0.72	0.70	4.23	–	–	–	–	–	–

in the Eifel the RLs are still underestimated. At 1.3 km (Figure 11f), the pattern slightly improves further, with RLs up to the observed 75 mm/day in the Hunsrück and higher RLs in the Eifel. The spatial correlations with the height of orography increase and range from $r = 0.54$ to $r = 0.72$ (Table 3). This indicates that winter precipitation extremes are caused by large-scale frontal systems, which produce orographic rainfall and thus show a dependency. Although it seems that the coarse pattern of extreme precipitation is not dependent on the resolution or treatment of convection, using the 1.3 km resolution improves the pattern as well as the absolute values considerably due to the better representation of the orography.

4.3 Return level estimates for maximum and minimum 2 m temperature

In case of temperature the SNR is about 8 (not shown), which indicates that the estimation of TX/TN RLs is much more robust compared to PR.

4.3.1 Spatial variability

The spatial variability of the 2, 5 and 10-year bootstrap-mean RLs of 2 m maximum temperature (TX) in summer (Figure 8c) increases distinctly with increasing horizontal resolution. The 5-year RLs from CCLM18 vary within 30.5–35.5 °C and are about 1 °C warmer (colder) for the 10 (2)-year RLs. In contrast, the 5-year RLs from CCLM1.3 vary within 25–42 °C with a slightly increased IQR. Thereby, the median remains nearly constant, e.g. at 32.5 °C for the 5-year RLs. The spatial variability of the RLs from CCLM1.3 is overestimated compared to the observations (see the discussion in Section 4.3.2). However, all mean/medians of the RL estimates from CCLM are about 1–1.5 °C colder than the corresponding RLs from IMET1.3. This cold bias is also visible in the QQ-plot (Figure 5c).

In winter, similar effects of the horizontal resolution on the RLs were found. While the spatial variability of 5-year RLs of 2 m minimum temperature (TN) from CCLM18 show a spatial variability of only 3 °C (Figure 8d), it increases again considerably in CCLM4.5 and CCLM1.3. The estimated RLs from CCLM1.3 are closest to those from IMET1.3. Considering the spatial mean

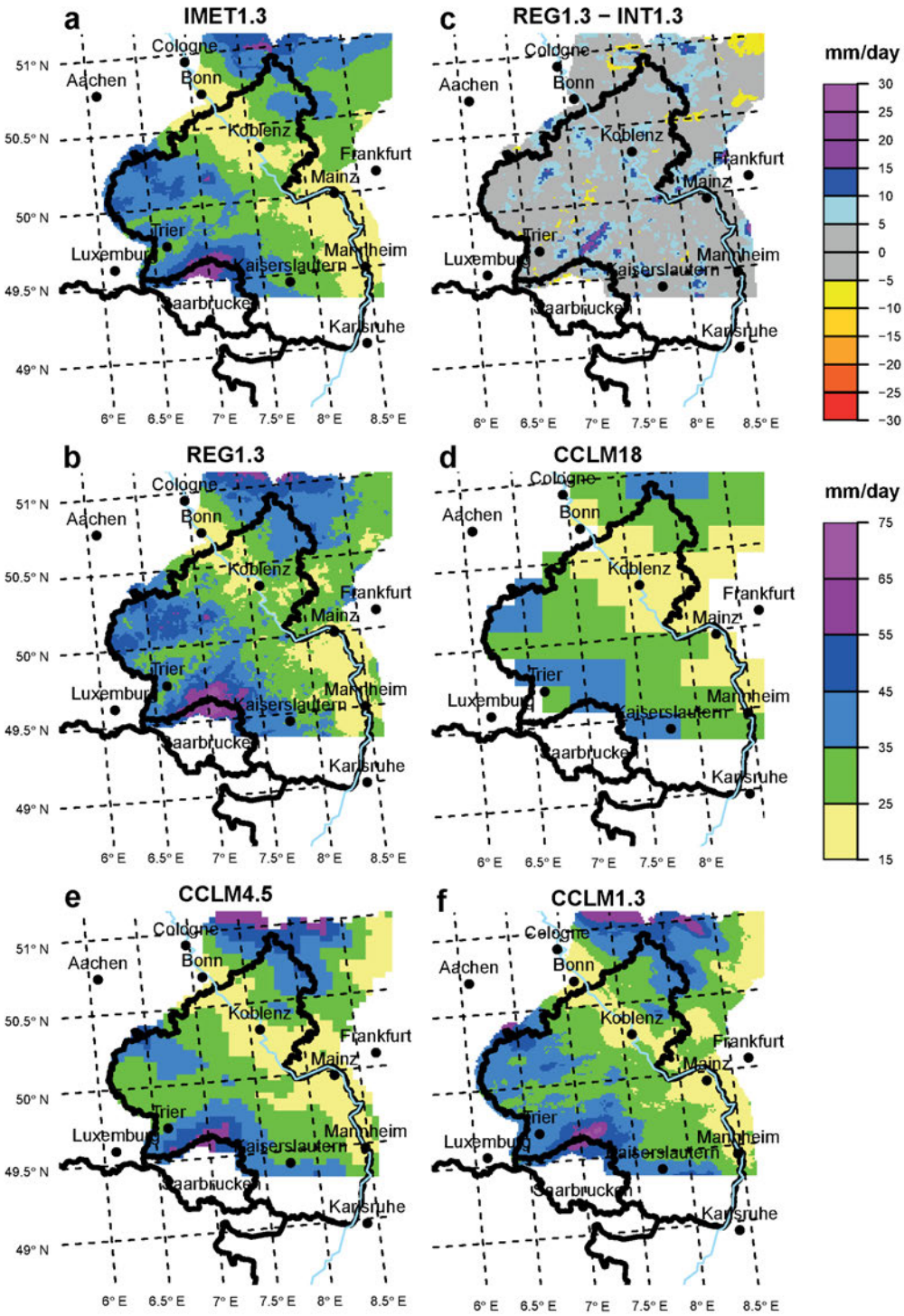

Figure 11: As in Figure 9 but for 5-year return levels of total precipitation (PR) in winter (DJF).

and bootstrap-median, the RLs are well in the order of the observations. In contrast to the summer extremes, the variability of winter RLs is much lower. Only the coldest RLs are not captured by the CCLMs.

4.3.2 Spatial patterns

The spatial maps of the return levels of TX (Figure 12) show the increased variance with increasing resolution and further reveal a high dependency on the orography. From IMET1.3, the highest RLs (> 36 °C) occur in the lowest regions of the domain, i.e. along the Rhine valley, the Moselle valley and in the Cologne Basin between Koblenz and Cologne.

Comparing the spatial patterns of CCLM18 (Figure 12b) to IMET1.3 (Figure 12a), CCLM18 again captures the large-scale pattern, but is not able to produce local features with much higher RLs, such as in the Rhine Valley or Moselle Valley. In particular the sig-

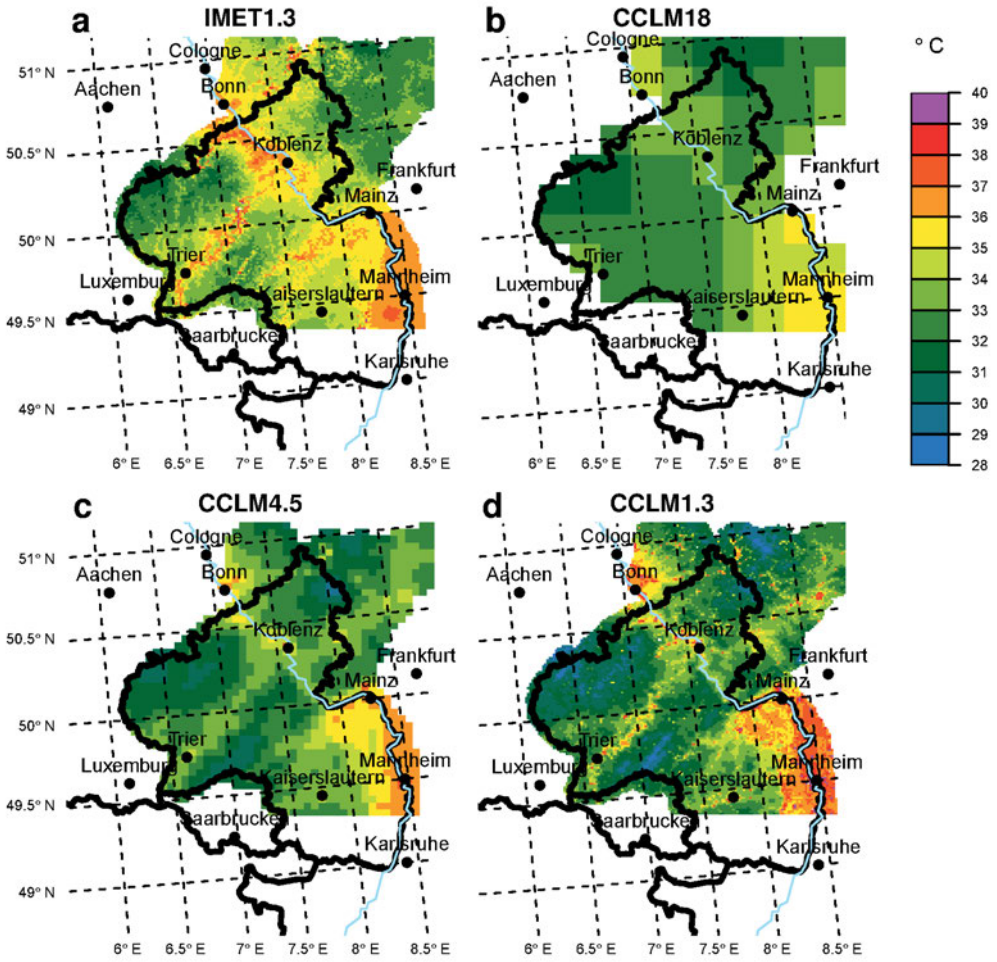

Figure 12: As in Figure 9 but for 5-year return levels of 2 m maximum temperature (TX) in summer (JJA).

nature of the latter is completely missing in the coarse model. By increasing the resolution to 4.5 km, the spatial structure of the RLs now resembles those of the observations but underestimates the RLs in the low areas. In the Moselle Valley and along the river Rhine between Koblenz and Bonn they are about 2–4 °C too cold in CCLM4.5. Only at a horizontal resolution of 1.3 km (Figure 12d) the range of the RLs is captured adequately.

In areas with high orography, the RLs of CCLM4.5 and CCLM1.3 are considerably colder than in IMET1.3. For example, in the Hunsrück the estimated 5-year RLs are about 28 °C in CCLM1.3, but about 32 °C in IMET1.3. In this region InterMet is affected by outliers in 2 m temperature reaching 40 °C (not shown), which is not realistic in this area with heights up to 800 m above sea level. Although the most severe outliers have been removed prior to the analyses, the remaining time series still causes an overestimation of RLs at the highest mountain ridges. Therefore we exclude this area from further analysis. Similar differences can be seen for the Eifel and Westerwald. Although we did not determine the reasons for these deviations, we want to point out that the temperature station density is generally sparse, in particular in high terrain, which might cause a too smooth gridded observational data set with too warm RLs in mountainous areas. On the other hand, in the Rhine Valley and around Bonn and Koblenz, the RLs are up to 4 °C warmer than in IMET1.3, reaching 40 °C. The overestimation in flat areas may be due to the warm bias of CCLM in these areas.

The spatial correlation with the orography is negative for TX in summer with $r = -0.67$ (IMET1.3), -0.88 (CCLM1.3) and -0.92 (CCLM18, CCLM4.5). However, this correlation is larger for CCLM than for IMET1.3. A possible reason for this seems to be an averaging effect of the orography that produces too high correlations in CCLM18 and CCLM4.5, which are not as high as in the observations.

Although the spatial maps of the 5-year RLs of TN (Figure 13) show again a similar effect of the horizontal resolution as for the TX extremes, the overall effect is smaller. From IMET1.3 (Figure 13a) the coldest RLs were estimated in the Eifel with about −22 °C, but also in the northern part of the domain. In general, mountainous areas show RLs below −15 °C. In flat and low areas, however, the RLs are not as cold and thus the warmest TN RLs occur in the Rhine Valley with only −12 °C. Another region with less severe RLs is the area around Trier, where RLs range between −13 °C and −15 °C. Comparing the 5-year RLs fields based on the CCLM

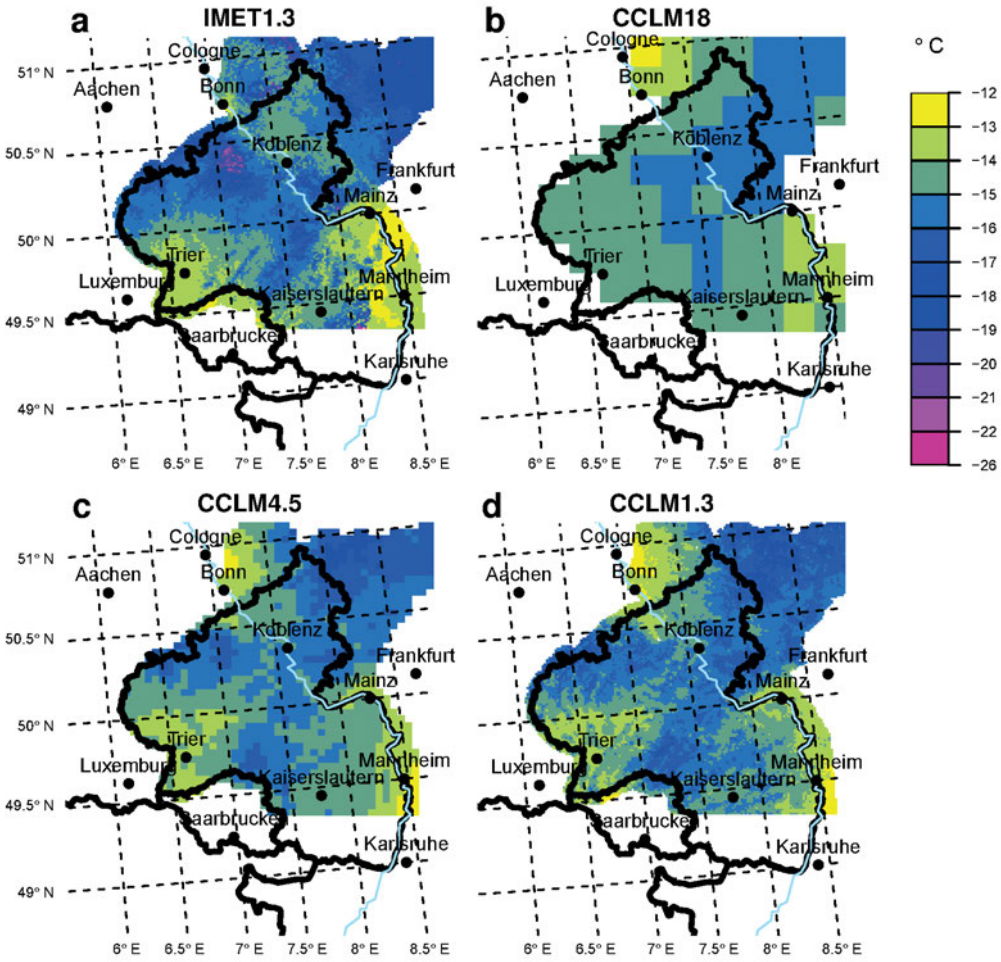

Figure 13: As in Figure 9 but for 5-year return levels of 2 m minimum temperature (TN) in winter (DJF).

simulations to Figure 13a, it is obvious that CCLM18 (Figure 13b) is not able to reproduce the structure of the RLs from IMET1.3, although the pattern resembles the large-scale gradients. At a resolution of 4.5 km (Figure 13c) the pattern is considerably improved, but is best captured by CCLM1.3 (Figure 13d), although differences are still visible. The coldest areas in the Eifel are not adequately captured by CCLM1.3 and the RLs in the area east of the city Bonn are considerably warmer in all models. The spatial correlation show a weak negative correlation ($r = -0.43$ to -0.53) with orography (Table 3).

5 Conclusions and outlook

In this study we investigated the effect of the horizontal resolution in the regional climate model COSMO-CLM on the estimated 2, 5, and 10-year return levels of daily precipitation and 2 m minimum/maximum temperature at resolutions ranging from 18 km over 4.5 km to 1.3 km. Furthermore, we validated these hindcast runs with regard to their performance to simulate extremes with the observational data sets InterMet and REGNIE. The comparison with observations and the extension of the analysis to the winter season constitutes a new aspect with respect to the previous study of Knote et al. (2010).

In general, the CCLM simulations were able to reproduce the main characteristics of the observed extremes. Summer precipitation extremes were systematically overestimated due to a very severe event on 7th August 1995, which have a return period of at least 18 years. By comparing RL estimations from an 18 year long simulation, we could show that the bias in the RL reduces considerably. This event, but also the higher percentiles of the intensity distribution, were overestimated in comparison to the observations. This tendency of overestimating in convection-resolving models has been also found for other models, e.g. by Kendon et al. (2012). Temperature extremes in summer were about 1 °C too cold in the CCLMs, whereas cold extremes in winter were about 1–3 °C too warm in the domain-average compared to InterMet.

Based on our limitations with respect to the length of the time series, we found high uncertainties for estimating return levels, in particular for return periods longer than 5 years. Furthermore, single events do have a severe impact on the magnitude of the return levels and hence the absolute values have to be treated with caution.

The most striking effect of an increased horizontal resolution was found to be an improvement of the spatial pattern and an increase in the spatial variability of extreme values with only minor changes in the median, although in case of maximum temperature the variability was found to be larger than for the RL estimates of the observations. In general, this effect is larger for temperature extremes than for precipitation, and in case of the latter, larger in winter than in summer. We performed spatial correlations of the estimated 5-year RLs and the height of orography in order to investigate the dependency of extremes on the spatial resolution of orography. The results showed that the dependency is highest for summer maximum temperature extremes, but also high for minimum temperature extremes and precipitation extremes in winter. Furthermore, these dependencies increase with the horizontal resolution of CCLM.

Regarding precipitation this can be explained by the scale on which the process of precipitation formation is acting: in winter, precipitation is mainly produced by orographic rainfall in connection with frontal systems (high correlation); whereas in summer, precipitation forms predominantly within convective cells, which depend only weakly on the orography in the region of investigation. However, the effect of better resolved processes, such as deep convection, was impeded by the dry bias of summer precipitation with regard to seasonal sums. Since the bias already occurred in CCLM18, it is likely that it was transferred to CCLM4.5 and CCLM1.3 and hence it might be a model specific problem.

We found that the intensity of RLs increases in small, local areas if the resolution is increased, which may be due to a better representation of deep convection. Thus CCLM1.3 simulates more convective summer precipitation extremes. This confirms the results of Prein et al. (2013), who also found that summer precipitation occurs in smaller areas with higher intensities. It further confirms the results of Ban et al. (2014), who found that heavy precipitation intensities associated with convection increase in CCLM at the convection-resolving scale increases.

However, this issue needs to be investigated more thoroughly in general, as there are only a few studies at the convection-resolving scale, and specifically with a focus on the dynamics of CCLM.

One aspect that arose during this investigation was the quality of gridded observational products. At the 1 km scale the quality of such data sets does have a severe influence on the comparison and on the assessment of extremes. This could cause problems in regions that have only a sparse station density, such as mountainous areas, where the quality of the observational data set depends heavily on the interpolation method. This is an issue that needs further addressing in order to find an adequate strategy how to deal with it, as the number of simulations at the convection-resolving scale is increasing due to the progress in computational resources. From this point of view, the paradigm to use observational data to validate model data could be reversed, so that high-resolution simulations might be used to judge the quality of observational products. However, therefore it has to be assumed that the model physics are correct and a single model is not sufficient. This issue might become of interest when ensemble projects, such as CORDEX reach the 1 km scale.

Our results show that the horizontal resolution does have a considerable effect on the simulation of precipitation and temperature extremes and that our high-resolution simulation adds details to the pattern and spatial variability of extreme values, which are in good agreement with observations.

This improvement can also be beneficial for impact models and assessment studies. In particular, because it is likely that climate change signals may be affected as well. This issue might be the topic of a forthcoming study.

Acknowledgments

The authors thank the CLM-Community, in particular Burkhardt Rockel (HZG Geesthacht), Andreas Will (BTU Cottbus) and Hans-Jürgen Panitz (KIT Karlsruhe), for giving advices on the model configuration and Klaus Keuler and Kai Radtke (BTU Cottbus) for providing the CCLM18 data, Merja Tölle (University of Gießen) for the REGNIE data set and Oliver Gronz (University of Trier) for preparing the InterMet data set. We wish to thank the DKRZ for providing the computational resources for our simulations. This study was funded by the State of Rhineland-Palatinate (Research Initiative Rhineland-Palatinate) and carried out within the project *Impacts of Global Change on biological resources, legislation and standards* (http://www.uni-trier.de/index.php?id=40193&L=2) of the University of Trier. Finally, we thank two anonymous reviewers who helped significantly to improve this manuscript.

References

Ban, N., J. Schmidli, C. Schär, 2014: Evaluation of the convection-resolving regional climate modeling approach in decade-long simulations. – J. Geophys. Res. Atmos. **119**, 7889–7907, DOI: 10.1002/2014JD021478.

Boberg, F., P. Berg, P. Thejll, W.J. Gutowski, J.H. Christensen, 2010: Improved confidence in climate change projections of precipitation further evaluated using daily statistics from ENSEMBLES models. – Climate Dyn. **35**, 1509–1520, DOI: 10.1007/s00382-009-0683-8.

Coles, S., 2001: An Introduction to Statistical Modeling of Extreme Values. – Springer, Berlin, 224.

De Haan, L., A. Ferreira, 2006: Extreme Value Theory: An Introduction. – Springer, New York, 436.

Dobler, L., A. Hinterding, N. Gerlach, 2004: INTERMET-Interpolation stündlicher und tagesbasierter meteorologischer Parameter. Gesamtdokumentation. – Institut für Geoinformatik der Westfälischen Wilhelms-Universität Münster i.A. Landesamt für Wasserwirtschaft Rheinland-Pfalz Manual.

Doksum, K.A., G.L. Sievers, 1976: Plotting with Confidence: Graphical Comparisons of Two Populations. – Biometrika **63**, 421–434, DOI: 10.2307/2335720.

Doms, G., M. Baldauf, 2015: A Description of the Nonhydrostatic Regional COSMO-Model Part I: Dynamics and Numerics. – Consortium for Small-scale Modelling, Deutscher Wetterdienst, Offenbach, Germany.

Doms, G., J. Förster, E. Heise, H. Herzog, M. Raschendorfer, T. Reinhardt, B. Ritter, R. Schrodin, J.-P. Schulz, G. Vogel, 2011: A Description of the Nonhydrostatic Regional COSMO-Model Part II: Physical Parameterization. Consortium for Small-scale Modelling, Deutscher Wetterdienst, Offenbach, Germany.

Feldmann, H., B. Früh, C. Kottmeier, H.-J. Panitz, G. Schädler, 2010: Hochauflösende regionale Simulationen künftiger Starkniederschlagsereignisse inBaden-Württemberg (ReSiPrec). – Technical report, IMK-TRO, KIT.

Früh, B., H. Feldmann, H.-J. Panitz, G. Schädler, D. Jacob, P. Lorenz, K. Keuler, 2010: Determination of precipitation return values in complex terrain and their evaluation. – J. Climate **23**, 2257–2274, DOI: 10.1175/2009JCLI2685.1.

Ghil, M., P. Yiou, S. Hallegatte, B.D. Malamud, P. Naveau, A. Soloviev, P. Friederichs, V. Keilis-Borok, D. Kondrashov, V. Kossobokov, O. Mestre, C. Nicolis, H.W. Rust, P. Shebalin, M. Vrac, A. Witt, I. Zaliapin, 2011: Extreme events: dynamics, statistics and prediction. – Nonlin. Processes Geophys. **18**, 295–350.

Giorgi, F., 2006: Regional climate modeling: Status and perspectives. – Journal de Physique IV (Proceedings) **139**, 101–118, DOI: 10.1051/jp4:2006139008.

Giorgi, F., C. Jones, G.R. Asrar, 2009: Addressing climate information needs at the regional level: the CORDEX framework. – Bull. World Meteor. Org. **58**, 175–183.

Gutjahr, O., G. Heinemann, 2013: Comparing precipitation bias correction methods for high-resolution regional climate simulations using COSMO-CLM. – Theo. Appl. Climatol.**114**, 511–529, DOI: 10.1007/s00704-013-0834-z.

Haylock, M.R., N. Hofstra, A.M.G.K. Tank, E.J. Klok, P.D. Jones, M. New, 2008: A european daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. – J. Geophys. Res. **113**, D20119, DOI: 10.1029/2008JD010201.

Heise, E., B. Ritter, R. Schrodin, 2006: Operational implementation of the multilayer soil mode. – Technical Report No. **9**, Consortium for Small-Scale Modelling (COSMO), Deutscher Wetterdienst, Offenbach, Germany.

Hohenegger, C., P. Brockhaus, C. Schär, 2008: Towards climate simulations at cloud-resolving scales. – Meteorol. Z. **17**, 383–394, DOI: 10.1127/0941-2948/2008/0303.

IPCC, 2012: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Group I and II of the Intergovernmental Panel on Climate Change. – Cambridge University Press, Cambridge, UK and New York, USA, 582 pp.

Jacob, D., J. Petersen, B. Eggert, A. Alias, O.B. Christensen, L.M. Bouwer, A. Braun, A. Colette, M. Déqué, G. Georgievski, E. Georgopoulou, A. Gobiet, L. Menut, G. Nikulin, A. Haensler, N. Hempelmann, C. Jones, K. Keuler, S. Kovats, N. Kröner, S. Kotlarski, A. Kriegsmann, E. Martin, E. v. Meijgaard, C. Moseley, S. Pfeifer, S. Preuschmann, C. Radermacher, K. Radtke, D. Rechid, M. Rounsevell, P. Samuelsson, S. Somot, J.-F. Soussana, C. Teichmann, R. Valentini, R. Vautard, B. Weber, P. Yiou, 2014: EURO-CORDEX: new high-resolution climate change projections for european impact research. – Reg. Env. Change **14**, 563–578, DOI: 10.1007/s10113-013-0499-2.

Kendon, E.J., N.M. Roberts, C.A. Senior, M.J. Roberts, 2012: Realism of rainfall in a very high-resolution regional climate model. – J. Climate **25**, 5791–5806, DOI: 10.1175/JCLI-D-11-00562.1.

Kessler, 1969: On the distribution and continuity of water substance in the atmospheric circulations. – Meteorological Monographs 10, American Meteorological Society, Boston.

Keuler, K., K. Radtke, G. Georgievski, 2012: Summary of evaluation results for COSMO-CLM version 4.8.CLM13 (CLM17): Comparison of three different configurations over europe driven by ECMWF reanalysis data ERA40 for the period 1979–2000. – Technical report, Brandenburg University of Technology, Cottbus.

Knote, C., G. Heinemann, B. Rockel, 2010: Changes in weather extremes: Assessment of return values using high resolution climate simulations at convection-resolving scale. – Meteorol. Z. **19**, 11–23.

Leadbetter, M.R., I. Weissmann, L. de Haan, H. Rootzén, 1989: On clustering of high values in statistically stationary series. – Technical Report **253**, Center for Stochastic Processes, University of North Carolina, Chapel Hill.

Martins, E.S., J.R. Stedinger, 2000: Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. – Water Resour. Res. **36**, 737–744, DOI: 10.1029/1999WR900330.

Mellor, G.L., T. Yamada, 1974: A hierarchy of turbulence closure models for planetary boundary layers. – J. Atmos. Sci. **31**, 1791–1806, DOI: 10.1175/1520-0469(1974)031<1791:AHOTCM>2.0.CO;2.

MUFV, 2007: Klimabericht Rheinland-Pfalz 2007. – Ministerium für Umwelt, Forsten und Verbraucherschutz Rheinland-Pfalz, 98 pp.

Pickands, J., 1975: Statistical inference using extreme order statistics. – The Annals of Statistics **3**, 119–131, DOI: 10.2307/2958083.

Prein, A.F., A. Gobiet, M. Suklitsch, H. Truhetz, N.K. Awan, K. Keuler, G. Georgievski, 2013: Added value of convection permitting seasonal simulations. – Climate Dyn. **41**, 2655–2677, DOI: 10.1007/s00382-013-1744-6.

Prein, A.F., W. Langhans, G. Fosser, A. Ferrone, N. Ban, K. Goergen, M. Keller, M. Tölle, O. Gutjahr, F. Feser, E. Brisson, S. Kollet, J. Schmidli, N.P.M. van Lipzig, R. Leung, 2015: A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges. – Rev.Geophys. 2014RG000475, DOI: 10.1002/2014RG000475.

Rauthe, M., H. Steiner, U. Riediger, A. Mazurkiewicz, A. Gratzki, 2013: A central european precipitation climatology – Part I: Generation and validation of a high-resolution gridded daily data set (HYRAS). – Meteorol. Z. **22**, 235–256, DOI: 10.1127/0941-2948/2013/0436.

Richter, D., 1995: Ergebnisse methodischer Untersuchungen zur Korrektur des systematischen Messfehlers des Hellmann-Niederschlagmessers. – Berichte des Deutschen Wetterdienstes, Offenbach am Main, 194 pp.

Ritter, B., J.-F. Geleyn, 1992: A comprehensive radiation scheme for numerical weather prediction models with potential applications in climate simulations. – Mon. Wea. Rev. **120**, 303–325, DOI: 10.1175/1520-0493(1992)120<0303:ACRSFN>2.0.CO;2.

Rockel, B., A. Will, A. Hense, 2008: The regional climate model COSMO-CLM(CCLM). – Meteorol. Zet **17**, 347–348.

Rummukainen, M., 2010: State-of-the-art with regional climate models. – Wiley Interdisciplinary Reviews: Climate Change **1**, 82–96, DOI: 10.1002/wcc.8.

Schättler, U., G. Doms, C. Schraff, 2009: A Description of the Nonhydrostatic Regional COSMO-Model Part VII: User's

Guide. – Consortium for Small-Scale Modelling Cosmo Model 4.11.

Schrodin, R., E. Heise, 2001: The Multi-Layer version of the DWD soil model TERRA-ML. – Technical Report No. **2**, Consortium for Small-Scale Modelling (COSMO), Deutscher Wetterdienst, Offenbach, Germany.

Seo, C., J.H. Thorne, L. Hannah, W. Thuiller, 2009: Scale effects in species distribution models: implications for conservation planning under climate change. – Biol. Lett. **5**, 39–43, DOI: 10.1098/rsbl.2008.0476 PMID: 18986960.

Steppeler, J., G. Doms, U. Schättler, H.W. Bitzer, A. Gassmann, U. Damrath, G. Gregoric, 2003: Meso-gamma scale forecasts using the nonhydrostatic model LM. – Meteor. Atmos. Phys. **82**, 75–96, DOI: 10.1007/s00703-001-0592-9.

Tiedtke, M., 1989: A comprehensive mass flux scheme for cumulus parameterization in large-scale models. – Mon. Wea. Rev. **117**, 1779–1800.

Tölle, M.H., O. Gutjahr, G. Busch, J.C. Thiele, 2014: Increasing bioenergy production on arable land: Does the regional and local climate respond? Germany as a case study. – J. Geophys. Res. Atmos. **119**, 2711–2724, DOI: 10.1002/2013JD020877.

Uppala, S.M., P.W. Kållberg, A.J. Simmons, U. Andrae, V.D.C. Bechtold, M. Fiorino, J.K. Gibson, J. Haseler, A. Hernandez, G.A. Kelly, X. Li, K. Onogi, S. Saarinen, N. Sokka, R.P. Allan, E. Andersson, K. Arpe, M.A. Balmaseda, A.C.M. Beljaars, L.V.D. Berg, J. Bidlot, N. Bormann, S. Caires, F. Chevallier, A. Dethof, M. Dragosavac, M. Fisher, M. Fuentes, S. Hagemann, E. Hólm, B.J. Hoskins, L. Isaksen, P.A.E.M. Janssen, R. Jenne, A.P. Mcnally, J.-F. Mahfouf, J.-J. Morcrette, N.A. Rayner, R.W. Saunders, P. Simon, A. Sterl, K.E. Trenberth, A. Untch, D. Vasiljevic, P. Viterbo, J. Woollen, 2005: The ERA-40 re-analysis. – Quart. J. Roy. Meteor. Soc. **131**, 2961–3012, DOI: 10.1256/qj.04.176.

Weisman, M.L., W.C. Skamarock, J.B. Klemp, 1997: The resolution dependence of explicitly modeled convective systems. – Mon. Wea. Rev. **125**, 527–548, DOI: 10.1175/1520-0493(1997)125<0527:TRDOEM>2.0.CO;2.

Wernli, H., M. Paulat, M. Hagen, C. Frei, 2008: SAL – a novel quality measure for the verification of quantitative precipitation forecast. – Mon. Wea. Rev. **136**, 4470–4487.

Wicker, L.J., W.C. Skamarock, 2002: Time-Splitting Methods for Elastic Models Using Forward Time Schemes. – Mon. Wea. Rev. **130**, 2088–2097, DOI: 10.1175/1520-0493(2002)130<2088:TSMFEM>2.0.CO;2.

Zwiers, F.W., V.V. Kharin, 1998: Changes in the extremes of the climate simulated by CCC GCM2 under CO_2 doubling. – J. Climate **11**, 2200–2222, DOI: 10.1175/1520-0442(1998)011<2200:CITEOT>2.0.CO;2.