FUTURE CHANGE OF PRECIPITATION EXTREMES IN EUROPE: AN INTERCOMPARISON OF SCENARIOS FROM REGIONAL CLIMATE MODELS

Christoph Frei, Regina Schöll, Sophie Fukutome, Jürg Schmidli and Pier Luigi Vidale

Atmospheric and Climate Science, ETH, Zürich, Switzerland

Journal of Geophysical Research - Atmospheres

submitted

10. March. 2005

corresponding author:
Christoph Frei, Atmospheric and Climate Science ETH,
Winterthurerstrasse 190, CH – 8057 Zürich, Switzerland
Tel: +41 1 635 52 32, Fax: +41 1 362 51 97,
email: christoph.frei@env.ethz.ch

ABSTRACT

An analysis is undertaken of the climate of precipitation extremes as simulated by six European regional climate models (RCMs) in order to describe/quantify future changes and to examine/interpret differences between models. Each model has adopted boundary conditions from the same ensemble of GCM integrations for present (1961-1990) and future (2071-2100) climate under the A2 emission scenario. The main diagnostics are multi-year return values of daily precipitation totals, estimated from extreme value analysis. An evaluation of the RCMs against observations in the Alpine region shows that model biases for extremes are comparable to, or even smaller than for wet-day intensity and mean precipitation.

In winter, precipitation extremes tend to increase north of about 45°N while there is an insignificant change or a decrease to the south. In Northern Europe the 20-year return value of future climate corresponds to the 40-100 year return value of present climate. There is a good agreement between the RCMs, and the simulated change is similar to a scaling of present-day extremes by the change in average events. In contrast, there are large model differences in summer when RCM formulation contributes significantly to scenario uncertainty. The model differences are well explained by differences in the basic frequency and intensity process, but the scaling of present-day extremes would yield much smaller increases or larger decreases than those simulated. There is evidence for a component of the change that affects extremes specifically and is consistent between models, despite the large variation in the total response.

1) Introduction

An accumulating body of evidence suggests that the increase of atmospheric greenhouse-gas concentrations could increase the frequency of heavy precipitation in many regions of the globe: Physical considerations imply that the sensitivity of heavy precipitation may be determined primarily by the change in the atmospheric moisture transport capacity (governed by the Clausius-Clapeyron relation) rather than the change in mean precipitation and evaporation (Trenberth 1999; Allen and Ingram 2002; Trenberth et al. 2003). The moistening of the atmosphere could result in progressively larger frequency increases at high precipitation intensities, and increases could even occur in regions where mean precipitation decreases (Katz and Acero 1994; Frei et al. 1998; Groisman et al. 1999). Consistent with these conceptual considerations, recent global warming experiments with general circulation models (GCMs) show an increase of precipitation extremes over many areas of the globe (e.g. Kharin and Zwiers 2000; Palmer and Räisänen 2000; Semenov and Bengtsson 2002; Voss et al. 2002; Kiktev et al. 2003, 2004; Watterson and Dix 2003; Wehner 2004). The details of the distribution and the magnitude of the change vary between models, but there is similarity in that increases are found predominantly over land areas of the mid and high latitudes.

Results from GCMs may be considered with some reservation as regards the subcontinental pattern and magnitude of the change. The coarse grid-spacing poses limitations to the explicit simulation of mesoscale processes and to the representation of topography and land-sea distribution. Regional climate models (RCMs) are promising tools, which, when nested into a GCM, permit the derivation of GCM-consistent climate change scenarios with more regional detail and a more trustworthy representation of processes active during heavy precipitation. Experiments with the perfect-model approach have demonstrated the ability of the one-way nesting technique in reproducing the fine-scale features of atmospheric fields in areas where surface forcing is strong (e.g. Denis et al. 2002; De Elia and Laprise 2003). Also, RCMs were found to reproduce the main characteristics of the

larger-scale hydroclimate during episodes of heavy precipitation (Anderson et al. 2003) and prominent patterns of precipitation extremes on scales not resolved by a current GCM (Frei et al. 2003; Fowler et al. 2005). However biases are quite large in some cases. (Note, that there are alternative approaches to climate change downscaling using statistical techniques (e.g. Wilby et al. 1998; Goodess et al. 2005), but we focus on dynamical downscaling, i.e. RCMs, in this study.)

For the European continent, a number of recent climate-change simulations with regional climate models have been analysed for future changes in precipitation extremes. One group of studies considers direct empirical diagnostics such as quantiles or seasonal/annual extremes: For example, Durman et al. (2001) find that the fraction of precipitation exceeding the 99-percentile of daily values increases in their RCM by several tens of percent by the end of the 21st century under a 1% per year CO_2 increase. In a different RCM, Christensen and Christensen (2003, 2004) find an increase of very high quantiles even in summer and for regions in Central Europe, where mean precipitation decreases. Pal et al. (2004) find similar results in their RCM. Räisänen et al. (2004) document an increase in annual precipitation extremes for two RCM experiments with different GCMs, but details of the geographical pattern of the change was different between the two experiments.

Another group of studies has adopted techniques of extreme value analysis to estimate the change in events with return periods of several years: Using data from two RCMs, Booij (2002) estimates a 25-60% increase (depending on model) in the 20-year return period one-day rainfall in an area of north-western continental Europe by the time of CO_2 doubling. For the area of the UK, several applications of extreme value statistics have been undertaken based on the Hadley Centre climate model chain (Jones and Reid 2001; Huntingford et al. 2003). In a recent version of this RCM, Ekström et al. (2005) find a 10% increase of one-day precipitation extremes with return periods of 10-50 years across the UK and more regionally variable changes for 10-day precipitation extremes. Finally, Semmler and Jacob (2004) report, with their RCM, an increase of annual rainfall extremes over most parts of the European continent with particularly large absolute changes over the Baltic Sea area and the central Mediterranean.

In all these published RCM results, there is a common tendency for increases in European precipitation extremes, but there appears to be considerable inter-model variation in the distribution and magnitude of the change. A quantitative comparison of the published results is however difficult because of differences in the diagnostics and the techniques with which they were estimated. In principle, inter-model differences in scenarios can arise from the use of different emission scenarios and GCMs and from differences in RCM formulation and technical specifications. The role of these factors for scenarios of mean surface climate is examined in Déqué et al. (2005).

The purpose of this study is to compare scenarios of European precipitation extremes for the late 21st century between six different RCMs using consistent diagnostics. The idea is to isolate the contribution to scenario uncertainty which is due to differences in the formulation of the regional models. Accordingly, all the RCM simulations being analysed here are based on the same emission scenario (SRES A2, Nakicenovic et al. 2000), are nested into the same global climate model (HadAM3H; Pope et al. 2000) and are operated at a comparable grid spacing. Clearly, our analysis satisfies some obvious interest in scenarios of precipitation extremes. But its results are also relevant for the design of multi-model ensembles, when it comes to estimating the full range of scenario uncertainty. High sensitivity of scenarios to RCM formulation may suggest the consideration of several different RCMs nested in the same GCM, whereas a low sensitivity may suggest that computational resources are used more efficiently in sampling GCM formulation, i.e. by nesting RCMs into several different GCMs.

The diagnostics of primary focus in this analysis are extremes of rainfall with return periods between 5 and 50 years. Their estimation is based on the technique of extreme value statistics (see e.g. Coles 2001; Katz et al. 2002) very similar to the studies mentioned above. Here, this method is applied consistently to all the RCMs and results are compared quantitatively for specific regions. In addition, we also consider more direct diagnostics of average or intense events, which allows us to describe a wider range of the frequency distribution and to employ a simple scaling concept to interpret changes for rare extremes.

All our analyses are carried out seasonally stratified in order to identify seasonal variations in scenarios and uncertainties (see also Wehner 2004).

One part of this study is also devoted to an evaluation of the RCMs with respect to their representation of precipitation extremes. For this purpose, the European Alps are used as a test ground. This region has at its disposal a very dense rain-gauge network from which an accurate observational dataset could be created, which is compatible with the grid-spacing of the models. Although the Alps cover only a limited part of the full model domains (typically 25x15 grid points), they are particularly interesting for assessing downscaling abilities because of the complex topography. Also the evaluation in the Alps complements previous evaluation studies that have focused on more northern parts of Europe (Booij 2002; Huntingford et al. 2003; Semmler and Jacob 2004; Fowler et al. 2005).

The RCM integrations considered in this study were derived as part of a larger multi-model ensemble in the frame of the European project PRUDENCE (Christensen et al. 2005). The present analysis forms part of an even broader intercomparison of downscaling methods for extremes, involving statistical and dynamical methods, in the frame of the European project STARDEX (Goodess 2003).

The outline of this paper is as follows: Section 2 describes the statistical procedures used in the analysis of the regional climate models, which are introduced in section 3. Results of the model evaluation in the Alpine region are discussed in section 4. The scenarios of precipitation extremes are presented and interpreted in section 5. Finally, section 6 summarizes the results and draws conclusions.

2) Statistical analysis

The statistical analysis of this study operates on datasets of 24-hour precipitation totals, simulated by six regional climate models, each for a time slice of present climate (1961-1990, also referred to as CTRL) and future climate (2071-2100, SCEN). (Details of the experiments and models are described in section 3.) For 3 of the models an ensemble of 3 integrations is

available for both time slices, and these are dealt with simply as a 90-year sample of the corresponding time slice. The ensembles help to reduce estimation errors due to inter-annual climate variability. All analyses were conducted directly with the data on the native model grids.

We consider a range of different diagnostics with the aim of sampling the frequency distribution of precipitation from moderate to extreme intensities (Table 1). The climatological mean precipitation, wet-day frequency, and mean wet-day intensity are basic diagnostics of the precipitation occurrence and intensity process. (1 mm per day is used as wet-day threshold.) They will be used primarily for comparison and interpretation of the results for extremes. In addition, several precipitation quantiles are considered (Table 1), describing the range from moderate to intense precipitation. Calculated for wet days only, these quantiles describe the precipitation intensity distribution, independently from the wet-day frequency. Basic diagnostics and quantiles were calculated using a modified version of the STARDEX diagnostic software tool (Haylock 2002; Schmidli and Frei 2005).

The diagnostics of primary focus in this study are return values of precipitation intensities with an average recurrence of 5, 10, 20 and 50 years (Table 1). Essentially these are quantiles of the far tail of the frequency distribution. The return value for a return period of T years is defined as the precipitation intensity that is exceeded with a probability of 1/T in one year (or season of the year). These diagnostics are examined for one-day (x1d) and for five-day (x5d) precipitation extremes, characteristic, respectively, for short-term heavy precipitation (possibly of a convective nature) and extended heavy rainfall periods (related to synoptic disturbances or persistent flow conditions). In Europe, impacts from heavy precipitation are mostly due to short period rainfalls in summer and multi-day episodes in winter, and this is why we focus on the results for one-day and five-day extremes in summer and winter respectively.

Like in several previous studies concerned with extremes in climate models (e.g. Zwiers and Kharin 1998; Arora and Boer 2001; Voss et al. 2002), we employ the technique of *extreme value analysis* in estimating return values for precipitation extremes. Briefly, the technique is based on an asymptotic theory on the statistics of the maximum (or minimum) of

a sample of independent, identically distributed variables (Fisher and Tippett 1928; Gnedenko 1943). Under fairly general conditions and in the limit of a large sample size, the extreme is distributed like the Generalized Extreme Value distribution (GEV). Gumbel (1958) pioneered the practical use of this theory, *the block-maximum technique*, which consists in (a) the extraction of extremes from large blocks of observations, (b) the fitting of an appropriate limit distribution (the GEV) and finally (c) the calculation of quantiles from that distribution. (An alternative approach, *the peaks-over-threshold technique*, is not considered in this study, because of difficulties anticipated with declustering and with the definition of thresholds that are suitable across models and regions with very different climates (e.g. Palutikof et al. 1999).) The GEV is a three-parameter distribution family with a location, scale and shape parameter. Positive (negative) shapes describe situations where extremes have an upper (lower) bound. (Note that we use the sign convention for shape as in Zwiers and Kharin (1998), which is opposite to that in Coles (2001).) More details about extreme value analysis (block-maximum and peak-over-threshold approaches) are described for example in Coles (2001). Palutikof et al. (1999) review applications for wind extremes, and Katz et al. (2002) discuss applications in hydrology using covariates.

The details of our specific application of the block-maximum approach are as follows:

Data selection:

Maxima of one-day and five-day precipitation intensities are extracted from each season of the 30 (90) years in the two time slices. The analysis is carried out independently for each model grid point and for each season of the year (winter: DJF, spring: MAM, summer: JJA, autumn: SON). For some areas in the Mediterranean region the regional models have simulated extended dry periods in summer, so that precipitation maxima were found to be zero or small in some years. Such 'maxima' cannot be considered as being taken from a large sample of data of the precipitation process, and return values estimated from such samples must be considered unreliable. Therefore, an extreme value analysis was only performed for grid-points where at least 15 seasonal maxima larger than the intensity of 5

mm per day were simulated in a time slice. Grid-points not meeting this criterion (mostly in Southern Europe in summer) will be masked out in our result plots.

Estimation:

Estimates of the GEV distribution parameters are calculated by the method of Maximum Likelihood. However, in this study we use a modified form of the classical GEV likelihood function (see e.g. Coles 2001), which includes a Bayesian prior distribution for the GEV shape parameter, exactly as is proposed by Martins and Stedinger (2000).

It has been demonstrated that Maximum Likelihood Estimation using the standard GEV likelihood function can be unstable for small samples (varying from 15 to 100) and produce absurd values of the shape parameter (Hosking et al. 1985; Martins and Stedinger 2000). Indeed, in our application (with sample sizes of 30 or 90) quite many incidents of unrealistic shape estimates were found when using the standard likelihood function. Two selected cases are illustrated in Fig. 1. In the first case (Fig. 1a), the standard method estimates a very heavy tail (shape value: –0.52) because of an outlier value in the sample. In the second case (Fig 1b), the GEV fit levels off and suggests an upper bound near 35 mm per day (shape value: +0.55). Both of these standard estimates appear unrealistic and are not supported by the distributions estimated at adjacent grid points.

The purpose of the geophysical prior distribution is to reduce the likelihood for estimating extreme shape values that are unrealistic in geophysical applications. Fig. 2 displays the prior distribution that Martins and Stedinger (2000) propose for hydrological applications and which is used throughout in this study. The distribution constrains values of the shape to essentially the range (–0.3, +0.15) and it totally prevents estimates outside (–0.5, +0.5). The prior distribution is biased towards negative shapes. Maximum density is obtained for a shape of –0.1. This is justified for hydrological applications where lower bounds of extremes and hence heavier tails than the Gumbel distribution are very common. (See also later.) Using sample sizes of 25 to 100, Martins and Stedinger (2000) show, that for GEVs with a negative shape, the modified likelihood function leads to much more accurate quantile

estimates (in terms of root mean square error) compared to the standard likelihood, the moment and the L-moment (Hosking 1990, 1992) estimators. Moreover, in many cases, the bias is smaller or at least comparable to that of other estimators.

In Figure 1, GEV distributions estimated from the modified likelihood function are also depicted. The resulting distributions seem physically more meaningful (shape parameters are –0.27 for case a, and 0.12 for case b). The modification influences, in particular, quantiles at return periods of 10 years and more. The stabilizing effect of the geophysical prior is also demonstrated in Figure 3, showing the spatial distribution of the shape parameter for one of the models' control integrations. When estimated with the standard likelihood function (Fig. 3a), the shape shows an irregular pattern, with many anomalous values even over the ocean. More than 5% of the grid points exhibit shape values outside the range (–0.4, +0.4) in this example. In contrast, estimates with the modified likelihood function (Fig. 3b) are much less variable and smoother in distribution. Note also, that in case of the standard estimation, negative shape values were estimated for more than 70% of the grid points and that the median across Europe is –0.09. This is close to the density maximum of the prior distribution and it justifies the choice of a prior distribution that is shifted to negative values.

Confidence intervals and statistical tests

Two different methods are used for statistical assessment of our estimates in return values: On a grid-point basis, likelihood confidence intervals were calculated directly from the observed information matrix of the optimization process using the modified likelihood function (see e.g. Coles 2001). A confidence interval for the difference in return values between the CTRL and SCEN sample is then derived from the standard errors in each sample, assuming normal distribution of errors. This provides a statistical test for the change. Note that the asymptotic properties of likelihood confidence intervals (symmetry and normal errors) may not be satisfied with the small samples considered. We therefore view the results only as an approximate indication of the gross pattern of statistical significance across the continent.

A more accurate, bootstrap based, estimation of confidence intervals is used for spatial averages of return values across selected subdomains. (The subdomains used are depicted in Fig. 4.) Bootstrap samples of domain mean return values were generated by resampling of years (non-parametric bootstrap). In order to preserve the spatial correlation of errors, all grid-points within the subdomain are sampled from the same years (see e.g. Wilks 1997). GEVs are then estimated for each grid-point sample and return values averaged over the subdomain. 50 bootstrap samples were generated for each time slice. Confidence intervals for the relative change in return values between CTRL and SCEN are then obtained by resampling between the pairs of bootstrap samples.

3) Models and experiment

The RCM integrations analyzed in this study were conducted by nesting into the atmosphere-only GCM (HADAM3H) of the Hadley Centre at the UK Meteorological Office. (One RCM is also nested into HADAM3P, a more recent version of the same GCM. See later.) HadAM3H was derived from the coupled atmosphere-ocean model HadCM3 (Gordon et al., 2000; Johns et al., 2003) and is described in Pope et al. (2001). The forcing fields for the RCMs came from GCM integrations at a resolution of about 150 km in mid-latitudes for the time-slices 1961-1990 (CTRL) and 2071-2100 (SCEN). For CTRL, the sea surface temperature and sea ice distributions for HADAM3H were prescribed from observations of the same period (i.e. an AMIP-type integration). For SCEN, seas surface conditions were constructed from observations and anomalies from a transient integration of HADCM3 using the IPCC SRES A2 emission scenario (Nakicenovic et al. 2000). With this scenario HADAM3H simulated a global mean surface temperature increase of 3.18 K between CTRL and SCEN (D. Rowell, personal communication, 2004). This is in the upper half of the warming range predicted by the IPCC (Cubasch et al. 2001) and corresponds approximately to the 55% quantile of the probabilistic prediction of Wigley and Raper (2001) for the warming between 1990 and the end of the 21st century.

Three integrations were carried out with HADAM3 for both time slices, starting from different initial conditions. Two of the considered RCMs were integrated from all six members. The remaining four RCMs ran for only one (but the same) pair of ensemble members. Consideration of ensemble integrations is particularly valuable in the analysis of extremes, because quantiles can be estimated from a larger sample and uncertainty arising from interannual variations is reduced.

The RCMs considered in this study are operated at a resolution of about 50 km, with a comparable domain, covering the European continent from the Mediterranean to Scandinavia and from Iceland to the Black Sea. (Domains for some of the models are depicted in Frei et al. (2003).) All six RCMs are state-of-the-art limited-area climate models with one-way nesting over a lateral boundary zone. (The boundary zone is excluded from all analyses and displays.) Table 2 gives a list of the RCMs with acronyms, basic characteristics and references to more technical descriptions. It may be of interest to mention that there are relationships between some of the models: CHRM and REMO share the same dynamical core. CHRM and GKSS have very similar physical parameterizations. The same is the case for HIRHAM and REMO.

For the regional model of the Hadley Centre, two model versions (HADAM3H and HADAM3P) will be considered in our analysis. The more recent model (HADRM3P) uses a newer version of physical parameterizations (mainly with effects on the vertical cloud profile). However the integration with the newer model is obtained with the corresponding new version of the atmospheric GCM (HADAM3P), and hence is not strictly in our common setting. Results of HADAM3P will therefore not be displayed as extensively as for the other models, but eventual comparison to its predecessor is interesting, primarily because it is the new version, which is used in current impact studies.

The RCM experiments used in this study form part of an even larger collection of downscaling experiments, conducted in EU project PRUDENCE (Christensen et al. 2005). The large effort required with the handling of daily datasets, with extreme value analysis and resampling experiments has prevented us from considering all available experiments. A

comparison of scenarios for mean seasonal surface climate with all PRUDENCE RCMs can, however, be found in Déqué et al. (2005).

4) Evaluation in the Alpine region

This section presents an evaluation of the climate of precipitation extremes as simulated by the CTRL integrations of the RCMs. Results are discussed/depicted only for a selection of the diagnostics considered in this study (Tab. 1), but the selection provides a representative picture of the models' behavior. The evaluation is conducted for the European Alps, a 1100 x 700 km^2 region, encompassing typically 25x15 model grid points. The Alps are located well in the interior of all the model domains, usually slightly south of the domain centers. This high-mountain area is an ambitious but interesting test ground as it illustrates the downscaling ability of RCMs. Some care should however be exercised in the interpretation of the results because discrepancies to observations may also be due to errors in the driving GCM, not only the RCMs themselves.

Evaluation Dataset

The observational reference for the present study is very similar to that used in a previous evaluation of reanalysis-driven RCMs in Frei et al. (2003). In summary, it is a gridded precipitation analysis for every day of the 20 years 1971–1990, based on data from the operational high-resolution rain-gauge networks, encompassing more than 6500 station records. The analysis grid has a resolution comparable to the grid spacing of the RCMs, and the analysis scheme estimates averages of the rainfall intensity over grid pixels, ensuring compatibility of the resulting statistics with the resolution of the models (see e.g. Osborn and Hulme 1997; Frei et al. 2003). 10–50 station values contribute to the analysis at each grid point. In deviation from Frei et al. (2003), the analysis for this study was performed with a climatological scaling, similar to Widmann and Bretherton (2000), using the climatology of Schwarb et al. (2001). This procedure reduces the sensitivity of the analysis to biases in

station distribution. The diagnostics of Table 1 were then determined from the observed daily analysis as for the models. Unfortunately, it was not possible to cover a 30-year period as in the models, because of limited data availability in the 1960ies and problems in data quality in the 1990ies. More details about the dataset and analysis technique are given in Frei and Schär (1998) and Frei et al. (2003).

Results:

In the Alps, the highest frequency of heavy precipitation occurs in autumn. It is therefore natural to have a special focus on this season first. Figure 5 compares the distribution of the 5-year return value of one-day precipitation extremes (x1d.5, see Table 1) between observations (top left) and models. In autumn, heavy precipitation is frequently associated with moist and weekly-stratified airflows from the south, often with embedded convection. Accordingly, large values of x1d.5 are observed along the southern rim of the ridge, with a characteristic mesoscale pattern (note areas exceeding 120 mm per day), reflecting topographic detail of the ridge and the proximity of the ocean. All RCMs reproduce the general southern rim pattern quite well and several models show features similar to the observed peaks, although eventually shifted by a few grid points (e.g. the Massif Central maximum in CHRM and GKSS). HADRM3H and HADRM3P show very similar distributions. Both tend to overestimate the topographic enhancement at the southern rim but underestimate the return values in the foreland (the Po valley). HIRHAM and SMHI show an overall underestimate and GKSS tends to overestimate return values.

It is interesting to note that, despite the rareness of events considered, the correspondence of the x1d.5 pattern with observations is only slightly lower than that found in a previous evaluation for the 90% quantile (event recurrence one month) and for reanalysis-driven RCMs (Frei et al. 2003). Apparently the regional pattern of the heavy precipitation climate in this season is primarily determined by the topography and land-sea distribution. This pattern is not trivial and it has a mesoscale dimension, which we would not expect to be resolved by the parent GCM. This evaluation demonstrates that RCMs have

considerable downscaling ability for precipitation extremes and, indirectly, it attests to the quality of the GCM (HadAM3H and HadAM3P) in reproducing, at least in this season, the observed climate of large-scale flow conditions relevant for precipitation extremes in the Alps.

In a visual comparison of x1d.5 for winter and spring (not shown), we found a similar skill of the RCMs to that found in autumn. But the biases and inter-model differences were largest in summer. Figure 6 shows the seasonal variation in some of the precipitation diagnostics averaged over 2 Alpine subdomains. (Domain definitions are displayed in Fig. 4.) For x1d.5 (Fig. 6e, f) the observed intra-season and across-ridge variations are reasonably reproduced by individual models although there are, in cases, substantial biases. An exception to this is summer in the southern Alpine region, where all models (except GKSS) simulate smaller values of x1d.5 than in winter and spring, but in the observations it is larger. In the other seasons, however, the biases are more model-specific, with both HADRM3 models and the GKSS tending to overestimate and CHRM, HIRHAM, SMHI and REMO tending to underestimate. Note that, even though the uncertainty in x1d.5 from interannual variations is quite large, most of the model biases well exceed the 90% confidence ranges, implying that the depicted model biases are not artifacts of random errors due to the short observation/simulation periods.

Interestingly, the model biases for the tail of the distribution (x1d.5) are, in relative terms, comparable to or smaller than the bias for wet-day intensity (int, Fig. 6c,d). Also the inter-model pattern of biases is very similar between int and x1d.5, suggesting that the model errors in precipitation extremes are primarily related to deficiencies in the intensity process rather than the occurrence process. Indeed, the bias pattern and seasonal variation of the wet-day frequency (fre, Fig. 6a,b) is quite different from int and x1d.5. All models overestimate wet-day frequency in the northern Alps from autumn to spring. This is likely due to errors in the driving GCM, because no similar bias was seen in reanalysis-driven integrations with the same RCMs (see Frei et al. 2003).

In summary, the present evaluation demonstrates that RCMs are capable of reproducing non-trivial mesoscale patterns of observed precipitation extremes at least during dynamically

active seasons. Nevertheless, there are model specific biases of up to several tens of percent, especially in summer. But the model performance for rare extremes is not worse than for less extreme quantiles or for mean wet-day intensity. The evaluation did not reveal previously undiscovered model deficiencies that are specific to rare extremes. Instead, it can be expected that future improvements in the modeling of the precipitation intensity process will also significantly reduce current biases for rare extremes.

5) Simulated change in precipitation extremes

This section describes the distribution and magnitude of the change in precipitation extremes as simulated by the six RCMs between the CTRL and SCEN time-slices. It discusses the coherence/variation of the change between the models, and an interpretation will be made of how the change in extreme events is related to the change in average events, which will help to identify origins of the difference between models. Detailed discussions will be presented in the first two subsections for winter and summer, the two seasons for which fundamentally different inter-model variations were found. A brief discussion of the results for the transition seasons is then provided in the last subsection.

Winter:

Figure 7 depicts the change in the 5-year return value of 5-day precipitation intensity in winter (x5d.5, DJF, see Table 1). The 5-day intensity was chosen because adverse impacts in winter are mostly due to multi-day heavy precipitation, but in fact, the result for extreme 1-day events (x1d.5, not shown) is very similar. Maps are displayed for those RCMs nested into the same GCM (HADAM3H). (The result for HADRM3P, which was nested in a more recent version of the GCM (see section 3), is very similar to that of HADRM3H.)

The continental-scale pattern of the change shows an increase over large areas north of about 45°N and small changes or a decrease south of it. Small changes of variable signs are also noted over the high latitudes of the northeastern Atlantic. For large areas of Central and

Northern Europe, the increase is statistically significant (at the 5% level) even at the scale of single model grid points. The larger-scale pattern is very similar between the models. This is also true when comparing models with three ensemble members (HADRM3H, HIRHAM) with those consisting of one member only. The longer simulation period of the former primarily results in larger areas with statistically significant changes but does not affect the overall pattern.

A quantitative comparison is provided in Table 3, where the relative change in the domain mean return value x5d.5 is listed for the three European subdomains (see Fig. 4a) and for all models (including HADRM3P). The sign of the change is consistent among the models in Scandinavia, with increases of 10-25%, and in Iberia with decreases of 7-12%. In Scandinavia, the increase is statistically significant (based on the 95% resampling confidence interval, see section 3) for all models. In Central Europe the model results vary in sign, which is also obvious from Figure 7. Here, the models with one ensemble member exhibit consistently smaller increases (even a slight decrease in the case of CHRM) compared to the three-member models (HADRM3H, HADRM3P and HIRHAM). Indeed, interannual variations mostly explain the inter-model variations in this region: The integrations of HADRM3H and HIRHAM for the particular ensemble member common to all the other models, exhibit a change of x5d.5 by 0% and +1% respectively in Central Europe, which is very similar to the results from the other models. It appears that the common pair of ensemble members had an anomalously small response in Central Europe and we could expect a fairly consistent increase in the order of about 10% if all RCMs had integrated all ensemble members.

What does the change in return values mean in terms of changes in the recurrence of precipitation extremes? This is addressed in Figure 8, which shows frequency distributions for both the CTRL and SCEN time slice for subdomain Scandinavia. The distributions are obtained by averaging the return values x5d.T over all grid points of the subdomain. Hence, Figure 8 can be considered as a Gumbel diagram for an average grid point in the subdomain. For a better visual comparison, the distributions have been standardized for each model (i.e. return values shifted and rescaled), in such a way that the CTRL distributions from all models

are approximately aligned (grey shaded area). The comparison between the two time slices reveals that the future 5-year return value corresponds to extremes with a return period of 8-18 years under present-day conditions, depending on the model. Similarly, the 20-year return value of future climate corresponds to events with a recurrence of 40-100 years under present climate. Hence, for very rare precipitation extremes the models simulate a frequency increase by a factor of 2 to 5 over Northern Europe in winter. The corresponding numbers for Central Europe are from *no change* to a frequency increase by a factor of about 2.

It is instructive to compare the change for rare extremes, obtained by means of extreme value analysis, to that for average or intense events as revealed by more basic diagnostics of the precipitation frequency distribution. Such a comparison is displayed in Figure 9 for Scandinavia and Central Europe. In each panel, the three columns to the left (fre, mea and int) describe basic properties of the precipitation occurrence and intensity process and the remaining columns (q40, ..., x5d.50) provide a section across the distribution from moderate to extreme precipitation events. (Note, that qXX are quantiles for wet days only while x5d.T are quantiles for wet and dry days (Table 1).)

Figure 9 reveals that, the structure of the change across the precipitation distribution is very coherent between the models in winter. For example, in Scandinavia and Central Europe, there is an approximately equal relative increase in wet-day frequency and intensity, which is mostly statistically significant and is reflected in an increase of mean winter precipitation. Moreover, the change in wet-day quantiles (q40 and q90) is very similar to the change in wet-day intensity. Also, the change for extreme quantiles (x5d.T), although different between models, varies very little between return periods and is similar in magnitude to the change in wet-day intensity of the corresponding model. Exceptions to this are CHRM, GKSS and REMO in Central Europe, which show a slight decrease in extreme quantiles contrasting with the increase in wet-day intensity. The general picture arising from this is that the precipitation frequency distribution has changed essentially like a simple rescaling, where the relative change in quantiles is independent of frequency. Thus the change for wet-day intensity (or that for moderate quantiles) is a reasonable estimate of the change in extreme return values.

The situation is somewhat more complex for subdomain Iberia (not shown), where extreme return values decrease more than the wet-day intensity. However, a clear interpretation of the changes for this region is difficult because of the large estimation uncertainty. In Iberia, the change in any of the basic diagnostics (fre, mea, int, qXX) is statistically not significant for all models, and only two models (HADRM3H and HADRM3P) show a significant decrease for extreme return values (see Table 3).

Let us examine to what extent the simulated change in extreme return levels can be quantitatively explained by a simple rescaling of the precipitation frequency distribution. For this purpose we derive, for each model, a hypothetical distribution function for the SCEN period by scaling the corresponding CTRL distribution. The scaling consists of the following steps: The distribution of return levels under CTRL is transformed into a conditional distribution for wet days only, using the simulated wet-day frequency of CTRL. The resulting distribution is scaled (multiplying quantiles) using the simulated relative change in wet-day intensity, and the result is considered as a conditional distribution (i.e. for wet days only) under SCEN. Finally an unconditional distribution for SCEN is determined using the simulated wet-day frequency for SCEN. This procedure is formalized in:

$$X'_{SCEN}(T) = \frac{i_{SCEN}}{i_{CTRL}} \cdot X_{CTRL}\left(\frac{f_{SCEN}}{f_{CTRL}} \cdot T\right)$$

where $X'_{SCEN}(T)$ is the scaled quantile function for SCEN, X_{CTRL} is the simulated quantile function under CTRL, f_{SCEN}/f_{CTRL} is the ratio of simulated wet-day frequencies and i_{SCEN}/i_{CTRL} the ratio of simulated wet-day intensities. Hence, the scaling exercise estimates future return levels purely from the simulated changes in wet-day frequency and intensity, assuming that these changes are independent of event intensity. A similar scaling and its effect on extreme return values was examined by Whetton et al. (1993) and by Fowler and Hennessy (1995).

Figure 10 compares the simulated change in return levels for winter with the change estimated from the simple scaling exercise. For both, the 5-year and 50-year return levels the dots are closely aligned along the diagonal line. There are two implications from this: Firstly, it appears that changes in precipitation extremes in Central and Northern Europe are

dominated by the effect of changes in wet-day frequency and intensity. Changes specific to extremes are of secondary importance in winter. Secondly, the scaling can accurately explain the variation of the change between the models. The inter-model variance explained by the scaling is larger than 90% for x1d.5 and larger than 66% for x1d.50 in both regions. This implies in turn, that inter-model differences in the change of extremes (though small anyway) are due to, primarily, differences in how individual models respond in wet-day frequency and intensity rather than in extremes per se. In fact, intensity change is the dominant factor at the tail (see also Fig. 9 and Fowler and Hennessy, 1995).

The generic picture emerging from these analyses for winter is that precipitation extremes increase over Central and Northern Europe primarily as a consequence of increases in wet-day intensity and frequency. There is a high coherency in the change between the models. Inter-model differences arise primarily from differences in the change of wet-day intensity and frequency and those are reasonably explained by interannual variations. The picture for summer is quite more complex as will be shown below.

Summer:

Figure 11 depicts the spatial distribution of the simulated change in the 5-year return value of one-day precipitation intensity. As for winter, the continental-scale pattern of the change has some similarity between the models: There is tendency for decreases over Southern Europe and increases over Northern and Eastern Europe. (Note that some areas of Southern Europe were masked out because of too few wet days. See section 3.) However, the details of the distribution are much more different between models than in winter: For example HADRM3H simulates insignificant changes over Central Europe and a decrease over Eastern Europe, whereas other models show a significant increase in these areas. Also, SMHI seems to have a generally smaller response compared to other models. Note that all models (except SMHI) exhibit a prominent increase over the Baltic Sea. This is an artifact resulting from an unrealistic increase in summer sea surface temperatures, which is due to the representation of the Baltic Sea as a lake in the coarse resolution of the coupled climate

model. This does not affect SMHI because it is coupled to a regional ocean model for the Baltic Sea (Döscher et al. 2002), and hence does not specify ocean surface conditions from the GCM there.

In contrast to winter, much larger inter-model differences are found in the simulated change of return values for summer (Table 3). In Central Europe for the 5-year return value (x1d.5), for example, results range from a statistically significant decrease by 13% (HADRM3P) to a statistically significant increase by 21% (CHRM). Similarly large variations are found for the basic diagnostics and moderate quantiles in Scandinavia and Central Europe (Fig. 12). The use of a different set of ensemble members does not explain these variations. SMHI and REMO, for example, show changes at opposite extremes in the model set, for several diagnostics in Scandinavia, although they are based on the same pair of ensemble members.

Compared to winter, summer exhibits a much more complex structure of the change across the precipitation distribution (Fig. 12). In both regions there is a decrease in wet-day frequency, mostly statistically significant, but this is compensated by an increase of wet-day intensity in many of the models. q40 decreases or changes very little in all models, but there are progressively larger increases or smaller decreases as one moves towards the tail of the distribution. In fact, for both regions, it is the return value for the longest return period (50 years) that displays the most significant increase (except HADRM3H and HADRM3P in Central Europe). Although the tendency for larger increases with longer return periods is a general behavior of all models, the critical frequency at which the decrease turns into an increase as one moves towards the tail, varies considerably between models (and also between the two regions). In Central Europe, for example, HIRHAM changes sign at moderate precipitation intensities (between q40 and q90), but HADRM3H has increases only for very rare events with a recurrence of more than 10 years. Clearly, in contrast to winter, the pattern of change does not look like a simple scaling of quantiles. For Scandinavia, for example, SMHI simulates a decrease in wet-day frequency and intensity, but extreme return values increase nevertheless.

More insight into the peculiar behavior at the tail can be gained when the simple scaling exercise of the previous subsection is applied to the results in summer (Fig. 13). In contrast to winter (Fig. 10), all symbols are located well above the diagonal line, but, similarly to winter, the symbols for the same return period are still arranged on a line. The implications thereof are twofold: Firstly, the simulated changes in wet-day frequency and intensity cannot explain the magnitude of the simulated change in extremes. They would suggest larger decreases or smaller increases than those simulated by the models. In the simulations, there appears to be a process which is specific to extreme events and which contributes an increase that is superimposed to the change extrapolated from the change in average events. The tail-specific component of the change is very coherent between models, as is implied by the systematic offset of symbols from the diagonal. The offset is larger for x1d.50 than for x1d.5 suggesting that the component is progressive with more rare extremes. This is also evident from the fact that the connecting lines between x1d.5 and x1d.50 symbols of the same model have a steeper slope than 1 for all models.

Secondly, even though not explaining the magnitude of change, the model specific changes in wet-day frequency and intensity do explain a large fraction of the inter-model variance, also in summer. R^2-values for all scatter plots in Fig. 13 exceed 0.9. Hence, the large difference in the response for extremes comes from differences in the way the models simulate changes in wet-day frequency and intensity (i.e. the change in average events), rather than the tail-specific component of the change.

In summary, for summer, there is a much larger variation in the change of precipitation extremes between the RCMs than in winter. All models show a tendency towards larger increases or smaller decreases (depending on the model) with more and more extreme events. This is primarily due to a component of change affecting extremes more than average events. This component is fairly coherent between the models. Model differences are more due to differences in the change simulated in average events, particularly the change in wet-day frequency and intensity.

Transition seasons:

Changes in one-day precipitation extremes show a very similar continental-scale pattern in spring and in autumn (not shown). In both seasons x1d.5 increases over most parts of the continent, especially over the northern, central and eastern parts. An exception is the Iberian Peninsula where x1d.5 decreases in spring and changes marginally in autumn. Several models exhibit particularly large increases in autumn over the Mediterranean Sea and Italy. The magnitude of the increase over Scandinavia and Central Europe is similar to that in winter and the coherence between models is similar or even better than in winter (see Table 3). As regards the variation of the change across the frequency distribution, both seasons tend to replicate the more simple structure of change found previously for winter. However there are individual models (e.g. CHRM and REMO in autumn) showing progressively larger increases at longer return periods in Central Europe. More detailed analysis for the transition seasons suggests that the simple scaling model reasonably explains the simulated changes at the tail in spring in Scandinavia, but there is evidence for a tail-specific component of the change, similar to but much smaller than in summer, for all regions in autumn.

6) Conclusions

In the present study we undertook an intercomparison of precipitation extremes as simulated by six different European regional climate models, all with comparable model settings and driven with boundary data from the same global climate model. An evaluation of the model simulations for present climate in the region of the European Alps shows that RCMs are capable of representing mesoscale spatial patterns in precipitation extremes that are not resolved by today's GCMs. However, model biases are large in some cases, in particular in summer. Even for rare extremes (five-year return values), these biases were nevertheless similar to or smaller than those for wet-day intensity or mean precipitation. We conclude that there is no evidence from this evaluation for model errors that are specific to precipitation extremes and that are not evident in more simple, average diagnostics.

Moreover, comparison to earlier evaluations of similar RCMs, but driven by reanalysis data (Frei et al. 2003), implies that errors inherited from the driving GCM were small and did not alter the RCM specific error characteristics, which, in retrospect, attests to the quality of the GCM's present-day climate.

The simulated future change in European precipitation extremes shows a seasonally very distinct pattern: In winter, land regions north of about 45°N experience an increase in multi-year return values while the Mediterranean region experiences small changes with a general tendency towards decreases. Results are very consistent between the six RCMs, with the change in 5-year return values increasing by 0-11% in Central Europe and by 10-22% in Scandinavia. The simulated increase of extremes is accurately explained by the increase in wet-day intensity and frequency in a simple re-scaling of the distribution for present climate.

The increase in wintertime precipitation extremes is a robust feature in RCM climate change experiments over Europe. Versions of the RCMs considered in this study, but driven by different GCMs, yield changes very similar to the geographical pattern found in this study (Durman et al. 2001; Räisänen et al. 2004). Also, the simulated response in winter is qualitatively conform with the observed trends in heavy precipitation over Europe, which shows an increase in winter primarily north of 45°N (e.g. Klein-Tank and Können 2003; Fowler and Kilsby 2003; Haylock and Goodess 2004; Brunetti et al. 2004; Schmidli and Frei 2005). These similarities are worth noting, but it is premature to infer the detection of an anthropogenic influence on heavy precipitation in Europe (see also Kiktev et al. 2003; Hegerl et al. 2004).

In summer, the character of change is more complex: The larger-scale pattern shows a gradient from increases in northern Scandinavia to decreases in the Mediterranean region and this is fairly similar between models. But the transition across the continent differs between models and the magnitude of the change in the 5-year return value varies considerably (–13% to +21% for Central Europe and +2 to +34% for Scandinavia). The large model differences are well explained by differences in the change of average precipitation events as represented by wet-day intensity and frequency. This suggests that it is primarily the response in the basic intensity and occurrence process of precipitation where models differ, and not the particular

response in the extremes themselves. However, the simulated change in wet-day intensity and frequency alone would have resulted in larger decreases or smaller increases of extremes than simulated. The change in summer is therefore also governed by a factor affecting the frequency distribution more fundamentally and specifically at the tail. This factor tends to increase extreme quantiles in the simulation for future climate, independently of the sign of the total change in the quantiles. This tail-specific component is seen consistently in all RCMs and its magnitude is considerable, capable of reversing a decrease that would be expected from changes in mean conditions alone, into an increase.

The present analysis offers a more in depth statistical interpretation of results from previous studies on the future change of European summer precipitation extremes. In their RCM experiments, Christensen and Christensen (2003, 2004) and Pal et al. (2004) found increases in high precipitation quantiles in Central Europe although mean precipitation was simulated to decrease. In this study, for Central Europe we find an increase in five but a decrease in 2 models. In principle, an increase of extremes would be possible even if the change at the tail was determined by the change in average conditions alone (see Fowler and Hennessy 1995): at sufficiently high return values (and for pdfs with a nearly exponential tail such as for precipitation) an increase in wet-day intensity always dominates a decrease in wet-day frequency, even if the frequency decrease is larger than the intensity increase, i.e. when mean precipitation decreases. The present analysis suggests that this simple picture is not a sufficient explanation for the model results. The change in summer extremes indeed reflects a non-trivial change at the tail of the distribution. This excessive response for extremes is found consistently in all RCMs, even in those two models where extreme return values actually decrease. Also, the same behavior is noted further north in Scandinavia, where a response more similar to winter could have been expected. However, there is a large inter-model difference in the sign and magnitude of the change at a particular frequency level (return period), which is primarily related to the different responses in average conditions, and contributes to considerable uncertainty about the future change in European summer precipitation extremes.

Clearly, more research will be needed to understand the physical nature of the tail-specific response and the reasons for the different model responses in summer. From the point of view of scenarios, the present analysis suggests that the formulation of regional models (e.g. the parameterization) contributes significantly to the uncertainty in scenarios of summer precipitation extremes. It is therefore not a waste of resources if multi-model ensemble systems, devoted to estimating scenario uncertainties, include a set of RCMs nested into the same GCM, alongside the nesting of RCMs in several different GCMs.

Acknowledgement

We are grateful to Christopher A. T. Ferro, Christoph Schär and David B. Stephensen for useful discussions on the subject and results of this study. We are also indebted to Ole B. Christensen for setting up the PRUDENCE database from which the RCM data was retrieved. The following institutes have kindly provided access to climate model data and to daily precipitation data in the Alps: DMI, Copenhagen, Denmark; GKSS, Geesthacht, Germany; Hadley Centre UK Met. Office, Exeter, UK; MPI, Hamburg, Germany; SMHI, Norrköpping, Sweden; DWD, Offenbach, Germany; Hydrogr. Zentralbüro, Vienna, Austria; ZAMG, Vienna, Austria; Météo France, Toulouse, France; UCEA, Rome, Italy; MeteoSwiss, Zürich, Switzerland; SIMN, Rome, Italy; Meteorological Service, Zagreb, Croatia; Hydrometeorological Institute, Ljubljana, Slovenia. This research was supported by the 5th Framework Program of the European Union (projects STARDEX, contract No. EVK2-2001-00115 and PRUDENCE contract No. EVK2-CT200100132), and by the Swiss Ministry for Education and Science (BBW contract No. 01.0265-2). Additional support was provided by the Swiss National Science Foundation (NCCR Climate).

REFERENCES

Allen, M.R. and W.J. Ingram, 2002: Constraints on future changes in climate and the hydrologic cycle. *Nature,* **419,** 224-232.

Anderson, C.J. and al. et, 2003: Hydrological processes in regional climate model simulations of the Central United States flood of June-July 1993. *J. Hydrometeorol.,* **4,** 584-598.

Arora, V.K. and G.J. Boer, 2001: Effects of simulated climate change on the hydrology of major river basins. *J. Geophys. Res.,* **106,** 3335-3348.

Booij, M.J., 2002: Extreme daily precipitation in western Europe with Climate Change at appropriate spatial scales. *Int. J. Climatol.,* **22,** 69-85.

Brunetti, M., M. Maugeri, F. Monti and T. Nanni, 2004: Changes in daily precipitation frequency and distribution in Italy over the last 120 years. *J. Geophys. Res.,* **109,** D05102, doi:10.1029/2003JD004296.

Christensen, J.H., T.R. Carter and M. Rummukainen, 2005: Evaluating the performance and utility of regional climate models: The PRUDENCE project. *Clim. Change,* (submitted).

Christensen, J.H. and O.B. Christensen, 2003: Severe summertime flooding in Europe. *Nature,* **421,** 805-806.

Christensen, J.H., O.B. Christensen, P. Lopez, E. Van Meljgaard and M. Botzet, 1996: The HIRHAM4 regional atmospheric climate model. Sci. Rep., 96-4, Danish Meteorological Institute, Copenhagen, DK., 51 pp..

Christensen, O.B. and J.H. Christensen, 2004: Intensification of extreme European summer precipitation in a warmer climate. *Global and Planetary Change,* **44,** 107-117.

Coles, S., 2001: *An introduction to statistical modeling of extreme values.* Springer, London, 208 pp.

Cubasch, U. and al. et, 2001: Projections of future climate change. In: *Climate Change 2001: The scientific basis. The Third Assessment Report of the Intergovernmental Panel on Climate Change (IPCC)* , Eds.: Houghton J.T. et al., 525-582.

De Elia R. and R. Laprise, 2003: Distribution-oriented verification of limited-area model forecasts in a perfect-model framework. *Mon. Wea. Rev.,* **131,** 2492-2509.

Denis, B., R. Laprise, D. Caya and J. Côté, 2002: Downscaling ability of one-way nested regional climate models: the Big-Brother experiment. *Climate Dyn.,* **18,** 627-646.

Déqué, M., D. Rowell, C. Schär, F. Giorgi, J.H. Christensen, B. Röckel, D. Jacob, E. Kjellstrom, M. de Castro and B. van den Hurk, 2005: An intercomparison of regional climate models for Europe: Assessing uncertainties in model projections. *Climate Dyn.,* (submitted).

Döscher, R., U. Willén, C. Jones, A. Rutgersson, H.E.M. Meier, U. Hansson and L.P. Graham, 2002: The development of the coupled regional ocean-atmosphere model RCAO. *Boreal Env. Res.,* **7,** 183-192.

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 a regional climate model for present and future climates. *Q. J. Roy. 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.

Fisher, R.A. and L.H.C. Tippett, 1928: Limiting forms of the frequency distribution of the largest or smallest members of a sample.. *Proc. Cambridge Philos. Soc.,* **24,** 180-190.

Fowler, A.M. and K.J. Hennessy, 1995: Potential Impacts of Global Warming on the Frequency and Magnitude of Heavy Precipitation. *Natural Hazards,* **11,** 282-303.

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.

Fowler, H.J. and C.G. Kilsby, 2003: A regional frequency analysis of United Kingdom extreme rainfall from 1961 to 2000. *Int. J. Climatol.,* **23,** 1313-1334.

Frei, C., J.H. Christensen, M. Déqué, D. Jacob, R.G. Jones and P.L. Vidale, 2003: Daily precipitation statistics in regional climate models: Evaluation and intercomparison for the European Alps. *J. Geophys. Res.,* **108 (D3),** 4124 doi:10.1029/2002JD002287.

Frei, C. and C. Schär, 1998: A precipitation climatology of the Alps from high-resolution rain-gauge observations. *Int. J. Climatol.,* **18,** 873-900.

Frei, C., C. Schär, D. Lüthi and H.C. Davies, 1998: Heavy precipitation processes in a warmer climate. *Geophys. Res. Lett.,* **25,** 1431-1434.

Gnedenko, B., 1943: Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.,* **44,** 423-453.

Goodess, C.M., 2003: Statistical and regional dynamical downscaling of extremes for European regions: STARDEX . *The Eggs,* **6**.

Goodess, C.M., C. Anagnostopoulou, A. Bárdossy, M.R. Haylock, Y. Hundecha, P. Maheras, J. Ribalaygua, J. Schmidli, T. Schmith and R. Tomozeiu, 2005: An intercomparison of statistical downscaling methods for Europe and European regions – assessing their performance with respect to extreme temperature and precipitation events. *Clim. Change,* (submitted).

Gordon, C., C. Cooper, C.A. Senior, H. Banks, H.M. Gregory, T.C. Johns, J.F.B. Mitchell and R.A. Wood, 2000: The simulation of SST, sea ice extent and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Climate Dyn.,* **16,** 147-168.

Groisman, P.Y.A., T.R. Karl, D.R. Easterling, R.W. Knight, P.F. Jamason, K.J. Hennessy and al. et, 1999: Changes in the probability of heavy precipitation: Important indicators of climatic change. *Clim. Change,* **42,** 243-283.

Gumbel, E.J., 1958: *Statistics of extremes*. Columbia University Press, New York, 375 pp.

Hagemann, S., M. Botzet, L. Dümenil and B. Machenhauer, 1999: Derivation of global GCM boundary conditions for 1 km land use satellite data. Rep. No. 2189, Max-Planck Inst. für Meteorol. (MPI), Hamburg.

Haylock, M.R., 2004: STARDEX Diagnostic Extremes Indices Software - User information. Available from http://www.cru.uea.ac.uk/projects/stardex/deis/Diagnostic_tool.pdf.

Haylock, M.R. and C.M. Goodess, 2004: Interannual variability of extreme European winter rainfall and links with mean large-scale circulation. *Int. J. Climatol.,* **24,** 759-776.

Hegerl, G.C., F.W. Zwiers, P.A. Stott and V.V. Kharin, 2004: Detectability of anthropogenic changes in temperature and precipitation extremes. *J. Climate,* (submitted).

Hosking, J.R.M., 1985: Algorithm AS 215: Maximum-likelihood estimation of the parameter of the generalized extreme-value distribution. *Appl. Stat.,* **34,** 301-310.

Hosking, J.R.M., 1990: L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *J. R. Statist. Soc. B,* **52,** 105-124.

Hosking, J.R.M., 1992: Moments or L-moments? An example comparing two measures of distributional shape. *The American Statistician,* **46,** 186-189.

Huntingford, C., R.G. Jones, C. Prudhomme, R. Lamb, H.H.C. Gash and D.A. Jones, 2003: Regional climate-model predictions of extreme rainfall for a changing climate. *Q. J. Roy. Meteorol. Soc.,* **129,** 1607-1621.

Jacob, D., 2001: A note to the simulation of the annual and inter-annual variability of the water budget over the Baltic Sea drainage basin. *Meteorol. Atmos. Phys.,* **77,** 66-73.

Johns, T.C., J.M. Gregory, W.J. Ingram, C.E. Johnson, A. Jones, J.A. Lowe, J.F.B. Mitchell, D.L. Roberts, D.M.H. Sexton, D.S. Stevenson, S.F.B. Tett and M.J. Woodage, 2003: Anthropogenic climate change for 1860 to 2100 simulated with the HadCM3 model under updated emission scenarios. *Climate Dyn.,* **20,** 583-612.

Jones, C., 2001: A brief description of RCA2, Rossby Centre Atmosphere Model Version 2. SWECLIM Newsletter 11, 9-14. Available from SMHI, S-60176 Norrköping, Sweden and online at http://www.smhi.se/sgn0106/rossby/sweclim/newsletter/newsletter_11.pdf.

Jones, P.D. and P.A. Reid, 2001: Assessing future changes in extreme precipitation over Britain using regional climate model integrations. *Int. J. Climatol.,* **21,** 1337-1356.

Jones, R.G., J.M. Murphy and M. Noguer, 1995: Simulation of climate change over Europe using a nested regional-climate model I: Assessment of control climate, including sensitivity to location of lateral boundaries. *Q. J. R. Meteorol. Soc.,* **121,** 1413-1449.

Jones, R.G., J.M. Murphy, M. Noguer and A.B. Keen, 1997: Simulation of climate change over Europe using a nested regional-climate model. II: Comparison of driving and regional model responses to a doubling of carbon dioxide. *Q. J. R. Meteorol. Soc.,* **123,** 265-292.

Jones, R.G., M. Noguer, D.C. Hassell, D. Hudson, S.S. Wilson, G.J. Jenkins and J.F.B. Mitchell, 2004: Generating high resolution climate change scenarios using PRECIS. Available from Met. Office Hadley Centre, Exeter, UK, 35 pp.

Katz, R.W. and J.G. Acero, 1994: Sensitivity of extreme precipitation events. *Int. J. Climatol.,* **14,** 985-999.

Katz, R.W., M.B. Parlange and P. Naveau, 2002: Statistics of extremes in hydrology. *Advances in Water Resources,* **25,** 1287–1304.

Kharin, V.V. and F.W. Zwiers, 2000: Changes in the extremes in an ensemble of transient climate simulations with a coupled Atmosphere-Ocean GCM. *J. Climate,* **13,** 3760-3788.

Kiktev, D., D.M.H. Sexton, L. Alexander and C.K. Folland, 2003: Comparison of modeled and observed trends in indices of daily climate extremes. *J. Climate,* **16,** 3560-3571.

Kiktev, D., D.M.H. Sexton, L. Alexander and C.K. Folland, 2004: Corrigendum: Comparison of modeled and observed trends in indices of daily climate extremes . *J. Climate,* **17,** 2489.

Klein Tank, A.M.G. and G.P. Können, 2003: Trends in indices of daily temperature and precipitation extremes in Europe, 1946-1999. *J. Climate,* **16,** 3665-3680.

Lüthi, D., A. Cress, C. Frei and C. Schär, 1996: Interannual variability and regional climate simulations. *Theor. Appl. Clim.,* **53,** 185-209.

Martins, E.S. and J.R. Stedinger, 2000: Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resourc. Res.,* **36,** 737-744.

Meier, H.E.M., R. Döscher, A.C. Coward and J. Nycander, 1999: Rossby Centre regional ocean climate mode: Model description (version 1.0) and first results from the hindcast period 1992/1993. SMHI Reports Oceanography, 26, 102 pp. Available from SMHI, S-60176 Norrköping, Sweden.

Nakicenovic, N. and al. et, 2000: *Special Report on Emission Scenarios*. A special report of Working Group III for the Integovernmental Panel on Climate Change, Cambridge University Press, 599 pp.

Osborn, T.J. and M. Hulme, 1997: Development of a relationship between station and grid-box rainday frequencies for climate model evaluation. *J. Climate,* **10,** 1885-1908.

Palmer, T.N. and J. Räisänen, 2002: Quantifying the risk of extreme seasonal precipitation events in a changing climate. *Nature,* **415,** 512-514.

Palutikof, J.P., B.B. Brabson, D.H. Lister and S.T. Adcock, 1999: A review of methods to calculate extreme wind speeds. *Meteorol. Apps.,* **6,** 119-132.

Pal, J.S., F. Giorgi and X. Bi, 2004: Consistency of recent European summer precipitation trends and extremes with future regional climate projections. *Geophys. Res. Lett.,* **31,** L13202, doi:10.1029/2004GL019836.

Pope, D. V, M. Gallani, R. Rowntree and A. Stratton, 2000: The impact of new physical parameterizations in the Hadley Centre climate model: HadAM3. *Climate Dyn.,* **16,** 123-146.

Räisänen, J., U. Hannson, A. Ullerstig, R. Döscher, L.P. Graham, C. Jones, H.E.M. Meier, P. Samuelsson and U. Willén, 2004: European climate in the late twenty-first century: regional simulations with two global models and two forcing scenarios. *Climate Dyn.,* **22,** 13-31.

Roeckner, E., K. Arpe, L. Bengtsson, M. Christoph, M. Claussen, L. Dümenil, M. Esch, M. Giorgetta, U. Schlese and U. Schulzweida, 1996: The atmospheric general circulation model ECHAM-4: Model description and simulation of present-day climate. Rep. No. 218, Max-Planck Inst. für Meteorol. (MPI), Hamburg. 90 pp.

Schär, C., P.L. Vidale, D. Lüthi, C. Frei, C. Häberli, M.A. Liniger and C. Appenzeller, 2004: The role of increasing temperature variability in European summer heatwaves. *Nature,* **427,** 332-336.

Schmidli, J. and C. Frei, 2005: Trends of heavy precipitation and wet and dry spells in Switzerland during the 20th century. *Int. J. Climatol.,* (in press).

Schwarb, M., C. Daly, C. Frei and C. Schär, 2001: Mean annual and seasonal precipitation in the European Alps 1971-1990. Hydrological Atlas of Switzerland, Landeshydrologie und Geologie, Bern, Plates 2.6 and 2.7

Semenov, V.A. and L. Bengtsson, 2002: Secular trend in daily precipitation characteristics: Greenhouse gas simulation with a coupled AOGCM. *Climate Dyn.,* **19,** 123-140.

Semmler, T. and D. Jacob, 2004: Modeling extreme precipitation events - a climate change simulation for Europe. *Global and Planetary Change,* **44,** 119-127.

Steppeler, J., G. Doms, U. Schättler, H.W. Bitzer, A. Gassmann, U. Damrath and G. Gregoric, 2003: Meso-gamma scale forecasts using the nonhydrostatic model LM. *Meteorol. Atmos. Phys.,* **82,** 75-96.

Trenberth, K.E., 1999: Conceptual framework for changes of extremes of the hydrological cycle with climate change. *Clim. Change,* **42,** 327-339.

Trenberth, K.E., A. Dai, R.M. Rasmussen and D.B. Parsons, 2003: The changing character of precipitation. *Bull. Amer. Meteorol. Soc.,* **84,** 1205-1217.

Vidale, P.L., D. Lüthi, C. Frei, S. Seneviratne and C. Schär, 2003: Predictability and uncertainty in a Regional Climate Model. *J. Geophys. Res.,* **108 (D18),** 4586, doi: 10.1029/2002JD002810.

Voss, R., W. May and E. Roeckner, 2002: Enhanced resolution modelling study on anthropogenic climate change: Changes in extremes of the hydrological cycle. *Int. J. Climatol.,* **22,** 755-777.

Watterson, I.G. and M.R. Dix, 2003: Simulated changes due to global warming in daily precipitation means and extremes and their interpretation using the gamma distribution. *J. Geophys. Res.,* **198,** 4379, doi:10.1029/2002JD002928.

Wehner, M.F., 2004: Predicted twenty-first-century changes in seasonal extreme precipitation events in the Parallel Climate Model. *J. Climate,* **17,** 4281-4290.

Whetton, P.H., A.M. Fowler, M.R. Haylock and A.B. Pittock, 1993: Implications of climate change due to the enhanced greenhouse effect on floods and droughts in Australia. *Clim. Change,* **25,** 289-317.

Widmann, M. and C.S. Bretherton, 2000: Validation of mesoscale precipitation in the NCEP reanalysis using a new gridpoint dataset for the northwestern US. *J. Climate,* **13,** 1936-1950.

Wigley, T.M.L. and S.C.B. Raper, 2001: Interpretation of high projections for global-mean warming. *Science,* **293,** 451-454.

Wilby, R.L., T.M.L. Wigley, D. Conway, P.D. Jones and B.C. Hewitson, 1998: Statistical downscaling of general circulation model output: A comparison of methods. *Water Resourc. Res.,* **34,** 2995-3008.

Wilks, D.S., 1997: Resampling hypothesis tests for autocorrelated fields. *J. Climate,* **10,** 65-82.

Woth, K., R. Weisse and H. von Storch, 2005: Climate change and North Sea storm surge extremes: An ensemble study of strom surge extremes expected in a changed climate projected by four different Regional Climate Models. *Ocean Dynamics,* (submitted).

Zwiers, F.W. and V.V. Kharin, 1998: Changes in the extremes of the climate simulated by CC GCM2 under CO2 doubling. *J. Climate,* **11,** 2200-2222.

Tables

Table 1: Diagnostics of daily precipitation used in this study.

Acronym	*Definition*	*unit*
mea	climatological mean precipitation	mm per day
fre	wet-day frequency, days with precipitation ≥1 mm	fraction
int	wet-day intensity, mean precipitation on days with ≥1 mm	mm per day
qXX	empirical XX% quantile of precipitation during wet days, (XX = 40%, 60%, 80%, 90%, 95%)	mm per day
x1d.TT	return value of *one*-day precipitation intensity with a return period of TT. (TT = 2, 5, 10, 20, 50, 100 years)	mm per day
x5d.TT	return value of *five*-day precipitation intensity with a return period of TT. (TT = 2, 5, 10, 20, 50, 100 years)	mm per day

Table 2: Regional climate models from which integrations are analysed in this study.

Acronym	*Institution and model origin* *Technical descriptions and other recent analyses of climate change experiments*	*nx* x *ny*, *dx* *levels*	*GCM* *# ens. membrs.* *(CTRL/SCEN)*
CHRM	Swiss Federal Institute of Technology (ETH), Zürich, Climate High-resolution Model. Climate version of 'Europamodell' of German and Swiss weather services. Model: Lüthi et al. (1996), Vidale et al. (2003) Temperature Extremes: Schär et al. (2004)	81x91 0.5 deg 20	HADAM3H 1/1
GKSS	GKSS, Institute for Coastal Research, Geesthacht, Germany Climate Version of 'Lokalmodell' of German Weather Service Model: Steppeler et al. (2003) Wind extremes: Rockel and Woth (2005), Woth et al. (2005)	101x107 0.5 deg 20	HADAM3H 1/1
HADRM3H	Hadley Centre, UK Meteorological Office, Exeter Regional model of climate model suite at the Hadley Centre. Model: Jones et al., (1995, 1997), see also Pope et al. (2001). Prec. Extremes: Huntingford et al. (2003); Fowler et al. (2005); Ekström et al. (2005)	106x111 0.44 deg 19	HADAM3H 3/3
HADRM3P	Hadley Centre, UK Meteorological Office, Exeter Regional model of climate model suite at the Hadley Centre. Updated version of HADRM3H Model: Jones et al. (2004)	106x111 0.44 deg 19	HADAM3P 3/3
HIRHAM	Danish Meteorological Institute, Copenhagen Dynamical core from HIRLAM, Parameterisations from ECHAM4 Model: Christensen et al. (1996); Roeckner et al. (1996) Physiographic data: Hagemann et al. (1999) Prec. extremes: Christensen and Christensen (2003, 2004)	110x104 0.44 deg 19	HADAM3H 3/3
REMO	Max Planck Institute for Meteorology, Hamburg, Germany Dynamical core from Europamodell of German Weather Service. Parameterisations from ECHAM4. Model: Jacob (2001); Roeckner et al. (1996) Prec. extremes: Semmler and Jacob (2004)	97x109 0.5 deg 20	HADAM3H 1/1
SMHI	Swedisch Meteorological Institute, Stockholm Rossby Centre Atmosphere Ocean Model Model: Jones et al. (2001); Meier et al. (1999) Surface climate change: Räisänen et al. (2004)	106x102 0.44 deg 24	HADAM3H 1/1

Table 3: Change (Ratio SCEN/CTRL) in the 5-year return value of one-day (five-day in DJF) precipitation intensity for subdomains of Fig. 4a. Bold numbers indicate changes that are statistically significant, i.e. the ratio of 1.0 (no change) is outside the 95% confidence interval obtained by resampling (see section 2). Results for Iberia in summer are omitted because extreme value analysis was not feasible for many gridpoints with the small number of rainy days.

	Scandinavia			
	DJF (x5d.5)	*MAM (x1d.5)*	*JJA (x1d.5)*	*SON (x1d.5)*
CHRM	**1.22**	**1.27**	**1.18**	**1.15**
GKSS	**1.25**	**1.19**	**1.21**	**1.15**
HADRM3H	**1.17**	**1.18**	1.03	**1.11**
HADRM3P	**1.10**	**1.16**	1.02	**1.10**
HIRHAM	**1.15**	**1.18**	**1.18**	**1.15**
REMO	**1.21**	**1.18**	**1.34**	**1.23**
SMHI	**1.16**	**1.19**	1.05	**1.15**
	Central Europe			
	DJF (x5d.5)	*MAM (x1d.5)*	*JJA (x1d.5)*	*SON (x1d.5)*
CHRM	0.98	**1.12**	**1.21**	**1.22**
GKSS	1.00	**1.12**	**1.12**	**1.15**
HADRM3H	**1.10**	**1.14**	0.95	**1.11**
HADRM3P	**1.11**	**1.16**	**0.87**	**1.09**
HIRHAM	1.07	**1.16**	**1.09**	1.06
REMO	1.00	**1.24**	**1.19**	**1.21**
SMHI	1.05	**1.23**	**1.13**	**1.18**
	Iberia			
	DJF (x5d.5)	*MAM (x1d.5)*	*JJA (x1d.5)*	*SON (x1d.5)*
CHRM	0.86	**0.86**	-	0.99
GKSS	0.90	**0.89**	-	1.00
HADRM3H	**0.86**	**0.88**	-	0.99
HADRM3P	**0.88**	**0.86**	-	0.95
HIRHAM	0.93	0.94	-	0.96
REMO	0.92	**0.87**	-	0.99
SMHI	0.91	**0.91**	-	1.02

Figures

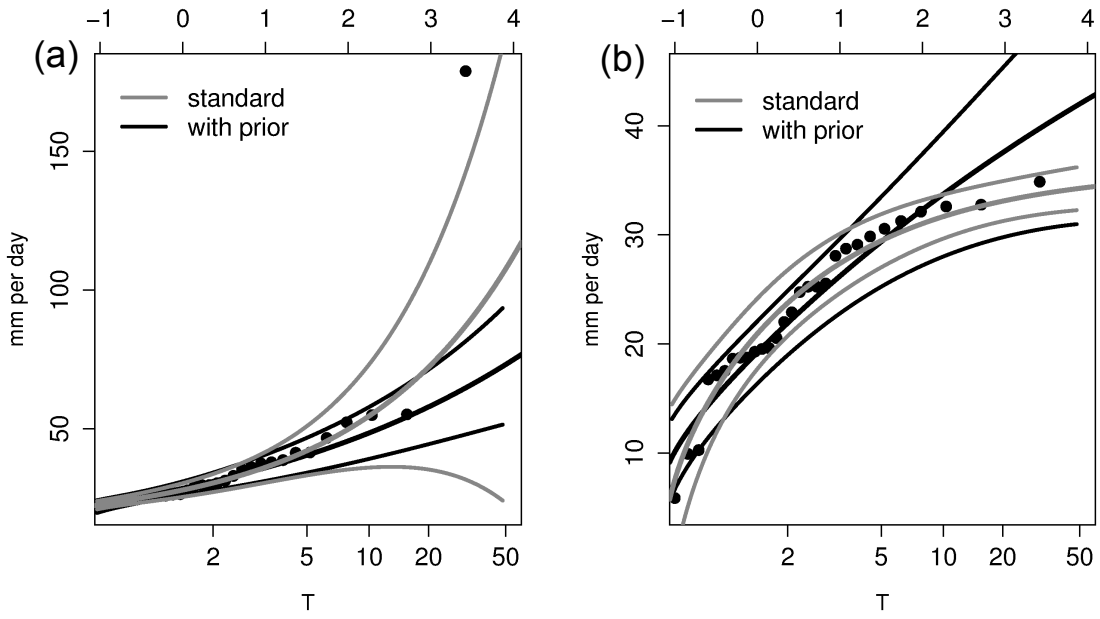

Figure 1: Gumbel diagrams (return value in mm per day as a function of return period *T* in years) for two grid-point samples of daily precipitation extremes, simulated by one of the RCMs (CTRL, winter). Sample extremes are depicted in dots and fitted GEV distributions (including pertinent maximum likelihood 95% confidence bands) in full lines. Maximum Likelihood Estimates of the GEV distribution are displayed for the standard likelihood function (grey) and for the likelihood function with the geophysical prior (black).

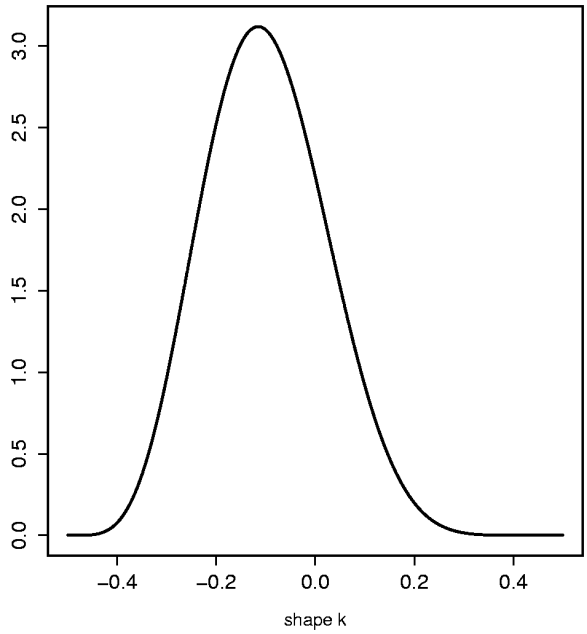

Figure 2: Probability density of the prior distribution for the GEV shape parameter. The prior distribution is used in this study to improve the robustness of Maximum Likelihood Estimates of GEV parameters and quantiles. Adapted from Martins and Stedinger (2000).

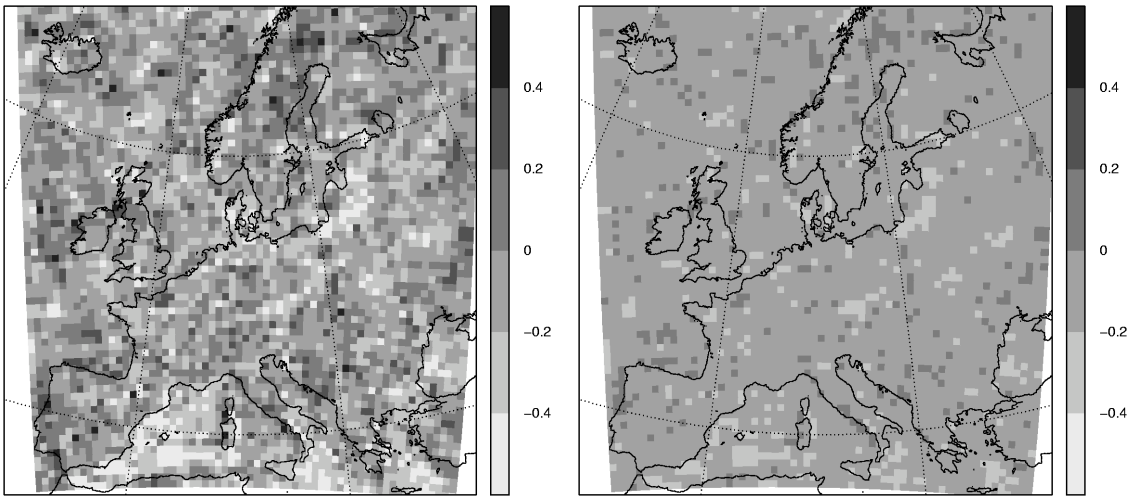

Figure 3: Distribution of the GEV shape parameter estimated from grid-point extremes without (a) and with (b) geophysical prior. Results are from a 30-year time slice of one of the RCMs (CHRM, CTRL).

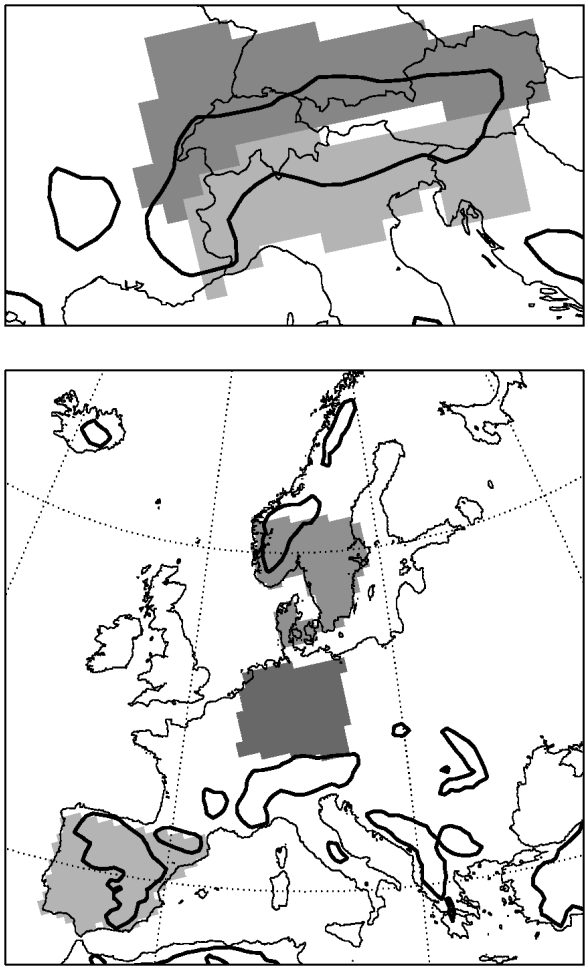

Figure 4: Subdomains used in the analysis of this study: (a) Northern and southern Alps, (b) Scandinavia, Central Europe and Iberia. Subdomains vary slightly between models due to different grids and land-sea mask. This figure is for model SMHI.

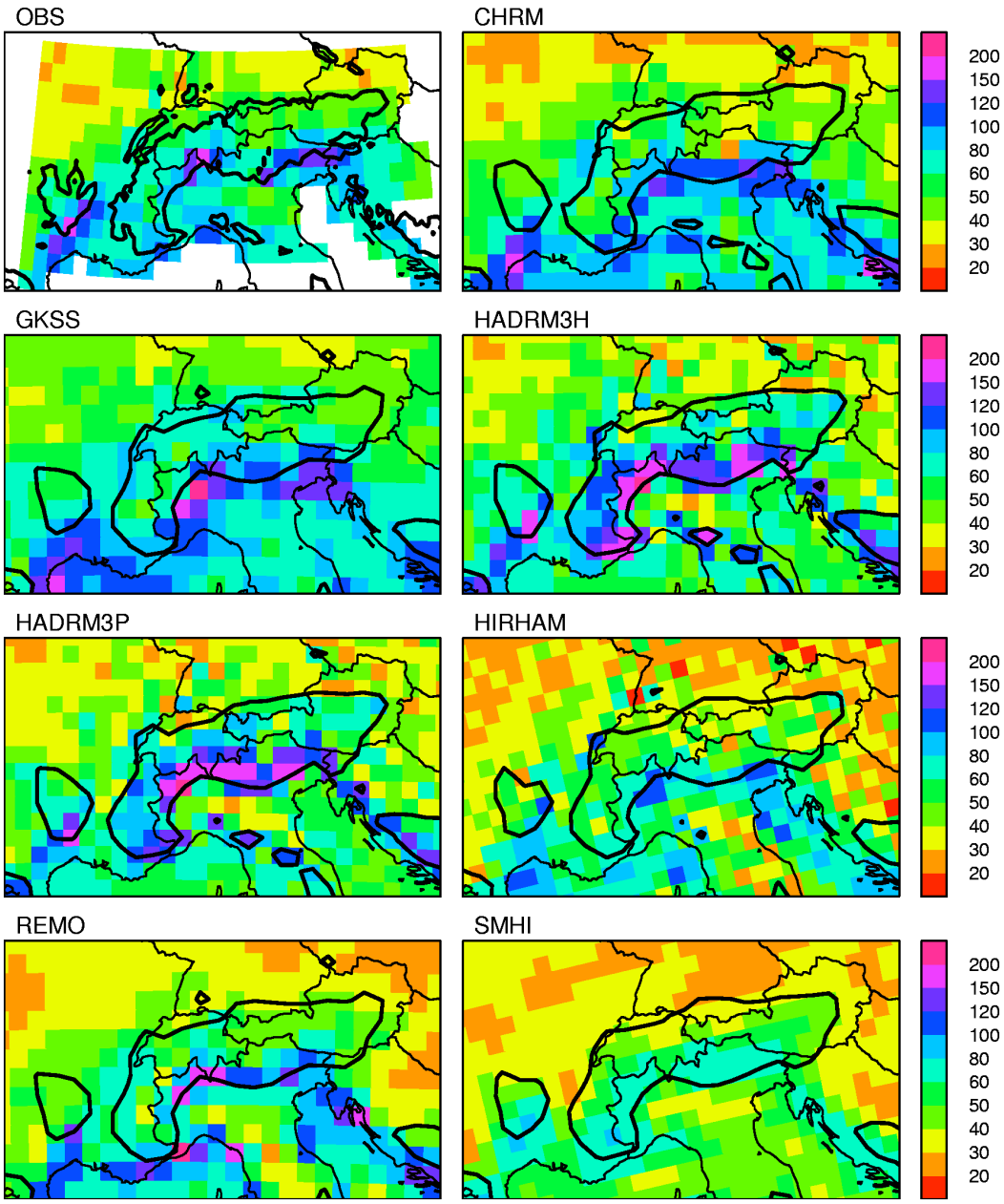

Figure 5: Five-year return value of one-day precipition extreme (x1d.5, mm per day) in autumn (SON). Observations (OBS, 1971-1990, top left) and RCMs (1961-1990). Model results are shown on original model grids. Thick lines are 700 m.a.s.l. contours of pertinent model topography. Topography in OBS is more detailed. In autumn there is the largest activity of heavy precipitation in the Alps.

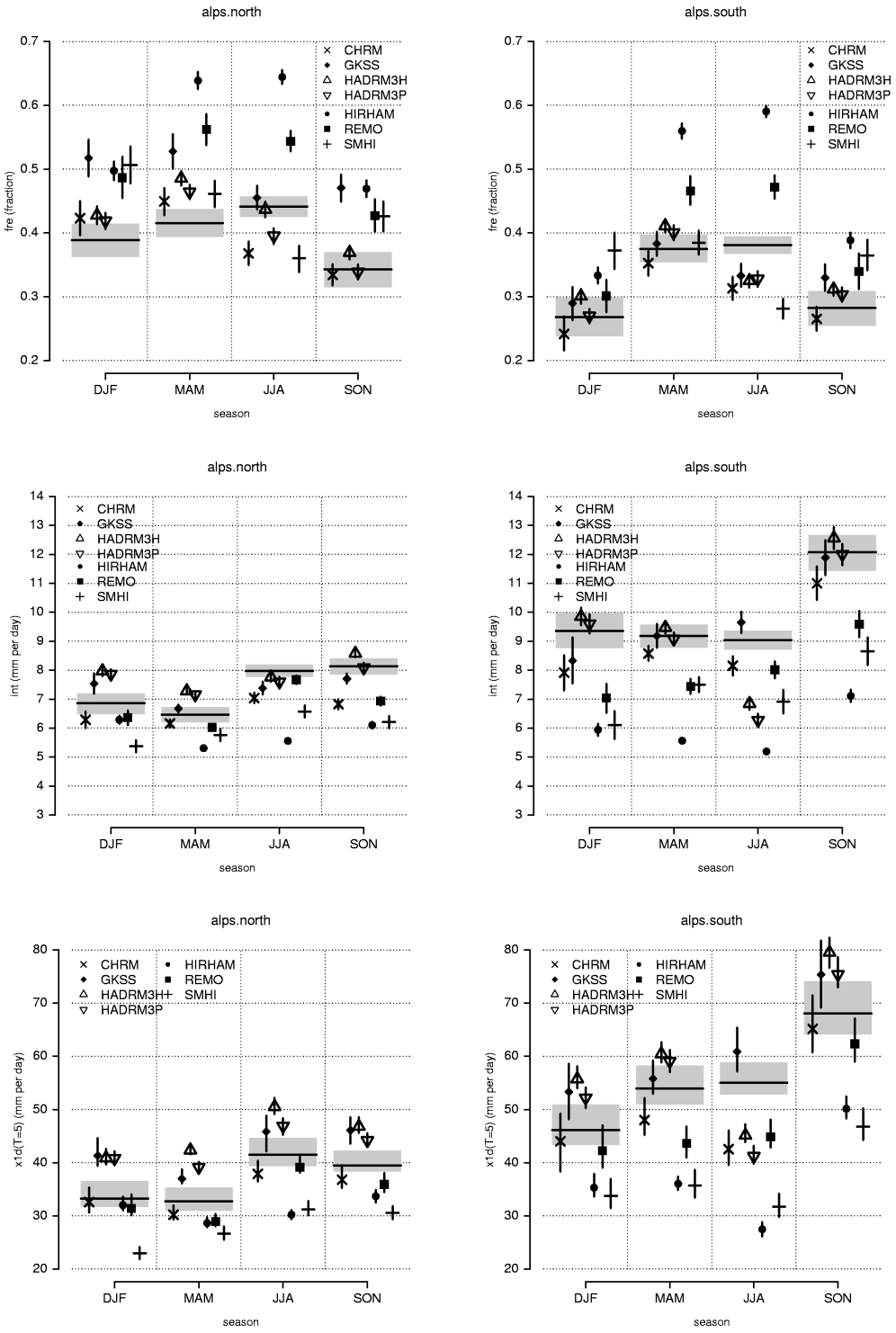

Figure 6: Best estimate and 90% bootstrap confidence interval of wet-day frequency (fre: a and b), wet-day intensity (int: c and d) and the 5-year return value of one-day precipitation (x1d.5: e and f) for RCMs (symbols, vertical lines) and observations (horizontal lines, shaded area). Results are shown for two subdomains (Northen Alps: left panels; Southern Alps: right panels, see Figure 4).

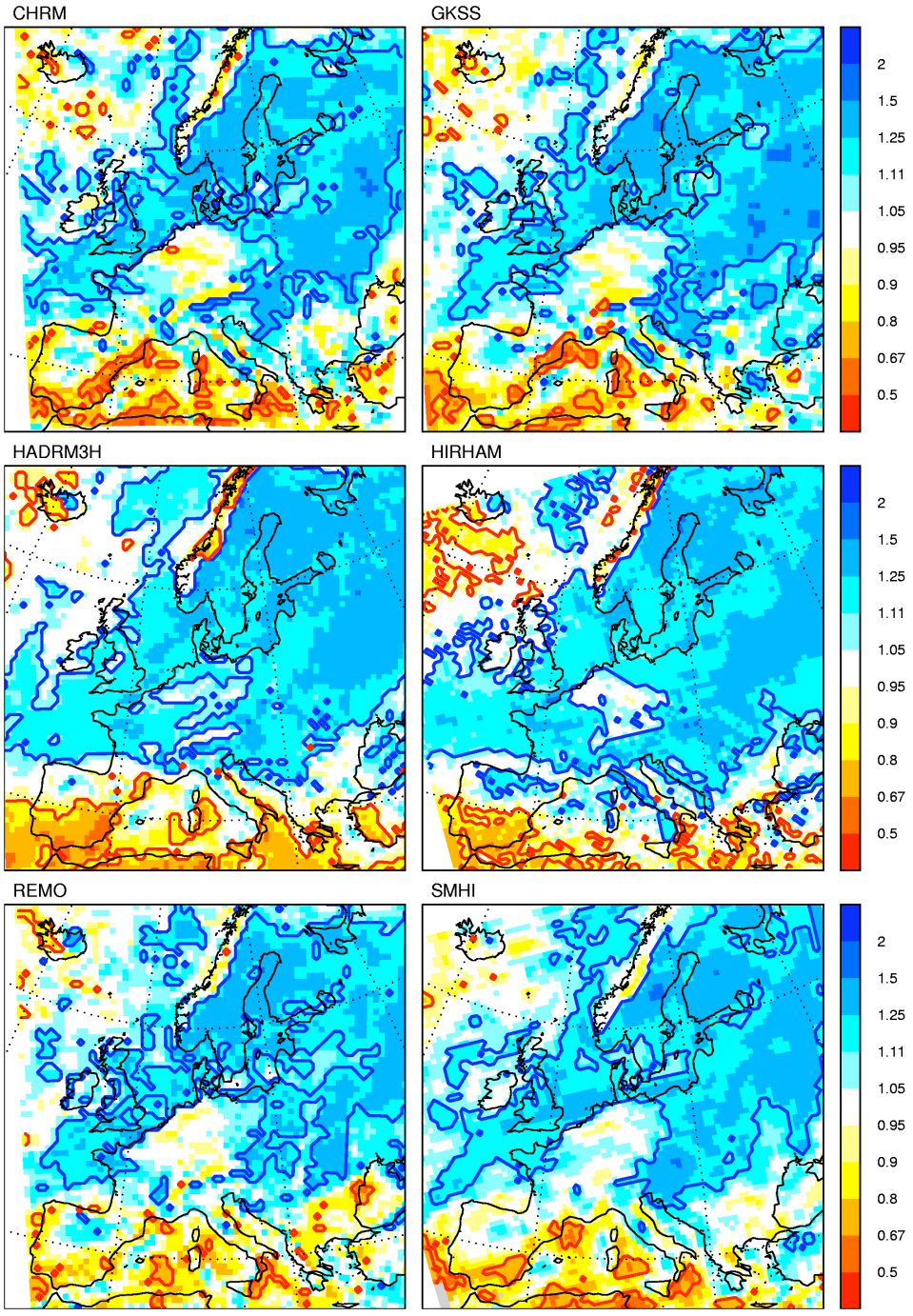

Figure 7: Ratio (SCEN: 2071-2100 / CTRL: 1961-1990) of the 5-year return value for 5-day precipitation intensity (x5d.5) in winter (DJF). Results from 6 RCMs under the A2 emission scenario (see Table 2). Note the log-scale in color coding. Full blue (red) lines indicate areas with statistically significant (5%) increase (decrease), in an independent test at each model grid-point.

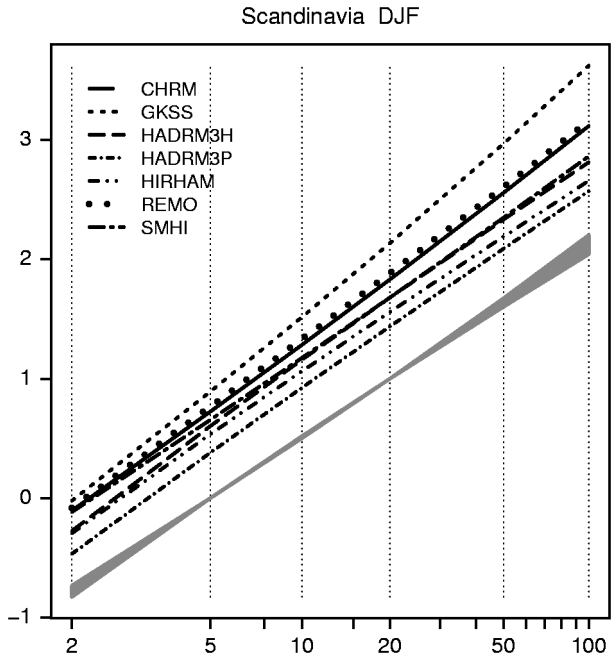

Figure 8: Standardized return value of the 5-day precipitation intensity as a function of return period in years (x5d.T, for winter in subdomain Scandinavia). The return period axis is transformed like in a Gumbel diagram. The standardized distributions are determined individually for each model by shifting CTRL and SCEN distributions so that the 5-year return value of CTRL is 0.0, and scaling the distributions so that the 20-year return value of CTRL is 1.0 (scaling the distributions). CTRL distributions are in grey, SCEN in full lines. The Figure illustrates the change in return period from the CTRL to the SCEN period.

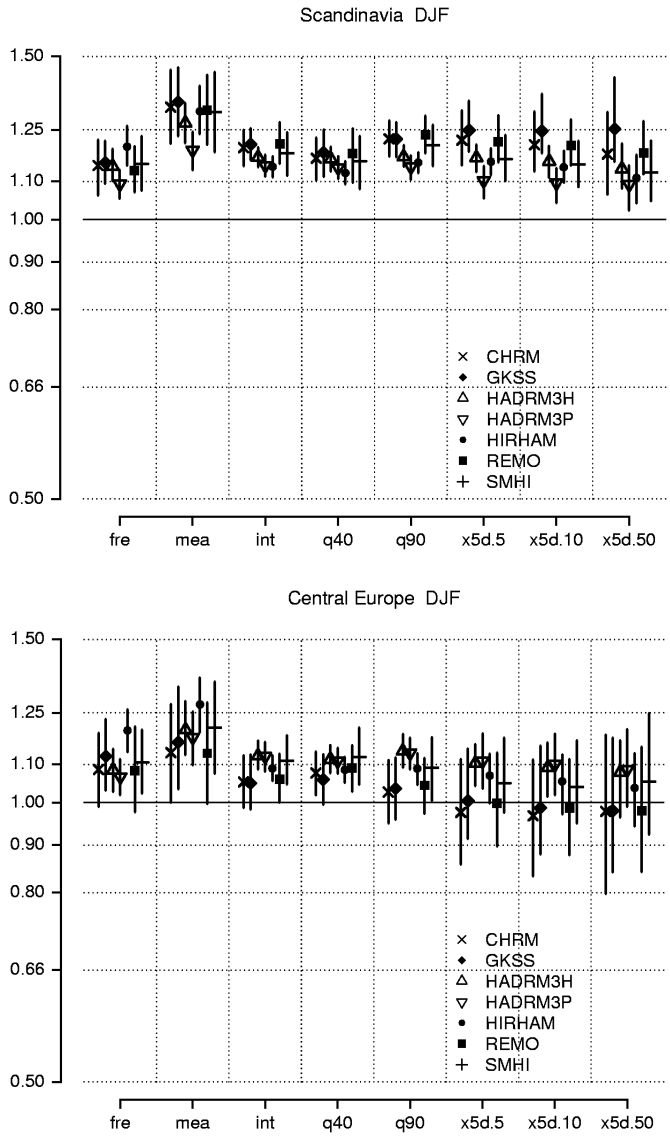

Figure 9: Simulated change (ratio SCEN/CTRL) in domain-mean precipitation diagnostics for Scandinavia (top) and Central Europe (bottom) in winter. See Table 1 for a description of diagnostics, Table 2 for model acronyms and Figure 4 for domain definitions. Vertical bars represent 95% confidence intervals of the change determined by resampling (see section 2).

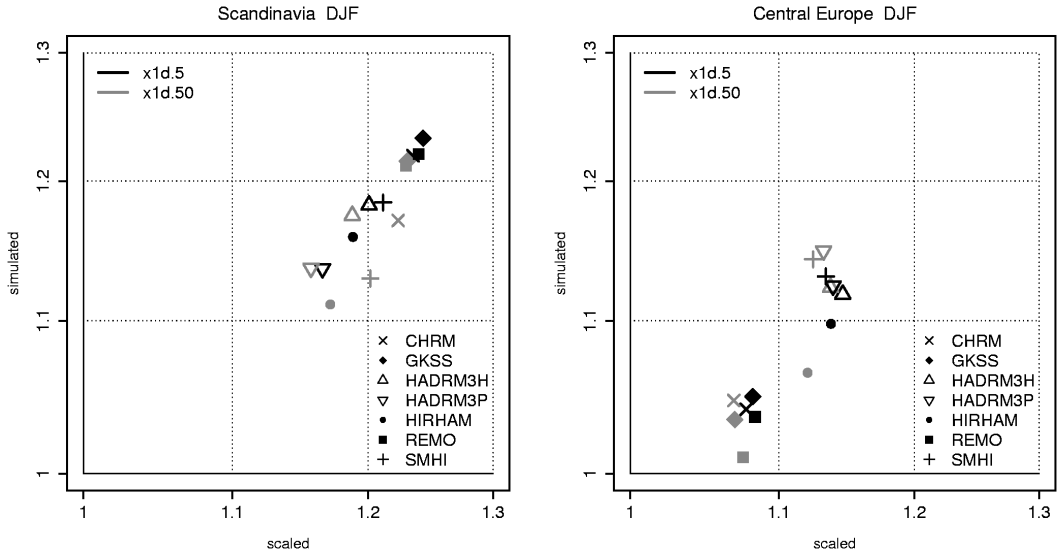

Figure 10: Simulated change in extreme return levels (ratio SCEN/CTRL, y-axis) against expected change from a simple scaling of the CTRL distribution based on changes in wet-day frequency and intensity (x-axis, see text for details). Results are shown for x1d.5 (black) and x1d.50 (grey) in winter. Both axes are log scaled.

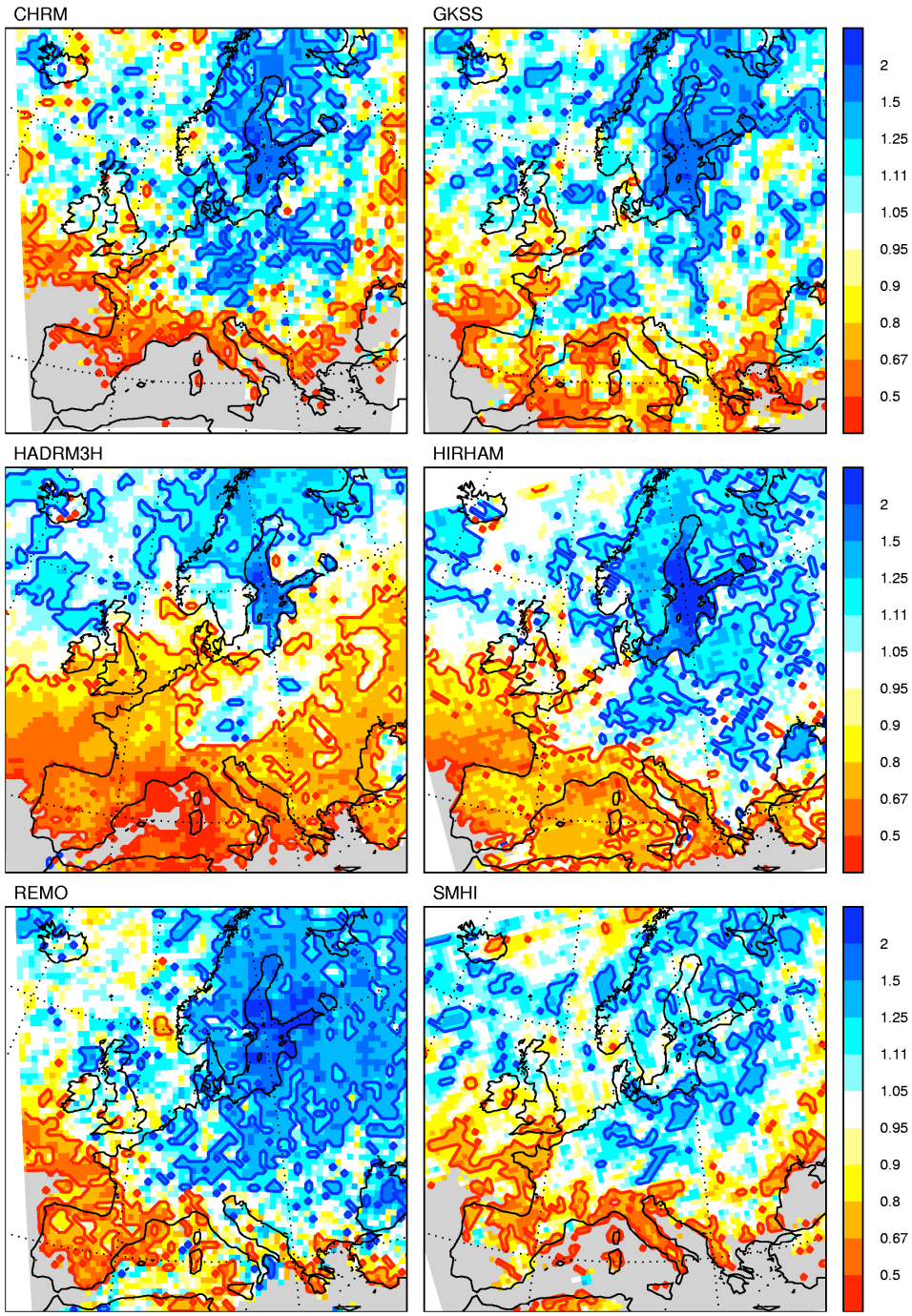

Figure 11: Same as Fig. 7 but for the 1-day precipitation total (x1d.5) in summer (JJA). Gray colors are for gridpoints where no extreme value analysis was feasible due to an insufficient number of rain days (see section 2).

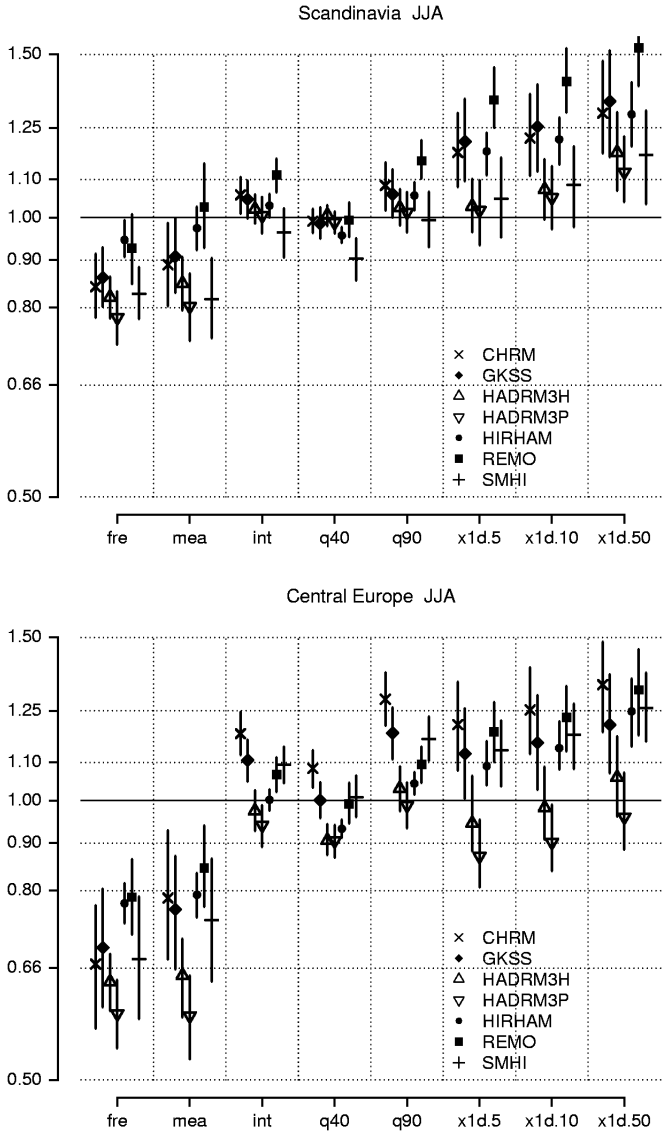

Figure 12: Same as Fig. 9 but for summer.

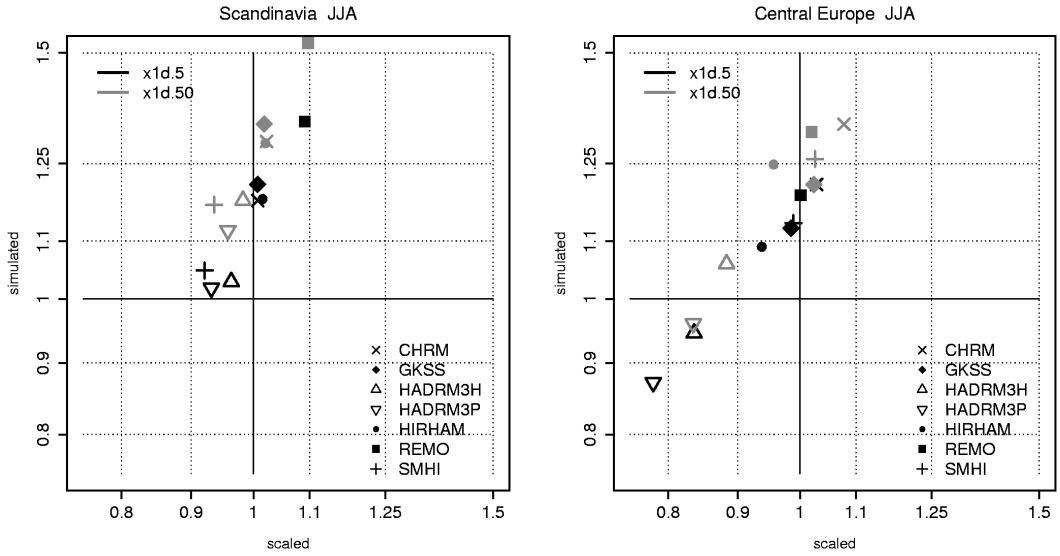

Figure 13: Same as Fig. 10 but for summer. Both axes are log scaled. Note the larger range compared to Fig. 10.