Bias Correction, Quantile Mapping, and Downscaling: Revisiting the Inflation Issue

DOUGLAS MARAUN

GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany

ABSTRACT

Quantile mapping is routinely applied to correct biases of regional climate model simulations compared to observational data. If the observations are of similar resolution as the regional climate model, quantile mapping is a feasible approach. However, if the observations are of much higher resolution, quantile mapping also attempts to bridge this scale mismatch. Here, it is shown for daily precipitation that such quantile mapping–based downscaling is not feasible but introduces similar problems as inflation of perfect prognosis (''prog'') downscaling: the spatial and temporal structure of the corrected time series is misrepresented, the drizzle effect for area means is overcorrected, area-mean extremes are overestimated, and trends are affected. To overcome these problems, stochastic bias correction is required.

1. Introduction

In the context of perfect prognosis (''prog''; PP) statistical downscaling, von Storch (1999) pointed out that the use of variance inflation or related approaches is not meaningful. PP approaches assume a relationship between large-scale predictors and local-scale predictands. As not all small-scale variability is explained by the large-scale predictors, the prediction of the local variable in general has lower variance than the observed local variable. Inflation aims to overcome this mismatch by scaling the predicted time series to match the observed variance. The fundamental misconception here is that inflation does not add any unexplained variability and therefore wrongly assumes that all local variance is indeed completely explained by the chosen large-scale predictors. A direct consequence of inflation is an increase in the root-mean-squared error. Instead of inflation, von Storch (1999) advocates randomization: that is, adding random small-scale variability. Refer to Maraun et al. (2010) for a recent review of such stochastic downscaling approaches.

Here, I show that inflation-related problems also occur for model output statistics (MOS): namely, if variance correction and quantile mapping are used to downscale simulated gridbox area averages to point values. The climate simulated by numerical models often shows a distinct systematic deviation from the true observed climate, limiting the usability of climate simulations for impact models. Therefore, it is often desired to post-process the climate model output to match the observed climate (Christensen et al. 2008). Bias correction methods are variants of MOS, a concept developed in weather forecasting and now commonly used in climate science (Maraun et al. 2010). The simplest methods correct the long-term climatological mean bias between simulations and observations; extensions also correct the variance. Quantile mapping even attempts to remove quantile-dependent biases.

Assume the meteorological variable of interest can, at a set of locations and days, be described by a time-stationary random process characterizing the spatial and temporal dependencies. For every location, the time-independent marginal density distribution describes the variable regardless of the spatiotemporal dependence.[1] Bias correction deterministically postprocesses the marginal distribution of the raw climate model data: a specific simulated value will always yield the same corrected value, and the spatiotemporal dependence is not explicitly altered. An implicit assumption of any bias correction adjusting more than climatological means is therefore that all local-scale spatiotemporal variability is completely determined and—apart from an

Corresponding author address: Douglas Maraun, GEOMAR Helmholtz Centre for Ocean Research Kiel, Düsternbrooker Weg 20, 24105 Kiel, Germany.
E-mail: dmaraun@geomar.de

---

[1] The empirical equivalent of the marginal density distribution is the histogram of the observations.

adjustment of the marginal distribution[2]—correctly represented by the simulated gridbox variability. This might in principle be a valid assumption for a pure bias correction: that is, if the model simulation is corrected against a gridded dataset of the same resolution as the climate model. If, however, the bias correction also attempts to downscale [i.e., if the correction is against station (or very-high-resolution gridded) data], deterministic variance correction and quantile mapping approaches are not feasible. In general, the spatiotemporal variability at the gridbox scale is much smoother than at the local scale. Yet as only the marginals are corrected and no additional local-scale variability is generated, the temporal dependence and the spatial dependence between locations across grid boxes are those of the gridbox scale. Even more, since the correction is a deterministic mapping, within a grid box the spatial dependence between locations is fully deterministic. Hence, in this downscaling setting also deterministic variance correction and quantile mapping rescale the simulated time series in an attempt to explain unexplained small-scale variability.[3] In other words, they inflate the simulated time series.

This study analyses potential consequences of inflation by quantile mapping[4] for a specific example. Consider a distributed hydrological model (e.g., Xu 1999; Das et al. 2008) that uses, among other variables, a high-resolution precipitation field (on the order of 1 km × 1 km) interpolated from gauge data as input. If such a model were to be used for climate change studies based on regional climate model (RCM) simulations, downscaling the RCM to the high-resolution precipitation field would be required. To assess the performance of quantile mapping for such a situation, I map RCM-simulated daily precipitation at one grid box to a set of observational rain gauge records within this grid box. I then consider the effect of quantile mapping on the spatiotemporal structure, the gridbox-aggregated daily precipitation series, and trends in seasonal total and seasonal maximum daily precipitation of the corrected RCM output.

2. Data and methods

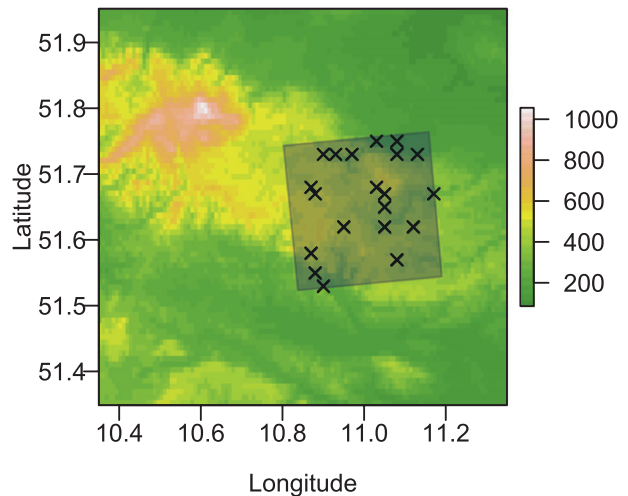As an RCM, I chose the Regional Model (REMO) from the Max Planck Institute of Meteorology (Jacob

FIG. 1. Map of the Harz Mountains with the selected gauges and the RCM grid box. Elevation is given in meters above mean sea level.

2001), operated on a 25-km rotated grid [available from the Ensemble-Based Predictions of Climate Changes and their Impacts (ENSEMBLES) project at http://ensemblesrt3.dmi.dk; van der Linden and Mitchell 2009]. The effect to be demonstrated occurs already in the calibration period; therefore, I deliberately do not choose a separate validation period. To avoid problems related to general circulation model (GCM) biases, the RCM is driven by 40-yr European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40) boundary conditions for the period 1961–2000. Such a perfect boundary condition setting roughly synchronizes simulated and observed precipitation but, as the quantile mapping "sees" only the marginal distributions, this temporal agreement is irrelevant for the analysis. The conclusions would be the same for an RCM driven by GCM boundary conditions. I selected the grid box centered on 11.00°N, 51.64°E in the eastern Harz Mountains in central northern Germany as a study area, mainly because of the high number of more than 20 rain gauges within its area. From all available rain gauges, a subset of 20 gauges with sufficiently long time series has been selected (see Fig. 1 and Table 1). The first three gauges (located in the southwest corner of the grid box) belong to the catchment of the Helme, and the other gauges belong to the Bode. Both rivers finally flow into the Saale, a tributary of the Elbe.

A simple empirical quantile mapping has been used: since observational and simulated time series were of equal length and no separate validation period has been considered, the simulated quantiles could be directly mapped onto the observed quantiles and no

---

[2] This includes a possible adjustment of wet day frequencies (Hay and Clark 2003; Piani et al. 2010).

[3] Recent covariate-dependent quantile mapping may be applied with randomization (Kallache et al. 2011).

[4] The variance correction can be seen as a special case.

TABLE 1. Chosen rain gauges. Elevation is provided in meters above mean sea level.

Gauge	Lat	Lon	Elevation	Period covered	Missing values DJF	Missing values JJA
Neustadt (Talsp.)	51.58°N	10.87°E	454 m	1 Jan 1961–31 Dec 2000	0	0
Herrmannsacker	51.55°N	10.88°E	315 m	1 Jan 1961–31 Dec 2000	0	0
Stempeda	51.53°N	10.90°E	242 m	1 Jan 1969–31 Dec 2000	0	0
Hayn (Harz)	51.57°N	11.08°E	435 m	1 Jan 1961–31 Dec 2000	0	0
Strassberg	51.62°N	11.05°E	400 m	1 Jan 1969–31 Dec 2000	0	0
Harzgerode (Schee.)	51.67°N	11.17°E	250 m	1 Jun 1969–31 Dec 2000	0	0
Neudorf	51.62°N	11.12°E	425 m	1 Jan 1969–31 Dec 2000	0	0
Siptenfelde (Uhl)	51.67°N	11.05°E	412 m	1 Jun 1969–31 Dec 2000	0	0
Siptenfelde	51.65°N	11.05°E	395 m	1 Jun 1969–31 Dec 2000	0	0
Altenbrak (Talsp.)	51.73°N	10.90°E	430 m	1 Jan 1969–31 Dec 2000	0	0
Altenbrak	51.73°N	10.93°E	300 m	1 Jan 1969–31 Dec 2000	5	0
Altenbrak-Todt.	51.73°N	10.97°E	425 m	1 Jan 1969–31 Dec 2000	242	246
Hasselfelde	51.68°N	10.87°E	461 m	31 Dec 1968–31 Dec 2000	0	0
Thale (Harz)	51.75°N	11.03°E	157 m	1 Jan 1969–31 Dec 2000	31	31
Neinstedt	51.75°N	11.08°E	140 m	1 Jan 1969–31 Dec 2000	0	0
Stiege	51.67°N	10.88°E	494 m	1 Jan 1961–31 Dec 2000	181	184
Friedrichsbrunn	51.68°N	11.03°E	523 m	31 Dec 1968–31 Dec 2000	0	0
Stecklenberg	51.73°N	11.08°E	160 m	1 Jan 1969–31 Dec 1999	0	0
Gernrode	51.73°N	11.13°E	210 m	1 Jan 1961–31 Dec 2000	182	184
Breitenstein	51.62°N	10.95°E	466 m	1 Jan 1969–31 Dec 2000	0	0

interpolation had to be carried out. In cases where the observational data had missing values, the corresponding simulated values were omitted to obtain time series of equal length. The drizzle effect was corrected based on a wet day threshold of 1 mm day^{-1} for the observations (e.g., Hay and Clark 2003; Piani et al. 2010). The mapping was carried out separately for winter and summer. Figure 2 shows quantile–quantile (Q–Q) plots of the raw and corrected RCM data against the observed precipitation for the rain gauge of Thale (Harz). In winter, the uncorrected RCM heavily underestimates observed precipitation, but produces too many drizzle days. These effects are well known (e.g., Maraun et al. 2010) and are at least partly caused by the scale mismatch between point observations and area-average simulations. For summer, the effect is similar, although the RCM produces some high rainfall events matching observed heavy precipitation. In both cases, by construction, the corrected RCM perfectly reproduces the marginal distribution of observed precipitation.
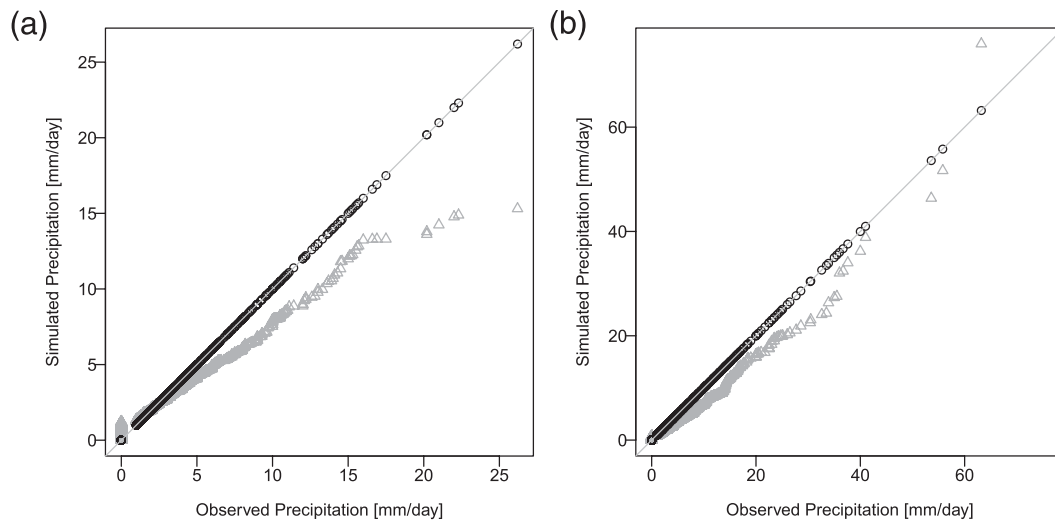

FIG. 2. Q–Q plot for Thale (Harz). Uncorrected (gray triangles) and corrected (black circles) simulated daily precipitation against observed daily precipitation: (a) December–February (DJF) and (b) June–August (JJA).
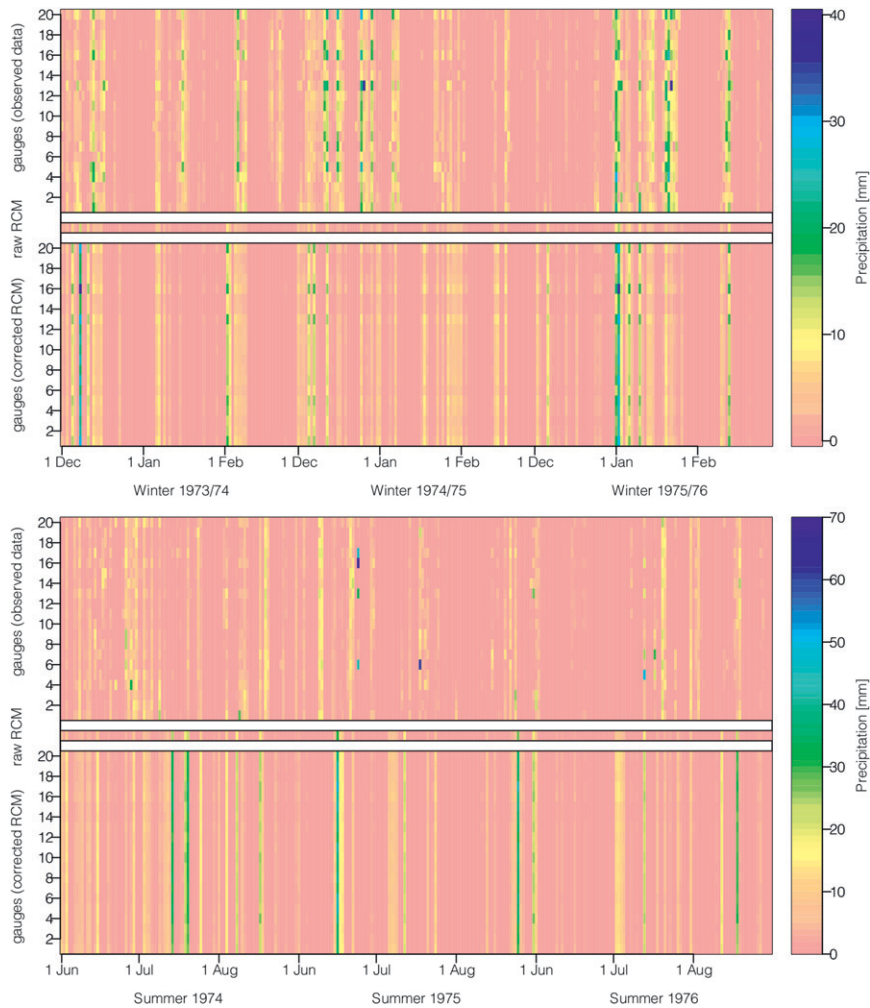
FIG. 3. Time series for selected seasons: (top) DJF and (bottom) JJA.

3. Results

Figure 3 shows the observed, simulated, and corrected time series for the 20 selected rain gauges during three example winters and summers. The synchronicity between the observed and modeled sequence of events (relatively high in winter and low in summer) will not be discussed here. An obvious difference between observations and corrected simulations becomes apparent: the spatial variability is quite high in the observations: even when it rains at some gauges, it might be dry at others; even when it rains heavily at some gauges, rainfall might be modest at others. In general, extreme events are spatially quite localized (more strongly in summer than in winter). This is different in the corrected RCM simulation: because quantile mapping is deterministic, a high (modest) RCM gridbox precipitation value is always transformed into a high (modest) local value. If the RCM simulates drizzle, the correction of the drizzle effect in

most cases leads to complete dryness across all gauges. In other words, on one hand, extreme events always cover the whole gridbox area, and their spatial extent should thus be heavily exaggerated. On the other hand, the drizzle effect is overcorrected and too many days with complete dryness in the grid box should thus occur. Finally, the ranking of precipitation across gauges can never change for a given quantile, and in most cases this ranking should be the same for all quantiles (only, if the quantile transfer function for one gauge intersects the transfer function from another gauge, the ranking for high and low quantiles might change). For instance, the gauge of Stiege is located on a hill and has on average higher precipitation than the rain gauge at Thale (Harz), which is located in a valley. However, whereas in reality on some days precipitation in the valley is higher than on the hill, this will never occur in the deterministic quantile mapping case.

The effect of these problems on the representation of area-mean precipitation is demonstrated in Fig. 4.

Shown are the Q–Q plots between the average of all 20 corrections of the simulated RCM time series against the average of all 20 observed time series. For both winter and summer the overcorrection of the drizzle effect as well as the exaggeration of extreme events becomes evident: whereas the corrected model simulates too many area-mean dry days, it strongly over-estimates area-mean extreme events by roughly 30%.

Finally, I analyze the effect of quantile mapping on trends in seasonal total and maximum precipitation. I consider both absolute trends (millimeters per decade) and trends relative to typical values (percent per decade relative to the expected value for the year 1985; for details, see appendix). Figure 5 shows an example, again for the gauge of Thale (Harz). The top panels depict seasonal total precipitation and its trends for winter (left) and summer (right); the bottom panels show the respective results for seasonal maximum precipitation. Observations are merely plotted to illustrate how the quantile mapping influences the amplitudes.[5] In this example, quantile mapping slightly deflates low values of winter total precipitation and inflates high values. However, as the trend is weak, inflation and deflation are evenly distributed in time. Hence, the trend is only marginally increased (absolute negative trend of 0.8 mm decade^{-1}; the increase of the relative trend by 27.7% is not relevant). For summer totals, the strong negative trend causes inflation mainly in the beginning of the series. As a result, the absolute negative trend increases by 3.9 mm decade^{-1} and the relative trend increases by 11.7%. The panel for winter maxima illustrates the effect of quantile mapping on heavy precipitation trends: the highest simulated values, occurring in the beginning, are strongly amplified by the quantile mapping (about 80%), whereas the amplification of the lower values toward the end of the series is weaker (about 30%). This asymmetric amplification increases the negative winter trend by 0.9 mm decade^{-1} (85.6% increase in the relative trend). The negative trend in summer maxima is weak and does not cause a time-dependent inflation. Thus, the effect on the resulting trend is negligible (0.28 mm decade^{-1}; the relative change of 30.1% is not relevant). The same analysis has been carried out for all rain gauges with similar results, suggesting that already strong trends (relative to the interannual
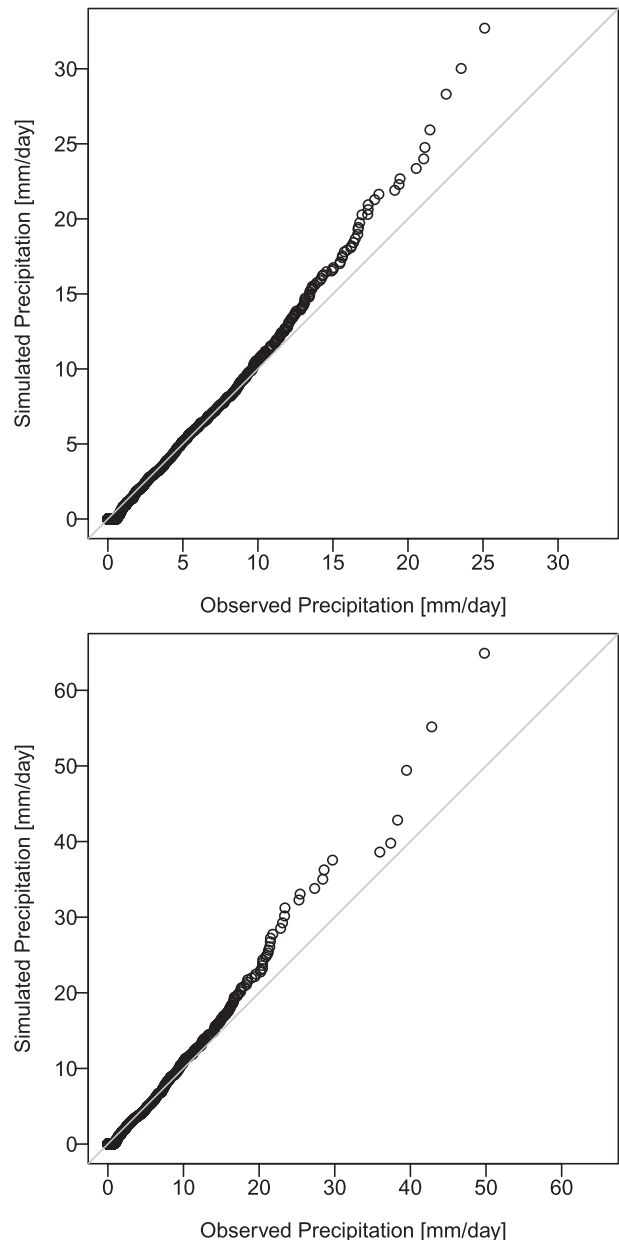

FIG. 4. Q–Q plot of area-mean precipitation for the chosen grid box. Corrected simulated daily precipitation against observed daily precipitation: (top) DJF and (bottom) JJA.

variability) tend to get amplified by quantile mapping, for both precipitation totals and heavy precipitation.

4. Conclusions

These findings clearly demonstrate the problems of inflation by quantile mapping (variance correction is a special case), when used to downscale from gridbox to local scales. Similar to the case of perfect prog statistical downscaling, the problems arise from the attempt to

---

[5] The fact that the perfect boundary-driven RCMs do not capture the observed trends is likely because of the driving reanalysis data (Bengtsson et al. 2004; Thorne and Voss 2010). Repeating the analysis with the Royal Netherlands Meteorological Office (Koninklijk Nederlands Meteorologisch Instituut; KNMI) Regional Atmospheric Climate Model, version 2 (RACMO2), yields similar results.
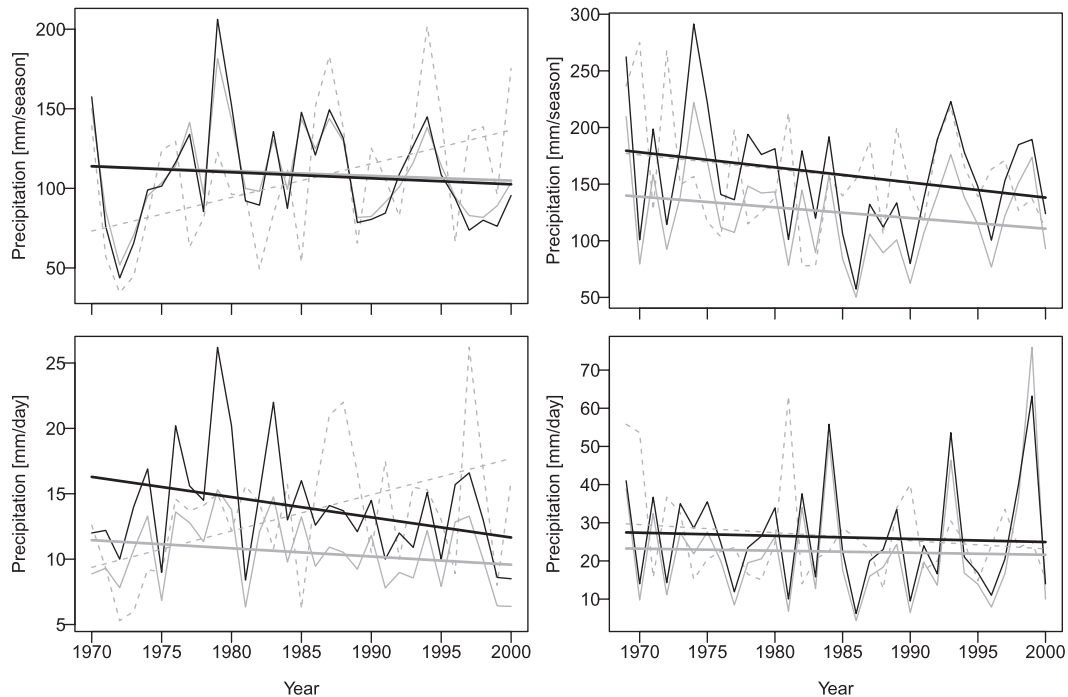
FIG. 5. Precipitation time series and trends for Thale (Harz): (top) seasonal total and (bottom) seasonal maxima for (left) DJF and (right) JJA. Dashed gray lines are observations, solid gray lines are uncorrected precipitation simulations, and black lines are corrected precipitation simulations.

explain local variability by gridbox variability. For local climate scenarios and impact modeling, the inflation effect may have severe consequences: as the quantile mapping does not introduce any small-scale variability, the temporal structure is still that of the gridbox and not the local scale (Fig. 3). If, in a particular application, the temporal structure is important, the results will most likely be misspecified. When used to provide local-scale input data for distributed hydrological models, flood risk (in particular for small rapidly responding catchments) might be heavily overestimated (Fig. 4). Finally, as trends are affected, changes in future mean and extreme precipitation, as well as any related impacts, are likely to be misrepresented. Equivalent analyses for other regions showed that these problems also occur in flat terrain. They join other problems of model output statistics such as bias nonstationarities (Christensen et al. 2008; Maraun 2012). To increase the signal-to-noise ratio, one often averages neighboring grid boxes. If the target resolution is of subgrid scale, this strategy increases the scale gap and thus exacerbates the inflation problem.

Eden et al. (2012) argue that model errors caused by parameterization and orography can reasonably be corrected by bias correction. If quantile mapping is used to downscale to local scales, an additional discrepancy—not error—between model and observations occurs because of unresolved small-scale variability. This study

shows for precipitation that quantile mapping cannot be used to bridge this scale mismatch. The effect might be less important for temperature, as this variable has a much higher spatial coherence and small-scale variations mostly stem from—correctable—orographic effects.

To avoid inflation, different strategies might be pursued: if one is not interested in the day-to-day variability, one should simply correct the mean to avoid effects on trends. If a single time series representing total catchment precipitation is required as input for a lumped hydrological model (e.g., Xu 1999; Das et al. 2008) with a large catchment size relative to a grid box, one should directly correct the required total precipitation and avoid downscaling to point sizes and averaging back to the catchment total. If, however, one is interested in the day-to-day variability at local scales, a solution is similar as in perfect prog downscaling (von Storch 1999). In a perfect boundary setting, a regression between modeled and observed precipitation with a suitable noise model describing the local spatiotemporal dependence should be carried out. The deterministic part of this regression would correct for systematic errors and realizations of the noise model would add the necessary small-scale variability. Hence, this study clearly demonstrates the need for stochastic bias correction.

APPENDIX

Trend Models

Trends in seasonal total precipitation y_i are modeled by linear regression: that is, for a year t_i, $i = 1 \ldots N$,

$$y_i \sim \mathcal{N}(\mu_i, \sigma) \quad \text{with} \quad \mu_i = a + bt_i, \qquad \text{(A1)}$$

where $\mathcal{N}(\mu_i, \sigma)$ denotes a normal distribution with time-dependent mean μ_i and constant width σ. Seasonal maxima are modeled by the generalized extreme value (GEV) distribution (Coles 2001),

$$y_i \sim \text{GEV}(\mu_i, \sigma_i, \xi), \qquad \text{(A2)}$$

with time-dependent location and scale parameters μ_i and σ_i and constant shape parameter ξ. The linear time dependence is modeled as

$$\mu_i = a_\mu + b_\mu \cdot t_i \quad \text{and} \quad \sigma_i = a_\sigma + b_\sigma \cdot t_i. \qquad \text{(A3)}$$

The expected seasonal maximum E_i, linearly depending on time, is then given by

$$E_i = \mu_i - \frac{\sigma_i}{\xi} + \frac{\sigma_i}{\xi} \cdot \Gamma(1 - \xi), \qquad \text{(A4)}$$

where $\Gamma(\cdot)$ denotes the gamma function.

REFERENCES

Bengtsson, L., S. Hagemann, and K. I. Hodges, 2004: Can climate trends be calculated from reanalysis data? *J. Geophys. Res.,* **109,** D11111, doi:10.1029/2004JD004536.

Christensen, J. H., F. Boberg, O. B. Christensen, and P. Lucas-Picher, 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.

Coles, S., 2001: *An Introduction to Statistical Modeling of Extreme Values.* Springer Series in Statistics, Springer, 228 pp.

Das, T., A. Bardossy, E. Zehe, and Y. He, 2008: Comparison of conceptual model performance using different representations of spatial variability. *J. Hydrol.,* **356,** 106–118.

Eden, J., M. Widmann, D. Grawe, and S. Rast, 2012: Skill, correction, and downscaling of GCM-simulated precipitation. *J. Climate,* **25,** 3970–3984.

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

Jacob, D., 2001: A note to the simulation of the annual and interannual variability of the water budget over the Baltic Sea drainage basin. *Meteor. Atmos. Phys.,* **77** (1–4), 61–73.

Kallache, M., M. Vrac, P. Naveau, and P.-A. Michelangeli, 2011: Nonstationary probabilistic downscaling of extreme precipitation. *J. Geophys. Res.,* **116,** D05113, doi:10.1029/2010JD014892.

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.

——, and Coauthors, 2010: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Rev. Geophys.,* **48,** RG3003, doi:10.1029/2009RG000314.

Piani, C., J. O. Haerter, and E. Coppola, 2010: Statistical bias correction for daily precipitation in regional climate models over Europe. *Theor. Appl. Climatol.,* **99** (1–2), 187–192.

Thorne, P. W., and R. S. Voss, 2010: Reanalysis suitable for characterizing long-term trends. Are they really achievable? *Bull. Amer. Meteor. Soc.,* **91,** 353–361.

van der Linden, P., and J. F. B. Mitchell, 2009: ENSEMBLES: Climate change and its impacts: Summary of research and results from the ENSEMBLES project. Met Office Hadley Centre Tech. Rep., 160 pp.

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

Xu, C., 1999: From GCMs to river flow: A review of downscaling methods and hydrologic modelling approaches. *Prog. Phys. Geogr.,* **23,** 229.