INTERNATIONAL JOURNAL OF CLIMATOLOGY
Int. J. Climatol. (2010)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/joc.2168

RMetS
Royal Meteorological Society

Empirical-statistical downscaling and error correction of daily precipitation from regional climate models

Matthias Jakob Themeßl,* Andreas Gobiet and Armin Leuprecht
Wegener Center for Climate and Global Change and Institute for Geophysics, Astrophysics, and Meteorology, University of Graz, Leechgasse 25, Graz 8010, Austria

ABSTRACT: Although regional climate models (RCMs) are powerful tools for describing regional and even smaller scale climate conditions, they still feature severe systematic errors. In order to provide optimized climate scenarios for climate change impact research, this study merges linear and nonlinear empirical-statistical downscaling techniques with bias correction methods and investigates their ability for reducing RCM error characteristics. An ensemble of seven empirical-statistical downscaling and error correction methods (DECMs) is applied to post-process daily precipitation sums of a high-resolution regional climate hindcast simulation over the Alpine region, their error characteristics are analysed and compared to the raw RCM results.

Drastic reductions in error characteristics due to application of DECMs are demonstrated. Direct point-wise methods like quantile mapping and local intensity scaling as well as indirect spatial methods as nonlinear analogue methods yield systematic improvements in median, variance, frequency, intensity and extremes of daily precipitation. Multiple linear regression methods, even if optimized by predictor selection, transformation and randomization, exhibit significant shortcomings for modelling daily precipitation due to their linear framework. Comparing the well-performing methods to each other, quantile mapping shows the best performance, particularly at high quantiles, which is advantageous for applications related to extreme precipitation events. The improvements are obtained regardless of season and region, which indicates the potential transferability of these methods to other regions. Copyright © 2010 Royal Meteorological Society

KEY WORDS empirical statistical downscaling; bias correction; error correction; Alpine region; precipitation modelling; regional climate modelling

Received 16 August 2009; Revised 7 April 2010; Accepted 15 April 2010

1. Introduction

General circulation models (GCMs) are established tools for estimating the large-scale evolution of the Earth's climate, but due to their relative coarse horizontal resolution, they are not suited to properly represent regional-scale climate characteristics. Therefore, dynamical downscaling techniques are often applied to derive regional-scale information from GCMs. Limited area regional climate models (RCMs) are forced by lateral boundary conditions of GCMs or reanalysis products and simulate the regional climate over a certain area on a finer grid (typical horizontal resolution 10–50 km; Giorgi and Mearns, 1991, 1999; Wang *et al*., 2004). RCMs have considerably advanced in reproducing regional climate, but are nevertheless known to feature systematic errors (e.g. Frei *et al*., 2003; Hagemann *et al*., 2004; Suklitsch *et al*., 2008, 2010). Particularly, small-scale patterns of daily precipitation are highly dependent on model resolution and parameterization and can often not be used directly in climate change impact assessment studies (Fowler *et al*., 2007). Statistical post-processing of RCMs, according to the concept of model output statistics (MOS; Wilks, 1995), may help to overcome these problems, leading to qualitatively enhanced climate information. Such statistical post-processing of RCMs is mostly neglected in climatological studies as traditional empirical-statistical downscaling methods (ESDMs) are preferably applied according to the concept of perfect prognosis (perfect prog; Wilks, 1995). Perfect prog downscaling determines a statistical model (transfer function) between suitable large-scale observation/reanalysis data and local observations (e.g. Wilby and Wigley, 1997; Murphy, 1999; Schmidli *et al*., 2006), which is applied directly to GCMs for generating regionalized climate scenarios (e.g. Schmidli *et al*., 2007) but without the intention of model error correction (compare Figure 1). RCMs are often favoured to traditional empirical-statistical downscaling because they are capable to simulate regional-scale climate feedback effects and were already shown to create added value compared to GCMs on the meso- and regional scale for surface variables (e.g. Wang *et al*., 2004; Feser, 2006).

Recently, the availability of regional RCM-based climate scenarios for Europe tremendously increased

* Correspondence to: Matthias Jakob Themeßl, Wegener Center for Climate and Global Change and Institute for Geophysics, Astrophysics, and Meteorology, University of Graz, Leechgasse 25, Graz 8010, Austria. E-mail: matthias.themessl@uni-graz.at

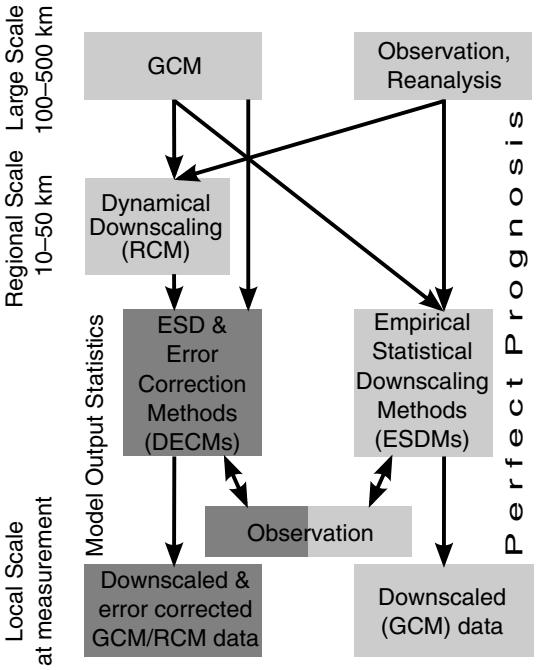

Figure 1. Scheme of different downscaling approaches. Darker grey colour indicates the applied approach for this study. Traditional empirical-statistical downscaling (right pathway) calibrates the statistical transfer function between large-scale observation/reanalysis data and local-scale observations. These empirical-statistical relationships can be used for downscaling of any GCM. DECMs (left pathway) are calibrated on RCM (or GCM) data and local observations, account for downscaling as well as model errors, but can only be applied to the model they are calibrated for.

due to projects like ENSEMBLES (http://ensembles-eu.metoffice.com/) or PRUDENCE (http://prudence.dmi.dk/). However, due to the error characteristics of RCMs and when climate information at the point scale is needed, statistical transfer functions are inevitable to provide suitable climate scenario data for climate change impact research.

Aiming at a reduction of RCM error characteristics as well as of resolution at the same time, this study follows the principles of MOS and compares direct ESDMs that solely rely on modelled precipitation (e.g. Wood *et al.*, 2004; Graham *et al.*, 2007; Dobler and Ahrens, 2008) to indirect ESDMs that derive fine-scale information by relating various model outputs (predictors–frequently upper air atmospheric data and not necessarily model precipitation) to observed surface variables (predictands–such as precipitation; e.g. Wilby and Wigley, 1997; Benestad *et al.*, 2009). All methods are applied to RCM results instead of their usual application to GCMs (compare Figure 1). In order to distinguish this application from perfect prog downscaling (which does not regard model errors), these methods are referred to as 'empirical-statistical downscaling and error correction methods' (DECMs) henceforth. Their skill is assessed by analysing their success in modelling daily precipitation on the station scale in the orographical complex Alpine region in Austria.

The study is organized as follows: Section 1 introduces the applied RCM and observational data as well as the study region. Section 2 describes the implemented DECMs, which are evaluated in Section 3. Finally, Section 4 summarizes the key findings of the study.

2. Data and study region

For this study the mesoscale limited area model MM5 (Dudhia *et al.*, 2005) from the Penn State University (PSU) and the National Center of Atmospheric Research (NCAR) is used to provide the predictor data. MM5 is a non-hydrostatic, terrain-following sigma-coordinate model designed to simulate mesoscale atmospheric circulation. The simulations used in this study originate from the Austrian project 'reclip : more–Research for Climate Protection: Model Run Evaluation' (Loibl *et al.*, 2007), in which parts of the ERA-40 reanalysis (Uppala *et al.*, 2005) were dynamically downscaled (hindcast simulation) in a two-step nesting approach (Gobiet *et al.*, 2006). The data cover the domain shown in Figure 2(a) with a horizontal grid spacing of 10 km. Temporally, the data are given in 6-h time steps for the time span from 1981 to 1990 and the single year 1999. The year 1999 is treated as any other year in the 11-year period. For all parameters, except precipitation, daily mean values are calculated from 00 : 00 UTC to 23 : 59 UTC. For comparison to the observational data, daily precipitation is summed up between 06 : 00 UTC and 05 : 59 UTC the following day. All parameters used in the analyses are listed in Table I.

The observed daily precipitation sums, provided by the Austrian Meteorological Service (ZAMG) and the Austrian Hydrological Service (HZB), are used as predictand for empirical-statistical error correction at 919 observation sites, which are evenly distributed across the entire area of Austria (Figure 2(b)). The data are quality checked, but not homogenized. The 919 observational stations fulfill the conditions of at least 80% of data availability and insignificantly changed data distributions after station replacements. The latter condition was tested by a nonparametric Wilcoxon rank-sum test (Wilks, 1995).

Although Austria is a rather small country, it features several climate provinces. These provinces originate from three main airflow directions (Atlantic, Mediterranean and continental Eastern Europe) and their interaction with the Eastern Alps which cover large parts of the country, and vertically range from basins and low-altitude regions of a few hundred metres in the eastern foothills of the Alps to the mountainous western parts of the Eastern Alps up to nearly 4000 m. The central, northern, as well as the southern parts of the Eastern Alps block the Atlantic and Mediterranean moist airflows, force them to raise and rain out, resulting in annual precipitation maxima in the northerly and southerly upslope regions of the Alpine crest. Due to strong advection of wet Adriatic air masses in summer months, southern and south-eastern parts of Austria frequently feature

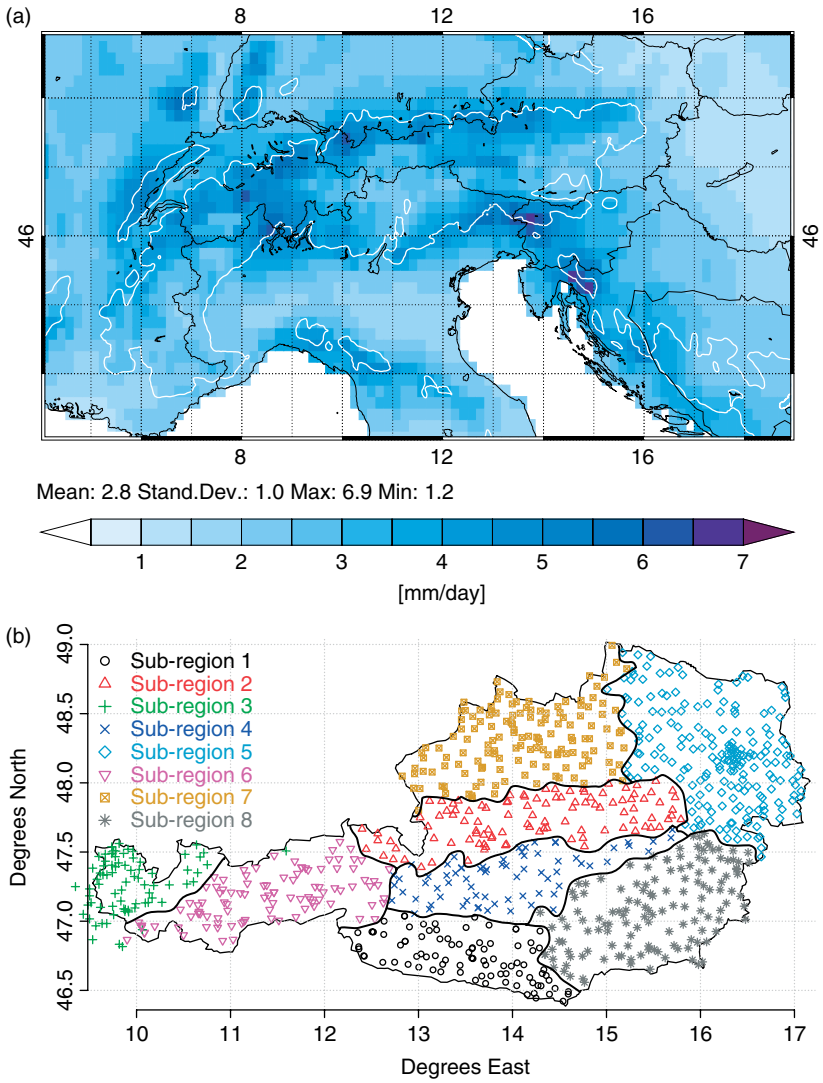

Figure 2. (a) Spatial domain of the RCM data. Additionally, climatological precipitation conditions for the Alpine region including Austria from 1981 to 1990 according to HISTALP observations. The bold, white line indicates the 800-m isoline for better orientation. (b) Location of the 919 observational stations within Austria and the clustered eight homogeneous precipitation regions. This figure is available in colour online at www.interscience.wiley.com/ijoc

severe thunderstorms as well as hail. Decreasing precipitation amounts towards the flatter eastern parts result from the predominating continental air masses there. Additionally, the inner alpine valleys feature reduced precipitation amounts due to the rain shadowing of the mountain ranges (compare Figure 2(a); Cebon *et al*., 1998; Matulla *et al*., 2003). The observed climatology given in Figure 2(a) covers the period 1981–1990 according to the focus period of this study, but the spatial patterns as well as the magnitudes are comparable to the climatic conditions between 1971 and 1990 shown in Frei and Schär (1998). Accounting for the regional climatological differences, the observational stations are clustered into eight sub-regions (Figure 2(b)), based on correlated daily precipitation. The clustering method is described in Suklitsch *et al*. (2008). The clusters represent spatially homogeneous regions, with only two displaced stations (both in sub-region 6). Despite their displacement, these two stations are allocated to their original cluster.

3. Methods

Seven statistical approaches are applied in this study. They are applied for each observational station separately. The approaches are selected to span indirect DECMs (Section 3.1) as well as direct DECMs (Section 3.2). The former comprise linear and nonlinear techniques. Furthermore, the selected methods cover point-wise approaches as well as spatial approaches that use distributed predictors. Point-wise approaches relate 3×3 adjacent RCM grid cells to each station. Spatial approaches are based on meteorological fields and use principle components (PCs) from the entire RCM domain shown in Figure 2(a) to build transfer functions to the station scale. The PCs originate from principle component analysis (PCA), which is also referred to as empirical orthogonal function (EOF) analysis in geophysics (e.g. von Storch and Zwiers, 1999; Zorita and von Storch, 1999).

Table I. The MM5 predictor variables used in this study.

Variable	Abbreviation	Level
Geopotential height	zg	[700; 500] hPa
Mixing ratio	q	[850; 700] hPa
Eastward wind	u	[850; 700] hPa
Northward wind	v	[850; 700] hPa
Upward air velocity	w	[700] hPa
Vapour pressure	e	[850; 700] hPa
Saturation vapour pressure	es	[850; 700] hPa
Total surface precipitation	pre	Surface (sfc)
Convective rain	accrcon	sfc
Advective rain	accrnon	sfc
Precipitable water	pwat	sfc
Atmosphere cloud condensed water content	iclc	sfc
Surface air pressure	psfc	sfc
Temperature	t2	2 m
Mixing ratio	q2	2 m
Vapour pressure	e2	2 m
Saturation vapour pressure	es2	2 m
Relative humidity	f2	2 m
Eastward wind	u10	10 m
Northward wind	v10	10 m
Sea level pressure	pslv	Sea level (slv)

3.1. Indirect DECMs

3.1.1. Multiple linear regression (MLR)

In addition to its application in MOS, MLR is frequently found in statistical downscaling of GCM data (e.g. Kilsby *et al*., 1998; Huth, 1999; Murphy, 1999; Schoof and Pryor, 2001; Hay and Clark, 2003) as well as in climate change impact analyses (e.g. Alexandrov and Hoogenboom, 2000). In general, linear regression models establish a linear transfer function between one or more predictors and the predictand such that

$$Y^{\mathrm{MLR}} = \alpha + \sum_{p=1}^{l} \beta_p X_p + \varepsilon \qquad (1)$$

with α being the intercept, β the regression coefficients, ε the error term, X_p the p predictor variables and Y^{MLR} the estimated predictand.

For this study, an MLR based on ordinary least squares (OLS) is applied on daily basis, for each season separately and point-wise since preliminary tests favoured point-wise to spatial application. In these tests, point-wise and spatial MLR (EOF-based) yielded similar results for seasonal means. However, the variability of daily errors was mostly smaller and the shapes of the modelled distributions were closer to the observed distributions in point-wise MLR than in EOF-based MLR. Furthermore, several data transformations were analysed in preliminary tests to account for daily precipitation's non-normal distribution and nonlinear predictors-predictand relationship (Wilks, 1995; Kidson and Thompson, 1998). Results are shown only for best performing cube root transformation of the predictand (denoted as MLRT; Helsel and Hirsch, 2002).

Predictors are chosen by a semi-objective procedure for each station. Physical meaningful variables are preselected empirically before a two-step objective predictor selection method is performed: Firstly, a stepwise regression, based on the Akaike Information Criteria (Wilks, 1995), is used to reduce the number of potential predictors; secondly, an all-subset regression (von Storch and Zwiers, 1999) selects the most influencing combinations of predictors, limited to a maximum number of four predictors per combination, according to the adjusted coefficient of determination (R^2; Helsel and Hirsch, 2002).

3.1.2. Multiple linear regression with randomization (MLRR)

Climate impact studies often need precipitation time series with realistic day-to-day variability. Since MLR models reduce variability because the regression line is fitted to pass through the centroid of the data (Helsel and Hirsch, 2002) and only a part of local climate variability is related to larger-scale variability in predictors (Fowler *et al*., 2007), von Storch (1999) proposed randomization of time series to recover their original variability. MLRR extends the MLR estimation $Y_{t,i}^{\mathrm{MLRR,val}}$ in a given validation period (val) at station i and day t by adding noise $R_i^{\mathrm{MLR,cal}}$ which represents the unexplained part of the regression model (compare Equation (1); Dehn and Buma, 1999) according to

$$Y_{t,i}^{\mathrm{MLRR,val}} = Y_{t,i}^{\mathrm{MLR,val}} + R_i^{\mathrm{MLR,cal}} \qquad (2)$$

$R^{\mathrm{MLR,cal}}$ is obtained from classified MLR residuals of the calibration period (cal) and grouped in four classes corresponding to the respective quantity of $Y^{\mathrm{MLR,cal}}$ between zero precipitation, the model (mod) wet-day threshold WT^{mod}, the 50th percentile of the time series, and the maximum estimated precipitation. WT^{mod} is defined after Schmidli *et al*. (2006) that the number of days greater or equal WT^{mod} in the calibration period equals the respective observed (obs) number of days greater or equal WT^{obs} in the calibration period. WT^{obs} is defined as 1 mm/day.

3.1.3. The analogue method (AM)

Resampling approaches can be classified as stochastic precipitation models (von Storch and Navarra, 1999). Their primary application area is hydrology where they are used to generate a large number of synthetic observations as input to hydrological models in order to assess their uncertainties (Mehrotra and Sharma, 2006). The AM represents a special case of resampling. In the context of AM, the resampling is conditioned on atmospheric states (predictors; e.g. Zorita and von Storch, 1999), which enables its application for downscaling purposes. Conceptually, AM compares the atmospheric state on the day under consideration (t) to an archive of historic atmospheric states and determines the most similar historic atmospheric state – the analogue – according to some measure of similarity. The local weather on this

analogue date u is then resampled as an estimate of the predictand on day t (Cubasch *et al*., 1996). Thus, for an adequate description of the local climate, particularly regarding extreme conditions, sufficiently long historic archives are necessary. The resampling limits AM to historical extremes, which can be considered as the method's main drawback. Further problems with consistency in the order of consecutive days may occur if atmospheric regimes are not well defined by the spatial predictors. In return, AM does not assume any particular probability distribution in the modelling process and enables to capture nonlinear predictors-predictand relationships (Fernández and Sáenz, 2003; Benestad *et al*., 2009).

The crucial point of AM is the definition of similarity of atmospheric states (e.g. Wetterhall *et al*., 2005; Matulla *et al*., 2008). A method based on PCs (e.g. Zorita and von Storch, 1999) is implemented in this study. PCs are derived from fields of all RCM variables given in Table I. Analogues are found for each season by minimizing a weighted Euclidean distance:

$$\sum_{k=1}^{n} \left\{ \boldsymbol{d}_i^k \left[R_{t,i}^{\mathrm{cal},k} - \boldsymbol{F}_{t,i}^{\mathrm{val},k} \right]^2 \right\} \rightarrow \min \qquad (3)$$

$\boldsymbol{F}$ represents the so-called feature vector at validation day t which takes into account k predictors. R corresponds to the respective historical archive in the same PC phase space as the predictors. The weighting vector $\boldsymbol{d}$ (Fernández and Sáenz, 2003; Imbert and Benestad, 2005) consists of normalized eigenvalues and reflects the importance of each considered predictor. If the same Euclidean distance is found several times in the historical sample, the temporally first condition is considered. Additionally, an unweighted distance was tested with slightly worse results (not shown).

Contrary to the standard AM application with multisite prediction, this study uses a site-specific AM. Such a site-specific approach weakens the merit of maintaining the spatial covariance structure of the predictand, but enables the selection of the most important predictors for each station separately. The first three PCs of each RCM variable are proposed as independent predictors to the automated predictor selection scheme (see MLR description). The PCs are standardized to unit variance. Limitations in AM's nonlinearity due to linearly selected predictors are accepted. The predictor selection results in the seasonally prevailing predictor combination of four predictors ($k = 4$).

3.1.4. The nearest neighbour analogue method (NNAM)

Extending the search for the analogue towards a probabilistic approach, the NNAM (e.g. Brandsma and Buishand, 1998; Mehrotra and Sharma, 2006; Moron *et al*., 2008) randomly chooses the analogue situation from the nn most similar historical conditions. Consequently, increased modelled variability can be expected with the drawback that equal predictor conditions on time t and $t + 1$ may result in different local predictands. The same predictors as applied to AM are considered. Equal to AM, a weighted Euclidean distance in PC phase space provides the measure of similarity. Instead of picking the most similar historic condition to prediction time t ($nn = 1$), the nn days with smallest Euclidean distances (nearest neighbours) are retained. The analogue is selected using a discrete probability distribution that weights the nn days according to

$$p_j = \frac{1/j}{\sum_{i=1}^{nn}(1/i)}, \; j = 1, \ldots, nn \qquad (4)$$

with p_j being the probability of the j closest neighbour (Lall and Sharma, 1996). By this means, higher weights are given to closer neighbours. Several tests with $nn = 5$, $nn = 10$ and $nn = \sqrt{n}$, with n being the calibration sample size, were performed (compare Beersma and Buishand, 2003). In this study $nn = 5$ is used. Unweighted random selection of the analogue was tested as well, but is outperformed by the presented approach (not shown).

3.2. Direct DECMs

3.2.1. Local intensity scaling (LOCI)

LOCI is a direct DECM and represents one traditional bias correction method (also compare Graham *et al*., 2007; Leander and Buishand, 2007), which is based on the work of Widmann *et al*. (2003), suggested by Schmidli *et al*. (2006) and successfully applied by, e.g., Salathé (2003), Dobler and Ahrens (2008) or Moron *et al*. (2008). The basic idea of direct DECMs is that climate model precipitation integrates all relevant predictors. Deviations between climate model precipitation and regional- or local-scale precipitation observations are in first order due to systematic climate model errors and an incomplete or inaccurate representation of the orography (Schmidli *et al*., 2006). Thus, instead of using various predictors to create local weather, LOCI applies a spatially varying scaling to climate model precipitation accounting for its long-term bias at the location of the observation.

Following the approach of Schmidli *et al*. (2006), a separate correction of wet-day frequency and wet-day intensity is applied point-wise and for each day of the year separately. With the definitions of WT^{mod} and WT^{obs} (see MLRR description), climate model precipitation X_t^{val} is corrected by Equation (5) using scaling factor S from Equation (6). $\overline{X}_t^{\mathrm{wet,cal}}$ and $\overline{Y}_t^{\mathrm{wet,cal}}$ represent climatological means on wet days (i.e. days with precipitation greater or equal to WT) of the modelled and the respective observed precipitation data over the calibration period at day t. Only pair-wise recorded modelled and observed data are used for calibration:

$$Y_{t,i}^{\mathrm{val}} = \max(S_{t,i}(X_{t,i}^{\mathrm{val}} - WT_{t,i}^{\mathrm{mod,cal}}) + WT_{t,i}^{\mathrm{obs,cal}}, 0) \qquad (5)$$

$$S_{t,i} = \frac{\overline{Y}_{t,i}^{\text{wet,cal}} - WT_{t,i}^{\text{obs,cal}}}{\overline{X}_{t,i}^{\text{wet,cal}} - WT_{t,i}^{\text{mod,cal}}} \quad (6)$$

Contrary to previous applications, this study uses a moving window approach centred over focus day t to calculate $\overline{X}_t^{\text{wet,cal}}$ and $\overline{Y}_t^{\text{wet,cal}}$. By this means, inhomogeneities at the end of fixed calibration periods (e.g. months, seasons) are avoided, and the dependence of model errors on the time of the year is included. Moving time windows between 15 and 61 days were investigated. A window size of 61 days is chosen to enable an annual-cycle sensitive correction as well as a sufficient large sample size. Further, setting WT^{obs} to 0.3 mm/day and 0.5 mm/day instead of the 1 mm/day standard value was investigated, but resulted in no significant differences (not shown).

3.2.2. *Quantile mapping (QM)*

Extending the correction from means (LOCI) to the entire distribution, QM corrects for errors in the shape of the distribution and is therefore capable to correct errors in variability as well. This quantile-based approach originates from the empirical transformation of Panofsky and Brier (1968) and was successfully implemented in hydrological applications (Dettinger *et al*., 2004; Wood *et al*., 2004; Boé *et al*., 2007) but recently also for error correction of RCMs (Dobler and Ahrens, 2008; Piani *et al*., 2009).

For this study, QM is based on point-wise and daily constructed empirical cumulative distribution functions (*ecdfs*; Wilks, 1995) of modelled and observed datasets in the calibration period. This is in contrast to other bias correction studies where theoretical *cdfs* are estimated only from wet days (e.g. Ines and Hansen, 2006; Dobler and Ahrens, 2008; Piani *et al*., 2009). Using *ecdfs*, QM is generally applicable to all possible meteorological parameters, whereas applications based on *cdfs* may become problematic for parameters that do not fit to theoretical functions such as global radiation, where the *ecdf*'s shape is changing with the season (compare Camuffo, 1978). Equal to LOCI, a 61-day moving window, centred over the focus day, is used for *ecdf* construction. The correction function transfers raw RCM output X^{val} to the corrected estimate Y^{val} such that

$$Y_{t,i}^{\text{val}} = ecdf_{t,i}^{\text{obs,cal}^{-1}}(ecdf_{t,i}^{\text{mod,cal}}(X_{t,i}^{\text{val}})) \quad (7)$$

with $ecdf^{-1}$ indicating the inverse *ecdf*, and thus a data quantile. As this purely empirical QM only maps modelled values to observed values, no new extremes (outside the observed range) can be obtained. This is a suitable approach for our study, since we apply the correction to a historical hindcast simulation. For applications to future climate simulations, however, some kind of extrapolation beyond the range of observations has to be added to allow for 'new extremes' (e.g. Boé *et al*., 2007).

4. Results and discussion

4.1. Validation framework

Statistical approaches implicitly assume stationarity in their transfer functions in the case of indirect DECMs or in model error characteristics in the case of direct DECMs (Wilby, 1997; Benestad *et al*., 2009). If this assumption is violated, statistical models cannot account for changes described by predictor forcings. As this assumption cannot be approved in advance, a temporal cross-validation framework is applied which repeatedly divides the data period into a calibration (10 years) and independent validation period (1 year). By this means, each year is estimated and evaluated independently with the remaining 10 years used for model calibration (sometimes denoted as 'leave one out' cross-validation; see Figure 3). For evaluation purpose, model skill scores as well as model error characteristics are used. The models' performances are analysed via mean skill scores and mean model error characteristics, averaged over all validation periods, as well as using the entire, not averaged, 11-year validation time series. Mean skill scores and model error characteristics are presented in Figures 5–9. The models' performances represent the station scale as each statistical model is calibrated and evaluated separately station by station. For graphical representation, the station-wise evaluation results are spatially averaged in sub-region 6, sub-region 8 (compare Figure 2(b)), and for entire Austria. The two sub-regions are selected because of their different climate characteristics. Sub-region 6 is mainly dominated by westerly flows from the Atlantic with high precipitation amounts, whereas sub-region 8 features more continental dry characteristics with additional influence from the Mediterranean Sea.

The results are divided in three parts. The first describes the general characteristics of the uncorrected RCM within all regions. The second focusses on the characteristics of each DECM and the third part analyses the effectiveness of DECMs compared to uncorrected RCM results.

4.2. RCM evaluation

Gobiet *et al*. (2006) already compared the MM5 precipitation data from 1981 to 1990 on monthly scale to HISTALP observations (Auer *et al*., 2007). The same comparison is shown in Figure 4 on seasonal basis. Regarding Austria, the RCM features seasonally and

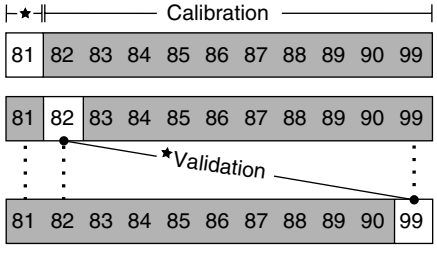

Figure 3. The 'leave one out' cross-validation scheme. Each of the 11 simulated years is post-processed once independently from the remaining 10 years used for calibration.

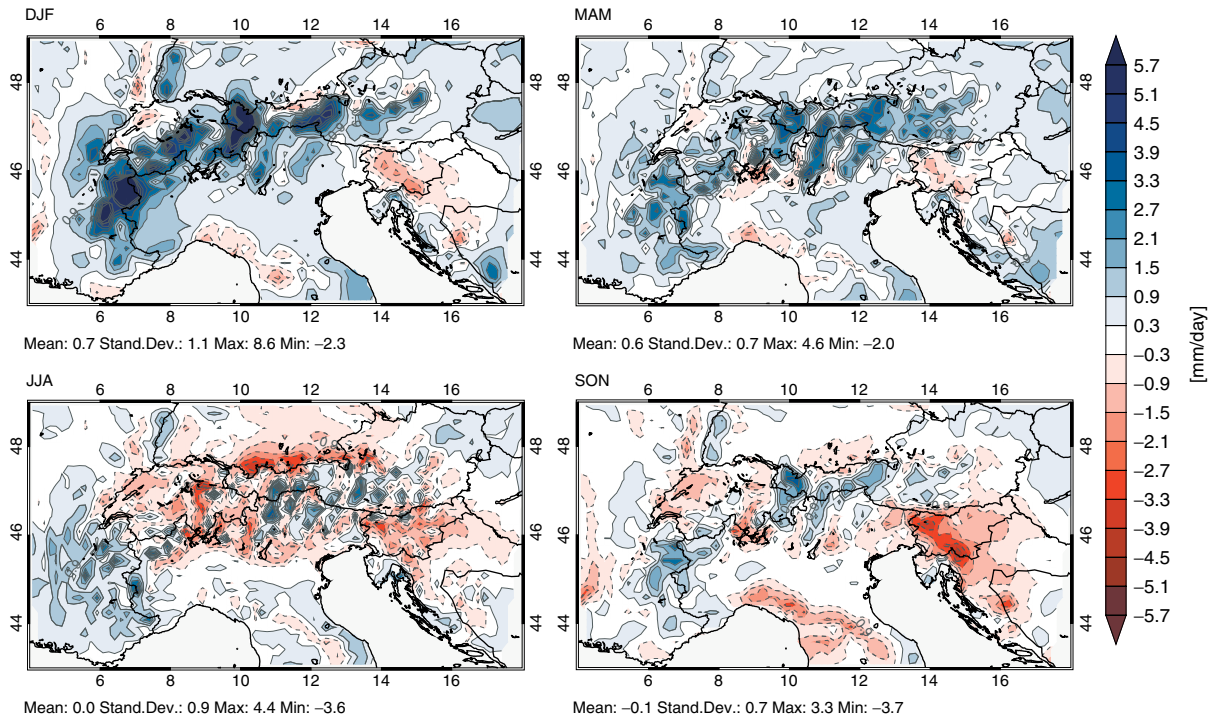

Figure 4. Mean seasonal differences of daily RCM precipitation sums compared to HISTALP observations. Continuous contours are positive; dashed contours are negative. This figure is available in colour online at www.interscience.wiley.com/ijoc

regionally varying error characteristics with strong precipitation overestimation along the Alpine crest in winter (DJF), an overall good performance in summer (JJA), and underestimation at the southern Austrian border in autumn (SON). Gobiet *et al.* (2006) argue that, besides possible model deficiencies, the well-known problematic precipitation measurement at high altitudes, especially in DJF, may partly cause the pronounced overestimation. Secondly, the reduced SON precipitation in south-eastern parts of Austria is probably related to an under-representation of northern Mediterranean cyclones and a consequent lack of humidity. These findings further motivate the selection of sub-regions 6 and 8 for evaluation as these regions cover the problematic areas.

4.3. Characteristics of the applied DECMs

Referring to the MLR predictor selection, Table II shows the most important seasonal predictors for the considered study regions. All three regions indicate precipitation (accrnon, accrcon, pre), humidity-related parameters at surface (q2, pwat), as well as eastward (u) and northward (v) wind at 10 m and 750 hPa, and surface vapour pressure (e2) to be the dominant predictors for local precipitation. The composition of the predictor set varies seasonally with increased importance of the convective precipitation (accrcon) and northward wind in summer months, which reasonably corresponds to the regional climate characteristics (see Section 2). The dominance of RCM precipitation as predictor supports the assumption that RCM precipitation integrates large parts of the relevant information for local precipitation. Further, the frequently claimed integration of humidity as predictor (e.g. Giorgi and Mearns, 1991; Wilby and Wigley, 2000; Fowler *et al*., 2007) is supported.

Similar to point-wise predictor selection for MLR, Table III indicates that PCs of precipitation fields are by far the most important ones for local precipitation for conditional resampling approaches. Further relevant predictors are pressure-related parameters at surface (es, psfc, pslv), geopotential height at 500 hPa (zg), and vertical velocity at 700 hPa (w).

Figure 5 illustrates the annual evolution of the wet-day thresholds WT^{mod} and the scaling factors S used in LOCI. Both parameters feature distinct annual cycles, which indicate frequency overestimation in winter and intensity underestimation in summer in the RCM. WT^{mod} ranges from 1.5 mm/day to 5 mm/day, which differs from the results of Schmidli *et al*. (2006), who found wet-day thresholds around 1 mm/day for the same region with LOCI, but applied to coarser ERA-40 reanalysis and calibrated on the entire year. S varies around one with a reversed pattern compared to WT^{mod} and shows comparable quantities to Schmidli *et al*. (2006). The ranges of the magnitudes of S and WT^{mod} indicate that, besides for the summer season, the RCM precipitation error is overall dominated by a frequency overestimation error.

The seasonal correction functions of QM in Figure 6 show differences of all percentiles between observed and modelled calibration *ecdfs* for all study regions. The respective precipitation quantities are indicated on the x-axes. Generally, and particularly in winter in sub-region 6, the RCM overestimates wet-day precipitation

Table II. Seasonal predictor variables for MLR approaches in sub-region 6, sub-region 8, and for entire Austria according to their occurrence probability (Prob) given in percent after objective predictor selection..

DJF						MAM					
Sub-region 6		Sub-region 8		Entire Austria		Sub-region 6		Sub-region 8		Entire Austria	
Predictor	Prob	Predictor	Prob	Predictor	Prob	Predictor	Prob	Predictor	Prob	Predictor	Prob
pre_sfc	19.5	q2_2m	16.2	pre_sfc	17.6	q2_2m	18.0	accrnon_sfc	19.6	accrnon_sfc	15.0
accrnon_sfc	17.0	accrnon_sfc	13.1	accrnon_sfc	12.1	accrnon_sfc	12.7	accrcon_sfc	14.4	q2_2m	13.3
u700hPa	10.0	e2_2m	13.1	q2_2m	10.8	pre_sfc	12.7	q2_2m	8.6	pre_sfc	10.0
iclc_sfc	8.2	pre_sfc	12.9	v_700hPa	9.8	e2_2m	11.7	e2_2m	7.5	e2_2m	9.4
e2_2m	6.0	v_700hPa	12.8	e2_2m	8.9	u10_10m	10.0	zg_700hPa	6.1	accrcon_sfc	6.6
JJA						**SON**					
Sub-region 6		Sub-region 8		Entire Austria		Sub-region 6		Sub-region 8		Entire Austria	
Predictor	Prob	Predictor	Prob	Predictor	Prob	Predictor	Prob	Predictor	Prob	Predictor	Prob
pre_sfc	14.5	v_700hPa	19.2	v_700hPa	17.4	pre_sfc	20.7	q2_2m	23.0	accrnon_sfc	16.6
v_700hPa	14.2	pre_sfc	17.6	pre_sfc	15.4	accrnon_sfc	16.0	e2_2m	22.8	pre_sfc	14.5
u_700hPa	10.7	accrcon_sfc	12.0	accrcon_sfc	7.7	q2_2m	11.0	accrnon_sfc	16.7	q2_2m	13.0
accrnon_sfc	10.2	iclc_sfc	8.1	accrnon_sfc	7.4	e2_2m	10.2	accrcon_sfc	9.7	e2_2m	12.0
pwat_sfc	8.5	u_700hPa	7.4	v_850hPa	6.3	v10_10m	7.2	pre_sfc	9.3	accrcon_sfc	8.6

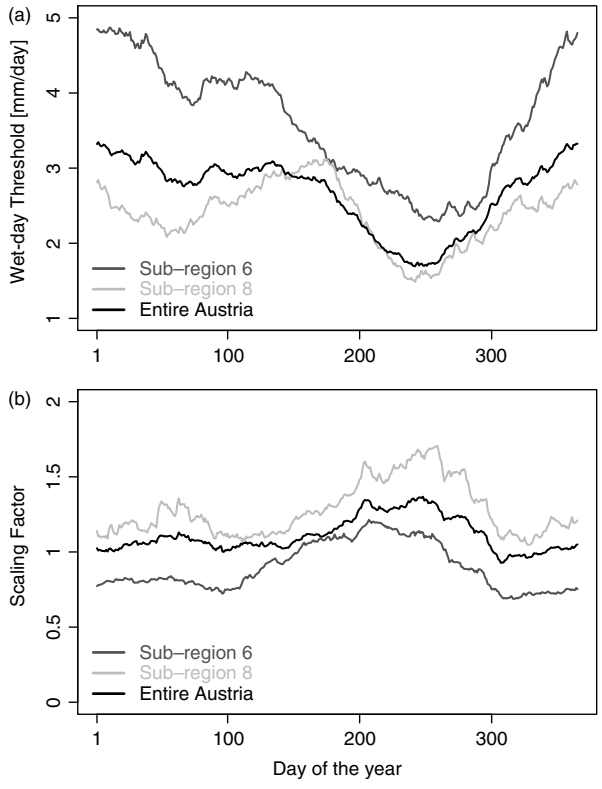

Figure 5. Annual cycles of wet-day threshold WT^{mod} (a) and scaling factor S (b) used for LOCI. Both parameters result from station-wise calibration and are spatially averaged.

intensities, which leads to partly significant negative correction values, especially at the highest precipitation intensities (i.e. at the highest percentiles). By contrast, particularly in summer in sub-region 8, significant positive correction values at the highest precipitation intensities indicate a lack of extreme precipitation events in the RCM data. The highest corrections are applied to the highest percentiles and range from −12 mm/day (in winter in sub-region 6) to +15 mm/day (in summer and autumn in sub-region 8). For entire Austria the correction function is strongly damped, which illustrates the importance of point-wise application where local error characteristics are taken into account instead of a broad spatial average. Abrupt changes of the correction function at highest modelled precipitation amounts, as illustrated in winter, spring and summer in sub-region 6, are more probably related to statistical noise at these percentiles than to RCM error characteristics.

4.4. DECM evaluation

For assessing the skill of the considered DECMs, their performances are evaluated regarding the median, variability, and indicators for extremes. Boxplots in Figure 7 display the median seasonal and annual differences between models and observations as lines in the middle of 25th and 75th quantile boxes derived from daily differences. Standardized Taylor diagrams (Figure 8; Taylor, 2001) show the normalized centred root-mean-square (RMS) difference of the different DECMs compared to observations as the distance to point 1 on the abscissa, the variance ratio between models and observations as the radial distance to the zero point, and the correlation between models and observations as the angle between the abscissa and the position vector (i.e. a perfect model would be displayed on point 1 of the abscissa). Error diagrams in Figure 9 illustrate the performances of the methods regarding precipitation intensity (SDII), wet-day frequency (Freq), the 95th percentile of all modelled days (Q95), and the 75th percentile on wet days (RQ75), where the latter two represent moderately extreme conditions.

Table III. As in Table II but with atmospheric predictor fields for the analogue methods. PC indicates the used principle component.

DJF						MAM					
Sub-region 6		Sub-region 8		Entire Austria		Sub-region 6		Sub-region 8		Entire Austria	
Predictor	Prob	Predictor	Prob	Predictor	Prob	Predictor	Prob	Predictor	Prob	Predictor	Prob
accrnon_sfc PC2	17.2	accrnon_sfc PC3	15.6	accrnon_sfc PC1	11.3	accrnon_sfc PC2	16.5	pre_sfc PC2	23	pre_sfc_D2 PC2	15.6
accrnon_sfc PC1	16.7	es_2m PC2	14.7	pre_sfc PC1	10.6	accrcon_sfc PC1	16.2	accrnon_sfc PC2	17.1	accrnon_sfc PC2	14.1
W_700hPa PC3	10.2	accrcon_sfc PC1	13.8	accrnon_sfc PC2	9.8	pre_sfc PC2	11.5	pre_sfc PC1	8.8	accrcon_sfc PC1	10.1
pre_sfc PC1	7.7	pre_sfc PC1	8.3	accrnon_sfc PC3	7.7	accrnon_sfc PC1	6.5	pre_sfc PC3	7.4	pre_sfc PC1	5.8
psfc_sfc PC2	6.7	zg_500hPa PC2	6.1	pre_sfc PC2	6.0			iclc_sfc PC1	6.3		
JJA						SON					
Sub-region 6		Sub-region 8		Entire Austria		Sub-region 6		Sub-region 8		Entire Austria	
Predictor	Prob	Predictor	Prob	Predictor	Prob	Predictor	Prob	Predictor	Prob	Predictor	Prob
accrnon_sfc PC1	18.0	pre_sfc PC2	18.9	pre_sfc PC1	13.0	accrnon_sfc PC3	22.5	accrcon_sfc PC3	20.9	accrnon_sfc PC3	22.1
zg_500hPa PC2	10.7	w_700hPa PC3	12.9	pre_sfc PC2	12.1	accrcon_sfc PC1	22	accrnon_sfc PC3	18.5	accrcon_sfc PC1	14.6
w_700hPa PC3	10.5	accrcon_sfc PC1	11.9	accrnon_sfc PC1	10.8	pre_sfc PC2	15.5	accrcon_sfc PC1	15.5	accrcon_sfc PC3	9.4
pre_sfc PC1	8.0	pre_sfc PC1	11.5	accrcon_sfc PC1	6.0			pre_sfc PC1	7.5	pre_sfc PC2	7.6
Pslv_slv PC3	5.2	accrnon_sfc PC1	7.9	w_700hPa PC3	5.9			pre_sfc PC2	6.6	pre_sfc PC1	7.0

The results in Figure 9 are colour-coded; lighter colours indicate smaller errors. Finally, a quantile-quantile plot in Figure 10 compares the 11-year seasonally and annually modelled to the observed distributions using all station time series within the respective region. This enables the analysis of the DECMs' performances for absolute extreme conditions. In the case of linear regression models, also negative precipitation values are produced. Though unphysical, we did not replace these negative values by zeros in order to avoid the introduction of biases or the reduction of variability in the evaluation statistics.

In Figure 7 the leftmost bars display the regional average RCM error characteristics. They indicate the largest error ranges in sub-region 6, as expected. The error range shows a high seasonality, which is related to overestimated temporal variability, shown in Figure 8. The results from Figures 5 and 6, showing that higher modelled precipitation sums are positively biased, can be identified by the positive skewness of the difference bars.

In comparison, all DECMs except MLR virtually correct the median error of daily precipitation to zero, independent of season and region. QM systematically yields the best results followed by LOCI, AM and NNAM. MLR partly even degrades error characteristics, which is probably related to nonlinear relations between predictors and local daily precipitation as well as to non-normally distributed and heteroscedastic residuals (compare Wilks, 1995). However, with the simple extension of MLR to MLRR this deficiency can be removed due to the incorporation of error residuals (Equation (2)). MLRT corrects the median difference to nearly zero, but shifts the error distribution to negative values, whereas all other statistical approaches show nearly equally distributed differences around the median. Though only two sub-regions are presented here in detail, all DECMs show similar performances in all sub-regions shown in Figure 2.

The effect of DECMs on variability is displayed in Figure 8. In general, the RCM tends to overestimate day-to-day variability, but also shows pronounced underestimation in sub-region 8. These deficiencies are removed by most DECMs. Major problems remain for MLR which strongly underestimates variability and MLRT which shows non-systematic errors in variability with the tendency to underestimation. However, by adding error residuals (MLRR) the variability is modelled adequately. Minor problems are shown for LOCI, where especially for entire Austria a tendency to variability overestimation

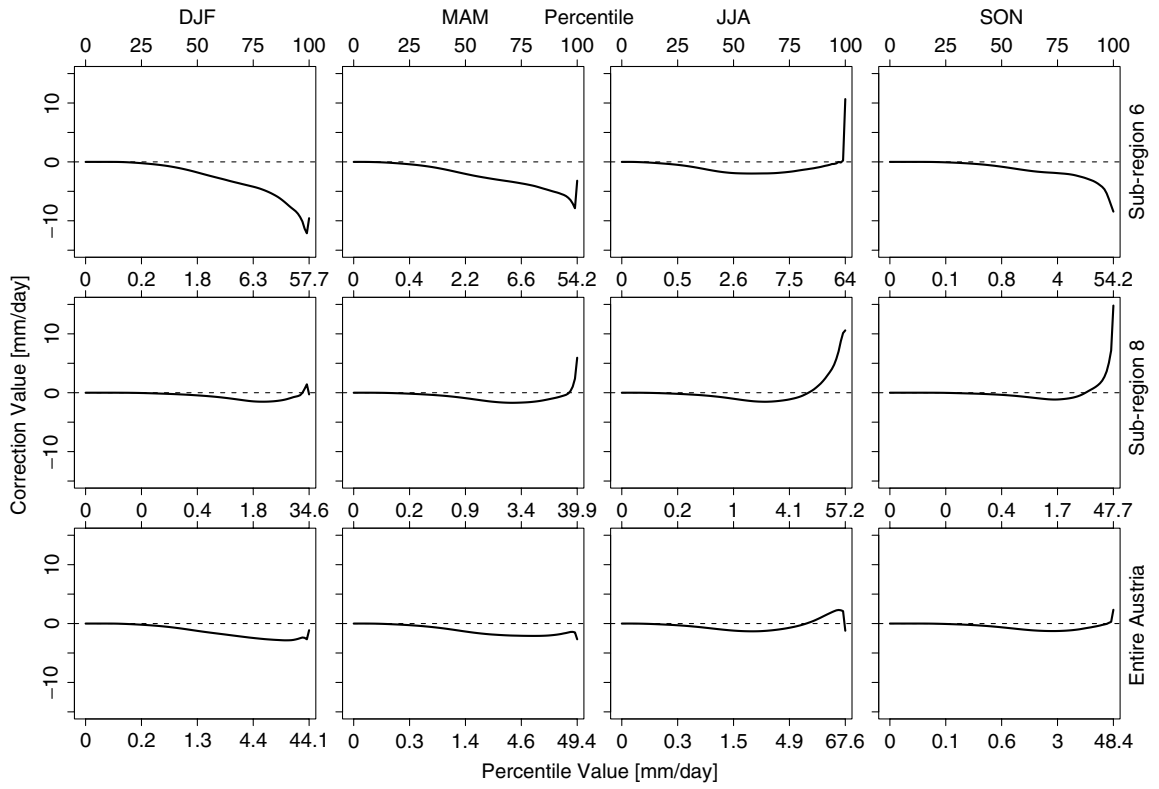

Figure 6. Seasonal correction functions derived from differences of all percentiles between observed and modelled *ecdfs* in sub-region 6 (upper panels), sub-region 8 (middle panels) and entire Austria (lower panels). The 0th percentile represents the difference between the *ecdfs*' minima, the 100th percentile represents the difference between the *ecdfs*' maxima. The precipitation quantities corresponding to these percentiles are indicated on the *x*-axes. The correction functions are obtained equally as described in Figure 5.

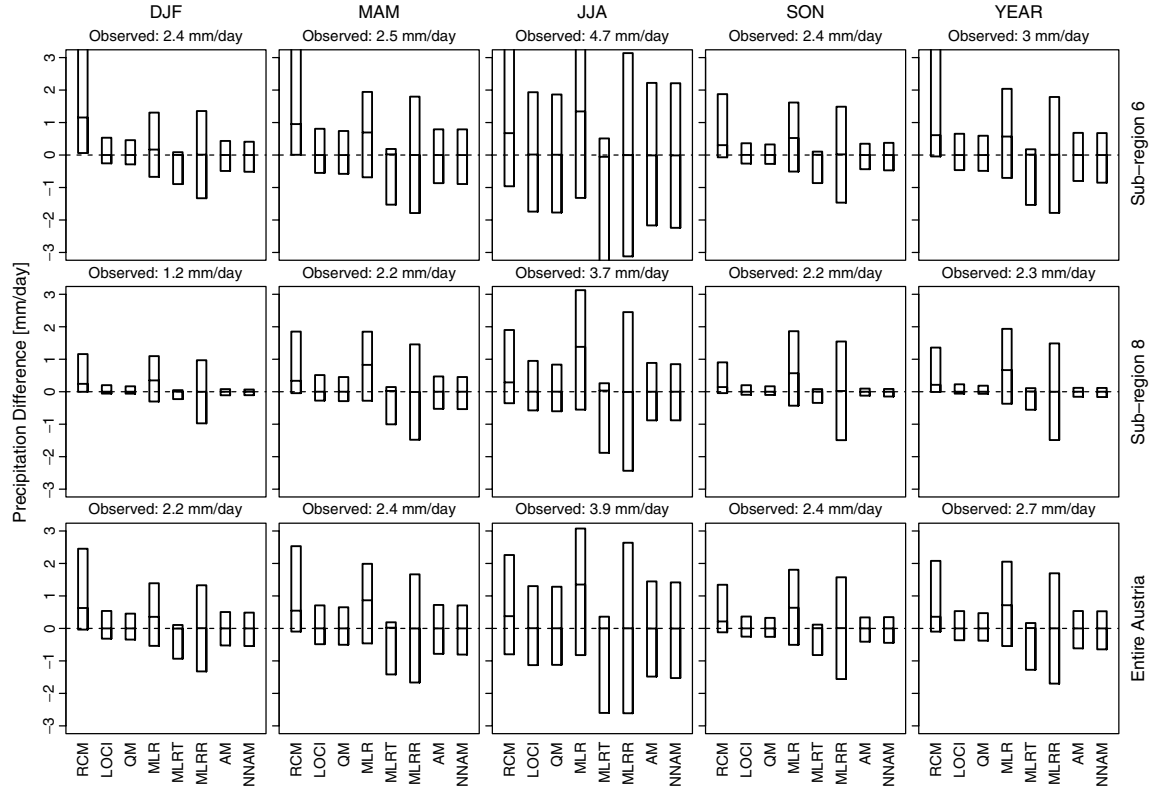

Figure 7. Seasonal and annual errors of the uncorrected RCM and the considered DECMs in sub-region 6 (upper panels), sub-region 8 (middle panels) and for entire Austria (lower panels). The boxes show the 75th percentile (upper limit), the median (line within the box) and the 25th (lower limit). The respective mean observed precipitation amount is given in the header of each panel. The statistics result from station-wise evaluation of daily precipitation data and are spatially averaged.

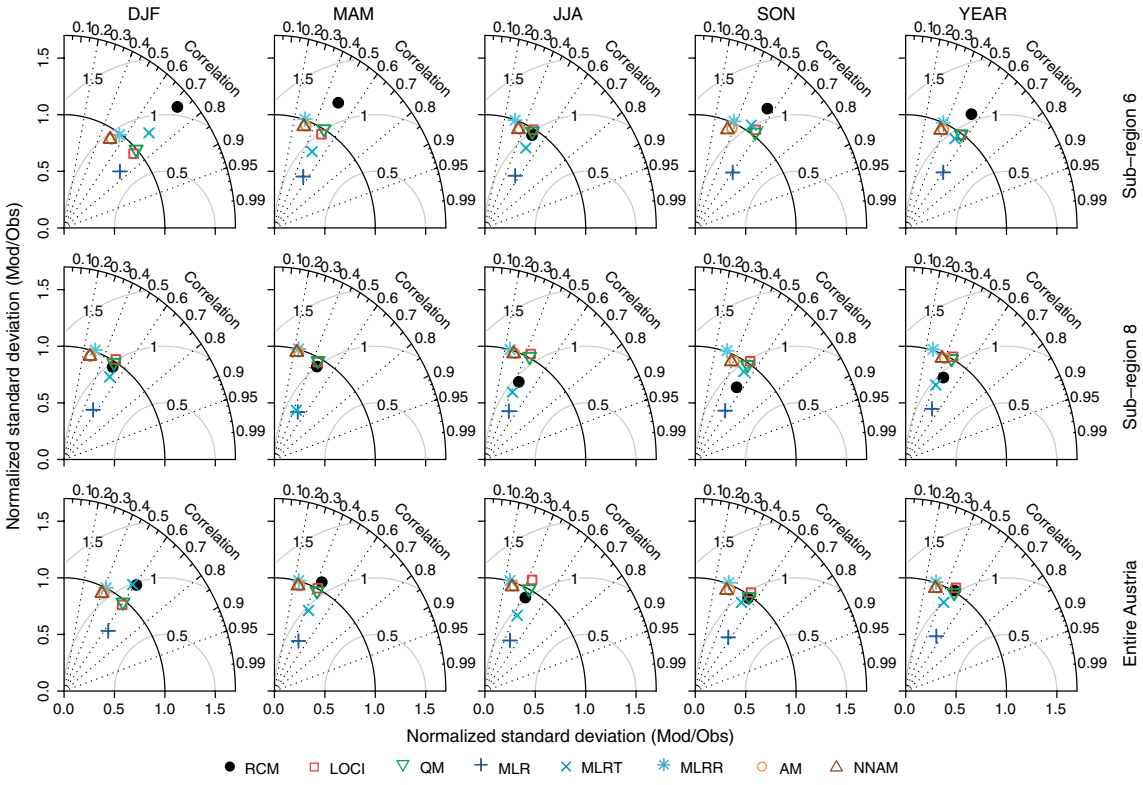

Figure 8. Seasonal and annual Taylor diagrams comparing the uncorrected RCM and the considered DECMs with observations in sub-region 6 (upper panels), sub-region 8 (middle panels) and for entire Austria (lower panels). The skill scores (normalized centred RMS, correlation as well as normalized variance ratio) are obtained equally as described in Figure 7. This figure is available in colour online at www.interscience.wiley.com/ijoc

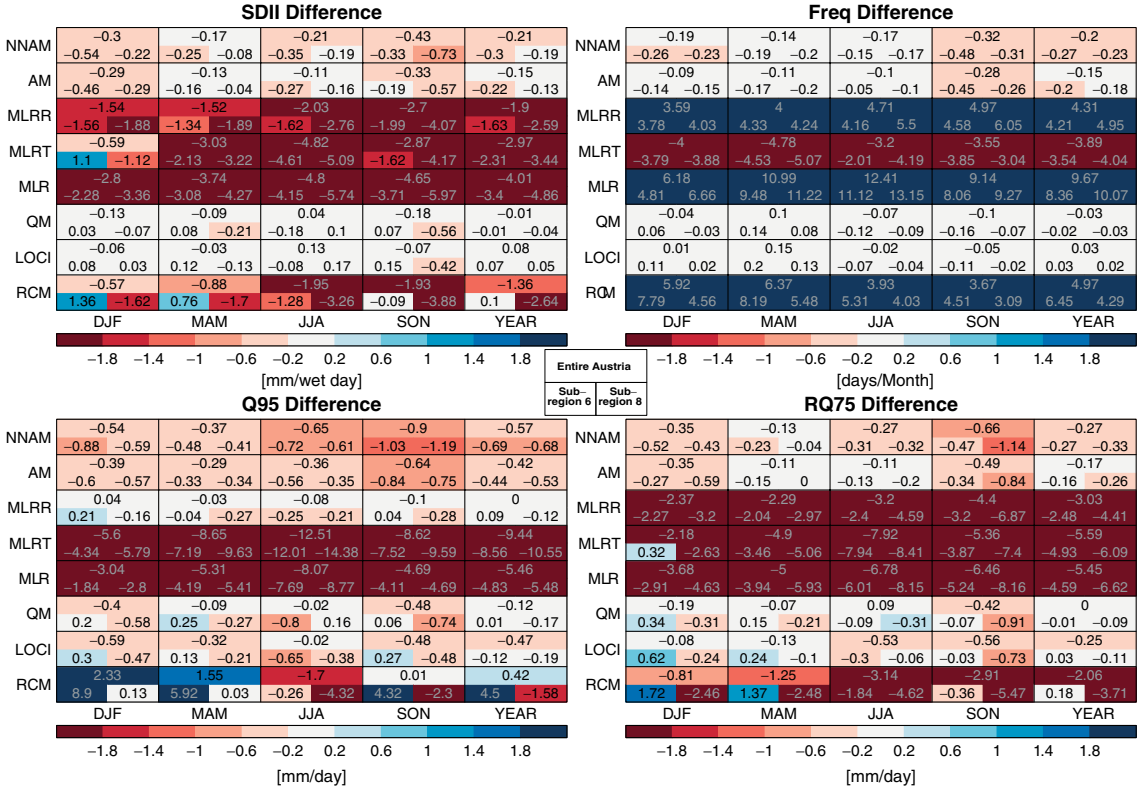

Figure 9. Seasonal and annual error portraits comparing the uncorrected RCM and the considered DECMs with observations. For each method and season, the results are given for entire Austria in the upper part, for sub-region 6 (left) and sub-region 8 (right) in the lower part of the respective box. SDII: precipitation intensity; FREQ: wet-day frequency; Q95: 95th percentile on all days; RQ75: 75th percentile only on wet days. The skill scores are obtained equally as described in Figure 7. This figure is available in colour online at www.interscience.wiley.com/ijoc

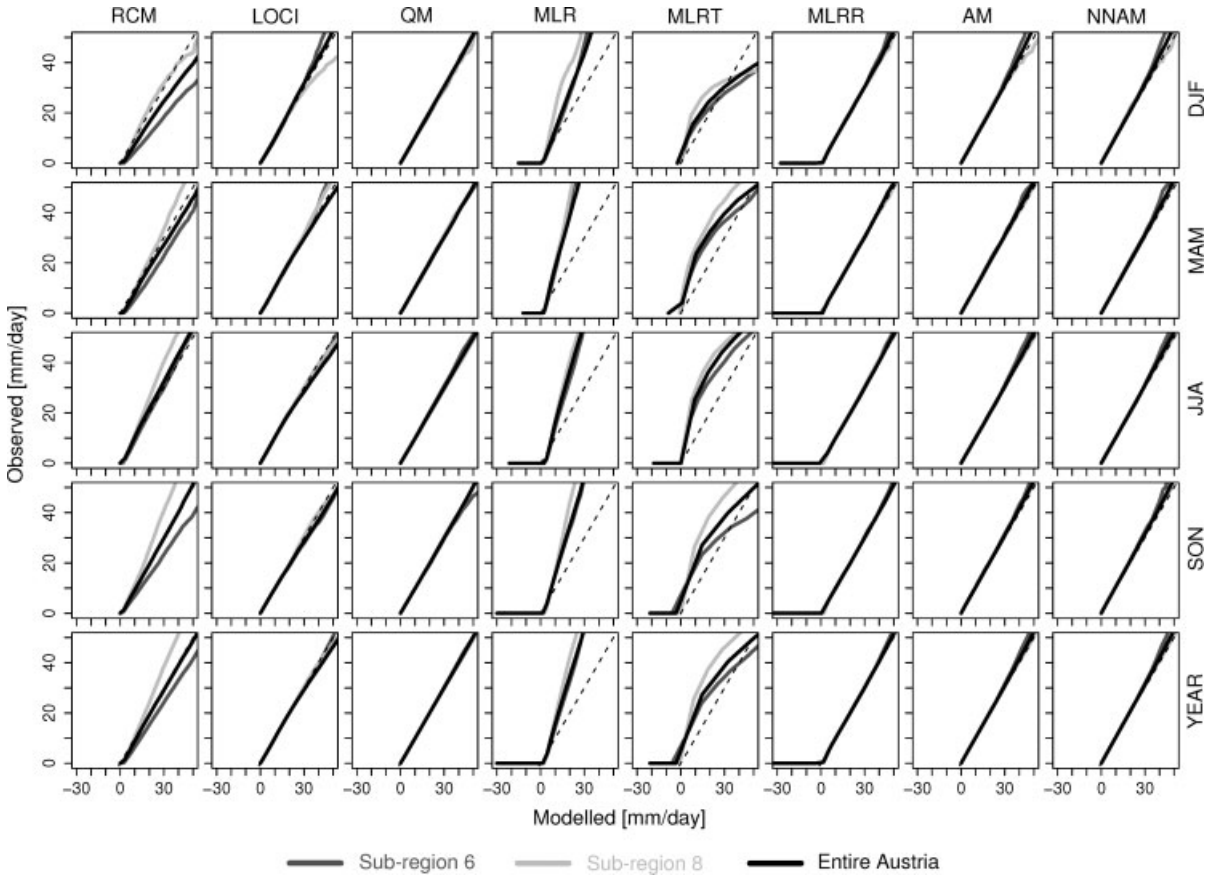

Figure 10. Seasonal and annual quantile-quantile (QQ) plots comparing the uncorrected RCM and the considered DECMs with observations. The QQ plots take into account the respective 11 years' seasonal or annual data of each individual station within the considered region. If the continuous curves equal the dashed line, the modelled data have the same distribution as the observed data. The limit of 50 mm/day at least represents the 99th percentile in the observed data.

is indicated. Additionally, LOCI was more sensitive to a reduced window size than QM concerning the variance ratio (not shown). None of the DECMs is able to increase correlation. This is expected for direct DECMs as they solely rely on temporal characteristics of climate model precipitation. AM, NNAM and MLRR even degrade correlation. In the case of MLRR this is caused by the random resampling of residuals, whereas concerning the conditional resampling methods this might be an indication that the mesoscale fields, used as predictors, do not fully explain local precipitation. Furthermore, with the exception of MLR, DECMs show no systematic reductions of the RMS, but even sometimes enlarge it. However, an increasing RMS does not indicate a worse model skill, as at low correlation levels an underestimated variance ratio lowers the RMS (compare MLR in Figure 8). In summary, most DECMs drastically reduce seasonal precipitation biases, some strongly improve the temporal variability, but many improve temporal correlation on a daily basis. However, since this study focusses on climate applications, the improvement of temporal correlation is not the objective.

Figure 9 depicts several further performance indices: the uncorrected RCM overestimates wet-day frequency (Freq), as already demonstrated. Daily precipitation intensity (SDII), in contrast, shows regional variations, but the tendency to be underestimated by the RCM. These RCM Freq and SDII behaviour are characteristic of the 'drizzle' problem in climate models (e.g. Gutowski *et al*., 2003; Fowler and Kilsby, 2007). LOCI and QM correct these errors to virtually zero. Resampling approaches, particularly NNAM, show significant skill, but slight systematic underestimation of the analysed indicators. Although MLRR improves MLR, both regression approaches fail in reproducing intensity and frequency, with drastic intensity underestimation (up to −4.6 mm per wet day) and overestimation of frequency (up to about 12 days per month). MLRT shows similar results for intensity, but underestimation of frequency.

Towards extreme precipitation (Q95, RQ75), the uncorrected RCM shows an inhomogeneous picture with overestimation in sub-region 6 and underestimation in sub-region 8. Only in summer all regions agree in underestimation of higher precipitation amounts. QM and LOCI but also AM as well as NNAM systematically reduce RCM error characteristics in these moderately extreme precipitation indices, which is also demonstrated by the quantile-quantile plots in Figure 10. MLR and MLRT underestimate Q95 and RQ75 significantly, demonstrating their deficiencies in estimating the

daily precipitation's distribution. MLRR captures Q95 surprisingly well, whereas RQ75 is heavily biased. This is related to MLRR resampling, which correctly broadens the entire distribution as seen in Figure 10, but does not correct the general MLR problem of estimating the right wet-day probability. The latter fact is confirmed by the underestimation of RQ75. The problematic characteristics of MLRT become obvious in Figure 10, which shows a significant curvature in the quantile-quantile relation. Figure 10 also confirms the superior performance of LOCI, AM, NNAM and particularly QM for higher quantiles. However, minor deficiencies still remain; e.g. in winter in sub-region 8, LOCI significantly overestimates heavy precipitation events greater or equal to 30 mm/day. This is caused by scaling factors which adequately correct for the mean, but fail to correct these extremer precipitation intensities in the RCM where the error characteristics change from under- to overestimation (compare Figure 10 upper leftmost panel).

5. Summary and conclusions

State-of-the-art RCMs feature significant errors and are therefore often not directly applicable to climate change impact research. This calls on the one hand for further RCM development and improvement, and on the other hand for the more pragmatic approach of empirical-statistical post-processing and correcting error in RCM results to create user-tailored datasets for climate change impact research.

This study evaluated and compared seven different empirical-statistical DECMs for daily precipitation sums from an RCM on the station scale. For this purpose, a cross-validation framework was used, where each evaluated year was post-processed independently from the calibration period.

Most DECMs show enormous potential for reducing RCM error characteristics, which underlines the advantages of combining RCMs and DECMs in climate change impact research. None of the DECMs is able to improve the modelled temporal correlation with observations, but this is of minor importance for climatological studies. Direct DECMs (QM, LOCI) and nonlinear indirect resampling methods (AM, NNAM) virtually remove RCM deficiencies in the entire precipitation distribution, i.e. they correct for mean, day-to-day variability, and extremes. The improvements are obtained regardless of season and region. This indicates the transferability of the presented DECMs from the Alpine area in this study to other regions and climates.

QM, LOCI, AM and NNAM result in nearly similar skills with slight advantages for QM. QM is also favourable due to its simplicity, nonparametric configuration, and consequent applicability to other parameters than daily precipitation. However, instabilities at the highest quantiles of the correction function and the possible extrapolation of the correction function beyond the range of observed values should be further investigated to optimize the applicability of QM to future climate scenarios, particularly with regard to the analysis of trends in extremes. LOCI is of comparable simplicity to QM, also independent of distribution, but features some instabilities in the estimation of temporal variability. Furthermore, as the LOCI scaling factors are calibrated on climatological mean values, LOCI does not adequately correct data which feature significantly curved, intensity-dependent error characteristics. The dominance of direct DECMs, which only build on model precipitation as predictor, is not surprising as the predictor selection for the indirect DECMs yielded RCM precipitation to be of major relevance for local precipitation. The resampling methods' performances are comparable to QM as well, though they tend to underestimate the highest quantiles as LOCI and slightly degrade temporal correlation. The linear MLR techniques, although optimized by randomization, power law transformation, and objective predictor selection, are defective in estimating non-normally distributed daily precipitation and thus cannot be recommended for RCM precipitation error correction. However, nonlinear regression techniques as the support vector regression approach (e.g. Hsieh, 2009) might be worth exploring.

The discussed improvements of DECMs are, in a strict sense, only valid for the MM5 mesoscale climate model used in this study and cannot directly be transferred to other RCMs. However, further works by Piani *et al.* (2009) or Dobler and Ahrens (2008) and our own experience give confidence in the robustness especially of direct DECMs applied to any RCM, if the general assumption of stationary error-characteristics is not violated.

Considering the application of the presented methods to climate scenarios, it has to be emphasized that this study does not take into account the effect of decadal climate variability on model error characteristics. Thus, the robustness of DECMs applied to long-term RCM simulations remains to be demonstrated in further investigations. Nevertheless, DECMs should be separated into two classes if applied to climate change assessment studies: firstly, methods that base their calibration on pairs of observed and modelled values are only applicable where the calibration-simulation is correlated with weather. This applies, e.g., to an RCM simulation driven by reanalysis boundary conditions. Such application would correct for the RCM error (assuming perfect boundary conditions), but if used for a future scenario which is driven by a GCM, not correct the GCM error. Secondly, 'climatological' methods that base their calibration on climatologies only, like QM and LOCI, do not require calibration-simulations that are correlated to actual weather. They can be, e.g., calibrated on RCM simulations driven by a GCM control run and would, if applied to future scenario simulations, correct for the combined GCM-RCM error, which is a clear advantage in climate applications.

Acknowledgements

The authors are grateful to the Austrian Meteorological Service (ZAMG), the Austrian Hydrological Service

(HZB), and the reclip : more project for the provision of observational and modelled data. This work was funded by the Munich Reinsurance Group and the European Union FP6 project CLAVIER (contract no. 037013, www.clavier-eu.org). The calculations and graphics were done using the open source software R (www.r-project.org) with supporting inputs from Rasmus Benestad.

References

Alexandrov VA, Hoogenboom G. 2000. The impact of climate variability and change on crop yield in Bulgaria. *Agricultural and Forest Meteorology* **104**: 315–327.

Auer I, Böhm R, Jurkovic A, Lipa W, Orlik A, Potzmann R, Schöner W, Ungersböck M, Matulla C, Briffa K, Jones P, Efthymiadis D, Brunetti M, Nanni T, Maugeri M, Mercalli L, Mestre O, Moisselin J-M, Begert M, Müller-Westermeier G, Kveton V, Bochnicek O, Stastny P, Lapin M, Szalai S, Szentimrey T, Cegnar T, Dolinar M, Gajic-Capka M, Zaninovic K, Majstorovicp Z, Nieplova E. 2007. HISTALP – historical instrumental climatological surface time series of the Greater Alpine Region. *International Journal of Climatology* **27**: 17–46.

Beersma JJ, Buishand TA. 2003. Multi-site simulation of daily precipitation and temperature conditional on the atmospheric circulation. *Climate Research* **25**: 121–133.

Benestad RE, Hanssen-Bauer I, Chen D. 2009. *Empirical Statistical Downscaling*. World Scientific Publishing Company: New Jersey, London.

Boé J, Terray L, Habets F, Martin E. 2007. Statistical and dynamical downscaling of the Seine basin climate for hydro-meteorological studies. *International Journal of Climatology* **27**: 1643–1655.

Brandsma T, Buishand TA. 1998. Simulation of extreme precipitation in the Rhine basin by nearest-neighbour resampling. *Hydrology and Earth System Sciences* **2**: 195–209.

Camuffo D. 1978. Cumulated frequency distribution of daily global solar radiation at Venice, Italy. *Archiv für Meteorologie Geophysik und Bioklimatologie* **26**: 45–50.

Cebon P, Dahinden U, Davies HC, Imboden D, Jaeger CC (eds). 1998. *Views from the Alps; Regional Perspectives on Climate Change*. The MIT Press: Cambridge.

Cubasch U, von Storch H, Waszkewitz J, Zorita E. 1996. Estimates of climate change in Southern Europe derived from dynamical climate model output. *Climate Research* **7**: 129–149.

Dehn M, Buma J. 1999. Modelling future landslide activity based on general circulation models. *Geomorphology* **30**: 175–187.

Dettinger MD, Cayan DR, Meyer MK, Jeton AE. 2004. Simulated hydrologic responses to climate variations and change in the Merced, Carson, and American river basins, Sierra Nevada, California, 1900–2099. *Climatic Change* **62**: 283–317.

Dobler A, Ahrens B. 2008. Precipitation by a regional climate model and bias correction in Europe and South Asia. *Meteorologische Zeitschrift* **17**: 499–509.

Dudhia J, Gill D, Manning K, Wang W, Bruyere C. 2005. PSU/NCAR Mesoscale Modeling System Tutorial-Class Notes and Users' Guide (MM5 Modeling System Version 3). <http://www.mmm.ucar.edu/mm5/documents/tutorial-v3-notes.html>. Accessed 2009 Aug 12.

Fernández J, Sáenz J. 2003. Improved field reconstruction with the analog method: searching the CCA space. *Climate Research* **24**: 199–213.

Feser F. 2006. Enhanced detectability of added value in limited-area model results separated into different spatial scales. *Monthly Weather Review* **134**: 2180–2190.

Fowler HJ, Blenkinsop S, Tebaldi C. 2007. Linking climate change modelling to impact studies: recent advances in downscaling techniques for hydrological modelling. *International Journal of Climatology* **27**: 1547–1578, DOI: 10.1002/joc.1556.

Fowler HJ, Kilsby CG. 2007. Using regional climate model data to simulate historical and future river flows in northwest England. *Climatic Change* **80**: 337–367.

Frei C, Schär C. 1998. A precipitation climatology of the Alps from high-resolution rain-gauge observations. *International Journal of Climatology* **18**: 873–900.

Frei C, Christensen JH, Déqué M, Jacob D, Jones RG, Vidale PL. 2003. Daily precipitation statistics in regional climate models: evaluation and intercomparison for the European Alps. *Journal of Geophysical Research* **108**(D3): 4124, DOI: 10.1029/2002JD002287.

Giorgi F, Mearns LO. 1991. Approaches to the simulation of regional climate change: a review. *Reviews of Geophysics* **29**: 191–216.

Giorgi F, Mearns LO. 1999. Introduction to special section: regional climate modelling revisited. *Journal of Geophysical Research* **104**: 6335–6352.

Gobiet A, Truhetz H, Riegler A. 2006. A climate scenario for the Alpine region, reclip : more project year 3. WegCenter progress report, Wegener Center for Climate and Global Change, University of Graz: Graz. <http://www.uni-graz.at/igam7www_agobietetal_wegcscireportfarc-reclippj3_2006.pdf>. Accessed 2009 Aug 12.

Graham LP, Andréasson J, Carlsson B. 2007. Assessing climate change impacts on hydrology from an ensemble of regional climate models, model scales and linking methods–a case study on the Lule River basin. *Climatic Change* **81**: 293–307.

Gutowski WJ, Decker SG, Donavon RA, Pan Z, Arritt RW, Takle ES. 2003. Temporal-spatial scales of observed and simulated precipitation in central U.S. climate. *Journal of Climate* **16**: 3841–3847.

Hagemann S, Machenhauer B, Jones R, Christensen OB, Déqué M, Jacob D, Vidale PL. 2004. Evaluation of water and energy budgets in regional climate models applied over Europe. *Climate Dynamics* **23**: 547–567.

Hay LE, Clark MP. 2003. Use of statistically and dynamically downscaled atmospheric model output for hydrologic simulations in three mountainous basins in the western United States. *Journal of Hydrology* **282**: 56–75.

Helsel DR, Hirsch RM. 2002. Statistical methods in water resources techniques of water resources investigations. U.S. Geological Survey. <http://pubs.usgs.gov/twri/twri4a3/html/pdf_new.html>. Accessed 2009 Aug 12.

Hsieh WW. 2009. *Machine Learning Methods in the Environmental Sciences*. Cambridge University Press: Cambridge.

Huth R. 1999. Statistical downscaling in central Europe: evaluation of methods and potential predictors. *Climate Research* **13**: 91–101.

Imbert A, Benestad RE. 2005. An improvement of analog model strategy for more reliable local climate change scenarios. *Theoretical and Applied Climatology* **82**: 245–255, DOI: 10.1007/s00704-005-0133-4.

Ines AVM, Hansen JW. 2006. Bias correction of daily GCM rainfall for crop simulation studies. *Agricultural and Forest Meteorology* **138**: 44–53.

Kidson JW, Thompson CS. 1998. A comparison of statistical and model-based downscaling techniques for estimating local climate variations. *Journal of Climate* **11**: 735–753.

Kilsby CG, Cowpertwait PSP, O'Conell PE, Jones PD. 1998. Predicting rainfall statistics in England and Wales using atmospheric circulation variables. *International Journal of Climatology* **18**: 523–539.

Lall U, Sharma A, 1996. A nearest neighbor bootstrap for resampling hydrologic time series. *Water Resources Research* **32**: 679–693.

Leander R, Buishand TA. 2007. Resampling of regional climate model output for the simulation of extreme river flows. *Journal of Hydrology* **332**: 487–496.

Loibl W, Beck A, Dorninger M, Formayer H, Gobiet A, Schöner W (eds). 2007. Kwiss-Programm reclip:more: research for climate protection:model run evaluation. Final Report, ARC-sys-0123. Austrian Research Centers–systems research, Vienna. <http://systemsresearch.arcs.ac.at/SE/projects/reclip/>. Accessed 2009 Aug 12.

Matulla C, Penlap EK, Haas P, Formayer H. 2003. Comparative analysis of spatial and seasonal variability: Austrian precipitation during the 20th century. *International Journal of Climatology* **23**: 1577–1588.

Matulla C, Zhang X, Wang XL, Wang J, Zorita E, Wagner S, von Storch H. 2008. Influence of similarity measures on the performance of the analog method for downscaling daily precipitation. *Climate Dynamics* **30**: 133–144, DOI: 10.1007/s00382-007-0277-2.

Mehrotra A, Sharma A. 2006. Conditional resampling of hydrologic time series using multiple predictor variables: a K-nearest neighbour approach. *Advances in Water Resources* **29**: 987–999.

Moron V, Robertson AW, Ward MN, Ndiaye O. 2008. Weather types and rainfall over Senegal. Part II: downscaling of GCM simulations. *Journal of Climate* **21**: 288–307.

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

Panofsky HW, Brier GW. 1968. *Some Applications of Statistics to Meteorology*. The Pennsylvania State University Press: Philadelphia.

Piani C, Haerter JO, Coppola E. 2010. Statistical bias correction for daily precipitation in regional climate models over Europe. *Theoretical and Applied Climatology* **99**: 187–192, DOI: 10.1007/s00704-009-0134-9.

Salathé E. 2003. Comparison of various precipitation downscaling methods for the simulation of streamflow in a rainshadow river basin. *International Journal of Climatology* **23**: 887–901.

Schmidli J, Frei C, Vidale PL. 2006. Downscaling from GCM precipitation: a benchmark for dynamical and statistical downscaling methods. *International Journal of Climatology* **26**: 679–689.

Schmidli J, Goodess CM, Frei C, Haylock MR, Hundecha Y, Ribalaygua J, Schmith T. 2007. Statistical and dynamical downscaling of precipitation: an evaluation and comparison of scenarios for the European Alps. *Journal of Geophysical Research* **112**: D04105, DOI: 10.1029/2005JD007026.

Schoof JT, Pryor SC. 2001. Downscaling temperature and precipitation: a comparison of regression-based methods and artificial neural networks. *International Journal of Climatology* **21**: 773–790.

Suklitsch M, Gobiet A, Leuprecht A, Frei C. 2008. High resolution sensitivity studies with the regional climate model CCLM in the Alpine Region. *Meteorologische Zeitschrift* **17**: 467–476.

Suklitsch M, Gobiet A, Truhetz H, Awan NK, Göttel H, Jacob D. 2010. Error characteristics of high resolution regional climate models over the Alpine Area. *Climate Dynamics* (in press).

Taylor KE. 2001. Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research* **106**: 7183–7192.

Uppala SM, Kållberg PW, Simmons AJ, Andrae U, Da Costa Bechtold V, Fiorino M, Gibson JK, Haseler J, Hernandez A, Kelly GA, Li X, Onogi K, Saarinen S, Sokka N, Allan RP, Andersson E, Arpe K, Balmaseda MA, Beljaars ACM, van De Berg L, Bidlot J, Bormann N, Caires S, Chevallier F, Dethof A, Dragosavac M, Fisher M, Fuentes M, Hagemann S, Hólm E, Hoskins BJ, Isaksen L, Janssen PAEM, Jenne R, Mcnally AP, Mahfouf JF, Morcrette JJ, Rayner NA, Saunders RW, Simon P, Sterl A, Trenberth KE, Untch A, Vasiljevic D, Viterbo P, Woollen J. 2005. The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society* **131**(612): 2961–3012.

von Storch H. 1999. On the use of "inflation" in statistical downscaling. *Journal of Climate* **12**: 3505–3506.

von Storch H, Navarra A (eds). 1999. *Analysis of Climate Variability: Applications of Statistical Techniques*. Springer: Berlin.

von Storch H, Zwiers FW. 1999. *Statistical Analysis in Climate Research*. Cambridge University Press: Cambridge.

Wang Y, Leung LR, McGregor JL, Lee DK, Wang WC, Ding Y, Kimura F. 2004. Regional climate modeling: progress, challenges, and prospects. *Journal of the Meteorological Society of Japan* **82**: 1599–1628.

Wetterhall F, Halldin S, Xu C. 2005. Statistical precipitation downscaling in central Sweden with the analogue method. *Journal of Hydrology* **306**: 174–190.

Widmann M, Bretherton CS, Salathé E. 2003. Statistical precipitation downscaling over the Northwestern United States using numerically simulated precipitation as predictor. *Journal of Climate* **16**: 799–816.

Wilby R. 1997. Non-stationarity in daily precipitation series: implications for GCM down-scaling using atmospheric circulation indices. *International Journal of Climatology* **17**: 439–454.

Wilby R, Wigley T. 1997. Downscaling general circulation model output: a review of methods and limitations. *Progress in Physical Geography* **21**: 530–548.

Wilby R, Wigley T. 2000. Precipitation predictors for downscaling: observed and general circulation model relationships. *International Journal of Climatology* **20**: 641–661.

Wilks DS. 1995. *Statistical Methods in Atmospheric Science, Volume 59 of International Geophysics Series*. Academic Press: San Diego, London.

Wood AW, Leung LR, Sridhar V, Lettenmaier DP. 2004. Hydrologic implications of dynamical and statistical approaches to downscale climate model outputs. *Climatic Change* **62**: 189–216.

Zorita E, von Storch H. 1999. The analog method as a simple statistical downscaling technique: comparison with more complicated methods. *Journal of Climate* **12**: 2474–2489.

