Multi-variable error correction of regional climate models

**Renate Anna Irma Wilcke · Thomas Mendlik ·
Andreas Gobiet**

Abstract Climate change impact research needs regional climate scenarios of multiple meteorological variables. Those variables are available from regional climate models (RCMs), but affected by considerable biases. We evaluate the application of an empirical-statistical error correction method, quantile mapping (QM), for a small ensemble of RCMs and six meteorological variables. Annual and monthly biases are reduced to close to zero by QM for all variables in most cases. Exceptions are found, if non-stationarity of the model's error characteristics occur. Even in the worst cases of non-stationarity, QM clearly improves the biases of raw RCMs. In addition, QM successfully adjusts the distributions of the analysed variables. To approach the question whether time series and inter-variable relationships are still plausible after correction, we evaluate the root-mean-square error (RMSE), autocorrelation and inter-variable correlation. We found improvement or no clear effect in RMSE and autocorrelation, and no clear effect on the correlation between meteorological variables. These results demonstrate that QM retains the quality of the temporal structure in time series and the inter-variable dependencies of RCMs. It has to be emphasised that this cannot be interpreted as an improvement and that deficiencies of the RCMs in those features are retained as well. Our results give some indication for the performance of QM applied to future scenarios, since our evaluation relies

R. A. I. Wilcke (✉) · T. Mendlik · A. Gobiet
Wegener Center for Climate and Global Change and Institute for Geophysics,
Astrophysics, and Meteorology, University of Graz,
Brandhofgasse 5, 8010 Graz, Austria
e-mail: renate.wilcke@uni-graz.at

on independent calibration and evaluation periods, which are affected by climate variability and change. The effect of non-stationarity, however, can be expected to be larger in far future. We demonstrate the retainment of the RCM's temporal structure and inter-variable dependencies, and large improvements in biases. This qualifies QM as a valuable, though not perfect, method in the interface between climate models and climate change impact research. Nonetheless, in case of no correlation between re-analysis driven RCM and observation, one should consider that QM does not correct this correlation.

1 Introduction

Climate change impact research needs regional climate scenarios of multiple meteorological variables. This typically includes temperature, precipitation, relative humidity, global radiation, and wind speed (e.g. Finger et al. 2012), which are the focus of this study. They are usually available from regional climate models (RCMs), but affected by considerable biases.

RCMs are common tools to regionalise general circulation model (GCM) results and currently simulate the climate on grid spacing between 50 and 12 km (e.g. ENSEMBLES project (http://ensembles-eu.metoffice.com), EURO-CORDEX (http://www.euro-cordex.net/)). For complex terrain like the Alps or Scandinavia, however, the biases of RCM results are considerable (Christensen et al. 2008; Suklitsch et al. 2011; Kjellström et al. 2010).

One way to mitigate these errors is a quantile based empirical-statistical error correction method, quantile-quantile mapping. It was introduced by Brier and Panofsky (1968) as empirical transformation and first used for downscaling and error correction by e.g. Déqué (2007) and Boé et al. (2007). Quantile based methods are getting more popular lately and have been applied to downscale and error-correct temperature and precipitation data from RCMs (e.g. Dobler and Ahrens 2008; Piani et al. 2009; Dosio and Paruolo 2011). Themeßl et al. (2011) compared various downscaling and error correction methods and showed that a quantile based method (quantile mapping; QM) performs best for daily precipitation. Räisänen and Räty (2012) compared 10 bias correction methods on mean temperature using a pseudo reality approach for far future evaluations and showed that QM performs best for most percentiles but the highest (98 %). The question, however, whether error correction degrades temporal characteristics and inter-variable dependencies is an open issue. As many climate impact models use muliple variables at the same time, not only consistent time series, but also inter-variable dependencies are of importance.

In this study we investigate the performance of QM applied to a multi-variable output of four RCMs with a grid-spacing of 25 km on daily basis. The particular focus of this study is on biases, frequency distributions, temporal structure of time series, and inter-variable dependencies. Comments on the application of QM to poorly performing RCMs can be found in the conclusions. Data and the implementation of QM are described in Sections 2 and 3. The results are discussed in Section 4, and in Section 5 we draw the conclusions.

2 Data

2.1 RCM data

Daily mean RCM data were derived from the multi-model data-set of the ENSEMBLES project. The RCMs have a horizontal grid-spacing of 25 km and cover entire Europe. Four simulations are analysed in detail: performed at C4I (Community Climate Consortium for Ireland) with the RCA3 RCM driven by the HadCM3Q16 (GCM), at the ICTP (Italian Centre for Theoretical Physics) with the REGCM3 RCM which was driven by the ECHAM5-r3 (GCM), at the METNO (Norwegian Meteorological Institute) with the HIRHAM RCM which was driven by the HadCM3Q0 (GCM), and at the SMHI (Sweden's Meteorological and Hydrological Institute) with the RCA RCM driven with the BCM (GCM), all for the SRES A1B scenario. Compared to the other ENSEMBLES simulations C4I-RCA3 shows a strong warming and wetter conditions in our study region, ICTP-REGCM3 shows little warming and drier conditions, METNO-HIRHAM shows moderate warming and moderate change in precipitation, and SMHI-RCA shows little warming and wetter conditions in the future (Wilcke et al. 2012). For the evaluation of temporal dependencies the same RCMs driven by re-analysis data (ERA40; Uppala et al. 2005) are analysed.

2.2 Observational data

Daily mean observational data were obtained from about 80 stations of the Austrian Central Institute for Meteorology and Geodynamics (ZAMG) and 18 stations of the Swiss Federal Office of Meteorology and Climatology (MeteoSwiss). The ZAMG data covers entire Austria for the period 1971 to 2010. The selected Swiss stations in the Rhone catchment covers the period 1981 to 2010. To ensure a sufficiently large sample size only stations with more than 80 % (compromise from experiences with missing values) data coverage are included.

Observations are error-prone (e.g. Auer et al. 2001; Frei and Schär 1998; Schmidli et al. 2002; Caussinus and Mestre 2004; Della-Marta and Wanner 2006) and particularly wind induces errors in precipitation measurements (e.g. Nešpor and Sevruk 1999; Frei et al. 2003). Consequently this influences model evaluation that strongly depends on the quality of the observations. Here, however, we neglect observational errors and the reader should keep in mind this limitation.

Four stations with different climatological characteristics were analysed in depth. Hohe Warte (lon 16.35638889 E, lat 48.24861111 N) in Vienna is located in a flat region between the north-eastern deviating veins of the Alps, in the north-western part of the Vienna basin. Sonnblick is a meteorological observatory on top of the mountain Hoher Sonnblick (lon 12.9575 E, lat 47.05416667 N) at 3,105 m above sea level and is exposed to the free atmosphere. Innsbruck University (lon 11.38416667 E, lat 47.26 N) maintains an observation station at 578 m height that represents a valley (Inn valley) with open ends on both sides. The valley lies in southwest-northeast direction and has a width of about 7 km. The station of Zermatt (lon 7.752468 E, lat 46.029282 N) at 1,608 m height lies at the inner end of the Matter valley, which extends in south-north direction and has a width of about 1.5 km.

3 Method

3.1 Error correction

Themeßl et al. (2011) proposed a QM method to correct RCM simulations based on Déqué (2007). Themeßl et al. (2012) demonstrated its successful application to future scenarios for precipitation. Such methods are often used in numerical weather prediction and belong to deterministic downscaling methods in the family of model output statistics (MOS) (Themeßl et al. 2011; Maraun et al. 2010). The process combines downscaling aspects with model error correction ("bias correction"). The correction of the altitude differences between the model and actual orography is implicitly included. The QM method used in our study is purely empirical (i.e. no assumption about the distributions of the meteorological variables are made) and is based on Themeßl et al. (2012). Adaptation to the specific requirements of different variables are described in this section.

Our implementation of QM fits modelled daily empirical cumulative distribution functions (ECDFs) to corresponding observational ECDFs of one station. For each day of year (DOY) in the calibration period, ECDFs are constructed using a sliding window of 31 days, which results in, e.g., 620 values for 20 years of calibration. This sensibilise the correction to varying error characteristics throughout the year. The ECDFs are calculated by sorting the values into bins with adjustable width. The bin width is set to the resolution of the observational data (mostly 0.1) and a linear interpolation is applied between two percentiles (bins) (Déqué 2007). In some cases, particularly in the case of low-quality observational data, this interpolation can lead to inadequate approximations, as will be demonstrated in Sections 4.2 and 4.4. In order to avoid the suppression of new extremes in the future periods (i.e. values outside the calibration range), our implementation extrapolates the correction by keeping the correction term of minimum and maximum values constant outside the observational range.

We want to emphasise that we calculate daily ECDFs and apply them on daily data, thus there is no temporal scale discrepancy. Spatially, we minimise the discrepancy by using RCMs instead of GCMs.

For the correction of temperature, which is described here exemplarily, the cumulative probability p for a modelled value is calculated at point i and day t (Eq. 1). The correction term $\Delta x_{t,i}$ is calculated as a difference of the inverse ECDFs (quantile) of the observation ($\text{ECDF}^{\text{obs,cal}^{-1}}$) and model ($\text{ECDF}^{\text{mod,cal}^{-1}}$) for probability p at a certain day of a station (Eq. 2). The ECDFs are created for a DOY in a 31-day moving window for the calibration period.

$$p_{t,i} = \text{ECDF}_{\text{DOY},i}^{\text{mod,cal}} \left(x_{t,i}^{\text{mod,raw}} \right) \tag{1}$$

$$\Delta x_{t,i} = \text{ECDF}_{\text{DOY},i}^{\text{obs,cal}^{-1}} (p_{t,i}) - \text{ECDF}_{\text{DOY},i}^{\text{mod,cal}^{-1}} (p_{t,i}) \tag{2}$$

$$x_{t,i}^{\text{mod,cor}} = x_{t,i}^{\text{mod,raw}} + \Delta x_{t,i} \tag{3}$$

The correction term $\Delta x_{t,i}$ is then added to the raw model value $x_{t,i}^{\text{mod,raw}}$ (Eq. 3).

Dealing with precipitation, a frequency adaptation is implemented to parry a deficiency of QM leading to a wet bias. That bias occurs if the dry day frequency in the raw model output ($ECDF^{mod,cal}$) is larger than in the observations ($ECDF^{obs,cal}$), which would lead to a strong positive bias after the correction (Themeßl et al. 2012). Thus, the model data below 0.1 mm/d is divided to finer bins with width of 0.01. Dry days are generated by randomly sampling the observational distribution into the first bin (0–0.01 mm/d). However, this bias is a rare case. More often the model overestimates the light precipitation frequency ("drizzling-effect"; e.g. Gutowski et al. 2003), which is caught by QM automatically.

The correction of relative humidity required minor adaptations, since the interpolation can lead to values outside the physically reasonable range, in particular to values above 100 %. These non-physical values were set to the maximum value observed on the corresponding ECDF. This adaptation, however, only takes effect in very rare cases and the overall effect can hardly be noticed in climatological evaluation.

For wind speed, global radiation, and surface pressure QM works straightforward, as it has been implemented for temperature.

3.2 Evaluation

Our evaluation focuses on biases, frequency distributions, temporal structure of time series, and inter-variable dependencies in present and future climate. We use a split sample evaluation approach to mimic the application to future climate as far as possible. The calibration period has no overlap with the evaluation period. The available observation period from 1971 to 2010 is divided in halves, taking 1971 to 1990 as calibration period and 1991 to 2010 as application period, and vice versa. Of course, the split sample approach can only give rough indication about the performance of QM in far future periods. However, since calibration and evaluation periods are independent and climate variability and change results in different climate characteristics in both periods, severe deficiencies can be expected to be detected. Wherever possible, i.e. in the analysis of the inter-variable evaluation, we additionally regard future periods.

In addition to the split sample approach described above, evaluation was done also with equal calibration and application period, which is further on denoted as technical evaluation. This allows judging the performance of QM in an idealised world by neglecting climate variability and climate change. To keep the comparability with the split sample approach the technical test was performed on 20-year periods.

The evaluation is designed with two sets of RCMs: GCM driven and re-analysis driven (perfect boundary). Most climate impact studies are interested in future climate, which requires GCM driven RCMs. Re-analysis driven RCMs are expected to simulate the past climate correctly, including the temporal structure. Those were used to investigate the effect of QM on the temporal structure of RCMs.

For the evaluation of the GCM-driven RCM simulations three basic statistics were inspected: the bias, the density distribution, and the inter-variable correlation. The bias and density distribution describe the performance of QM on single variables and are analysed on the past only, since observations are indispensable. The inter-variable correlation—where we mainly focus on comparing raw and corrected RCM results—is analysed additionaly in near- and far-future periods. For the re-analysis

driven RCM simulations, the temporal structure of time-series was analysed by the root-means-square error (RMSE) and autocorrelation.

4 Results and discussion

4.1 Bias

The bias is defined as long term average difference between model and observation. For the ICTP-REGCM3 model the biases of temperature, precipitation, relative humidity, and wind speed are shown in Figs. 1 and 2, exemplarily. Figure 1 shows the annual mean bias of the raw and split sample corrected model for stations in Austria. The biases for the other three models are presented in Figs. S1, S2, and S3 (the "S" indicates supplementary material).

Figure 2 shows the annual cycle of monthly biases. In this case, the spatial error variability is indicated by box-and-whisker plots. This analysis shows both versions of the split sample approach, one evaluated in the period 1991 to 2010 and calibrated in the period 1971 to 1990 (blue), the other one vice-versa (green), together with the raw RCM bias of each period (red and orange). The results for the other models are presented in Figs. S4, S5, and S6.

4.1.1 Temperature

QM performs very well in removing the annual mean temperature bias from the ICTP-REGCM3 model. In the split sample evaluation (Fig. 1a), the mean bias over all stations is reduced from $-1.5\,^{\circ}$C to $-0.3\,^{\circ}$C and the biases of the individual stations range to below $1.1\,^{\circ}$C at maximum, compared to $3.5\,^{\circ}$C before correction. The results for the other models confirm these results.

The monthly temperature biases of the ICTP-REGCM3 model are shown in Fig. 2a. The biases and their spatial variability are generally strongly reduced. In some months, however, considerable errors remain after the correction (e.g., in summer), which is caused by different model error characteristics in the calibration and the application period (i.e. by non-stationarity). Similar analysis for other models (c.f. Figs. S4–S6) show that ICTP-REGCM3 is rather extreme in this respect and that bias correction of other models is mostly less affected by non-stationarity. Non-stationarity, however, is a limitation of bias correction (and empirical-statistical methods in general) and narrows their application to periods not too far in the future. Maraun (2012) investigated non-stationarities of a very simple bias correction method on a seasonal time scale and found that at the end of the 21st century bias correction only partly improves RCM results. Here we demonstrate that the improvement in case of non-stationarity is still large for near future (20 years) applications, particularly when averaged over the year.

4.1.2 Precipitation

The ICTP-REGCM3 bias of precipitation is strongly improved in the split sample evaluation with a bias reduction from $+0.8$ mm/d to -0.1 mm/d (Fig. 1b). The results for other RCMs are very similar. On the monthly timescale (Fig. 2b), the
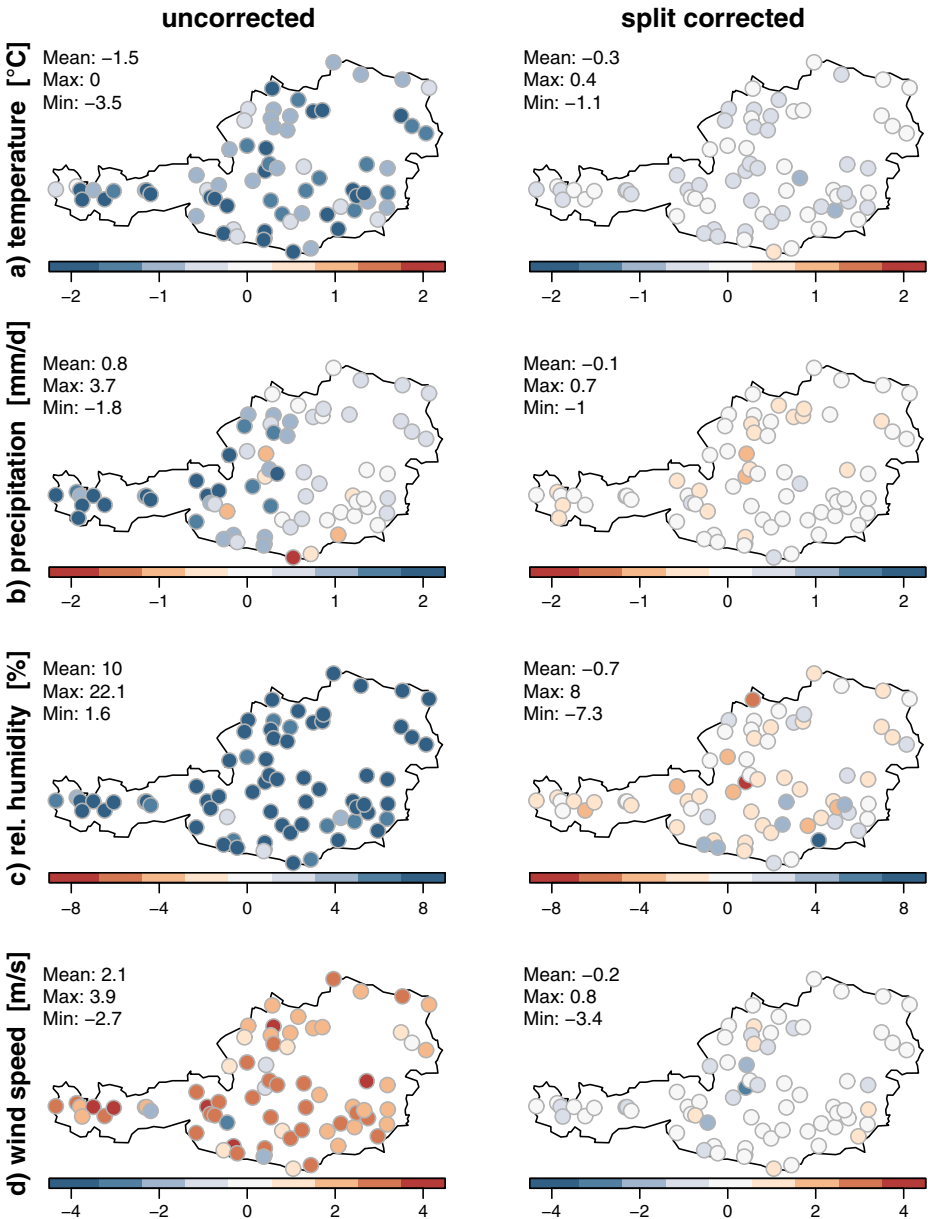
uncorrected   **split corrected**

Fig. 1 Annual mean RCM (**ICTP-REGCM3**) bias at observation stations (1991–2010) for **a** temperature, **b** precipitation, **c** relative humidity, and **d** wind speed (*top down*) for the raw RCM (for temperature altitude corrected), the error-corrected RCM with split sample evaluation

split sample evaluation generally results in smaller biases than the raw model and in a smaller bias range. The influence of non-stationarity is smaller than for temperature in this case. Bias correction of other models yields very similar results (Figs. S1 to S6).
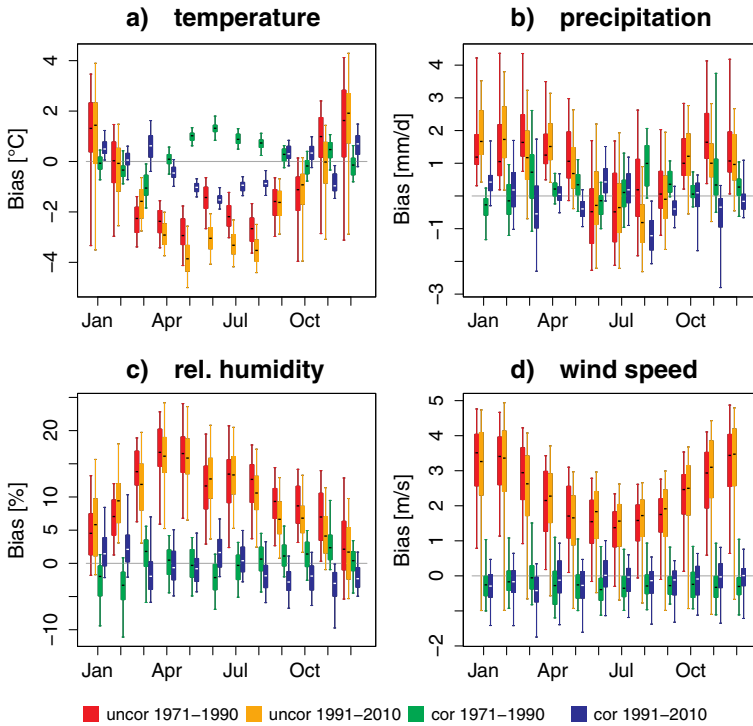
Fig. 2 Monthly bias of **a** temperature, **b** precipitation, **c** relative humidity, and **d** wind speed as box-whisker plots for the uncorrected RCM (for temperature altitude corrected) split in two periods (*red* and *orange*), the error-corrected RCM (**ICTP-REGCM3**) with split sample evaluation (*green* and *blue*, see Section 4.1 for detailed description). *Box* and *whiskers* indicate the (spatial) variability of errors at the different Austrian stations. *Boxes* indicate the first (q25) and the third (q75) quantile, the *whiskers* extend to q5 and q95, and the *black horizontal line* indicates the median

4.1.3 Relative humidity

With regard to relative humidity the mean bias of ICTP-REGCM3 is reduced to close zero; the maximum of biases over the stations is reduced from 22 % to 8 % in the split sample approach (Fig. 1c). Figure 2c reflects similar results on the monthly scale. The median biases in both periods are close to zero in all months and the error range over the stations is smaller than for the raw RCM. The analysis of the other RCMs confirm these results, with some additional indication for non-stationarity in the SMHI-RCA model.

4.1.4 Wind speed

The error correction of wind speed leads to a reduction of the mean bias from 2.1 m/s to −0.2 m/s in the split sample approach (Fig. 1d). The range of biases over the stations is only reduced from 6.6 m/s to 4.2 m/s. However, this small reduction is caused only by one single station. The mean bias of 0.2 m/s also is visible in the monthly bias of the split sample analysis (Fig. 2d) and the technical evaluation (not shown). A similarly small remaining bias can also be found in the other three RCMs. These

remaining biases are mainly caused by lacking quality of the observational data: As mentioned in Sections 2.2 and 3.1, an interpolation error occurs if the resolution of the observational data is much lower than the resolution of the RCM. Wind measurements sometimes have gaps in the data distribution of up to 0.6 m/s, already at small velocities. The vertical stripes in the scatter-plots of Fig. 6a (see Section 4.4) for wind speed demonstrates that. For testing, random noise was added to fill those gaps, which results in a mean bias of zero. Nevertheless, also with a remaining bias of 0.2 m/s after correction, QM strongly improves raw RCM output.

4.1.5 Surface air pressure

The main bias of surface air pressure results from the different altitudes of the model grid and the station. This effect is fully corrected by QM, and results in a mean bias of zero for the technical as well as for split sample evaluation, for the entire year as well as for single months (not shown).

4.2 Density distribution

The correction of density distributions from all four RCMs is presented for the period from 1991–2010 (calibration: 1971–1990) on seasonal scale (Fig. 3 for summer), and the entire year in Fig. S7. Four selected stations are evaluated (see Section 2.2). We compare the density distributions of observation (black fat curve) with raw (red thin lines) and error-corrected (green thin lines) RCMs for temperature, precipitation, relative humidity, wind speed, and global radiation.

Figure 3 demonstrates that the distributions of all models and variables are nicely adjusted to the observed distribution. Various distortions of the temperature distributions are corrected and the over-pronounced frequency of light and medium precipitation ("drizzling effect") is adjusted. More details on the performance of QM for daily precipitation including extremes are discussed in Themeßl et al. (2012). With regard to relative humidity, the overestimated frequency of near-100 % values and with regard to wind speed and global radiation, underestimated frequency of higher values are corrected.

Such adjustment of the distributions would be trivial if the calibration and evaluation periods were the same, but is quite remarkable in the split sample analysis with two independent periods. In particular, some variables (e.g., relative humidity) and models (e.g., ICTP-REGCM3 for wind speed) feature distributions that are very different from the observation. Such strong modification of the distribution by error correction raises the question, whether the corresponding time series and inter-variable relationships are still plausible after correction. These issues are analysed in the following Sections 4.3 and 4.4.

4.3 Temporal structure

In order to analyse a potential distortion of the RCM's temporal structure by QM, we regard the autocorrelation and the RMSE of corrected and uncorrected time series from re-analysis driven RCM simulations. In Fig. 4 the RMSE for the four RCMs averaged over the period 1981–2000 (calibration period 1961–1980) are displayed.
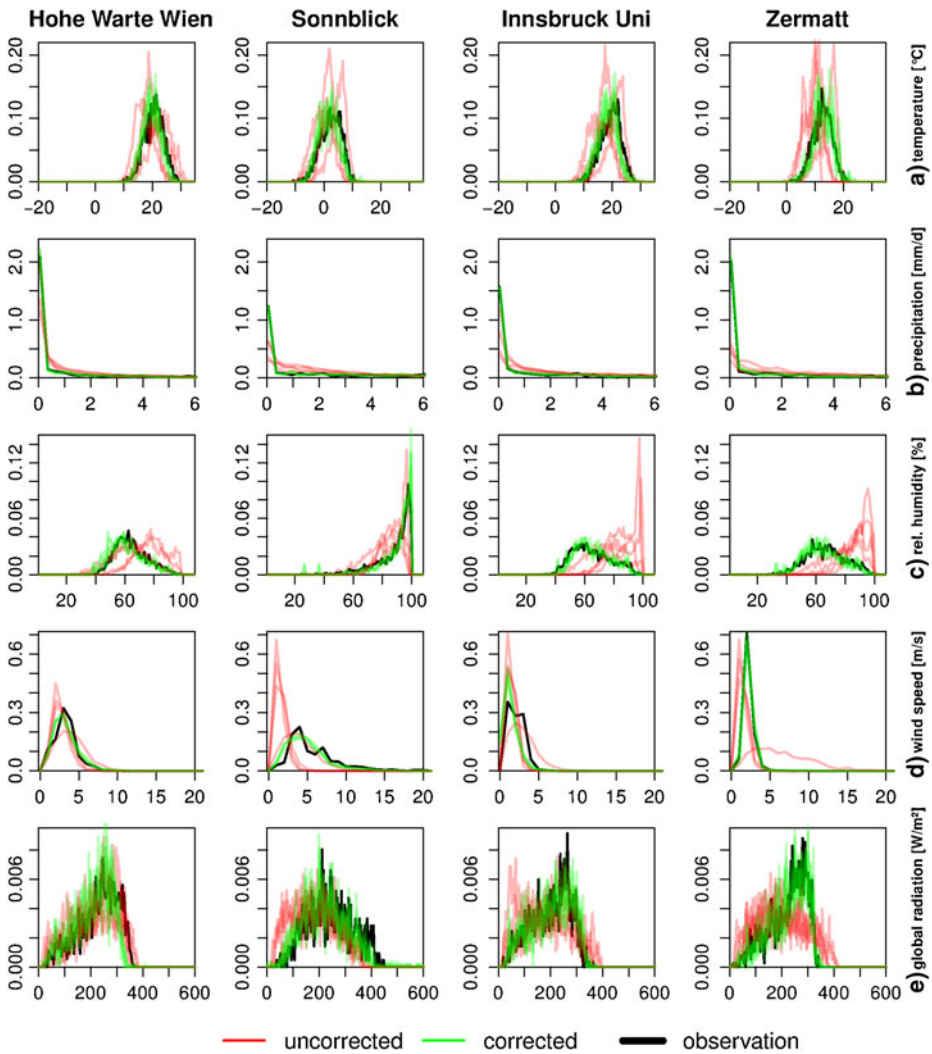
Fig. 3 Density distributions of the 4 raw RCMs (*red thin lines*), the split sampled error-corrected RCMs (*green thin lines*), and the observations (*black fat line*) for summers (JJA) for the period 1991–2010 of Hohe Warte Vienna, Sonnblick, Innsbruck University, Zermatt for **a** temperature, **b** precipitation, **c** relative humidity, **d** wind speed, and **e** global radiation

The box-and-whisker plots show spatial variability. The RMSE of temperature is generally improved by QM, with stronger improvement for models with larger RMSE (e.g., C4I-RCA3-ERA40). For precipitation, QM has no clear effect on the RMSE. For relative humidity the RMSE is improving, comparable to temperature. For wind speed, the RMSE is only clearly affected for the model with the largest RMSE and the worst distribution (ICTP-REGCM3-ERA40). Improvements in the RMSE are related to the correction of strong biases and shifts in distributions.

Figure 5 shows the autocorrelation of precipitation of the observation (black), the raw (red), and corrected (green) ICTP-REGCM3-ERA40 model for lags of up to
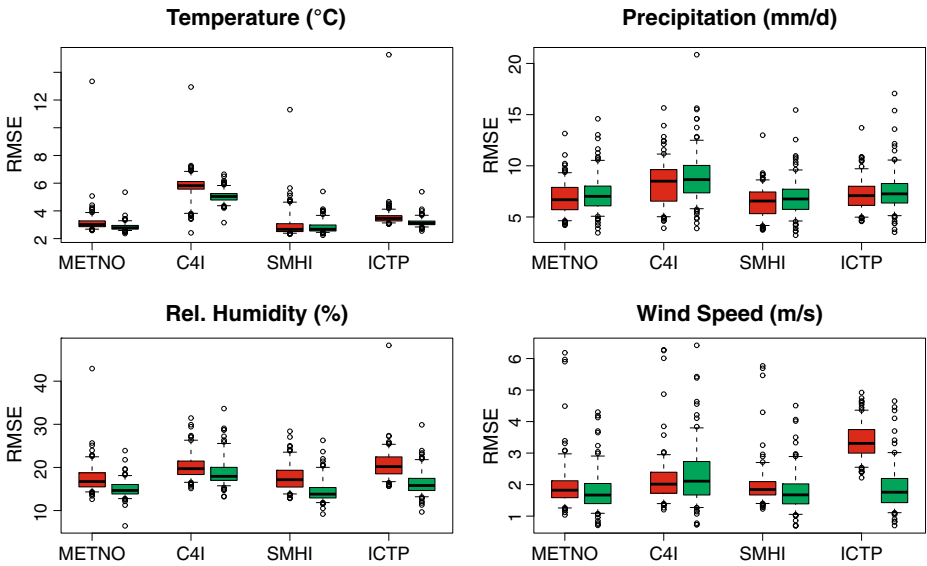
Fig. 4 RMSE of ERA40 driven RCMs of raw model (*red*) and split sample corrected (val: 1981–2000) (*green*) for whole years. Boxes contain RMSE of about 60 stations in Austria. *Box* and *whiskers* indicate the (spatial) variability of errors at the different Austrian stations. *Boxes* indicate the first (q25) and the third (q75) quantile, the *whiskers* extend to q5 and q95, and the *black horizontal line* indicates the median

6 days (see Figs. S8 to S11 for the other variables and models). Autocorrelation of the precipitation time-series is predominantly visible at lag-1 (around 0.3) with very small values after that. The RCMs generally feature larger autocorrelation than the observation, with slightly lower values of the corrected than the raw
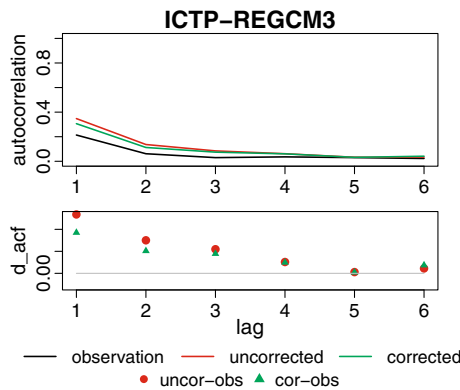

Fig. 5 *Upper panel* shows autocorrelation for 20-year period (validation period 1981–2000, calibration period 1961–1980) of precipitation for observations (*black*), raw ICTP-REGCM3-ERA40 (*red*), and split sample corrected ICTP-REGCM3-ERA40 (*green*). The difference in autocorrelations of raw (*red*) and corrected (*green*) RCM to observations is shown in the lower panels. Here the *grey line* represents the observation (zero)

model. To emphasise the day-to-day structure of temperature we removed the seasonal correlation by removing the annual cycle (Fig. S8). All four models catch this autocorrelation well, which is not seriously disturbed by error correction. The differences in autocorrelation coefficients (lower panels) between corrected and uncorrected RCM are very small (about 0.01 to 0.15). With regard to relative humidity (Fig. S10), the autocorrelation is partly over- and partly underestimated by the RCMs, depending on the model. For wind speed, like for precipitation, the inter-daily dependency is weak (Fig. S11). However, the RCMs show a stronger autocorrelation than the observations do, and the corrected RCMs are always closer to the observation than the raw ones.

In summary, we found improvement or no change in RMSE and autocorrelation due to error correction. We want to emphasise, that this improvement cannot be interpreted as an improvement of the temporal structure of the time series in a strict sense, but is rather caused by correction of intensity and distribution. An actual improvement of the temporal structure is out of the scope of the presented error correction method. Our results mainly demonstrate, that QM conserves the temporal structure of RCM time series, including their strengths and weaknesses.

4.4 Inter-variable relation

QM acts on each variable separately, so concern exists, whether inter-variable dependencies are distorted by QM. Our focus here is not on the physical consistency of raw RCMs, but on changes of inter-variable relations RCMs due to QM.

The correlations of five variables (omitting pressure here) are analysed pairwise before and after correction. As additional information, the observed correlation is shown. Figure 6 illustrates scatter-plots and correlation coefficients of each pair of variables on the four selected stations. The correlations are discussed exemplarily for the ICTP-REGCM3 model in summer. The other three models show the same results (not shown). Results for further seasons are shown in Figs. S12 to S26.

Since most variables are not linearly correlated, we choose the Spearman rank correlation coefficient, which is based on the ranks and not on linear relation like the Pearson correlation coefficient (Wilks 1995). QM is a transformation that conserves ranks. This, however, is only valid for a specific DOY in our implementation of QM, since for each day of the year, single ECDFs are created. The Spearman coefficient regards ranks of the entire time series under consideration which can be modified by QM indeed.

The coefficients are shown as pie-charts in Fig. 6 and additionally in Table S1. The correlation coefficients are calculated for period 1971–2010 (1981–2010 for Zermatt). For the historical analysis, the technical approach (c.f. Section 3.2, calibration period equals evaluation period) has been used. We do not focus on the comparison with observations, but on the inter-comparison of the raw and corrected RCMs. In addition, a similar analysis for future scenarios is shown in Fig. 7 for near (2021–2050) and a further future (2069–2098) for the station of Innsbruck (for other stations see Figs. S12 to S26).

Comparing the scatter-plots and the Spearman coefficients of the raw and error-corrected model, no significant differences are visible (Fig. 6b and c). Table S1 underlines this for the station of Zermatt. The correlation of temperature with global radiation shows small differences, same as precipitation with wind speed,
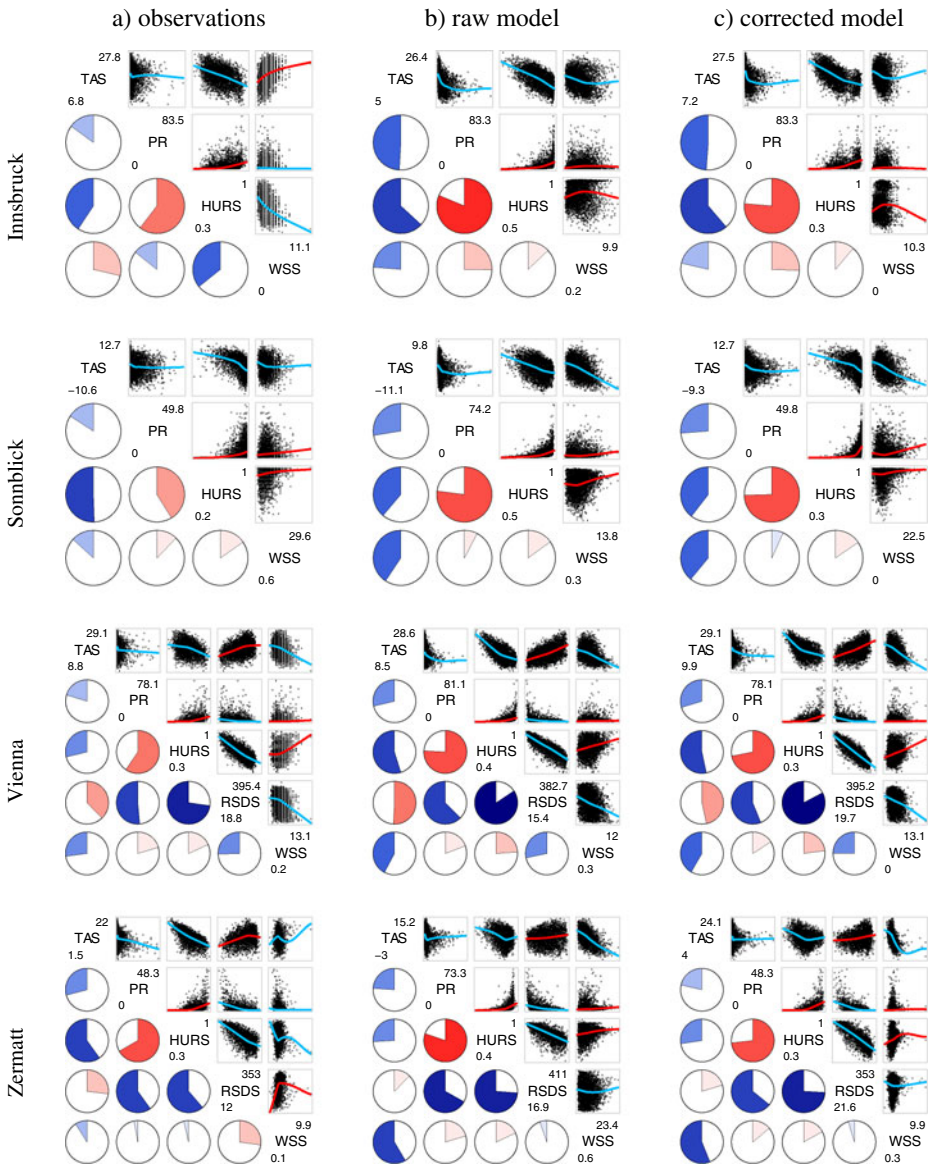
Fig. 6 Correlation matrices for the period 1971–2010 (Zermatt 1981–2010) including temperature (tas), precipitation (pr), relative humidity (hurs), global radiation (rsds), and wind speed (wss) for selected stations in Austria and Switzerland for **a** observed, **b** modelled, **c** error-corrected modelled data in summer (JJA). Pie charts show Spearman correlation coefficients, indicated with counterclockwise (negative correlation) and clockwise (positive correlation) pie slices. *Lines* in scatter plots are the Loess fit. The values above and below the variable names give the range of the data. The model shown here is the **ICTP-REGCM3**

and precipitation with relative humidity. This counts for the other seasons as well. Nevertheless, no systematic degradation of the RCM's correlation by applying QM can be detected. Differences in the scatter-plots are related to the mapping of the
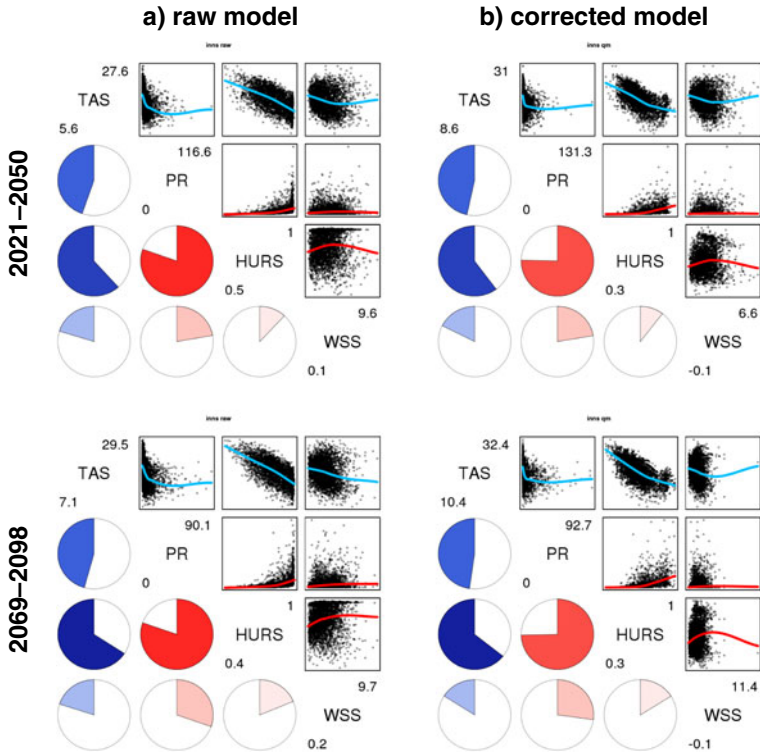
Fig. 7 Correlation matrices for **Innsbruck** for summer (JJA) of the periods 2021–2050 and 2069–2098. Same as Fig. 6

value range towards the observations. This compresses or stretches the scatter-plots without changing the correlation itself.

Comparing the correlations of observation (Fig. 6a) with those of raw RCMs, considerable differences are visible. E.g., the observed correlations of wind speed and relative humidity for Innsbruck and Sonnblick show opposite signs and different shapes of scatter-plots. The correlation coefficient of wind speed with temperature is much higher in the RCM than in the observation. Those differences can be caused by model parametrisation, further model deficiencies, observational errors, or local effects which RCMs cannot capture due scale discrepancies. QM does not correct such effects, it rather conserves the inter-variable correlations of the RCMs.

For future periods, as for past periods, the correlation given by the RCM is not systematically changed by QM.

5 Conclusions

We evaluate the application of an empirical-statistical error correction method, quantile mapping (QM), for a small ensemble of RCMs and six meteorological variables: temperature, precipitation, relative humidity, wind speed, global radiation, and surface air pressure. The evaluation includes biases and measures for temporal

and inter-variable consistency and is based on a split sample approach with strictly independent calibration and evaluation periods.

Annual and monthly biases are reduced by QM to close to zero for all variables in most cases. Exceptions are found, when non-stationarities of the model's error characteristics occur. Those non-stationarities are not restricted to highly variable variables like precipitation and one particularly prominent case is found for temperature. Even in the worst cases of non-stationarity, QM still clearly improves the biases of the raw RCM. We use independent calibration and evaluation periods, which are affected by climate variability and change. Thus, these results give some indication for the performance of QM applied to future scenarios. However, the effect of non-stationarity can be expected to be larger in far future, which limits the scope of our results. Maraun (2012) demonstrated that for periods at the end of the 21st century, a simple bias correction method only partially improves the raw RCM results.

Our purely empirical implementation of QM successfully corrects variables with very different density distributions, which makes it highly flexible and applicable to various meteorological variables and regions. The drawback is the necessity to interpolate between values of the empirical cumulative distribution function (ECDF), which leads to small systematic errors in some cases with low-quality observational data (in our case wind speed). For the proper representation of new extremes, an extrapolation of the ECDF outside the calibration range is necessary, which was not the focus of our study. Themeßl et al. (2012) found that a simple constant extrapolation leads to satisfying results also for precipitation extremes that are outside of the calibration range. One might circumvent these issues by fitting theoretical distributions or some functions to the data (e.g. Piani et al. 2009; Rojas et al. 2011), but this would lead to less flexibility and the need for a specific implementation for each variable and probably also for each climate regime.

We find considerable differences between the distributions of the uncorrected RCMs and observations in some variables (e.g., relative humidity) and models (e.g., ICTP-REGCM3 for wind speed). QM successfully adjusts also these distributions. Such strong modifications raise the question whether the time series and inter-variable relationships are still plausible after correction. This question is examined by analysing the autocorrelation and root-mean-square error (RMSE) of raw and corrected hindcast simulations and the inter-variable correlations of historical and future simulations. When applying QM to RCM output, we found improvement or no clear effect in RMSE and autocorrelation, and no clear effect in the correlation between meteorological variables.

These results demonstrate that QM retains the quality of the temporal structure of the time series and the inter-variable dependencies of RCMs. We emphasise that this is not an improvement and that deficiencies of the RCMs in those features are retained as well. A similar situation arises regarding fine-scale spatial variability (which was outside the scope of our study). Maraun (2013) demonstrated that spatial and temporal variability show considerable deficiencies after applying QM compared to observations. Those limitations are important to be aware of for the application of error-corrected model results in climate change impact studies.

QM can, by construction, map any distribution onto an arbitrary other distribution. This, however, does not necessarily indicate that the mapping is sensible in a physical way. The overall assumption for error-correcting RCMs is that RCMs represents the regional climate in a physical correct way over space and time. The open discussion if this assumption is justified is not a topic in this article. It is

known that the performance of RCMs depends on region, season and meteorological variable (e.g. Christensen et al. 2008). Nonetheless, in case of no correlation between re-analysis driven RCM and observation (e.g. Fig. S27), one should re-consider or at least be aware of it when applying QM to this RCM (Widmann et al. 2003; Eden et al. 2012). If simulation and observation are not correlated, there is also no confidence about possible future trends in the observations. Further studies will investigate the consistencies in RCMs regarding inter-variable relations and correlations.

However, the retainment of the RCM's temporal structure and inter-variable-dependencies together with large improvements with regard to biases qualifies QM as a valuable, though not perfect, method in the interface between climate models and climate change impact research.

Future improvements of QM with regard to multi-variable error correction can be particularly expected from multi-variate approaches, which might lead to improved inter-variable dependencies. However, such approaches are not straightforward to implement due limitations in the sample sizes usually available to build or estimate the distributions. In addition, more sophisticated inter- and extrapolation techniques could mitigate the effect of low-quality observational data and improve the representation of extremes. Particularly promising are also stochastic approaches, which could be implemented as add-on to QM and could lead to improvements with regard to spatial and temporal variability.

References

Auer I, Böhm R, Schöner W (2001) Austrian long-term climate 1767–2000. Zentralanst. für Meteorologie u. Geodynamik

Boé J, Terray L, Habets F, Martin E (2007) Statistical and dynamical downscaling of the seine basin climate for hydro-meteorological studies. Int J Climatol 27(12):1643–1655

Brier W, Panofsky H (1968) Some applications of statistics to meteorology. Mineral Industries Extension Services, School of Mineral Industries, Pennsylvania State College

Caussinus H, Mestre O (2004) Detection and correction of artificial shifts in climate series. J R Stat Soc Ser C Appl Stat 53(3):405–425

Christensen JH, Boberg F, Christensen OB, Lucas-Picher P (2008) On the need for bias correction of regional climate change projections of temperature and precipitation. Geophys Res Lett 35:L20709. doi:10.1029/2008GL035694

Della-Marta PM, Wanner H (2006) A method of homogenizing the extremes and mean of daily temperature measurements. J Clim 19(17):4179–4197

Déqué M (2007) Frequency of precipitation and temperature extremes over france in an anthropogenic scenario: model results and statistical correction according to observed values. Glob Planet Change 57(1–2):16–26

Dobler A, Ahrens B (2008) Precipitation by a regional climate model and bias correction in Europe and South Asia. Meteorol Z 17(4):499–509

Dosio A, Paruolo P (2011) Bias correction of the ensembles high-resolution climate change projections for use by impact models: evaluation on the present climate. J Geophys Res 116(D16):1–22

Eden J, Widmann M, Grawe D, Rast S (2012) Skill, correction, and downscaling of gcm-simulated precipitation. J Clim 25(11):3970–3984

Finger D, Heinrich G, Gobiet A, Bauder A (2012) Assessing future water resources and its uncertainty in a glaciated alpine catchment and its subsequent effects on hydropower operations during the 21st century. Water Resour Res 48:20. doi:10.1029/2011WR010733

Frei C, Schär C (1998) A precipitation climatology of the alps from high-resolution rain-gauge observations. Int J Climatol 18(8):873–900

Frei C, Christensen H Jens, Déqué M, Jacob D, Jones G Richard, Vidale L Pier (2003) Daily precipitation statistics in regional climate models: evaluation and intercomparison for the european alps. J Geophys Res 108(D3):1–19

Gutowski WJ Jr, Decker SG, Donavon RA, Pan Z, Arritt RW, Takle ES (2003) Temporal spatial scales of observed and simulated precipitation in central US climate. J Clim 16:3841–3847

Kjellström E, Boberg F, Castro M, Christensen JH, Nikulin G, Sánchez E (2010) Daily and monthly temperature and precipitation statistics as performance indicators for regional climate models. Clim Res 44(2–3):135–150

Maraun D (2012) Nonstationarities of regional climate model biases in European seasonal mean temperature and precipitation sums. Geophys Res Lett 39:L06706. doi:10.1029/2012GL051210

Maraun D (2013) Bias correction, quantile mapping and downscaling. Revisting the inflation issue. J Clim 26:2137–2143

Maraun D, Wetterhall F, Ireson AM, Chandler R, Kendon E, Widmann M, Brienen S, Rust H, Sauter T, Themeßl MJ et al (2010) Precipitation downscaling under climate change: recent developments to bridge the gap between dynamical models and the end user. Rev Geophys 48(3):1–34

Nešpor V, Sevruk B (1999) Estimation of wind-induced error of rainfall gauge measurements using a numerical simulation. J Atmos Oceanic Technol 16(4):450–464

Piani C, Haerter JO, Coppola E (2009) Statistical bias correction for daily precipitation in regional climate models over europe. Theor Appl Climatol 99(1–2):187–192

Räisänen J, Räty O (2012) Projections of daily mean temperature variability in the future: cross-validation tests with ensembles regional climate simulations. Clim Dyn 1–16. doi:10.1007/s00382-012-1515-9

Rojas R, Feyen L, Dosio A, Bavera D (2011) Improving pan-european hydrological simulation of extreme events through statistical bias correction of rcm-driven climate simulations. Hydrol Earth Syst Sci 15(8):2599–2620

Schmidli J, Schmutz C, Frei C, Wanner H, Schär C (2002) Mesoscale precipitation variability in the region of the european alps during the 20th century. Int J Climatol 22(9):1049–1074

Suklitsch M, Gobiet A, Truhetz H, Awan N, Gttel H, Jacob D (2011) Error characteristics of high resolution regional climate models over the alpine area. Clim Dyn 37:377–390

Themeßl MJ, Gobiet A, Leuprecht A (2011) Empirical-statistical downscaling and error correction of daily precipitation from regional climate models. Int J Climatol 31(10):1530–1544

Themeßl MJ, Gobiet A, Heinrich G (2012) Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal. Clim Change 112(2):449–468

Uppala SM, Kållberg PW, Simmons AJ, Andrae U, Bechtold VDC, Fiorino M, Gibson JK, Haseler J, Hernandez A, Kelly GA, Li X, Onoi K, Saarinen S, Sokka N, Allan RP, Andersson E, Arpe K, Balmaseda MA, Beljaars ACM, Berg LVD, 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. Q J R Meteorolog Soc 131(612):2961–3012

Widmann M, Bretherton C, Salathé Jr E (2003) Statistical precipitation downscaling over the northwestern United States using numerically simulated precipitation as a predictor. J Clim 16(5):799–816

Wilcke RAI, Leuprecht A, Andreas G (2012) Effects of climate change on future snow conditions in tyrol and styria (cc-snow), final report – climate. http://www.uni-graz.at/igam7www-wcv-scirep-no48-wilcke-etal-2012_compressed.pdf. Wegener Center Verlag, Graz, Austria

Wilks DS (1995) Statistical methods in the atmospheric sciences, vol 91. Academic, New York