JOURNAL OF GEOPHYSICAL RESEARCH, VOL. ???, XXXX, DOI:10.1029/,

Statistical and Dynamical Downscaling of Precipitation: An Evaluation and Comparison of Scenarios for the European Alps

J. Schmidli,[1] C. M. Goodess,[2] C. Frei,[1,3] M. R. Haylock,[2] Y. Hundecha,[4] J. Ribalaygua,[5] T. Schmith[6]

J. Schmidli, Atmospheric and Climate Science, ETH Zürich, Universitätsstrasse 16, 8092 Zürich, Switzerland. (schmidli@env.ethz.ch)

[1]Atmospheric and Climate Science, ETH Zürich, Switzerland.

[2]Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich, UK.

[3]Federal Office of Meteorology and Climatology (MeteoSwiss), Zürich, Switzerland.

[4]Institut für Wasserbau, University of Stuttgart, Germany.

[5]Fundación para la Investigación del Clima, Spain.

Abstract. This paper compares six statistical downscaling models (SDMs) and three regional climate models (RCMs) in their ability to downscale daily precipitation statistics in a region of complex topography. The six SDMs include regression methods, weather typing methods, a conditional weather generator, and a bias-correction and spatial disaggregation approach. The comparison is carried out over the European Alps for current and future (2070–2100) climate. The evaluation for the current climate shows that the SDMs and RCMs tend to have comparable biases, but that the SDMs strongly underestimate the magnitude of interannual variations. Clear differences emerge also with respect to the reproduction of year-to-year variations. In winter, over complex terrain, the better RCMs achieve significantly higher skills than the SDMs. Over flat terrain and in summer, the differences are smaller. Scenario results using A2 emissions show that in winter mean precipitation tends to increase north of about 45°N and insignificant or opposite changes are found to the south. There is good agreement between the downscaling models. In summer, there is still good qualitative agreement between the RCMs, but large differences between the SDMs and between the SDMs and the RCMs. According to the RCMs, there is a strong trend towards drier conditions including longer periods of drought. The SDMs, on the other hand, show mostly

[6]Danish Meteorological Institute,

Denmark.

no significant or even opposite changes. Overall the present analysis suggests that downscaling does significantly contribute to the uncertainty in future scenarios, especially for the summer precipitation climate.

1. Introduction

Precipitation is a key component of the hydrological cycle and one of the most important parameters for a range of natural and socio-economic systems: Water resources management, agriculture and forestry, tourism, flood protection, to name just a few. The study of consequences of global climate change on these systems requires scenarios of future precipitation change as input to climate impact models. Direct application of output from General Circulation Models (GCMs) is often inadequate because of the limited representation of mesoscale atmospheric processes, topography and land-sea distribution in GCMs [e.g. *Cohen*, 1990; *von Storch et al.*, 1993]. Moreover, and of particular concern with precipitation, GCMs exhibit a much larger spatial scale (grid-point area) than is usually needed in impact studies and this leads to inconsistencies in frequency statistics, such as the exceedance of a threshold for heavy precipitation [e.g. *Osborn and Hulme*, 1997; *Mearns et al.*, 1997].

Techniques have been developed to downscale information from GCMs to regional scales. These can be categorized into two approaches: "Dynamical downscaling" uses regional climate models (RCMs) to simulate finer-scale physical processes consistent with the large-scale weather evolution prescribed from a GCM [c.f. *Giorgi et al.*, 2001; *Mearns et al.*, 2004]. "Statistical downscaling", on the other hand, adopts statistical relationships between the regional climate and carefully selected large-scale parameters [c.f *von Storch et al.*, 1993; *Wilby et al.*, 2004; *Goodess et al.*, 2005]. These relationships are empirical (i.e. calibrated from observations) and they are applied using the predictor fields from GCMs in order to construct scenarios.

There are a number of application related criteria that contribute to an appropriate choice of downscaling method in a particular context [c.f. *Mearns et al.*, 2004; *Wilby et al.*, 2004]. But there are assumptions involved in both techniques [see also *Giorgi et al.*, 2001] which are difficult to verify a priori and contribute to the uncertainty of results. Rather than developing a downscaling scheme for a particular application, the purpose of the present study is to examine uncertainty in the procedure of downscaling, by comparing several different downscaling models from both approaches.

Several previous studies have compared dynamical and statistical downscaling methods. For example, *Kidson and Thompson* [1998] considered a regression-based statistical model and a RCM integration for present-day climate and found that both methods gave similar levels of skill in the representation of observed temperature and precipitation anomalies for stations in New Zealand. Similarly for stations in Europe, *Murphy* [1999] finds that a regression model for monthly temperature and precipitation anomalies has a comparable performance to a RCM, but scenarios developed from these methods differed substantially [*Murphy*, 2000]. Similarly, large differences were found in precipitation scenarios between a RCM and a weather typing technique over eastern Nebraska [*Mearns et al.*, 1999] and between a RCM and a multivariate regression model in Scandinavia [*Hellström et al.*, 2001]. Using a RCM climate change integration, *Charles et al.* [1999] have tested the stationarity of a statistical downscaling method and found that a relative humidity predictor is required for the reproduction of RCM simulated changes in precipitation occurrence in a global warming experiment. Finally, several inter-comparison studies have adopted dynamical and statistical downscaling for hydrological impact models and find, in part,

considerable differences between downscaling methods [*Wilby et al.*, 2000; *Hay and Clark*, 2003; *Wood et al.*, 2004].

The comparison of downscaling methods in the present study differs in several respects from such previous studies, which makes it particularly informative:

1. It encompasses several models in each category — three regional climate models and six statistical models. This permits the comparison of variability within and between categories of models. Also, we include fundamentally different methodologies in the group of statistical models (daily and seasonal models, single-site and multi-site models). This allows us to study the role of different statistical approaches.

2. The comparison is conducted for the mountain range of the European Alps. Here, numerous mesoscale flow features and precipitation processes shape a complex and regionally variable precipitation climate [e.g. *Frei and Schär*, 1998; *Schär et al.*, 1998]. For example, in winter, it is dominated by the regional response to synoptic disturbances, whereas convection processes contribute in summer. Hence, the Alps constitute an ambitious test ground for downscaling methods, but it is in such regions that downscaling is needed most, because processes are hardly resolved in current GCMs.

3. All downscaling methods are applied for a target resolution of 50x50 km grid boxes - the nominal resolution of the considered RCMs. This procedure avoids inconsistencies between dynamical and statistical methods that arise with the more common application of statistical models to the site scale (i.e. models calibrated with station data).

4. We consider a range of statistics of the day-to-day precipitation variability, including separate measures for precipitation occurrence and ntensity and measures for heavy

precipitation and long dry periods. In addition to seasonal means, statistics on variability and extreme events are relevant for many impacts.

5. This comparison examines the performance of methods for present-day climate (using several different skill measures) and illustrates similarities/differences in the scenarios obtained when all methods are applied to a GCM climate change integration.

The present study makes use of models and data derived in a series of independent but inter-related scientific projects of the European Union: The statistical downscaling methods and the schemes adopted for model intercomparison were developed and applied in the STARDEX project [*Goodess*, 2003]. The RCM integrations driven with perfect boundary conditions were taken from the MERCURE project and RCM climate change integrations from the PRUDENCE project [*Christensen et al.*, 2005]. Several intercomparison studies with a different focus have already been published on these downscaling methods [e.g. *Déqué et al.*, 2005; *Frei et al.*, 2003, 2005; *Goodess et al.*, 2005; *Haylock et al.*, 2005; *van den Hurk et al.*, 2005; *Vidale et al.*, 2005].

The outline of the paper is as follows: Section 2 introduces the precipitation predictands used and the procedures adopted in the comparison of downscaling models. The downscaling models are described in section 3, together with the adopted model chains. Section 4 compares results obtained for present-day climate to observations and section 5 discusses regionalized precipitation scenarios from a climate change integration. Finally, section 6 summarizes the results and draws some conclusions.

2. Study region, predictand and analysis procedure

The study region encompasses the region of the European Alps (geographical area defined by 43.3°–49°N, 2.1°–16.2°E). Its topographic structure is displayed in Fig. 1.

The main feature is the arc-shaped mountain range of the Alps, extending in a west-east direction over a distance of 800 km. The ridge has a width of 100–300 km and a typical crest height of 2500 m. The adjacent low-land regions are interspersed by various hill ranges with spatial scales of 50–200 km and typical elevations of 1000 m.

As predictands we consider selected summary statistics of daily precipitation (see Table 1), with the aim of sampling the precipitation occurrence (FRE, XCDD) and intensity process (INT, Q90, X1D, X5D). The diagnostics are calculated seasonally for each grid point of an Alpine mesoscale grid (see later). In addition, mean values for selected subdomains (see Fig. 1) are obtained by averaging the diagnostics over all grid points in the subdomain. These seasonal diagnostics are referred to as seasonal indices (SI) in the text.

SI from all downscaling models will be determined for a regular lat-lon grid over the Alpine region. The grid spacing is 0.5° (approximately 50 km) and it resolves the major climatic precipitation patterns of the Alpine region.

The observational reference, used for the evaluation of all methods and for the calibration of the statistical downscaling methods, consists of daily precipitation analyses on the above grid for the period 1966–1999. It has been derived by spatial aggregation of rain gauge observations into estimates of mean values for each grid pixel [*Frei and Schär*, 1998]. On average 10 to 50 station observations contribute to the analysis at each grid point. The dataset is very similar to that used in a previous evaluation of RCMs in *Frei et al.* [2003] and is referred to as OBS in the text.

The analysis procedure for the evaluation and comparison of the downscaling methods is based on the SI. Particular attention is given to the representation of interannual variability, as measured by the correlation between downscaled and observed interannual

anomalies of the SI. The evaluation of the year-to-year natural variability, that is of a range of realized climatic states, better captures the potential of a downscaling model for climate change studies. This is so because both dynamical and statistical models have been tuned specifically with reference to present-day climatic conditions and hence to the reproduction of the current climatological mean [see also *Lüthi et al.*, 1996; *Vidale et al.*, 2003]. The analysis is undertaken for each grid point and also for the area-mean SI of the subdomains defined in Fig. 1.

3. Downscaling Procedures

3.1. Techniques

The downscaling models include six statistical downscaling methods (SDMs) and three state-of-the-art regional climate models (RCMs, dynamical downscaling). Table 2 gives an overview of the basic features of the SDMs. They can be grouped into single- and multi-site methods and into daily and seasonal methods. In single-site methods the statistical models are separately calibrated and adopted for each grid point. Whereas multi-site methods are for spatial fields and hence take account of inter-site correlations. The daily methods operate on the daily time scale with daily precipitation series as output. The statistical model undergoes one calibration process and the SI are derived from the daily output. The seasonal methods predict directly series of SI and are therefore calibrated individually for each index.

A large number of potential predictors were considered for the development and calibration of the SDMs. These include sea-level pressure (SLP) and geopotential height (Z), temperature (T), relative (RH) and specific humidity (SH), divergence (DIV), vorticity (VOR), and geostrophic velocity (VG) at different pressure levels (see Table 2). In

addition to these more conventional predictors, further predictors such as moisture flux at 700 hPa (MF700), objective circulation patterns (CPs), and raw GCM precipitation (PRE) were considered for some of the SDMs. Some SDMs use a fixed set of predictors, while others select the predictors from a larger set of potential predictors using automatic or semi-automatic procedures. Often some form of cross-validation is used for predictor selection. Details of SDMs and the selection of predictors are described in the subsections below.

The dynamical downscaling methods (RCMs, 3) encompass three classical limited-area climate models, all with full packages of physical parameterizations. The domains and grids are very similar between the three RCMs. They cover the European continent and parts of the North-western Atlantic, with the Alpine region located near the domain centers. The grid spacing of the models is about 50 km. The integrations used in this study were conducted in recent European climate modelling projects (MERCURE and PRUDENCE). The selected models span the range of behavior found for a larger set of European RCMs participating in these projects [*Frei et al.*, 2003, 2005]. The SI for the RCMs were calculated on the respective native model grids and then interpolated to the common latitude-longitude grid of our intercomparison [similar to *Frei et al.*, 2003]. Details of individual models are described in the subsections below.

3.2. Experiments

A consistent set of downscaling experiments is undertaken with all the methods (Fig. 2). These encompass one experiment using large-scale predictors/lateral boundary forcing from reanalysis data and one experiment with predictors from climate change simulations with a global climate model (GCM). The former experiment is used for calibration and

evaluation purposes. Particular focus with these experiments will be given to the representation of year-to-year anomalies in SI. Note that the forcing of the methods by observed large-scale conditions allows a comparison of downscaled results to observed anomalies [*Lüthi et al.*, 1996; *Vidale et al.*, 2003].

In the case of RCMs, reanalysis driven downscaling experiments are based on the 15-year ECMWF reanalysis [ERA15, *Gibson et al.*, 1999] for 1979–1993. Note that these experiments originate from project MERCURE, as the newer 40-year reanalysis (ERA40) was not yet available. As for the SDMs, reanalysis driven experiments are based on the National Center for Environmental Prediction reanalysis [NCEP *Kalnay et al.*, 1996; *Kistler et al.*, 2001]. Again, ERA40 was not available at the time these experiments were undertaken in STARDEX. To enable an independent evaluation of the SDMs and a comparison to the RCMs, the 15 years 1979–1993 are taken for evaluation. The SDMs were calibrated over the remaining available period of NCEP and OBS (1966–1978, 1994–1999). Note that all potential predictor variables were interpolated to a standard 2.5° latitude/longitude grid.

The climate change experiment with all the downscaling models was conducted with predictors / boundary-forcing from the atmosphere-only GCM (HadAM3H/P) of the Hadley Centre at the UK Met. Office. HadAM3 was derived from the coupled atmosphere-ocean model HadCM3 [*Gordon et al.*, 2000; *Johns et al.*, 2003] and is described in *Pope et al.* [2000] (HadAM3H) and in *Jones et al.* [2005] (HadAM3P). The forcing fields for the downscaling models came from GCM integrations for the time-slices 1961–1990 (CTRL) and 2071–2100 (SCEN). For CTRL, the sea surface temperature and sea ice distributions for HadAM3 were prescribed from observations of the same period. For SCEN, sea surface

conditions were constructed from observations and anomalies from a transient integration of HadCM3 using the IPCC SRES A2 emission scenario [*Nakicenovic and et al.*, 2000]. With this scenario, HadAM3 simulates a global mean surface temperature increase of 3.18 K between CTRL and SCEN (D. Rowell, personal communication, 2004). HadAM3 has a grid spacing of about 150 km. But again only the information interpolated onto the standard 2.5° grid is used.

Three ensemble integrations were carried out with HadAM3 for both time slices, starting from different initial conditions. All statistical models and two of the three RCMs were driven by all six ensemble members.

For historical reasons two different GCM simulations have been used with the SDMs and the RCMs, which differ in the GCM model version (Fig. 2). The newer version, HadAM3P, is used with the SDMs, and the older version, HadAM3H, with the RCMs. We do not expect this difference to disturb the comparability of our results because the climate change patterns for precipitation over Europe are quite similar between the two GCMs and between RCMs forced with either model version [see also *Frei et al.*, 2005].

3.3. The statistical downscaling methods (SDMs)

3.3.1. Canonical correlation analysis (CCA)

The canonical correlation analysis [CCA, 2; *Barnett and Preisendorfer*, 1987] models the SI directly using seasonal means of circulation variables. For each season and precipitation index a CCA was carried out using all 15 possible combinations of four potential predictors. The best set of predictors was selected using cross validation (see Table 4). The skill measure was the average Spearman correlation over all grid points. Note that the predictor set varies between indices and seasons but is the same for all grid points. The CCA was

performed on the cross-covariance matrix of the leading principal components (PCs) of the predictor and predictand field. Only statistically significant PCs were retained [*Barnett and Preisendorfer*, 1987]. Therefore the number of eigenvectors retained was different for each predictor, predictand and season (see Table 2).

3.3.2. Multiple linear regression (MLR)

Like the CCA, the multiple linear regression model (MLR) downscales the SI directly from seasonal measures of the large-scale circulation, but unlike CCA, it establishes a separate model for each grid point. Each index is expressed as a linear function of a set of potential predictors (see Table 2), which were selected using correlation analysis between the indices and all the available predictors. In addition to the seasonal mean values of the predictor variables, their seasonal 90th and 10th percentiles were considered as potential predictor variables. Predictors for each index are then selected from the potential predictors using the forward selection method. The predictor values in the regression equation are taken as the average over the nearest four grid points to the target location. Note that apart from the more common predictors this method also uses objective circulation patterns [CPs; *Bárdossy and Plate*, 1992] and moisture flux at 700 hPa (MF700).

The selected predictors for the indices vary from season to season and from index to index. However, for a given season the tendency is that the leading predictors for most of the indices are similar (Table 5).

3.3.3. Multivariate auto-regressive model (MAR)

This is a classification based downscaling approach based on the modified version of the space-time model described in *Bárdossy and Plate* [1992]. The model is used to generate

daily series of precipitation at multiple locations simultaneously by taking into account the spatial correlation of the observed series. Objective circulation patterns defined by classifying the distribution of anomalies of sea level pressure using a fuzzy rule-based classification scheme [*Bárdossy et al.*, 1995, 2002] are used to condition the model parameters.

The distribution of the daily precipitation amount at a given location and day is modelled by a random variable with a mixed discrete-continuous distribution. The expected value is modelled as a function of the moisture flux at 700 hPa and the circulation pattern type. For further details see *Stehlík and Bárdossy* [2002].

3.3.4. Conditional weather generator (CWG)

A conditional weather generator (CWG) [e.g *Goodess and Palutikof*, 1998] is implemented as follows. First, a surface pressure pattern is obtained as the average pressure difference between wet and dry days observed at a given station. Second, a circulation index is obtained by regressing the daily surface pressure field onto this pattern. The circulation index is divided into a number of quantiles, usually between 5 and 10. Third, for each quantile the following precipitation quantities are calculated: the probability for wet/dry days, the probabilities for a wet/dry day following a dry/wet day, and the two gamma distribution parameters for precipitation amount. And finally, a two-state Markov Chain process combined with random sampling from the gamma distribution [*Wilks and Wilby*, 1999] is used to generate the daily precipitation series. Note that the CWG was applied only for winter and summer.

3.3.5. Two-step analog method (ANA)

In the first step, a set of analogs (the 30 most similar days) is selected from a reference dataset based on the similarity of the geostrophic wind (direction and velocity) at 1000

and 500 hPa. In the second step, based on the 30 analogs for each day of the season, a probabilistic model for precipitation is built. The probabilistic model gives better skill than using the average precipitation of the analog days. Lower tropospheric humidity was tested as an additional predictor, but it was found to give no additional skill.

3.3.6. Local intensity scaling (LOCI)

The local intensity scaling [LOCI; *Schmidli et al.*, 2005] uses GCM precipitation as a predictor, as proposed by *Widmann et al.* [2003], in contrast to most statistical downscaling methods which use circulation-based predictors [e.g. *Wilby and Wigley*, 2000]. The idea is that GCM precipitation, in some sense, integrates all relevant large-scale predictors. Thus deviations between the large-scale GCM precipitation and regional precipitation are to first order due to biases from systematic GCM errors and the lack in surface orography. Because GCM biases are less variable than relationships with circulation indices, it is expected that the GCM precipitation predictor should be less vulnerable to non-stationarities in the predictor-predictand relationship. In essence, LOCI compensates for biases in wet-day frequency and intensity of GCM precipitation by applying local corrections to the precipitation frequency distribution at each predictand grid point. A detailed description of LOCI is given in *Schmidli et al.* [2005]. LOCI can be regarded as a correction of GCM output which serves as a benchmark for more sophisticated downscaling methods.

3.4. The Regional Climate Models (RCMs)

3.4.1. CHRM

CHRM originates from the operational weather forecasting model of the German and Swiss meteorological services [*Majewski*, 1991], from which it was adapted into a climate

version at ETH Zürich [*Lüthi et al.*, 1996; *Vidale et al.*, 2003]. The model has a resolution of 0.5° (about 55 km) in a rotated pole coordinate system and 20 vertical levels in hybrid coordinates. Modifications for the climate version were made, among others, in the soil-atmosphere-vegetation transfers, the physiographic and biophysical parameters, the soil profiles and the convection scheme [see *Vidale et al.*, 2003].

3.4.2. HadRM3

HadRM3 is the regional climate model of the Hadley Centre of the UK Meteorological Office [*Jones et al.*, 1995; *Noguer et al.*, 1998]. It is operated at a resolution of 0.44° (about 50 km) and with 19 vertical levels. Its dynamics and physical parameterizations are similar to HadAM3, the atmosphere-only GCM from which the climate change integration is downscaled in this study. HadRM3 and HadAM3 are described in *Jones et al.* [2004] and details of their physical parameterizations in *Pope et al.* [2000]. Two different model versions were used for the integrations driven by reanalysis and GCM (HadRM3H and HadRM3P, respectively). The difference between the two versions for precipitation statistics in the Alps is small [*Frei et al.*, 2005].

3.4.3. HIRHAM

HIRHAM is the RCM of the Danish Meteorological Office. It is operated at a resolution of 0.44° (about 50 km) and with 19 vertical levels. Its dynamical part originates from HIRLAM [*Källén*, 1996] and the physical part from ECHAM4 [*Roeckner et al.*, 1996]. In this study, we use HIRHAM integrations from an updated version of HIRHAM4 [*Christensen et al.*, 1996], using high-resolution datasets of land surface characteristics [*Christensen et al.*, 2001] and a cyclic repetition for soil moisture initialization [*Chris-*

tensen, 1999]. Results on European precipitation statistics for the HIRHAM integrations used in this study are also described in *Christensen and Christensen* [2003, 2004].

4. Reproduction of present climate

This section focuses on the evaluation of the present-day precipitation climate as downscaled from the reanalysis runs (NCEP for the SDMs, ERA15 for the RCMs, see 2). The use of reanalysis predictors / boundary fields allows a direct comparison of the downscaled and observed climate including the year-to-year variations of the SI. Also in reanalysis mode there are almost no biases in the predictors in contrast to downscaling from GCM control runs. The evaluation results for autumn are presented first and in more detail (section 4.1), as autumn is the most important season for heavy precipitation in the Alpine region. Thereafter we continue with a systematic evaluation including further seasons and indices (section 4.2). Evaluation criteria include biases and standard deviation of interannual anomalies, and the correlation between downscaled and observed interannual anomalies of the SI. Finally, the reanalysis and HadAM3 control driven downscaling results are compared (section 4.3) and the main findings are briefly summarized (section 4.4).

4.1. Autumn heavy precipitation

Figure 3 compares the spatial distribution of the 90% quantile (Q90) for autumn (SON) of the downscaling models and the observations. Only a representative sample of downscaling models is depicted for reasons of space. The RCMs and the seasonal SDMs (CCA and MLR) show good qualitative correspondence with the gross regional distribution. They reproduce the higher Q90 values along the southern rim of the Alps and the three

embedded maxima exceeding 40 mm day^{-1} (SE of the Massif Central, south central Alps and south-eastern Alps). The daily downscaling models (MAR, ANA) and the benchmark (LOCI) also capture the gross regional distribution but they considerably underestimate the orographic amplification of Q90. It is not surprising that larger biases are found for the daily SDMs in comparison to the seasonal SDMs. While the former are calibrated to reproduce mean precipitation, the latter methods directly use the observed Q90 values for model calibration. Despite considerable differences for some of the downscaling models, all of them are much closer to the observations than ERA40, which has a large dry bias. A similar result was found for NCEP, not shown. As would be expected, the SDMs reproduce the correct locations of the maxima, while the RCM-simulated maxima are occasionally be shifted by a few grid points relative to the observed maxima.

Figure 4 compares the magnitude of the interannual variations of Q90. The gross patterns are again well simulated by the RCMs, with low variability north of the Alpine crest and higher variability to the south. Some discrepancies are found for the finer details of the patterns and the exact location of the maxima. All SDMs, on the other hand, strongly underestimate the observed variability. The domain-average ratio of downscaled to observed standard deviation of Q90 varies between about 2/3 for MLR and 1/4 for ANA. Note also the substantial differences between the two methodologically related methods CCA and MLR. The underestimation is larger for the CCA model which uses principal components of the predictands and predictors instead of grid-point values. As precipitation is a relative quantity, meaning that the standard deviation is larger where the mean is larger, it is expected that a model with a negative bias will underestimate the variance. However, the variance underestimation by the SDMs is substantially larger than would

be expected from the model bias alone (c.f. Figs 3, 4). This variance underestimation is a well-known problem of SDMs [*von Storch*, 1999].

Of particular interest for climate change applications is the ability of a downscaling model to capture interannual variations [e.g. *Lüthi et al.*, 1996]. Figure 5 depicts the spatial pattern of the anomaly correlation between downscaled and observed time series of SI. Only a representative sample of downscaling models are included in order to illustrate the differences between occurrence (FRE) and intensity (INT) related indices. With respect to FRE, the downscaling models are generally very skillful. The average correlation varies from about 0.6 for MAR to 0.8 for LOCI and ANA. Note, however, the large regional differences in skill. For CHRM, for instance, the skill varies from values below 0.4 in eastern parts of the domain, to values larger than 0.9 in western parts of the domain. A similar behavior was found for HADRM and HIRHAM, not shown. With respect to INT, the downscaling skill is generally much lower and spatially even more variable. This is not surprising, as, in general, FRE is strongly dependent on the large-scale atmospheric circulation forcings, whereas INT depends more on local processes and moisture fluxes. Similarities in the patterns for the different downscaling models indicates that at least part of the large spatial variability in skill is due to real differences in predictability from region to region, and not only to model deficiencies. As for instance the tendency to higher correlations along parts of the southern Alpine rim and the tendency to lower correlations in the eastern part of the domain. The generally higher skill for FRE, in comparison to INT, is found also in the other seasons and it is representative of the generally higher skill for occurrence related indices (FRE, XCDD, MEA), in comparison to intensity related

indices (INT, Q90, X1D, X5D). It should, however, be noted that autumn is not the season with the highest skills (see below).

4.2. Systematic evaluation

How general are the results obtained for the autumn season? In this section, we present a systematic evaluation and comparison of the winter and summer season for three representative subdomains (see Figure 1). These subdomains cover the variability of the Alpine region with flat areas (region WEST), the northern rim of the main ridge (NALP) and a region with frequent heavy precipitation in Ticino, Southern Switzerland (TIC). The three regions consist of 45, 27, and 15 grid points (0.5° grid), and they cover an area of 1.5, 1 and 0.5 GCM grid points (2.5° grid), respectively.

4.2.1. Bias and Standard Deviation

Figure 6 summarizes the biases for winter and summer for four SI and the three subdomains. The most striking feature is the large difference in bias from region to region. The difference between the regions is often larger than the difference between the downscaling methods.

In winter, the smallest biases are generally found for the region WEST. Typically, the bias is less than 10%. Larger discrepancies are found for HIRHAM, which has a wet bias resulting from too frequent precipitation events leading to too high FRE values and too low XCDD values [see also *Frei et al.*, 2003], and for CCA and CWG with too short dry periods (XCDD too small). For the smaller and more mountainous regions the biases are generally larger, especially for TIC. The largest biases, typically around 30%, are found for HIRHAM, HADRM3 and MAR for some of the indices/regions.

In summer, the biases tend to be somewhat larger. For the region WEST, the RCMs have substantial biases for the occurrence related indices. (CHRM, for instance, has a (dry) frequency bias of about 10–20% and resulting bias in XCDD of 30–50%.) The SDM biases in FRE and XCDD, on the other hand, are smaller than 10%. For the intensity related indices (INT, Q90), the RCMs and SDMs have similar biases. Note that for these indices, the region TIC is often the region with the smallest bias, at least for the SDMs.

In summary, the biases are of comparable magnitude for the better RCMs and the better SDMs, with the exception of the occurrence-related indices (FRE, XCDD) in summer, for which the RCMs tend to have larger biases. Further conclusions, with respect to model differences, are difficult to draw, due to the large variability of the bias from region to region. However, the differences between the regions appear to be quite systematic. In winter, the smallest biases tend to be found for the region WEST, which is the largest and least mountainous region. In summer, however, the results are more variable.

Figure 7 displays the ratio of downscaled to observed standard deviation for winter and summer for the four SI and the three subdomains. Now the differences between the methods is larger than the differences between the regions. The figure corroborates our previous finding of large underestimation of interannual variability by the SDMs. For both winter and summer, for most indices and SDMs, the downscaled standard deviation is smaller than half of the observed value. For CWG it is often even less than 25% of the observed value (which explains the missing bar for CWG). Relatively good results are obtained for ANA in winter for the occurrence indices (FRE, XCDD), and for MAR for the region WEST in winter for the intensity indices (INT, Q90). In comparison, the RCMs simulate about the correct amount of variability.

4.2.2. Interannual variations

Figure 8 depicts the correlation skill of the SI for winter and summer for the three subdomains. In order to reduce the influence of stochastic / local forcings the SI are aggregated over the respective subdomains prior to calculating the correlations with the observed SI. The results confirm the tendency to higher correlations for the SI related to the occurrence process (FRE, XCDD) than for the SI related to the intensity process (INT, Q90). The difference between the two categories is especially pronounced for the SDMs in winter, with correlations between 0.6 and 0.9 for FRE, but typically below 0.5 for INT and Q90. Note, however, the very good skill for the intensity indices (INT, Q90) for some of the models (CHRM, MAR, ANA) for the region TIC in summer. For both seasons, the skill of the benchmark (LOCI) is comparable to the skill of the best downscaling models. The good results for LOCI, reflect the generally good quality of the reanalysis precipitation with respect to temporal variations, quite in contrast to its large biases. It should however be noted, that even a perfect downscaling model would not obtain a correlation of 1.0 due to the limited predictability of the interannual variations, especially for the summer season [cf. *Vidale et al.*, 2003].

Comparison of the correlation results for all seasons and all indices shows that the analysis of just FRE (occurrence process) and INT (intensity process) gives a good overall picture of the characteristics of a specific downscaling model. The differences in skill between downscaling models are very similar for indices from the same group. Thus precipitation intensity (INT) is a good proxy, with respect to skill, for the more extreme indices such as Q90, X1D, and X5D. It is therefore sufficient to concentrate in the following on the downscaling skill for FRE and INT.

A compact overall comparison of downscaling methods, regions, and seasons is given in Table 6. With respect to the methods, the best overall skill in terms of anomaly correlation is obtained for LOCI, the bias-corrected ERA40 reanalysis, followed by the RCMs (CHRM, HADRM3, HIRHAM) and the daily SDMs (ANA and MAR). The lowest overall skill is found for the seasonal SDMs (MLR and CCA) and CWG. For FRE, good skill (correlation $r > 0.6$) is found for all methods and all regions in winter and autumn. For INT, on the other hand, good skill is obtained only for LOCI and the better RCMs in winter and autumn, and for some regions also for MAR and ANA. With respect to the seasons, the highest skill is obtained usually in winter, followed by autumn and summer. For the region TIC, however, the highest skill is observed in autumn, the most important season for heavy precipitation and a time of high synoptic activity. Even in summer, TIC has a relatively high skill for INT. With respect to the regions, the ranking depends on the season. In winter, the highest average skill is obtained for the region WEST. Whereas in summer and autumn, the highest average skill is obtained for the region TIC.

4.2.3. Dependance of downscaling skill on the spatial scale

It has been shown, that there are large variations in skill from region-to-region and from season-to-season. How large are the variations in skill within a climatologically relatively uniform region? In order to investigate this question, we focus on precipitation intensity (INT), as this is the more challenging parameter for downscaling methods, but also the parameter which is more relevant for precipitation extremes.

Figure 9 depicts the correlation skill for the region-mean INT and the range of correlations obtained for INT for individual grid points. The former corresponds to the values in Figure 8. The less condensed display, however, makes the differences between the down-

scaling methods, the regions, and the seasons more clearly visible. In winter, the RCMs are clearly superior to the SDMs for the regions NALP and TIC, but of comparable skill to the SDMs MAR and ANA for the region WEST. In summer, differences between the RCMs and the SDMs are less systematic. The outstanding result in this season is the relatively good skill for the region TIC and the very low skill for NALP.

Comparing the average correlations obtained for the individual grid point series (dashed line) to those for the region-mean series (bold line) shows that the former are typically lower. This is no surprise. It is expected that spatial aggregation increases the predictability by averaging out local stochastic influences on the predictand. The range of correlations obtained for the individual grid point series (shaded band) illustrates the large spatial variablity in skill, also within climatologically relatively uniform regions. The criss-cross of the lines for selected grid points implies that an intercomparison of methods/regions based on single grid points, or even worse, on single stations, would yield rather random results, especially in summer. Note that the estimation error of the correlation coefficient from relatively short time series (15 years) is a major contributor to the uncertainty (Figure 10). Much longer time series would be required for clearer results based on results for single grid points or stations. Therefore spatial aggregation is essential in order to detect significant differences when comparing methods and regions.

4.3. The HadAM3 control run

In this section, results for the GCM chain for present conditions (see Figure 2) are briefly presented. Figure 11 depicts the same biases as in Figure 6, but now for the GCM-driven downscaling models. Comparison of the two Figures shows that the biases are mostly similar, especially the relative differences between the methods. This indicates that for

most models the biases are not overtly sensitive to the transition from reanalysis to GCM driving. Larger differences are found for MLR which uses local grid-point predictors, and as expected for LOCI which has to be recalibrated for the GCM. In general, however, the biases and especially the mesoscale bias patterns (not shown) are determined mainly by the downscaling model and not by the driving GCM or reanalysis. This indirectly attests to the quality of the GCM.

4.4. Summary

• Performance is generally quite similar for indices related to the occurrence process (FRE, XCDD) and for those related to the intensity process (INT, Q90, X1D, X5D); the skill for MEA is comparable to that for FRE. Therefore results for just FRE and INT provide a good characterization of a downscaling method. Typically, the performance is best for FRE and MEA, a little lower for XCDD, and substantially lower for the intensity indices. (An exception to this rule is the summer season in region TIC.)

• There are large differences in performance from region-to-region and from season-to-season. The ranking of the seasons depends on the region. The performance is best in winter and spring for the region WEST, in autumn and winter for NALP, and in autumn and spring for TIC. On average, summer is the season with the lowest skill in all regions, but there are also exceptions to this rule.

• The variation of the skill from grid point to grid point within a given region can be very large, due partly to random sampling errors [see also *Goodess et al.*, 2005]. Thus considerable aggregation, as has been undertaken in this study, is required in order to detect systematic differences when comparing methods and regions.

• All downscaling models are able to reproduce mesoscale patterns in the climatology (mean conditions in SI) not resolved by the driving model. The spatial congruence tends to be better for the SDMs than for the RCMs, for which the patterns may be shifted by a few grid points. The seasonal SDMs tend to have the smallest biases. The magnitude of the biases for the daily methods and the RCMs (CRHM and HADRM3) are comparable.

• All SDMs underestimate the magnitude of the interannual variations, especially for the intensity indices and the smaller regions. This underestimation is particularly large for CWG and CCA. Relatively good results were obtained for MAR and the intensity indices and for ANA and the occurrence indices, in winter. The RCMs produce about the right amount of interannual variability.

• Significant differences are found with respect to the reproduction of interannual variations, in particular of the intensity indices. In winter, the better RCMs (CHRM and HADRM3) are clearly superior to the SDMs for the two mountainous regions (NALP and TIC). The differences are smaller over relatively flat terrain (WEST). In summer, the better RCMs (CHRM and HADRM3) and the better SDMs (MAR and ANA) tend to have similar skill. In general, the daily SDMs (MAR and ANA) tend to have higher skills than the seasonal SDMs (CCA and MLR).

• The performance of the LOCI benchmark is in most cases comparable to the best downscaling models. With respect to SI, the better RCMs tend to show added value only for the region TIC. More generally, however, it can be expected that the RCMs produce more realistic daily fields and heavy precipitation events than the LOCI benchmark. Even higher skills can be expected by applying LOCI to RCM output, but this application was not examined systematically in the present study.

• For a given method and season, the bias patterns are often very similar for indices from the same group (e.g. intensity indices). For the RCMs, in particular, the patterns are often also similar for different seasons (e.g. winter and autumn) and for different models. Thus there appear to be regions for which downscaling is intrinsically more skillful and others for which it is less skillful.

5. Simulated change in daily precipitation statistics

This section compares the simulated change of the daily precipitation statistics, the SI, as downscaled by the RCMs and the SDMs. All downscaling models were forced by the HadAM3 integrations for the IPCC SRES A2 emission scenario (see section 3). Results are presented for winter, summer, and autumn.

5.1. Winter

Figure 12 shows the change in mean precipitation in winter (MEA, DJF). Mean precipitation was chosen because it is expected to be one of the easier parameters to downscale and nevertheless it is important for hydrological applications. Most models (GCM, LOCI, the RCMs, and ANA) show an increase north of the Alpine ridge and a transition to small changes or decreases near the Mediterranean. The two linear downscaling methods (CCA and MLR) differ considerably, especially in the southern parts of the domain, despite having a similar evaluation skill under current climate conditions. According to the CCA method, for which the main significant predictor is SLP, the scenario conditions imply an enhanced WNW flow over most of central Europe and the Alps leading to increases in mean precipitation.

A quantitative comparison of the downscaling models for the region WEST is provided in Figure 13. With respect to the simulated change of MEA the downscaling models fall into three groups. The RCMs, ANA and CCA together with the GCM and LOCI show increases of 20–30%, the two SDMs MLR, and MAR show increases of 40–60%, and CWG exhibits no changes. How are the changes in MEA related the precipitation frequency and intensity? For the RCMs the increase in MEA originates from about equal increases in FRE and INT. For the SDMs the relative contributions of FRE and INT are much more variable between the methods. Overall the coherence between the RCMs is quite good. The changes obtained by the SDMs, however, vary considerably from model to model, even for similar downscaling methods (e.g. CCA and MLR).

5.2. Summer

Figure 14 compares the relative change in maximum length of dry spells in summer (XCDD, JJA). All RCMs show an increase of XCDD, whereas the SDMs show strong decreases (CCA and MLR) or no change (MAR and ANA). The simulated increase in XCDD is 50–100% for the GCM, LOCI, CHRM, and HadRM3P, and 25–50% for HIRHAM. The results for the SDMs are much less uniform and range from no change for MAR, ANA, and CWG, to large decreases of XCDD, by more than a factor of two, for CCA and MLR. In view of the low evaluation skill, these results are not interpreted any further. Similar patterns of change are found also for FRE, that is strong decreases of FRE for the RCMs, and large increases or no change for most SDMs (not shown).

The quantitative comparison of the indices for summer (Figure 15) reveals much larger differences between methods than in winter. Note that for MLR there is an inconsistency in the changes for FRE which decreases and for XCDD also decreasing (by a factor of 2

and more). Such inconsistencies can arise in seasonal methods when different empirical models are developed for individual indices. These inconsistencies are potentially a serious drawback of the seasonal methods. The daily downscaling methods (MAR and ANA), on the other hand, provide time series of daily precipitation and therefore the SI will implicitly be more consistent.

5.3. Autumn

Figure 16 depicts the relative change of the 90% quantile in autumn (Q90, SON). The RCMs and MAR show increases in Q90. Again, the other SDMs show no changes (CCA and ANA) or even decreases (MLR). According to the CCA method, for which the main significant predictor in autumn is SLP, the scenario conditions imply a reduction of low pressure conditions in the Alpine region leading to lower precipitation frequency.

Figure 17 reveals a coherent picture for the RCMs: A moderate decrease in MEA resulting from a strong decrease in FRE which is partially compensated by an increase in INT. Consistent with this we found also an increase in Q90 and XCDD. Note from Figure 16 that the differences between the three subregions are smaller than for any of the other seasons. The SDMs, on the other hand, exhibit mainly smaller or even no changes.

6. Summary and conclusions

6.1. Evaluation

In the present study we undertook an intercomparison of daily precipitation statistics as downscaled by nine different downscaling models, six statistical and three dynamical, for the region of the European Alps. An evaluation of the downscaling models for the present climate shows that:

• All models are capable of representing mesoscale spatial patterns not resolved by today's GCMs.

• While the seasonal SDMs tend to have smaller biases, the RCMs and the daily SDMs tend to better capture the interannual variability. However, all SDMs tend to strongly underestimate the magnitude of the interannual variations, especially in summer and for the indices related to precipitation intensity.

• No method is always superior. Nevertheless a clear pattern emerges with respect to the reproduction of interannual variations. In winter, the better RCMs (CHRM and HADRM3) are clearly superior to the SDMs for the two mountainous regions (NALP and TIC). The differences are smaller for the less mountainous region (WEST). In summer, the better RCMs (CHRM and HADRM3) and the better SDMs (MAR and ANA) tend to have similar skill.

• Performance varies substantially from region to region and from season to season. Performance is generally better for the indices related to precipitation occurrence than for the indices related to precipitation intensity. Within each group, the performance and the spatial variation of the skill tend to be similar.

• The variation of the skill from grid point to grid point within a given region can be very large, due partly to random sampling errors [see also *Goodess et al.*, 2005]. Thus considerable aggregation is required in order to identify the reliability of and assess differences between methods. Thus analyses based on single grid points or even single stations would be of very limited use in a highly complex region such as the European Alps.

Comparison of the downscaling results from the reanalysis with those from the GCM present-day control run shows that the errors inherited from the driving GCM were rela-

tively small and did not alter the specific error characteristics of the downscaling models. This attests to the quality of the GCM's present-day climate and to the ability of the downscaling models to extract the large-scale, more or less unbiased, information. However, one of the models, MLR, was found to have substantially larger biases when driven by GCM predictors. This might be due to the use of predictors based on local grid point values.

6.2. Scenario

The RCM simulated future change in European precipitation climate shows a seasonally very distinct pattern: In winter, regions north of about 45°N experience an increase in mean precipitation while in the Mediterranean region there is a tendency towards decreases [see also *Frei et al.*, 2005]. Results are very consistent between the three RCMs. All three RCMs attribute the increase in mean precipitation (MEA) about equally to an increase in wet-day frequency (FRE) and precipitation intensity (INT). In addition the spatial patterns of relative change are quite similar. Most of the SDMs produce an increase in mean precipitation similar to that of the RCMs. However, the partition of the increase between FRE and INT varies considerably between the SDMs. Nevertheless, the generally good agreement between the downscaling models suggests that the downscaled scenario for winter is fairly reliable and robust.

In summer, the RCMs simulate a strong decrease in mean precipitation in the entire Alpine region. This decrease is mainly due to a substantial reduction of the wet-day frequency. The smaller number of wet days results in a large increase, 50–100%, of the maximum length of dry spells (XCDD). In comparison to winter, the differences between the downscaling models, especially between the RCMs and the SDMs, but also between

the RCMs, are much larger. Even the two daily SDMs with good evaluation skill (MAR and ANA), produce almost no changes or decreases. This suggests that the RCM simulated changes for summer are not primarily related to large-scale circulation changes. Possibly, physical feedback processes with, for instance, the land surface [e.g. *Wetherald and Manabe*, 1995; *Seneviratne et al.*, 2002; *Schär et al.*, 2004] may contribute to the scenario. Overall the differences between the RCMs and SDMs, and the substantial biases of the RCMs in summer highlight the still large uncertainties of the scenario results for the summer season.

In autumn, the region experiences a decrease in mean precipitation resulting from a strong decrease in wet-day frequency and moderate increase in precipitation intensity. Again the results are very similar for the three RCMs.

It is interesting to compare the scenario changes for winter and autumn. In winter, the simulated changes in FRE and INT have the same sign, both indices increase by about 10%. In autumn, on the other hand, the simulated changes are of opposite sign. The similar changes of INT in autumn and winter (and also spring) suggests that the increase might be related to a common cause, such as an increased moisture content of the atmosphere resulting from higher mean temperatures [e.g. *Frei et al.*, 1998]. Note that this pattern, same sign of FRE and INT in winter and opposite sign in autumn is also found in the observed trends for the 20th century [*Schmidli and Frei*, 2005].

From the many sources of uncertainty associated with scenarios for climate change impacts, the present study has focused entirely on uncertainties related to the derivation of regional climate information, that is to statistical and dynamical downscaling. The present analysis suggests that the contribution to uncertainty from downscaling is relatively small

in winter and autumn, but very significant in summer. Clearly, more research will be needed to understand the different model responses and eventually reduce the spread in the projections.

Acknowledgments. We are indebted to the hydrological and meteorological services in the Alpine region [see *Frei and Schär*, 1998] for providing access to daily precipitation data. Reanalysis data was provided by the NOAA-CIRES, Boulder, USA, through their Web site (www.cdc.noaa.gov) and by the ECMWF, Reading, UK (www.ecmwf.int) through MeteoSwiss, Zürich. Data from regional climate models were kindly provided through the PRUDENCE data archive, funded by the EU through contract EVK2-CT2001-00132, and by the Hadley Centre UK Met. Office, the Danish Meteorological Institute and the Institute for Atmospheric and Climate Science ETH through the EU-project MERCURE. Analyses and graphics were produced with the open source software package R and NCAR's NCL. This work was funded by the Commission of the European Union under project STARDEX (STAtistical and Regional dynamical Downscaling of EXtremes for European regions), contract EVK2-CT-2001-00115. The Swiss contribution was funded by the Swiss Ministry for Education and Research (contract 01.0265-2) and by the Swiss National Science Foundation (NCCR Climate).

References

Barnett, T., and R. Preisendorfer (1987), Origins and levels of monthly and seasonal forecase skill for United-States surface air temperatures dtermined by canonical correlation analysis, *Mon. Wea. Rev.*, *115*, 1825–1850.

Bárdossy, A., and E. J. Plate (1992), Space-time model of daily rainfall using atmospheric circulation patterns, *Water Resour. Res.*, *28*, 1247–1259.

Bárdossy, A., L. Duckstein, and I. Bogardi (1995), Fuzzy rule-based classification of atmospheric circulation patterns, *Int. J. Climatol.*, *15*, 1087–1097.

Bárdossy, A., J. Stehlík, and H.-J. Caspary (2002), Automated objective classification of daily circulation patterns for precipitation and temperature downscaling based on optimized fuzzy rules, *Clim. Res.*, *23*, 11–22.

Charles, S. P., B. C. Bates, P. H. Whetton, and J. P. Hughes (1999), Validation of downscaling models for changed climate conditions: case study of southwestern Australia, *Clim. Res.*, *12*, 1–14.

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.

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

Christensen, O. B. (1999), Relaxation of soil variables in a regional climate model, *Tellus*, *51A*, 674–685.

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.

Christensen, O. B., J. H. Christensen, B. Machenhauer, and M. Botzet (2001), Very high-resolution regional climate simulations over Scandinavia - Present climate, *J. Climate*, *11*, 3204–3229.

Cohen, S. J. (1990), Bringing the global warming issue close to home: the challenge of regional impact studies, *Bull. Amer. Meteor. Soc.*, *71*, 520–526.

Déqué, M., et al. (2005), An intercomparison of regional climate models for Europe: Assessing uncertainties in model projections, *Climate Dyn.*, submitted.

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. Letters*, *25*, 1431–1434.

Frei, C., J. H. Christensen, M. Deque, 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-Atmos.*, *108 (D3)*, Art. No. 4124.

Frei, C., R. Schöll, S. Fukutome, J. Schmidli, and P. L. Vidale (2005), Future change of precipitation extremes in Europe: An intercomparison of scenarios from regional climate models, *J. Geophys. Res-Atmos.*, accepted.

Gibson, J. K., P. Kallberg, S. Uppala, A. Hernandey, A. Nomura, and E. Serano (1999), ERA-15 description (version 2), *ECMWF Reanalysis Proj. Rep. Series (Reading UK)*, *1*, 74.

Giorgi, F., B. Hewitson, J. Christensen, C. Fu, R. Jones, M. Hulme, L. Mearns, H. Von Storch, and P. Whetton (2001), Regional climate information — evaluation and projections, in *Climate Change 2001: The scientific basis*, p. 944pp, J. T. Houghton et al.,

Eds., Cambridge University Press.

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

Goodess, C. M., and J. P. Palutikof (1998), Development of daily rainfall scenarios for southeast spain using a circulation-type approach to downscaling, *Int. J. Climatol.*, *18*, 1051–1083.

Goodess, C. M., et al. (2005), An intercomparison of statistical downscaling methods for europe and european regions — assessing their performance with respect to extreme temperature and precipitation events, *Climatic 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.

Hay, L. E., and M. P. Clark (2003), Use of statistically and dynamically downscaled atmospheric model output for hydrologic simulations in three mountainous basins in the western United Stated, *J. Hydrology*, *282*, 56–75.

Haylock, M. R., G. C. Cawley, C. Harpham, R. L. Wilby, and C. M. Goodess (2005), Downscaling heavy precipitation over the UK: a comparison of dynamical and statisticalmethods and their future scenarios, *Int. J. Climatol.*, submitted.

Hellström, C., D. Chen, C. Achberger, and J. Räisänen (2001), Comparison of climate change scenarios for sweden based on statistical and dynamical downscaling of monthly precipitation, *Clim. Res.*, *19*, 45–55.

Johns, T. C., et al. (2003), Anthropogenic climate change for 1860 to 2100 simulated with the hadcm3 model under updated emission scenarios, *Climate Dyn.*, *20*, 583–612.

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, *Quart. J. Roy. Meteor. 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 reg ional climate model. II: Comparison of driving and regional model responses to a doubling of carbon dioxide, *Quart. J. Roy. Meteor. 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, *Tech. rep.*, available from Met. Office Hadley Centre, Exeter, UK, 35pp.

Jones, R. G., J. M. Murphy, D. C. Hassell, and M. J. Woodage (2005), A high resolution atmospheric gcm for the generation of regional climate scenarios, to be submitted April 2005.

Kalnay, E., et al. (1996), The ncep/ncar reanalysis project, *Bull. Amer. Meteor. Soc.*, *77*, 437–471.

Kidson, J. W., and C. S. Thompson (1998), A comparison of statistical and model-based downscaling techniques for estimating local climate variat ions, *J. Climate*, *11*, 735–753.

Kistler, R., et al. (2001), The ncep-ncar 50-year reanalysis: Monthly means cd-rom and documentation, *Bull. Amer. Meteor. Soc.*, *82*, 247–267.

Källén, E. (1996), Hirlam documentation manual, system 2.5, *Tech. rep.*, Swed. Meteorol. and Hydrol. Inst. (SMHI), S-60176 Norrköping, Sweden.

Lüthi, D., A. Cress, H. C. Davies, C. Frei, and C. Schär (1996), Interannual variability and regional climate simulations, *Theor. Appl. Climatol.*, *53*, 185–209.

Majewski, D. (1991), The Europa Modell of the Deutscher Wetterdienst, in *ECMWF Seminar on Numerical Methods in Atmospheric Models*, *Proc.*, vol. II, pp. 147–191, Reading UK.

Mearns, L. O., C. Rosenzweig, and R. Goldberg (1997), Mean and variance change in climate scenarios: Methods, agricultural applications, and measures of uncertainty, *Climatic Change*, *35*, 367–396.

Mearns, L. O., I. Bogardi, F. Giorgi, I. Matyasovszky, and M. Palecki (1999), Comparison of climate change scenarios generated from regional climate model experiments and statistical downscaling, *J. Geophys. Res-Atmos.*, *104*, 6603–6621.

Mearns, L. O., F. Giorgi, P. Whetton, D. Pabon, M. Hulme, and M. Lal (2004), Guidelines for use of climate scenarios developed from regional climate model experiments, *Tech. rep.*, Data Distribution Centre of the IPCC.

Murphy, J. (1999), An evaluation of statistical and dynamical techniques for downscaling local climate, *J. Climate*, *12*, 2256–2284.

Murphy, J. (2000), Predictions of climate change over Europe using statistical and dynamical downscaling techniques, *Int. J. Climatol.*, *20*, 489–501.

Nakicenovic, N., and et al. (2000), *Special Report on Emission Scenarios*, 599 pp., Cambridge University Press.

Noguer, M., R. Jones, and J. Murphy (1998), Sources of systematic errors in the climatology of a regional climate model over Europe, *Climate Dyn.*, *14*, 691–712.

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.

Pope, V. D., M. L. Gallani, P. R. Rowntree, and R. A. Stratton (2000), The impact of new physical parametrizations in the hadley centre climate model: Hadam3, *Climate Dyn.*, *16*, 123–146.

Roeckner, E., et al. (1996), The atmospheric general circulation model echam-4: Model description and simulation of present-day climate, *Tech. Rep. 218*, Max-Planck Inst. für Meteorol. (MPI), Hamburg.

Schmidli, J., and C. Frei (2005), Trends of heavy precipitation and wet and dry spells in switzerland during the 20th century, *Int. J. Climatol.*, *25*, 753–771.

Schmidli, J., C. Frei, and P. L. Vidale (2005), Downscaling from gcm precipitation: A benchmark for dynamical and statistical downscaling methods, *Int. J. Climatol.*, accepted.

Schär, C., T. D. Davies, C. Frei, H. Wanner, M. Widmann, M. Wild, and H. C. Davies (1998), Current alpine climate, in *A view from the Alps: Regional perspectives on climate change*, edited by P. Cebon and et al., pp. 21–72, MIT Press.

Schär, C., P. L. Vidale, D. Lüthi, C. Frei, C. Haberli, M. A. Liniger, and C. Appenzeller (2004), The role of increasing temperature variability in european summer heatwaves, *Nature*, *427*, 332–336.

Seneviratne, S. I., J. S. Pal, E. A. B. Eltahir, and C. Schär (2002), Summer dryness in a warmer climate: a process study with a regional climate model, *Climate Dyn.*, *20*,

69–85.

Stehlík, J., and A. Bárdossy (2002), Multivariate stochastic downscaling model for generating daily precipitation series based on atmospheric circulation, *J. Hydrology*, *256*, 120–141.

van den Hurk, B., et al. (2005), Soil control on runoff response toclimate change in regional climate model simulations, *J. Climate*, *18*, 3536–3551.

Vidale, P. L., D. Lüthi, C. Frei, S. I. Seneviratne, and C. Schär (2003), Predictability and uncertainty in a regional climate model, *J. Geophys. Res-Atmos.*, *108 (D18)*, art. no. 4586.

Vidale, P. L., D. Lüthi, R. Wegmann, and C. Schär (2005), Variability of European climate in a heterogeneous multi-model ensemble, *Climatic Change*, submitted.

von Storch, H. (1999), On the use of "inflation" in statistical downscaling, *J. Climate*, *12*, 3505–3506.

von Storch, H., E. Zorita, and U. Cubasch (1993), Downscaling of global climate change estimates to regional scales: An application to Iberian rainfall in wintertime, *J. Climate*, *6*, 1161–1171.

Wetherald, R. T., and S. Manabe (1995), The mechanisms of summer dryness induced by greenhouse warming, *J. Climate*, *8*, 3096–3108.

Widmann, M. L., C. S. Bretherton, and E. P. Salathé Jr. (2003), Statistical precipitation downscaling over the Northwestern United States using numerically simulated precipitation as a predictor, *J. Climate*, *16*, 799–816.

Wilby, R. L., and T. M. L. Wigley (2000), Precipitation predictors for downscaling: Observed and General Circulation Model relationships, *Int. J. Climatol.*, *20*, 641–661.

Wilby, R. L., L. E. Hay, W. J. Gutowski, R. W. Arritt, E. S. Tackle, G. H. Leavesley, and M. Clark (2000), Hydrological responses to dynamically and statistically downscaled general circulation model output, *Geophys. Res. Letters*, *27*, 1199–1202.

Wilby, R. L., S. P. Charles, E. Zorita, B. Timbal, P. Whetton, and L. O. Mearns (2004), Guidelines for use of climate scenarios developed from statistical downscaling methods, *Tech. rep.*, Data Distribution Centre of the IPCC.

Wilks, D. S., and R. L. Wilby (1999), The weather generation game: a review of stochastic weather models, *Prog. Phys. Geog.*, *23*, 329–357.

Wood, A. W., L. R. Leung, V. Sridhar, and D. P. Lettenmaier (2004), Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs, *Climatic Change*, *62*, 189–216.

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
Q90	empirical 90% quantile of precipitation on wet days	mm per day
XCDD	maximum number of consecutive dry days	day
XND	maximum n-day precipitation total (n=1,5)	mm

Table 2. Overview of the statistical downscaling methods (P=predictand, d=daily, s=seasonal; CP=circulation pattern, CI=circulation index; PC=principal component; S=space, si=single-site, mu=multi-site; ETH=ETH Zürich, UEA=University of East Anglia, USTUT=University of Stuttgart, DMI=Danish Meteorological Institute, FIC=Fundación para la Investigatción del Clima). Further acronyms see lead of section 3.

Acron	Institution	P	Predictor(s)	S	Description
LOCI	ETH	d	PRE	si	Local scaling of GCM precipitation with correction of frequency and intensity bias.
CCA	UEA	s	PCs of SLP, RH700, SH700, T700	mu	Canonical correlation analysis. 4-7 PCs per predictor, 4-14 PCs per SI.
MLR	USTUT	s	ZX, RHX, TX, DIVX and VORX with X=500, 700, 850; MF700, CPs	si	Multiple linear regression. Predictor values are averaged over 4 nearest grid points.
MAR	USTUT	d	CPs, MF700	mu	Multivariate auto-regressive model.
CWG	DMI	d	CI based on SLP	si	Conditional weather generator, conditional on quantiles of a CI. (transition probabilities, scale and location parameter)
ANA	FIC	d	VG1000, VG500	mu	Two-step analogue method. (1) determine the 30 most similar days; (2) determine pdf of daily precipitation from all days within a season

Table 3. The regional climate models from which results are analyzed in this study. The last column (#) indicates the number of ensemble members available for the climate change scenario.

Acronym	Institution and model origin	#
CHRM	Swiss Federal Institute of Technology (ETH), Zürich Climate version of 'Europamodell' of German and Swiss weather services *Lüthi et al.* [1996]; *Vidale et al.* [2003]	1
HADRM3	Hadley Centre, UK Meteorological Office, Exeter Regional model of climate model suite at the Hadley Centre *Jones et al.* [1995, 1997]; *Pope et al.* [2000]	3
HIRHAM	Danish Meteorological Institute, Copenhagen Dynamical core from HIRLAM, Parametrizations from ECHAM4 *Christensen et al.* [1996]	3

Table 4. Seasonal variation of the selected predictors for the CCA method.

Predictor	fre		int/q90		xcdd	
	DJF	JJA	DJF	JJA	DJF	JJA
SLP	x	x		x	x	x
RH700/SH700	x	x	x	x	x	x
T700			x		x	

Table 5. Seasonal variation of the common leading predictors for the MLR method.

Season	Leading predictors
Winter	Z850, DIV850, MF700, RH700
Spring	Frequency of wet CPs, DIV850
Summer	Frequency of wet CPs, RH700, VOR500
Autumn	Z850, DIV500

Table 6. Summary evaluation based on correlation skill r for region-mean indices. Each pair of symbols represents the skill for FRE and INT, respectively. A plus denotes $r > 0.6$. The second last column lists the number of pluses, that is $r > 0.6$, for each method, and last column lists the correlation averaged over the regions and seasons. The last row lists the correlation averaged over the downscaling methods (negative correlations were set to zero prior to averaging). Note that typically $r_{\mathrm{FRE}} \approx \bar{r} + (0.1 - 0.2)$, $r_{\mathrm{INT}} \approx \bar{r} - (0.1 - 0.2)$. However, for the region TIC in summer the skill for INT is higher than the skill for FRE (not shown).

	WEST			NALP			TIC				
method	wi	su	au	wi	su	au	wi	su	au	#+	$\bar{r}$
loci-e40	+/+	+/+	+/+	+/+	+/+	+/+	+/.	+/.	+/+	16	0.81
chrm	+/+	+/.	+/.	+/+	./.	+/+	+/+	./+	+/+	13	0.70
hadrm3h	+/+	./+	+/.	+/+	./.	+/+	+/+	./.	+/.	11	0.66
hirham	+/.	./.	+/.	+/+	./.	+/.	+/.	./.	+/+	8	0.59
cca	+/.	./.	+/.	+/.	./.	+/.	+/.	./+	+/.	7	0.39
mlr	+/.	+/.	+/.	+/.	+/.	+/.	+/.	+/.	+/.	9	0.54
mar	+/+	+/.	+/.	+/.	+/.	+/.	+/.	./+	+/+	11	0.56
cwg	+/.	./.		+/.	./.		+/.	./.		3	0.25
ana	+/+	+/.	+/.	+/.	+/.	+/.	+/.	./+	+/+	11	0.59
$\bar{r}$	0.75	0.42	0.61	0.62	0.34	0.62	0.60	0.50	0.75		0.58

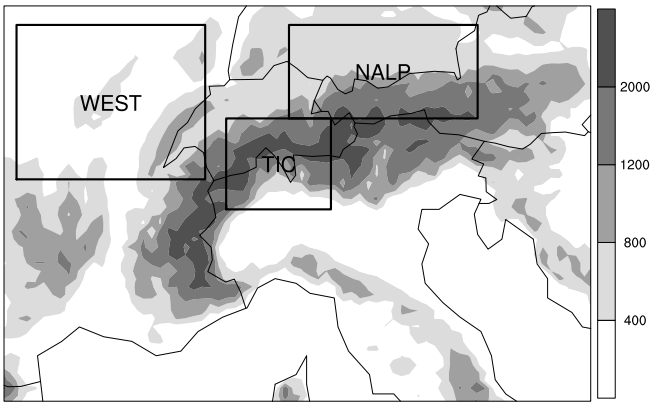

Figure 1. Study region for model evaluation and intercomparison. Shading represents topographic height (m) above MSL. The boxes indicate subregions used for specific analyses.

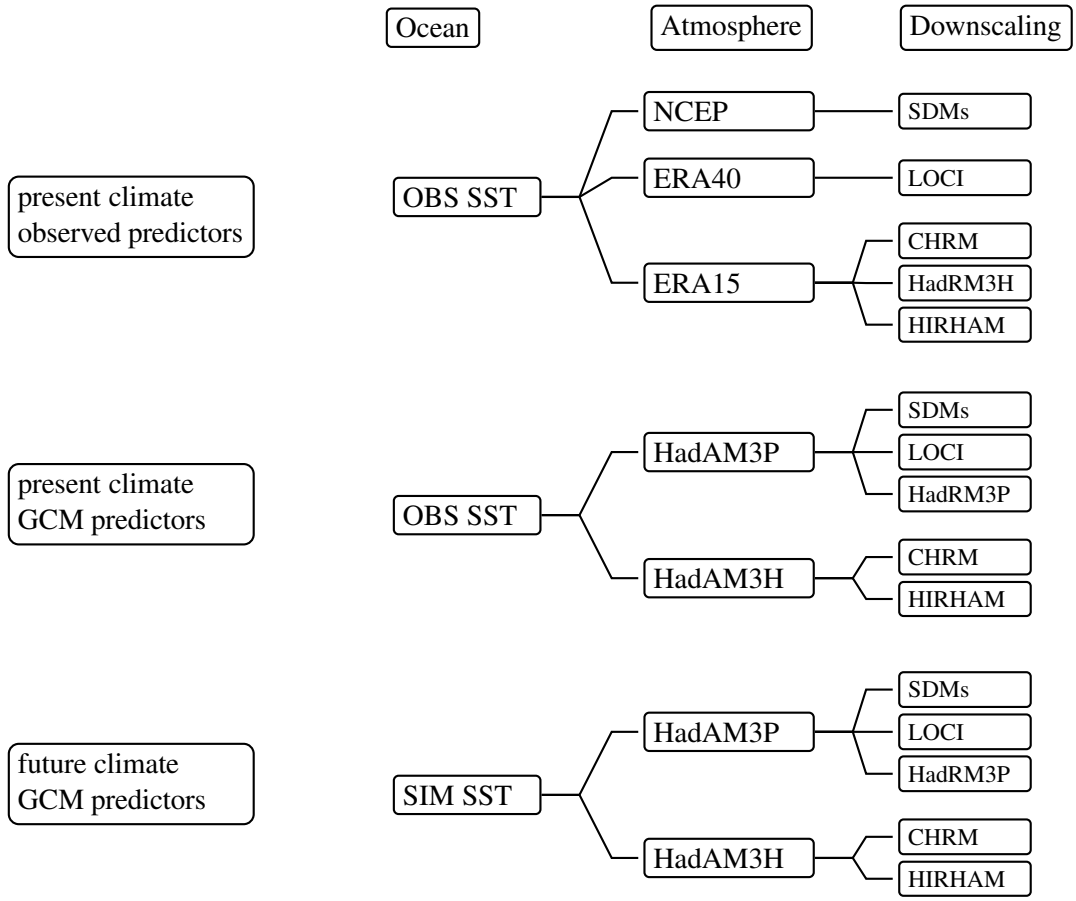

Figure 2. Overview of the model chains: reanalysis chain (top), present climate GCM (middle), future climate GCM (bottom). See section 3 for acronyms.

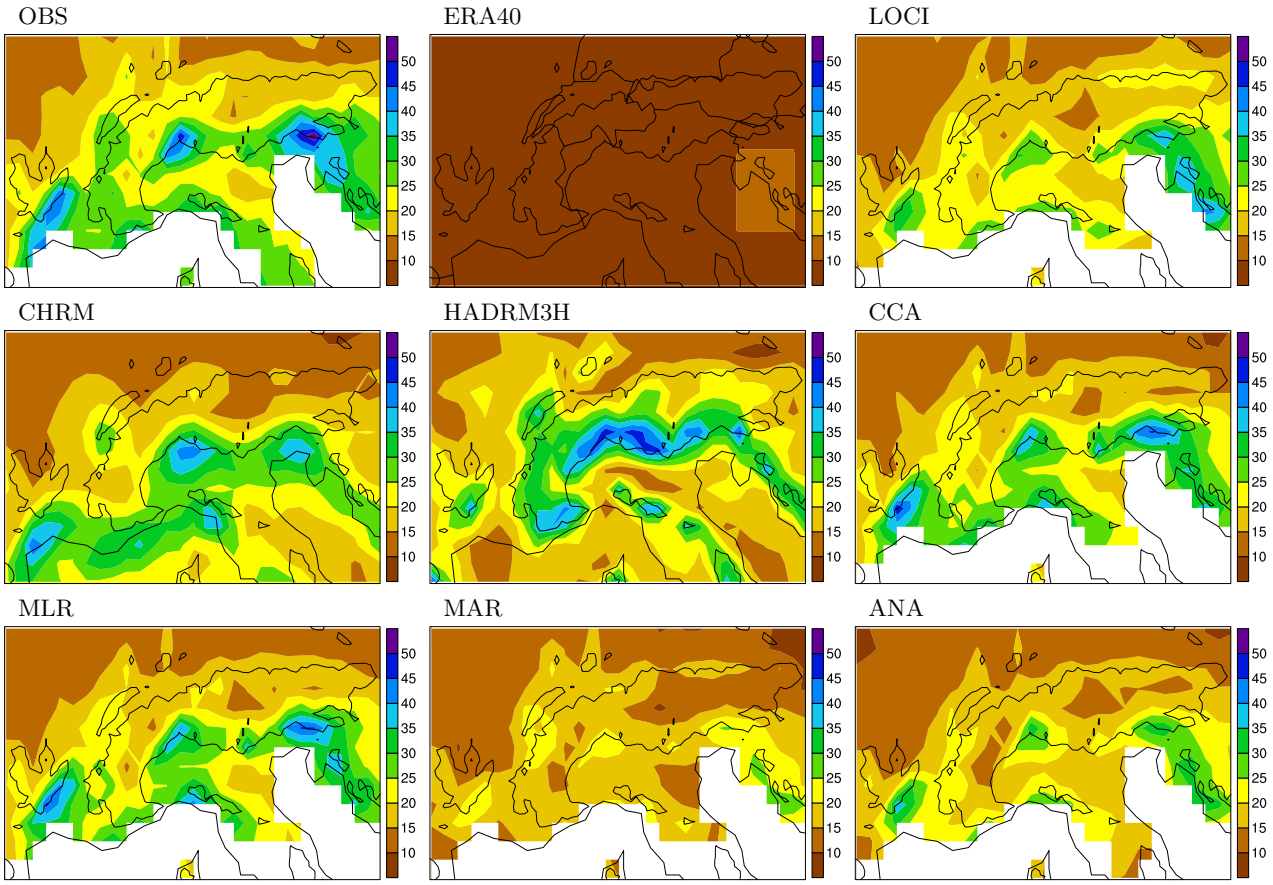

Figure 3. Q90 of daily precipitation (mm/d) in autumn (SON) for OBS (top left) and the models.

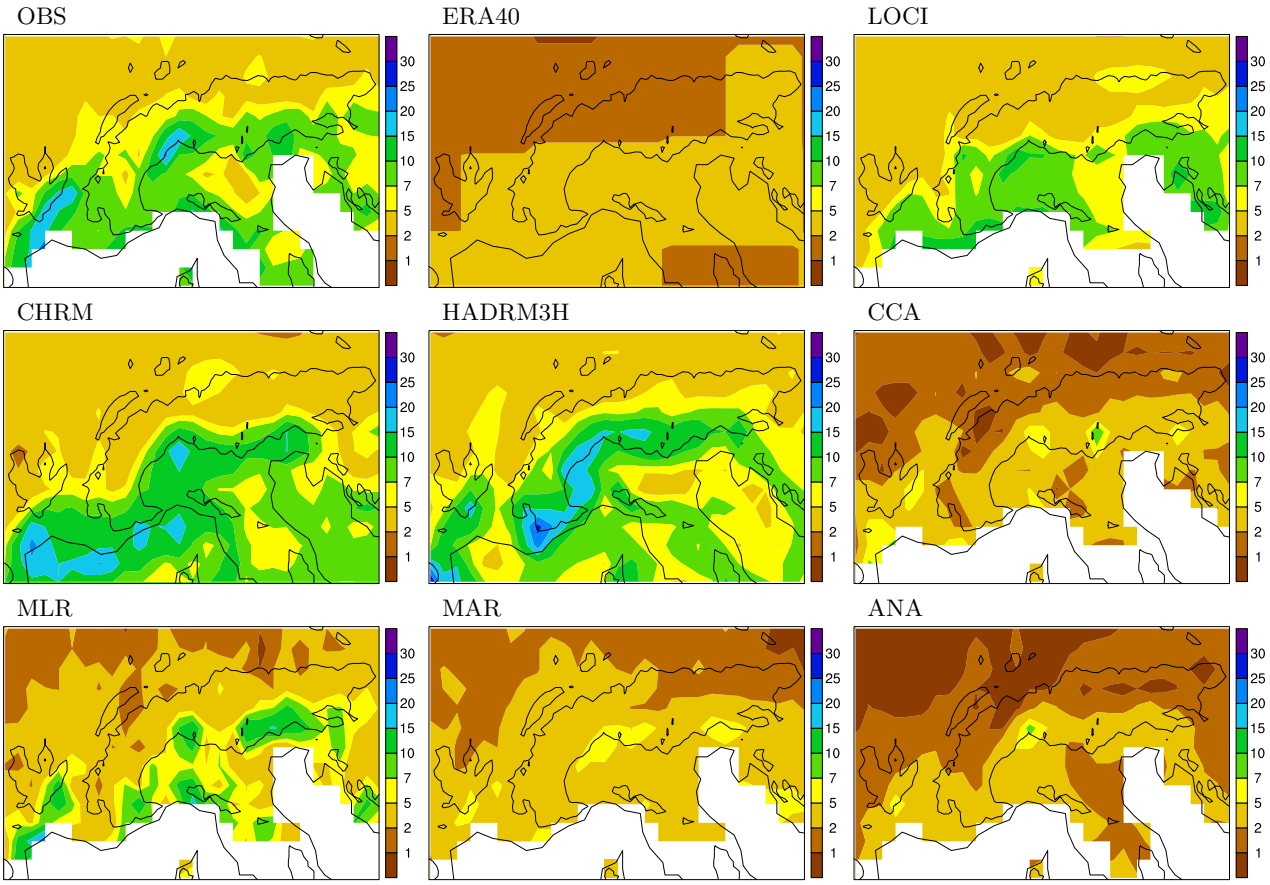

Figure 4. Interannual variability of Q90 (standard deviation) in autumn (SON) for OBS (top left) and the models.

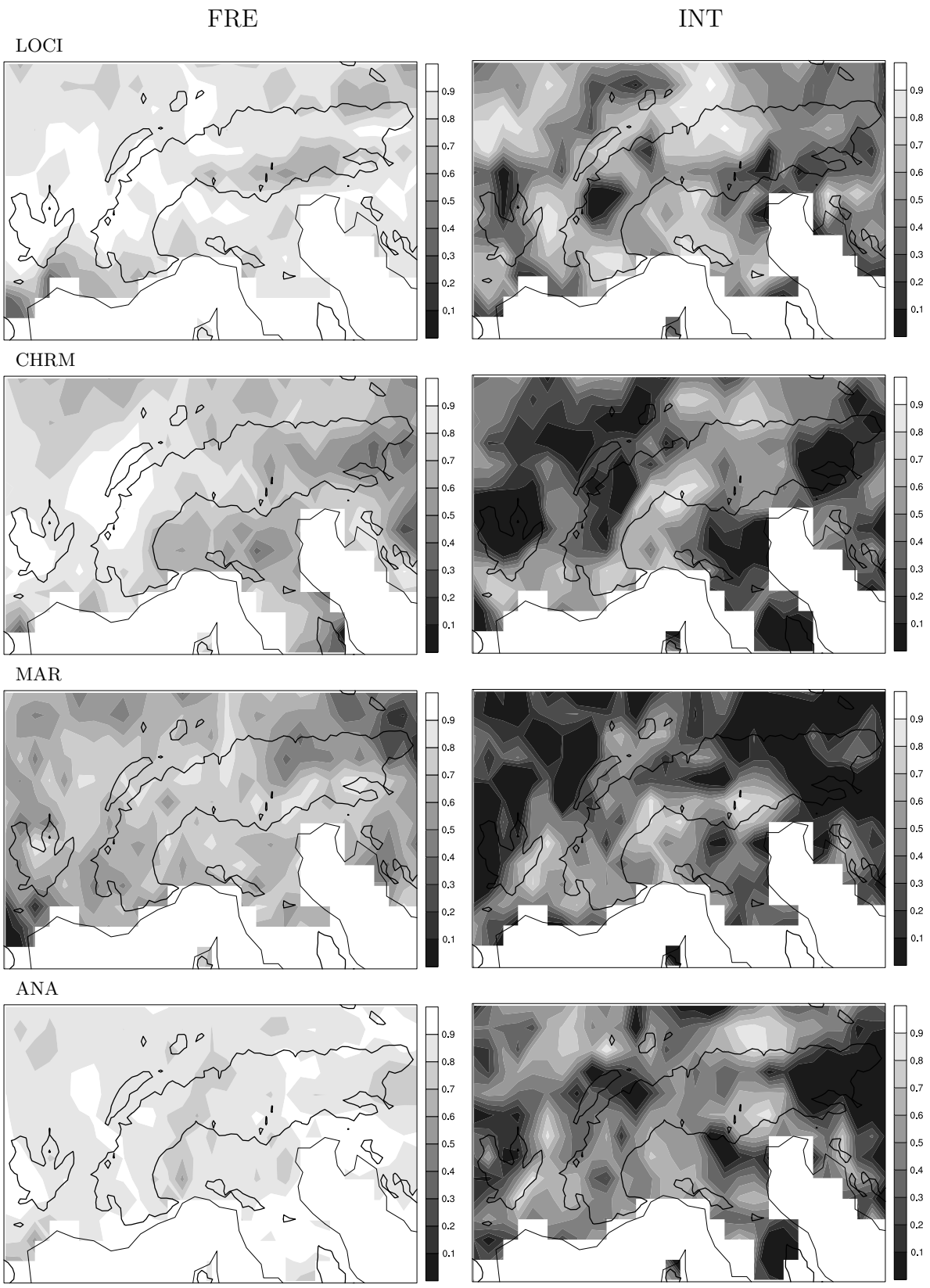

Figure 5. Correlation between downscaled and observed interannual anomalies for autumn (SON) FRE and INT for the ERA15 validation period (1979–1993).

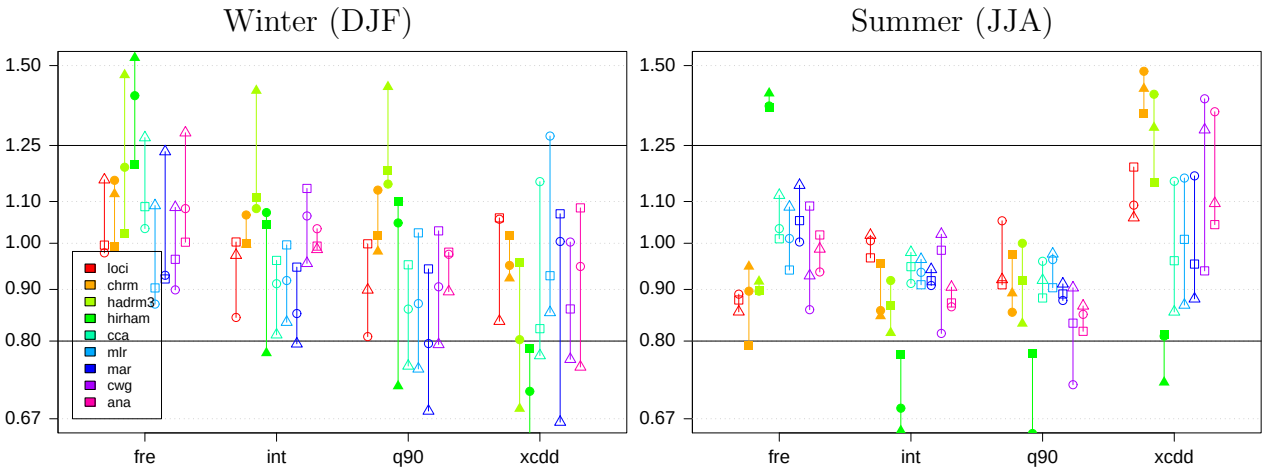

Figure 6. Relative bias (MOD/OBS) in precipitation indices for winter and summer for the three regions (WEST: squares, NALP: circles, TIC: triangles; filled symbols = RCMs) for the validation period 1979–1993.

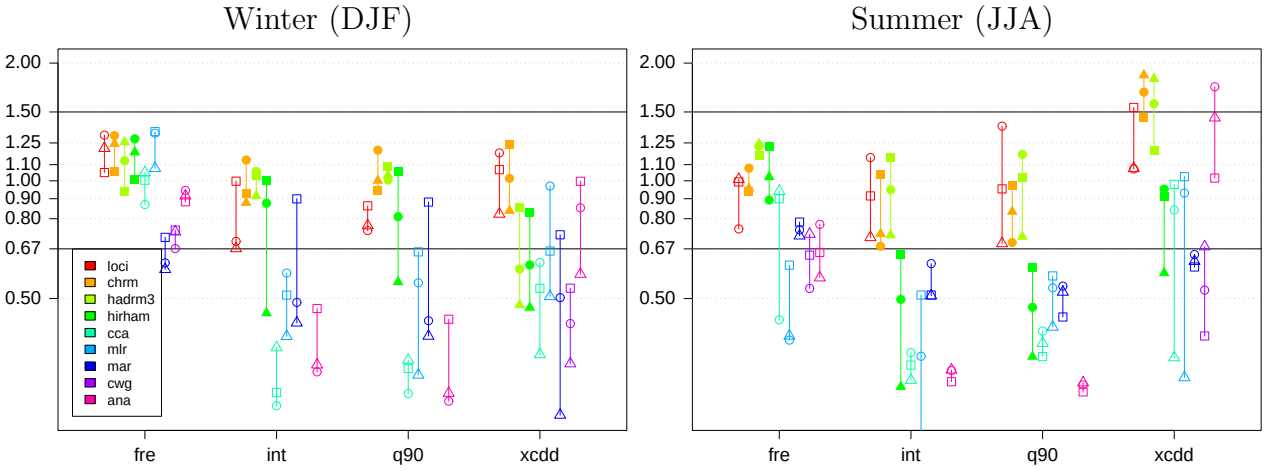

Figure 7. As in Fig. 6, but for the ratio of standard deviations between models and OBS.

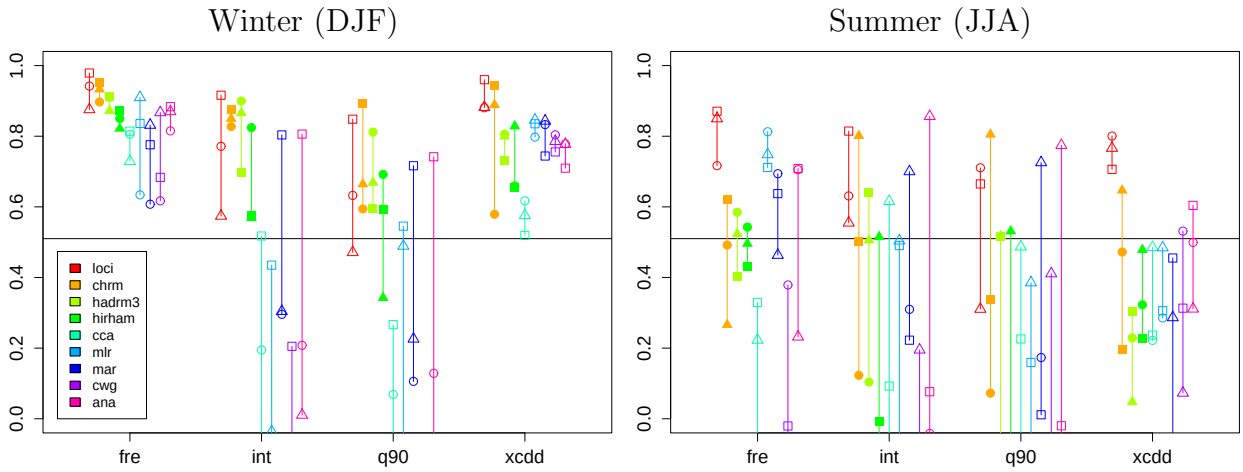

Figure 8. As in Fig. 6, but for the correlation skill of the *region-mean* precipitation index. The solid black line denotes the 5 % significance level of the null hypothesis of no correlation.

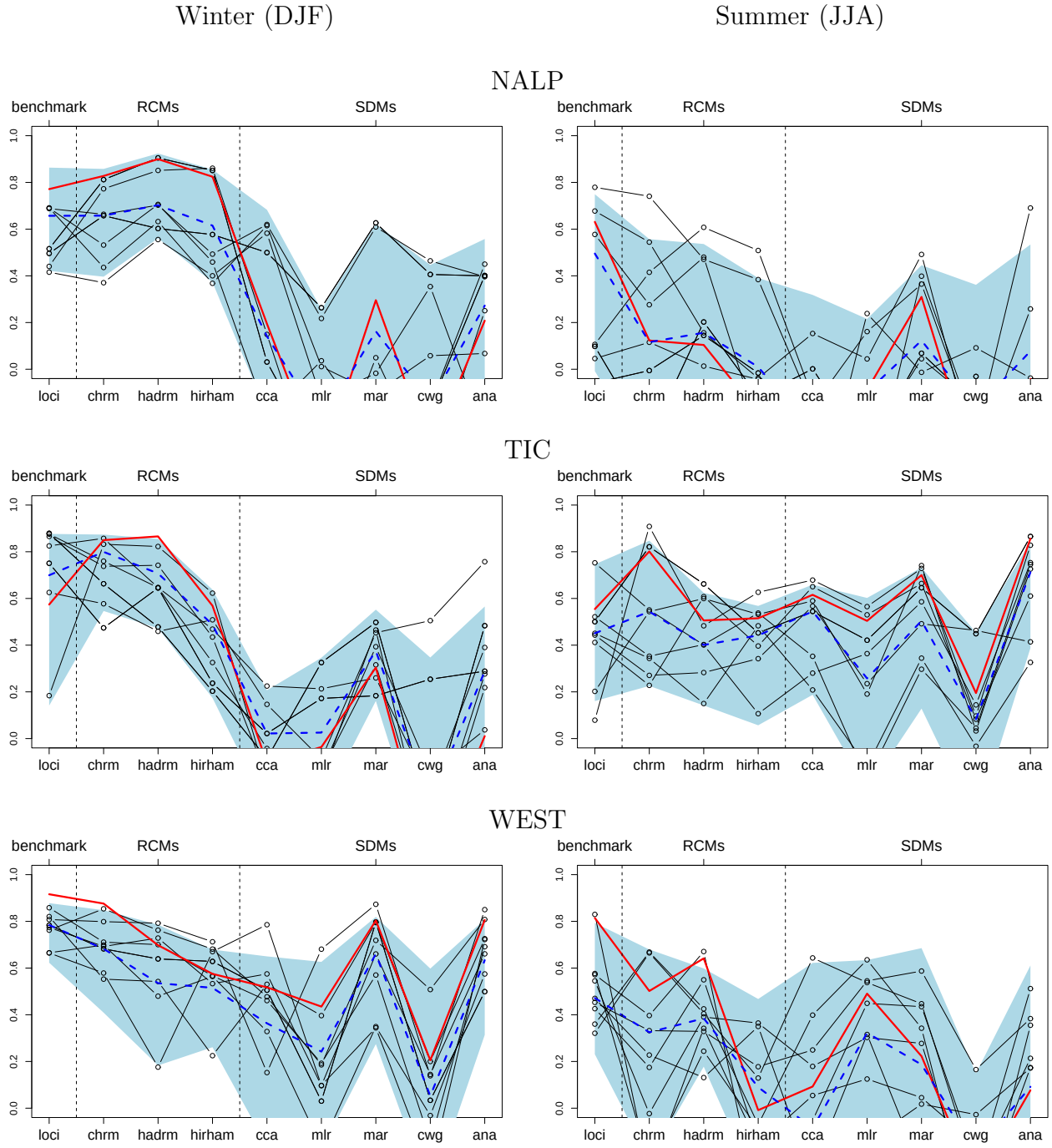

Figure 9. Grid-point correlations for INT for winter and summer for the three evaluation regions. The shaded area indicates the range of correlations (90% interval) obtained for individual grid points (results for 9 randomly selected grid points are denoted by thin lines). The dashed line denotes the median of the grid-point values, and the bold line denotes the correlation for the region-mean INT (cf. Fig. 8).

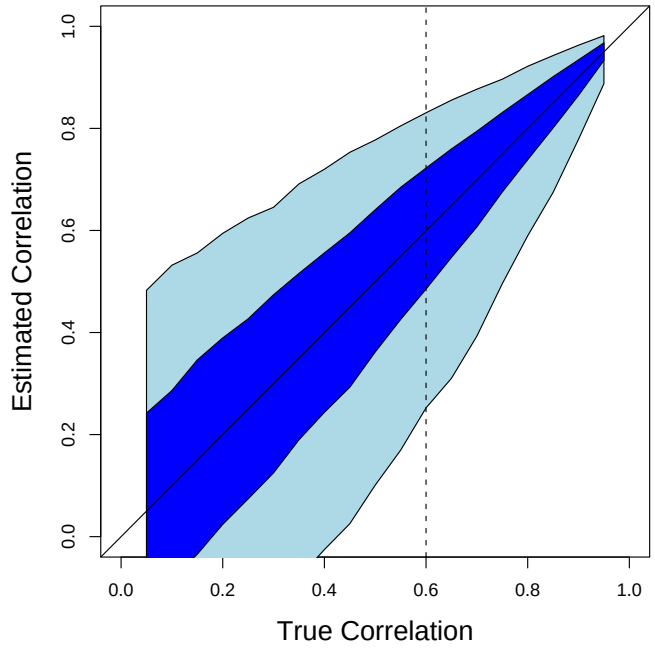

Figure 10. Illustration of the random sampling error of the correlation coefficient estimated from time series of length 15. Pairs of time series are generated from normally distributed random variables with a given cross-correlation (x-axis). The shading indicates the 50% and 90%-range of the obtained estimates, determined from Monte Carlo simulation with 5000 repetitions. The dashed line indicates that for a true correlation of 0.6, for example, the interval required to cover 90% of the estimates extends from 0.24 to 0.83.

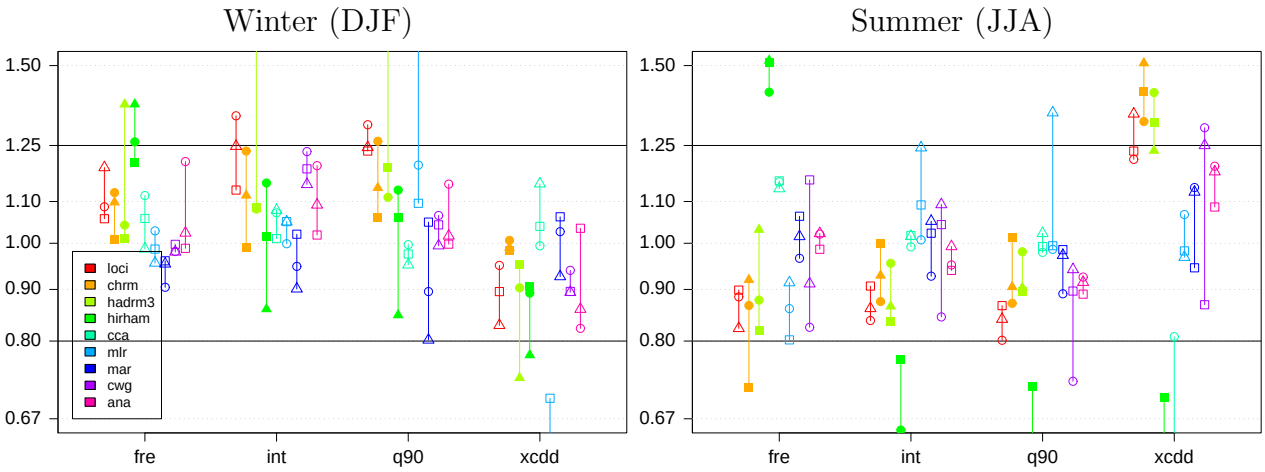

Figure 11. As in Fig. 6, but for the HadAM control run.

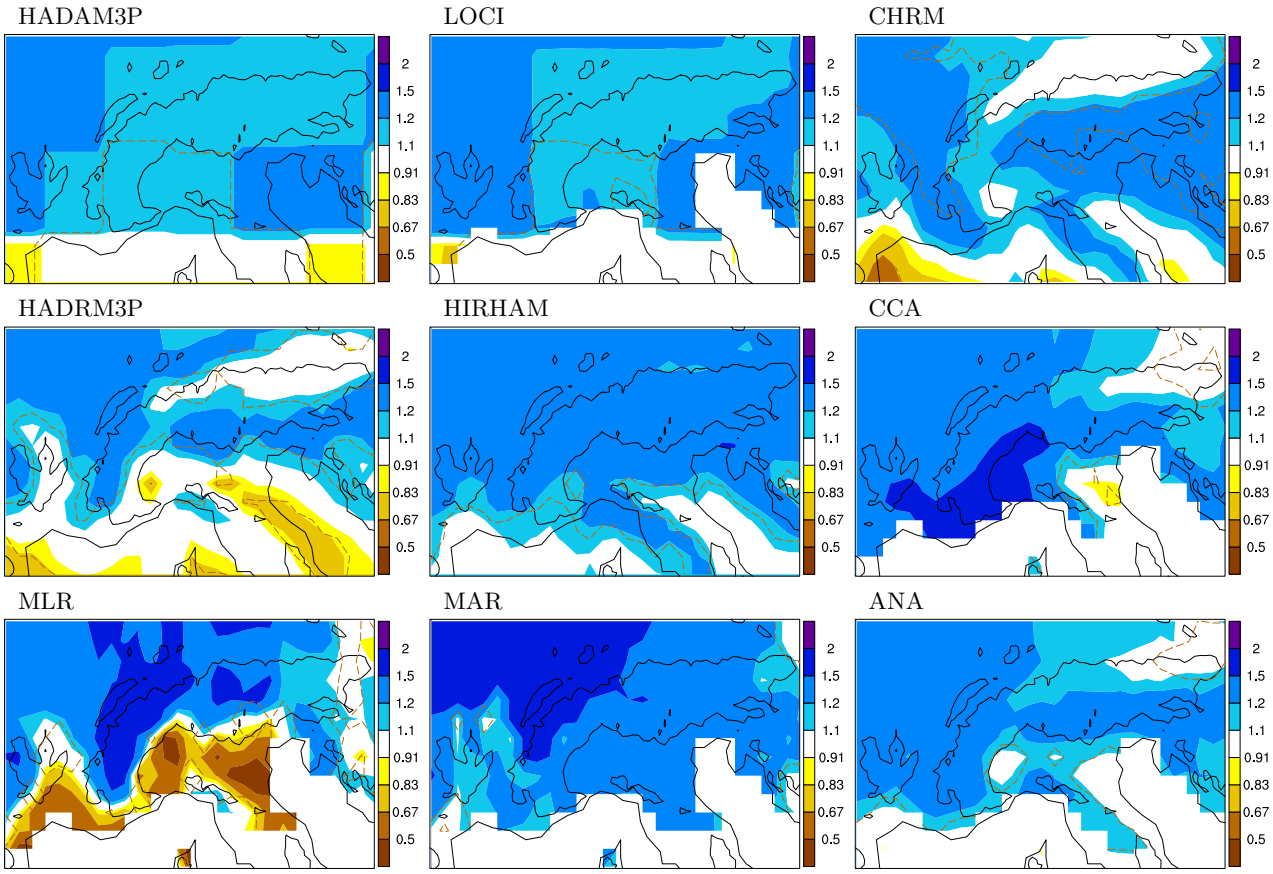

Figure 12. Ratio (SCEN: 2071–2100 / CTRL: 1961–1990) of MEA in winter (DJF). Results from the GCM, 3 RCMs, and 5 SDMs under the A2 emission scenario. Note the log-scale in the color coding. The dashed line (red) indicates areas with statistically significant (5%) change, in an independent (Mann-Whitney) test at each model grid point. Note that the increases are statistically significant for changes smaller than 10% for the SDMs, due to their smaller interannual variability, for changes of about 15% for HADRM3P and HIRHAM (3 ensemble members), and for changes of 20–30% for CHRM (1 ensemble member).

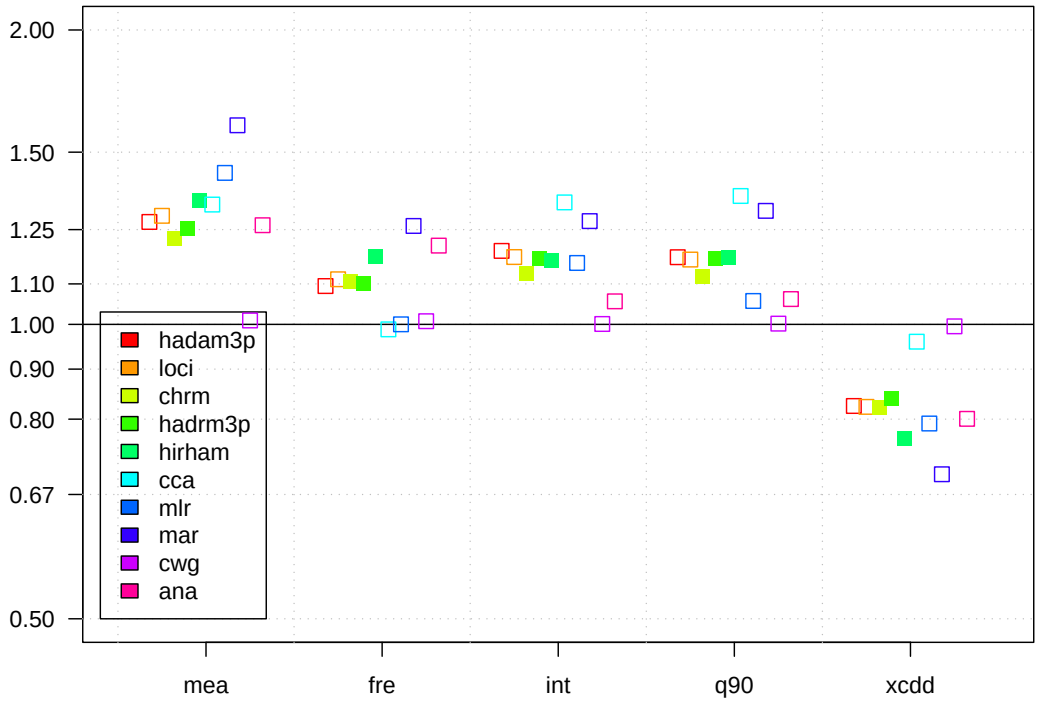

Figure 13. Simulated change (ratio SCEN/CTRL) in region-mean precipitation diagnostics for the region WEST for winter (DJF).

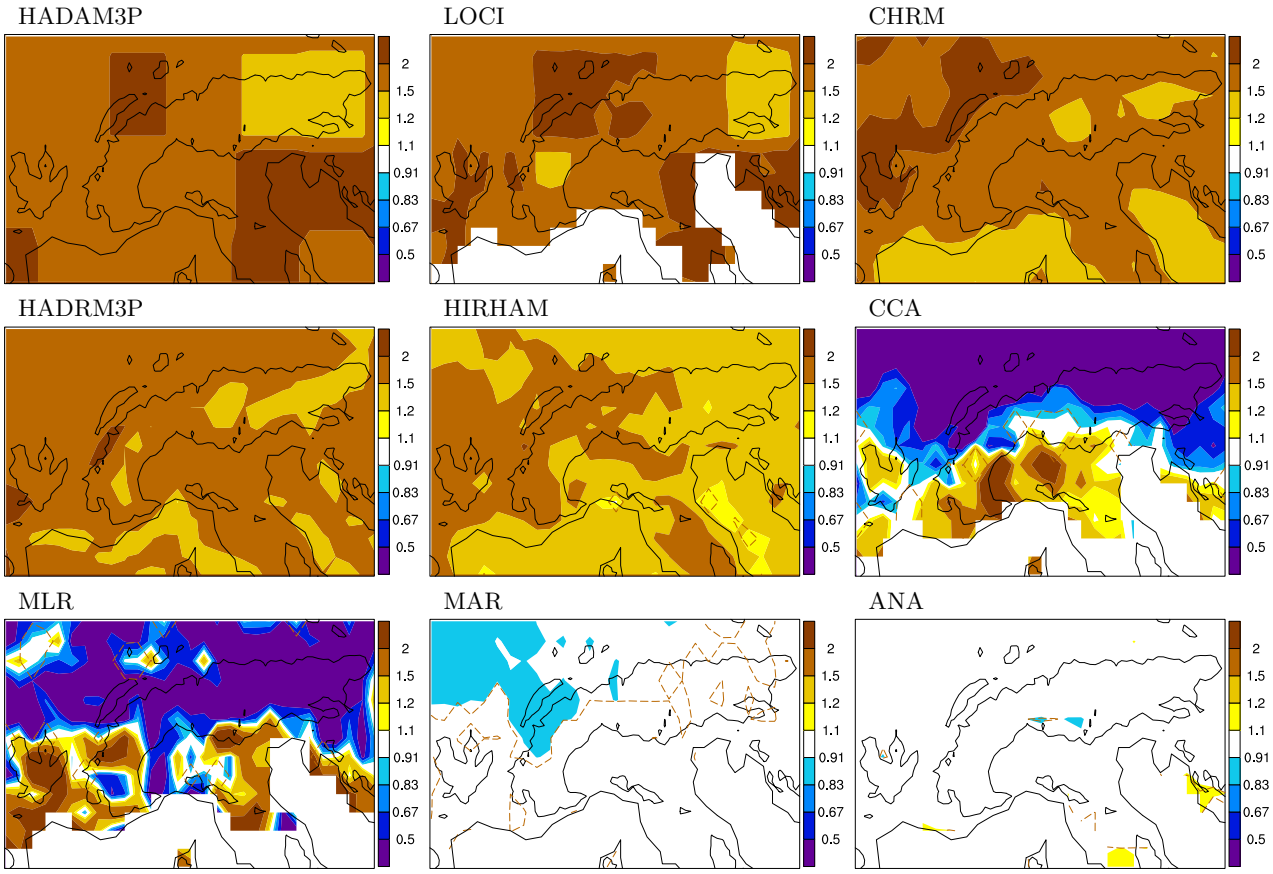

Figure 14. As in Fig. 12, but for XCDD in summer (JJA).

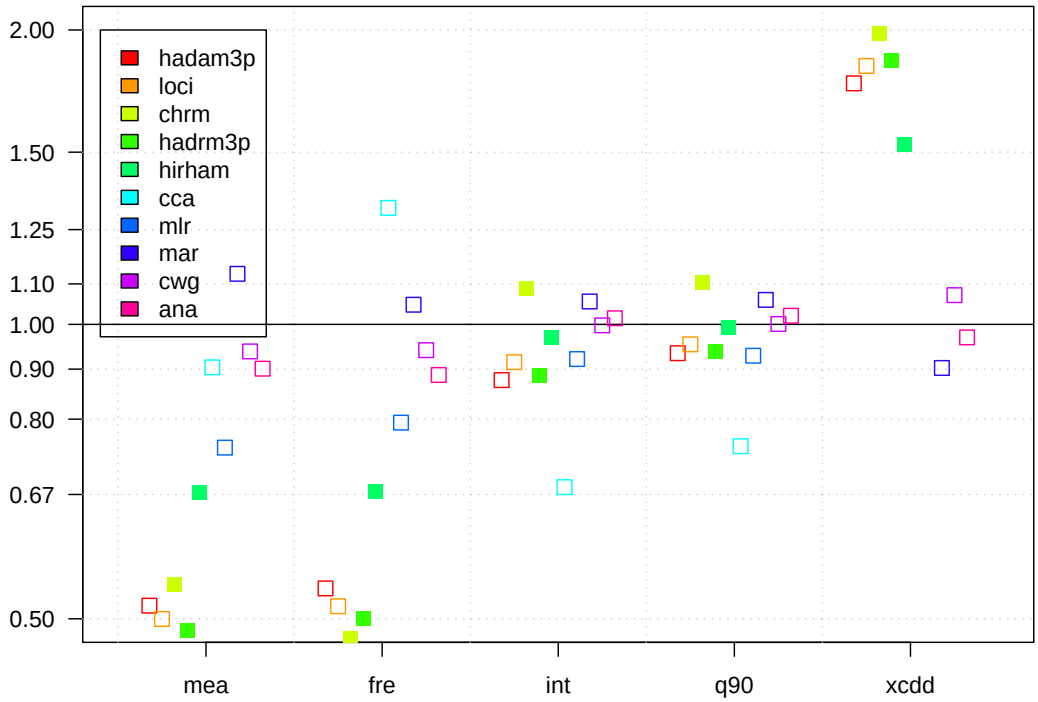

Figure 15. As in Fig. 13, but for summer. Note that the symbols for XCDD for CCA and MLR are not visible, due to values below 0.5.

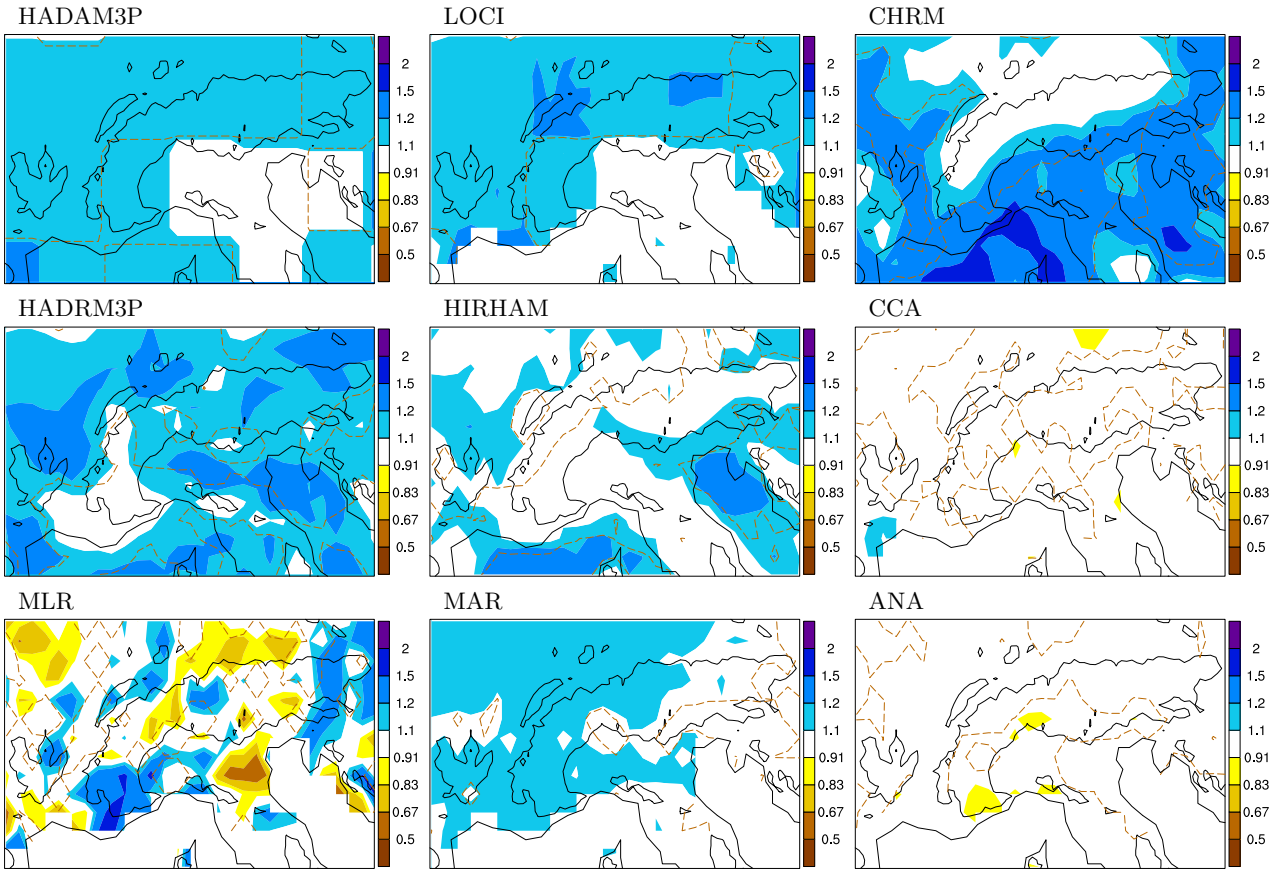

Figure 16. As in Fig. 12, but for Q90 in autumn (SON).

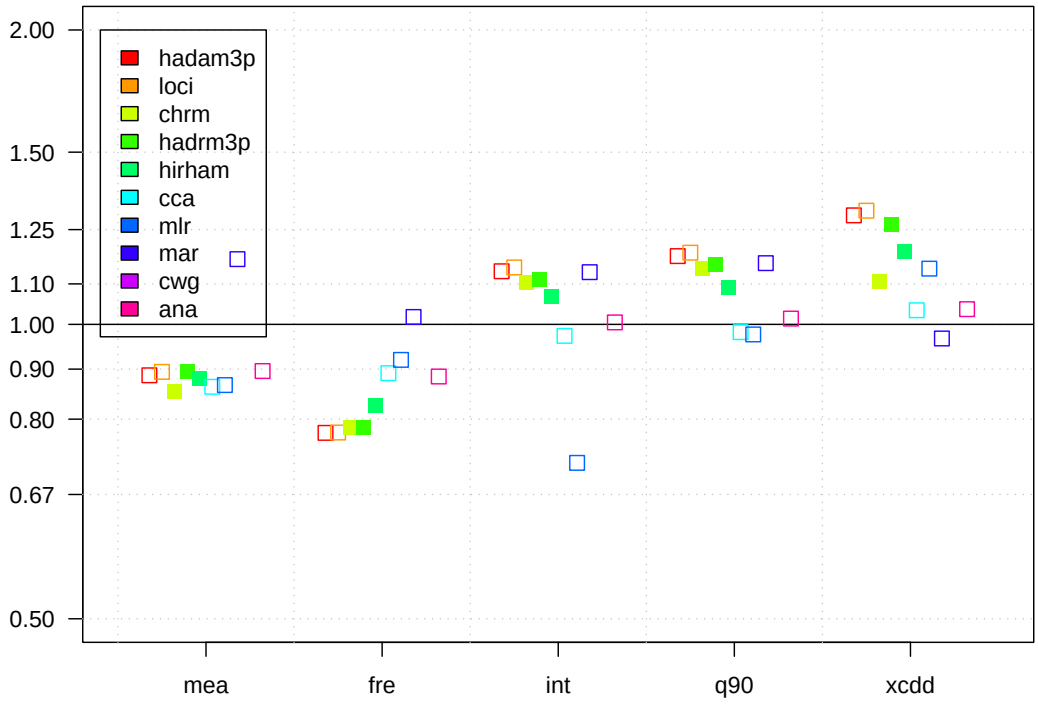

Figure 17. As in Fig. 13, but for autumn.