Original

Haerter, J.O.; Eggert, B.; Moseley, C.; Piani, C.; Berg, P.:

Statistical precipitation bias correction of gridded model data using point measurements

In: Geophysical Research Letters (2015) AGU

DOI: 10.1002/2015GL063188

AGU PUBLICATIONS

Geophysical Research Letters

RESEARCH LETTER
10.1002/2015GL063188

Key Points:
- Statistical bias correction using station data
- Improved corrections through scale adaptation
- Additional applications when comparing to station data for extreme events

Correspondence to:
J. O. Haerter,
haerter@nbi.dk

Citation:
Haerter, J. O., B. Eggert, C. Moseley, C. Piani and P. Berg (2015), Statistical precipitation bias correction of gridded model data using point measurements, *Geophys. Res. Lett.*, *42*, 1919–1929, doi:10.1002/2015GL063188.

Received 21 JAN 2015
Accepted 25 FEB 2015
Accepted article online 16 MAR 2015
Published online 24 MAR 2015

Statistical precipitation bias correction of gridded model data using point measurements

Jan O. Haerter[1], Bastian Eggert[2], Christopher Moseley[3], Claudio Piani[4], and Peter Berg[5]

[1]Niels Bohr Institute, Copenhagen, Denmark, [2]Helmholtz Zentrum Geesthacht, Climate Service Center 2.0, Hamburg, Germany, [3]Max Planck Institute for Meteorology, Hamburg, Germany, [4]Computer Science, Mathematics and Environmental Sciences, American University of Paris, Paris, France, [5]Hydrology Research Unit, Swedish Meteorological and Hydrological Institute, Norrköping, Sweden

Abstract It is well known that climate model output data cannot be used directly as input to impact models, e.g., hydrology models, due to climate model errors. Recently, it has become customary to apply statistical bias correction to achieve better statistical correspondence to observational data. As climate model output should be interpreted as the space-time average over a given model grid box and output time step, the status quo in bias correction is to employ matching gridded observational data to yield optimal results. Here we show that when gridded observational data are not available, statistical bias correction can be carried out using point measurements, e.g., rain gauges. Our nonparametric method, which we call scale-adapted statistical bias correction (SABC), is achieved by data aggregation of either the available modeled or gauge data. SABC is a straightforward application of the well-known Taylor hypothesis of frozen turbulence. Using climate model and rain gauge data, we show that SABC performs significantly better than equal-time period statistical bias correction.

1. Introduction

Statistical bias correction, as a field, has received substantial attention in recent years as it is a simple tool that makes impact studies possible in situations where climate model data are available but are subject to inherent biases. Since the early approaches [*Wood et al.*, 2004], statistical bias correction techniques have now diversified considerably [*Maraun et al.*, 2010; *Piani et al.*, 2010a, 2010b; *Haerter et al.*, 2011; *Teutschbein and Seibert*, 2012; *Piani and Haerter*, 2012] and have been widely applied to a range of global [*Piani et al.*, 2010b; *Li et al.*, 2010] and regional climate model (RCM) data sets [*Berg et al.*, 2012a; *Gudmundsson et al.*, 2012; *Teutschbein and Seibert*, 2012].

Most current statistical bias correction techniques have in common that some form of quantile mapping is applied to match the probability distribution function of climate model output to that of observed climate data. It is generally necessary that the spatial and temporal resolution of modeled and observed data match as closely as possible—e.g., in order to avoid the so-called inflation or deflation issue [*von Storch*, 1999; *Maraun*, 2013]. However, in many practical situations, observed spatial data may not be available. Some type of derived data set, such as reanalysis data [*Dee et al.*, 2011] or data products combining data from multiple sources [*Weedon et al.*, 2011; *Berg et al.*, 2015], must be used. The former, as a model interpolation technique, still suffers from error and bias, while in the latter, due to the blend of data sources, little is sometimes known on the exact details of the underlying data resolution in specific regions or time periods. Further, both model and observations must typically be regridded to obtain matching resolutions. For station data (point measurements), regridding to a common grid, i.e., finding an area representation of observations, is impossible from the outset—especially when few stations are available in the region of interest. One possible option are then stochastic methods to account for variability at small scales [*Eden et al.*, 2014; *Wong et al.*, 2014], which, however, entail assumptions on the distribution functions.

In this study we take a different approach that may be suitable when gridded observations of dynamical variables are not available, but reliable station measurements are. The approach is a simple application of the Taylor hypothesis of frozen turbulence. The original Taylor hypothesis states that as the mean atmospheric flow advects eddies past a station, the properties of the eddies remain unaltered [*Taylor*, 1938]. The hypothesis has previously been used for precipitation disaggregation [*Deidda*, 2000]. Recognizing this, mapping to a common grid is no longer necessary as lacking spatial information can be compensated by increased temporal information.

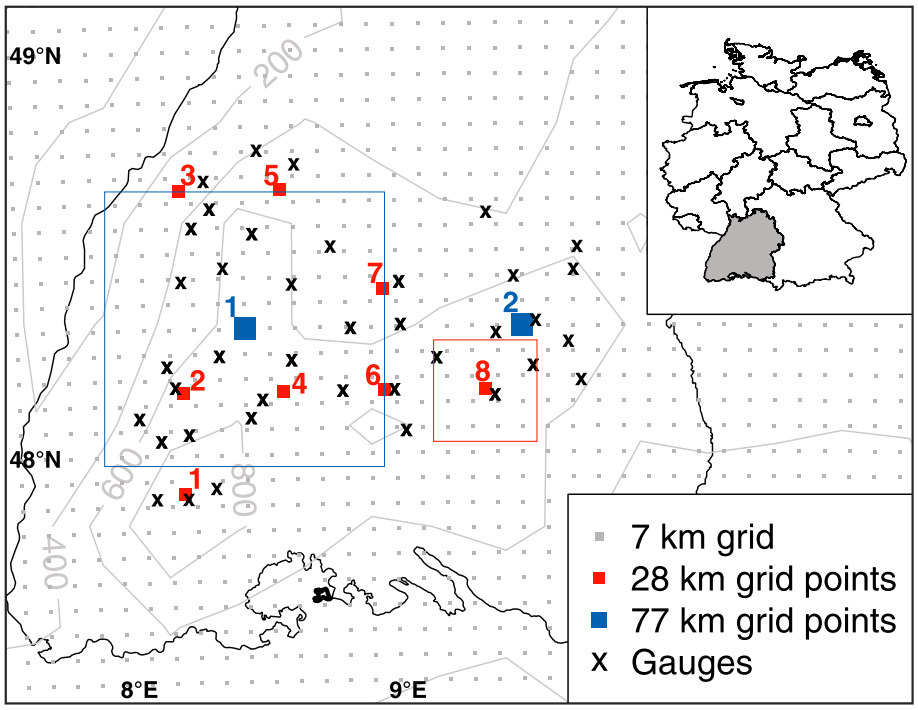

Figure 1. Map of the data used. Figure shows available rain gauge stations, the model grid at 7 km, and grid points (red and blue numbers) used for the 28 km and 77 km correction. Red and blue boxes exemplify areas of respective grid boxes. Also shown: topography in meters, coordinates, and state boundary (black) of Baden-Württemberg, Germany.

Here we will consider simulated precipitation versus rain gauge station observations. Each value of simulated precipitation represents a space and time average determined by the grid size $\delta_{x,\mathrm{mod}}$ and time step $\delta_{t,\mathrm{mod}}$ of the output. By contrast, station precipitation data represent a spatial point measurement ($\delta_{x,\mathrm{obs}}=0$) and a time average determined by the measurement time step ($\delta_{t,\mathrm{obs}}$). In general, we will have $\delta_{t,\mathrm{mod}} \neq \delta_{t,\mathrm{obs}}$. In the case of high-frequency observations, one might be tempted to simply aggregate station data to a coarser time interval so that $\delta_{t,\mathrm{mod}}=\delta_{t,\mathrm{obs}}$ and use the resulting data set to perform a statistical bias correction of the simulated precipitation. For example, if correcting simulated daily values of precipitation with hourly station data, one might simply derive daily station data and proceed as usual. We will show that further aggregation onto longer time intervals of the observed station data may lead to better quality bias correction, while also avoiding the inflation issue [*Maraun*, 2013]. We also show the inverse, when model data have comparably high resolution. In that case, moderate coarse graining of the model data improves the statistical bias correction.

2. Data

We use a fine-resolution (1 km and 5 min) composite of radar images from the RADOLAN-RY product of the German Weather Service (DWD) for Southern Germany, which was aggregated for varying resolutions [*Eggert et al.*, 2015]. Rainfall rates were calculated from radar echoes with the *Z-R* relationship [*Steiner et al.*, 2004] and are available for the 2 years 2007–2008. Additionally, a set of 1 h resolution rain gauge data from Baden-Württemberg in southwestern Germany was used for the bias correction experiments. The stations constitute a relatively dense network covering both the Black Forest mountain range and the topographically less variable Rhine Valley (Figure 1). The station network is most dense in the period 1997–2004, which is used here. Stations with more than 10% missing data were discarded from the analysis, and for the remaining stations NaNs were set to zero precipitation to simplify. This has little impact on the results presented here.

The model data are taken from a simulation with the COSMO-CLM (COnsortium for Small scale Modelling-CLimate Mode) [*Doms and Schättler*, 2002]. The current simulation uses ERA-Interim reanalysis [*Dee et al.*, 2011] as driving data in a double nesting setup with a second 7 km nest domain covering all

of Germany, with the lateral boundaries well outside Germany's borders. The output time step is 1 h for precipitation. (Simulation details: *Berg et al.* [2012b].) For the current investigations, the model results are applied at the original 7 km resolution, and also as representatives for coarser model resolution by remapping to 28 km and 77 km. As shown in *Berg et al.* [2012b] and *Fosser et al.* [2014], the model has a strong bias in mean precipitation amounts but performs well regarding the intensity distribution for both daily and hourly timescales. Equally long periods of 8 years are used for the model and station data.

3. Methodology

The Taylor hypothesis [*Taylor*, 1938] states that observations at a given spatial and temporal resolution should be similar to those at lower spatial but higher temporal resolution. Our method is hence based on the idea that lacking spatial resolution can be compensated by increased temporal resolution. Point measurements, i.e., data at high spatial resolution, should therefore be compared to spatially averaged data at correspondingly higher temporal resolution.

3.1. Frozen Turbulence and Comparison of PDFs

The probability density function (PDF) of precipitation intensity I depends on the resolution in space and time, expressed by the size of the grid box or time step (δ_x or δ_t), of the precipitation data. Our methodology is based on the assumption that PDFs of precipitation intensity obtained with different spatial and temporal data resolutions (δ_x, δ_t) can have similar, albeit not identical, features. For example, should the Taylor hypothesis hold perfectly, then PDF(δ_x, 0, I) = PDF(0, $\delta_t = \delta_x/v$, I), where v is an advection speed.

Using observed or modeled precipitation data (section 2), we produce PDFs of precipitation intensity at varying spatial and temporal resolutions. We then compare the PDFs at different resolutions (δ_x, δ_t): The *PDF agreement*

$$S\left(\delta_x, \delta_t; \delta'_x, \delta'_t\right) \equiv 1 - \int_{I_0}^{\infty} \mathrm{d}I \left| w(\delta_x, \delta_t, I) - w\left(\delta'_x, \delta'_t, I\right) \right| \tag{1}$$

measures the similarity of the PDFs corresponding to different resolutions [compare *Perkins et al.*, 2007; *Eggert et al.*, 2015]. In equation (1), $w(\delta_x, \delta_t, I)$ is some function of the probability density function of precipitation intensity and I_0 defines a possible low-intensity cutoff. Note that for most regions of the globe, zero intensity constitutes the bulk of the probability weight in the PDF, i.e., dry periods outnumber the wet. Intensity weighting gives more emphasis to nonzero intensities. For this study, we therefore choose $w(\delta_x, \delta_t, I)$ as the intensity-weighted PDF of precipitation, i.e., $w(\delta_x, \delta_t, I) \equiv I \cdot \mathrm{PDF}(\delta_x, \delta_t, I) / \int_0^\infty \mathrm{d}I'\ I' \cdot \mathrm{PDF}(\delta_x, \delta_t, I')$ and $I_0 = 0$. Other reasonable choices are the Kolmogoroff-Smirnov statistics or the use of the bare PDF with a nonzero intensity cutoff $I_0 > 0$. We found all to give similar results, albeit with varying degree of noise (not shown).

3.1.1. Observed Data

For observed high-resolution data (section 2) and for a given reference resolution (e.g., δ_{x0} = 25 km, δ_{t0} = 5 min, shown as a large white circle in Figures 2a and 2b), we now determine all equivalent resolutions. This can be done in two equivalent ways leading to similar, albeit not identical, results:

1. Consider Figure 2a, where we directly compare PDFs of all resolutions with that of the reference resolution. The PDF agreement is therefore maximal, i.e., unity, not only for $(\delta_x, \delta_t) = (\delta_{x0}, \delta_{t0})$ but also for other pairs of resolutions the figure shows high-agreement values along a "ridge" with increasing space and decreasing time resolution. To the extent that the ridge top can be approximated by a line, this indicates that PDFs for resolutions $\left(\delta_x, \delta_t^* - r\delta_x\right)$, with δ_t^* the ridge intersect with the vertical axis, are all very similar. Defining, analogously, δ_x^* as the ridge intersect with the horizontal axis and $r^* \equiv \delta_x^*/\delta_t^*$, $v^* = 1/r^*$ has units of a velocity and gives the effective speed of advection [*Taylor*, 1938]. We note that as r^* is the slope of the ridge, it can also be estimated when any two points (δ_t, δ_x) and $\left(\delta'_t, \delta'_x\right)$ on this line are known, i.e., $r^* \approx \left(\delta'_t - \delta_t\right) / \left(\delta'_x - \delta_x\right)$. This will then also deliver δ_t^*, e.g.,

$$\delta_t^* = \delta_t + r^*\delta_x \left(= \delta'_t + r^*\delta'_x\right) \ . \tag{2}$$

   We will refer to the line defined by r^* and δ_t^* as the relevant ridge. In short, scale-adapted statistical bias correction (SABC) demands finding this relevant ridge. This ridge will then be used to identify the proper scale adaptation.

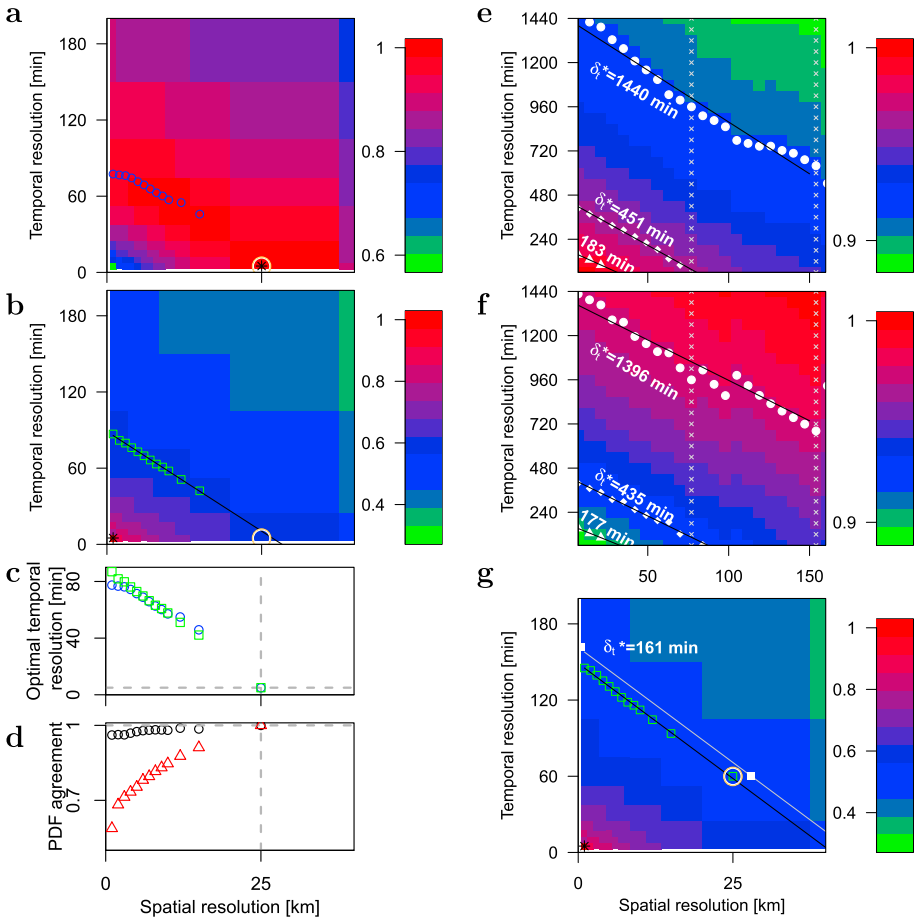

Figure 2. Mapping of distribution functions. (a) PDF agreement (equation (1)) of radar precipitation PDFs of different spatial and temporal resolution to the origin radar precipitation PDF with 25 km and 5 min resolution (shown as black star symbol). Resolutions equivalent to the reference resolution (white circle) shown as blue circles. (b) Similar to Figure 2a but using contours of similar PDF agreement for extrapolation. The origin distribution is now chosen at 1 km and 5 min (star symbol). Optimal resolutions shown as green squares. (c) Comparison of equivalent temporal resolutions from Figures 2a and 2b as function of spatial resolution. Dashed lines indicate reference resolution. (d) PDF agreement as function of spatial resolution: black circles indicate maximal PDF agreement derived in Figure 2a; red triangles denote PDF agreement when keeping temporal resolution fixed to 5 min. (e) PDF agreement with an origin resolution chosen at (7 km; 60 min) for model data. White triangles/diamonds/circles are computed ridges for reference resolutions (28 km; 60 min) and (77 km; 60 min), as well as $\delta_t^* = 1440$ min, respectively. (f) Similar to Figure 2c but for an origin resolution chosen at (154 km; 1440 min). Note the similar results for the ridges in Figures 2c and 2d. (g) Comparison to radar data. Color bars denote respective PDF agreement.

2. Another approach is to use a contour line of PDF agreement, not the maximum, to obtain the optimal resolutions. To obtain contours of PDF agreement, we specify a different origin resolution, that is, a resolution that serves as a comparison. In (1), this origin had been set to the same value as the reference resolution, but we now show that it can also be moved to another point in the plane, i.e., away from the reference resolution. Figure 2b exemplifies this for the case where the origin is set to (1 km; 5 min). Using now the PDF agreement corresponding to the reference resolution (i.e., 25 km and 5 min), i.e., comparing PDF(1 km, 5 min) to PDF(25 km, 5 min), we find the line of similar resolutions, this time by comparing with the respective PDF agreement (compare Figure 2b, green line).

Not surprisingly, the contour line (symbols in Figure 2c) differs very little from the ridge found in method (1). This is clear when noting that all PDFs along the ridge are very similar. Hence, also the agreement of any of them with an origin resolution away from the ridge should yield comparable values. Methods (1) and (2) are two equivalent ways to obtain the value of δ_t^*, i.e., the desired temporal resolution of the rain gauge. In practice (section 3.3) we use method (2).

We can compare the PDF agreement $S\left(0, \delta_t^*, \delta_{x0}; \delta_{t0}\right)$ of PDF $\left(0, \delta_t^*\right)$, obtained with method (1) with the standard approach of simply taking the time resolutions to match, i.e., maintaining a temporal gauge resolution of 5 min (Figure 2d). The values obtained for the ridge top are reasonably good (black circles), and much better than when using unchanged temporal resolution (red triangles). From the example we see that when the gauge data are first aggregated to approximately 80 min, it constitutes a much better representation of the distribution function of spatial data, i.e., PDF(0, 80 min) ≈ PDF(25 km, 5 min), while PDF(0, 5 min) ≠ PDF(25 km, 5 min).

3.1.2. Climate Model Data

In Figures 2e and 2f we repeat the analysis of PDF agreement for high-resolution climate model data (7 km spatially and 60 min temporally). Setting the origin resolution first to the highest model resolution (lower left corner in Figure 2e), we compute the contour lines for several choices of reference resolution (solid white symbols). These choices are made to reflect the cases of (i) typical high-resolution RCM output resolution (28 km, 60 min), (ii) high-resolution global climate model output (77 km, 60 min), and (iii) daily gauge temporal resolution (1440 min). Fitting linear functions, we extrapolate the corresponding values of δ_t^* (compare equation (2)). The fits show the following: For (i), the equivalent gauge resolution would be approximately 3 h. For (ii), gauge resolution should be chosen at 7.5 h. For (iii), daily gauge resolution requires model resolution given by any of the symbols (white circles) in Figure 2e. One possible choice is (150 km, ≈ 12 h). That is, if a model of 150 km spatial resolution was available, optimal temporal resolution would be approximately 12 hourly.

We repeat the analysis for another origin resolution (154 km spatially, and 1440 min, temporally, upper right corner in Figure 2f). Producing again the contour lines and corresponding fits, we obtain very similar values of δ_t^*, confirming that the choice of origin resolution has little impact on the resulting contour lines.

To compare these results with the observational data, we also extrapolate the contour line corresponding to one of the reference resolutions (28 km, 60 min) in the radar data (Figure 2g). In the observational data, the closest corresponding spatial resolution is 25 km. We use this to obtain the slope r. Together with the reference resolution (28 km, 60 min) we yield the extrapolated value of $\delta_t^* \approx 161$ min. This value lies somewhat lower than those from the model data (there, $\delta_t^* \approx 180\ min$). However, given the shortcomings of both observational and modeled data and limitations of available resolutions, the agreement of the results is remarkably good.

The previous analysis shows that the patterns obtained with observational data can approximately be reproduced using model data. However, typically, the model output resolution will lie at lower resolutions than the one used for Figures 2e and 2f. Consider, e.g., the resolution (77 km; 1 h); available spatial and temporal model resolutions are hence 77, 154, 231, … , km, respectively 60, 120, 180, … , min (shown as white crosses in Figures 2e and 2f). Desired contour lines must hence be determined using exclusively these discrete combinations of resolutions. We will show how to obtain these in the following.

3.2. Bias Correction Methodology

We distinguish the two cases resulting from comparably low model resolution (model limited) and comparably low gauge resolution (gauge limited).

3.2.1. Model-Limited Correction

Consider a climate model with a given output resolution (blue cross, Figure 3a). We assume that observations are only available through a rain gauge (red cross, Figure 3a), with a relatively high gauge resolution. The goal is now to estimate the contour line that is defined by the model output resolution, i.e., the blue cross symbol must lie on the contour line. The offset δ_t^* for this line will define the required coarsening, i.e., scale adaptation, of the gauge data. Once the scale adaptation has been produced, standard statistical bias correction [e.g., *Piani et al.*, 2010b] can be performed.

3.2.2. Gauge-Limited Correction

In many practical situations, it is also possible that rain gauge resolution is poor compared to the model resolution (Figure 3b, *gauge-limited* case). This may especially be the case in very data sparse regions, such as areas of the globe with little infrastructure (deserts and glaciated regions), but even in developed areas, complete spatial coverage by subdaily precipitation gauges is by no means standard and daily temporal resolution is usually the best available data. Under such circumstances, scale adaptation can only be performed by aggregating the *model* data. It must thereby be assured that the available gauge resolution lies on top of the contour line. Available model resolutions are shown in Figure 3b (small cross symbols). It is

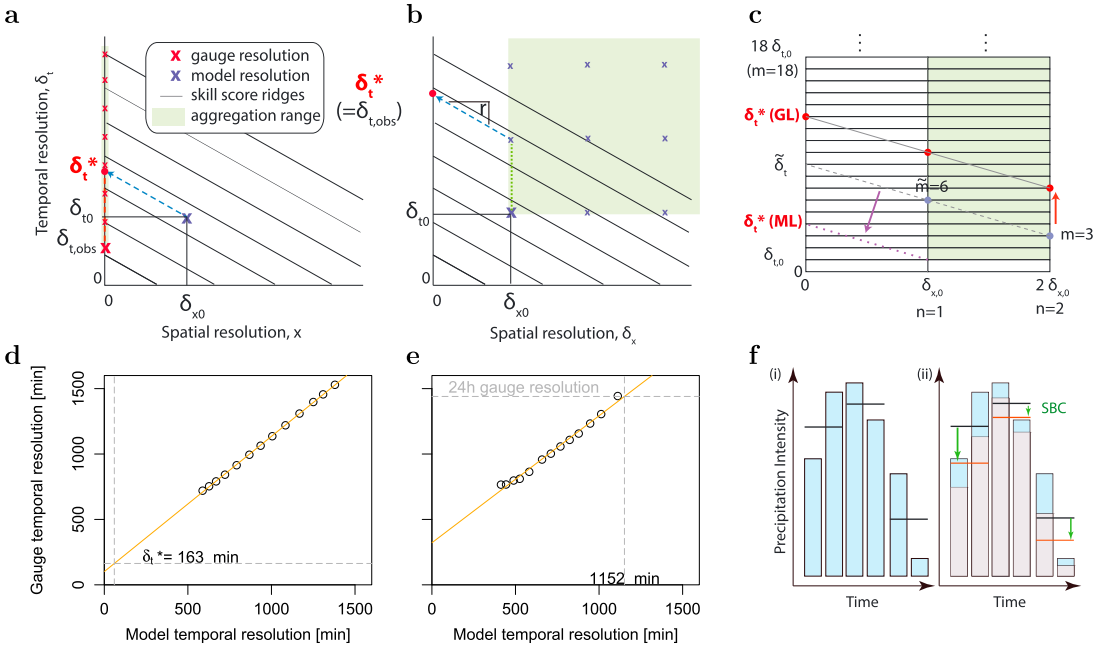

Figure 3. Scale-adapted statistical bias correction. (a) Model-limited correction. Two-dimensional schematic of spatial and temporal averaging intervals and PDF agreement ridges (black lines). Red and blue cross: symbols indicate available gauge and model resolutions, respectively. (b) Gauge-limited correction. Similar to Figure 3a but also including the range of resolutions where aggregation could be performed (green shades). Dotted green line indicates a possible aggregation choice. (c) Schematic showing the construction of equivalent resolutions. Corners of green-shaded region show available aggregated model resolutions. Dashed gray line indicates choice of $(n=2, m=3)$ and $(n=1, \tilde{m}=6)$, yielding $r=3\delta_{t0}/\delta_{x0}$ and $\tilde{\delta}_t=\tilde{m}\delta_{t0}+r\delta_{x0}=9\delta_{t0}$. Solid gray line shows equivalent resolutions yielding $\tilde{\delta}_t=\delta_t^*$ for the gauge-limited case (GL). Dotted purple line shows equivalent resolutions yielding $\tilde{\delta}_t=\delta_t^*$ for a model-limited case (ML); this line must be extrapolated (purple arrow indicates extrapolation). (d) Model-limited, extrapolation of $\tilde{t}$ as $\delta_x \to 0$ for an assumed model of resolution (28 km; 1 h). (e) Gauge-limited case, interpolation of $\tilde{\delta}_t$ as $\delta_x \to 0$ for an assumed model of resolution (77 km; 1 h) and gauge of 24 h resolution. (f) Schematic of data aggregation and disaggregation: (i) aggregation of pairs of two subsequent precipitation measurements (blue bars) to respective averages (black lines). (ii) Statistical bias correction (SBC) of aggregated data (black lines converted to red lines) and subsequent disaggregation of corrected data (gray bars).

now possible that multiple model resolutions can in principle be used for scale-adapted correction, i.e., any that are compatible with a contour line through $(0,\delta_t^*)$. Note that once the correction has been performed, the model data can be disaggregated to its original resolution. For each aggregated data point, we simply separate the data by applying the appropriate correction to each of its individual contributions (Figure 3f).

3.3. Practical Implementation

In practice, the resolution of the model output defines the reference resolution $(\delta_{x0}, \delta_{t0})$ and we set the origin resolution equal to this. Available model resolutions are all combinations of integer multiples $(n\ \delta_{x0}; m\ \delta_{t0})$ of the reference resolution, with n and m integers (Figure 3c). Implementation of scale-adapted bias correction follows four steps:

1. Set δ_t to δ_{t0}, i.e., $m=1$. Observe all PDFs for several accessible $n\ \delta_{x0}$, e.g., δ_{x0} and $2\delta_{x0}$.
2. Fix a specific PDF in (1) by choosing a multiplication $\tilde{n}$, e.g., choose the resolution $(2\delta_{x0}, \delta_{t0})$. Now obtain the value $\tilde{m}$ where $\tilde{m} \cdot \delta_{t0}$ yields maximum PDF agreement between PDF$(\tilde{n}\delta_{x0}, \delta_{t0})$ and PDF$(\delta_{x0}, \tilde{m}\delta_{t0})$, i.e., both will lie on the same contour line. The pairs $(\tilde{n}\delta_{x0}, \delta_{t0})$ and $(\delta_{x0}, \tilde{m}\delta_{t0})$ define lines which can be used to extrapolate to $\delta_x=0$, yielding a value $\tilde{\delta}_t$ (compare equation (2) and Figure 3c).
3. Repeat steps (1) and (2) starting with $m>1$ in (1).
4. Extrapolate or interpolate the resolution for required δ_t^*, i.e., achieving $\tilde{\delta}_t \approx \delta_t^*$.

Note: For better results, in step (2) a fit with respect to PDF agreement should be used to obtain $\tilde{m}$, thereby allowing noninteger $\tilde{m}$ and yielding more exact estimation of $\tilde{\delta}_t$.

Once δ_t^* has been obtained, in the ML case the gauge data will be aggregated to the resolution δ_t^*. In the GL case, the model data will be aggregated to a resolution closest to the line defined by the contour through δ_t^*. Once the data have been aggregated, standard statistical bias correction will be performed.

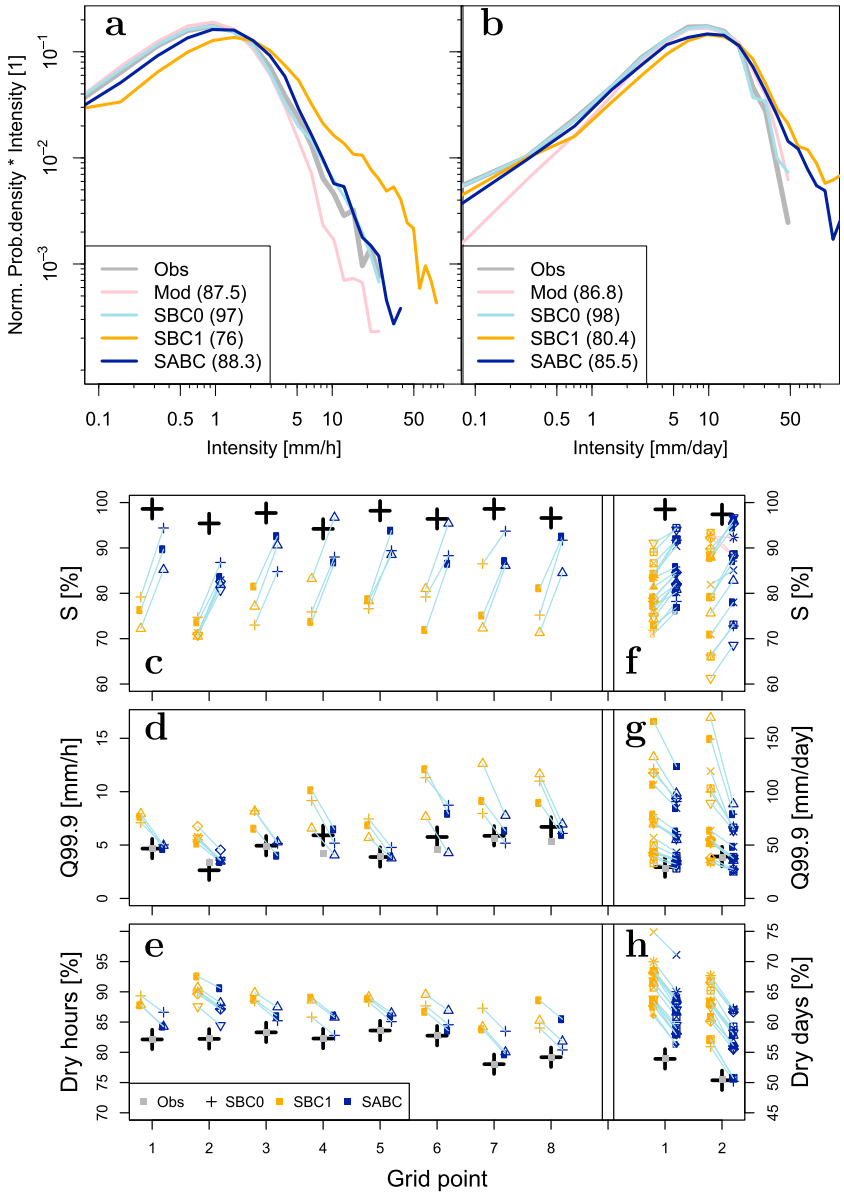

Figure 4. Bias correction results. (a) Model-limited bias correction; intensity-weighted PDFs for the estimated observations (Obs, gray), i.e., the average of all stations in a grid box, 28 km model data (Mod, pink), SBC1 using a single gauge at a 1 h resolution (orange), and SABC using a single gauge at 3 h temporal data aggregation (dark blue). PDFs shown are aggregate distributions for all stations and grid boxes. Values marked in parentheses in legend are PDF agreement values with ρ_{obs} in percent. (b) Gauge-limited bias correction for 77 km model; curves and colors analogous to Figure 4a. SBC1 is now for a single station and model data aggregated to 24 h; SABC uses 19 h temporal aggregation of model data. Note the different horizontal axes and units in Figures 4a and 4b. Results for individual grid boxes and stations for the ML case shown in Figure 4a: (c) PDF agreement , (d) 99.9th intensity percentile, and (e) dry period fraction. Colors as in Figures 4a and 4b. Symbols for each grid box are the different available stations (Figure 1), thin blue (pink) lines are guides to the eye, linking same stations for increased (decreased) performance. (f–h) Similar to Figures 4c–4e but corresponding to the GL case shown in Figure 4b.

4. Results

Does scale-adapted bias correction yield measurable improvement in practice? We carry out both types of corrections (Figures 3d, 3e, and 4) and compare each to the correction where temporal resolutions are simply matched for model and observations. Specifically, the tests are as follows.

(ML) Model-limited correction. We employ the same model used in Figures 2e and 2f but start from a minimal resolution of $\delta_{x0} = 28$ km spatially and 60 min temporally. This is to mimic output resolution typical of current RCMs. We assume station data to be available at 1 h resolution.

(GL) Gauge-limited correction. Again, using the same model and hourly temporal resolution, we now use coarse spatial resolution (77 km), mimicking state-of-the-art global climate model output. The station resolution is assumed to be daily.

Following the steps in section 3.3, we now obtain an estimate for the corresponding contour line (white triangles and circles in Figure 2f for the ML and GL cases) without the knowledge of the data corresponding to spatial scales below δ_{x0}. Specifically, we compute various curves (such as those shown in Figure 3c) and for each determine the corresponding $\tilde{\delta}_t$. This procedure yields a set of corresponding temporal resolutions $(\tilde{m}\delta_{t0}, \tilde{\delta}_t)$, which is plotted in Figures 3d and 3e. For the ML case, these points allow us to extrapolate the δ_t^* corresponding to model temporal resolution δ_{t0}, yielding $\delta_t^* \approx 163$ min. The best equivalent gauge aggregation (integer multiple of 1 h) is hence 3 h. We proceed analogously for the GL case, where we determine the value of $\tilde{m}$ that corresponds to $\delta_t^* = 1440$ min, yielding $\tilde{m} \approx 19$, hence a model aggregation to 19 h (compare Figure 2e, where the value is ≈18 h).

To obtain a proxy for spatially averaged observations, we first group rain gauge stations into 28 km (respectively 77 km) grid boxes and simply average their precipitation intensities. We discard grid boxes with fewer than three stations. For the ML case, we maintain a temporal resolution of 1 h, and for the GL case, we aggregate temporally to 24 h. The resulting spatially averaged signals will be less variable than that of any station by itself, but, due to the finite number of stations per grid box, somewhat more variable than the actual spatial average (i.e., that corresponding to an infinite number of stations per grid box). For each grid box, the averaged station data serve as the "ground truth" for our procedure (we call its probability density function ρ_{obs} in the following).

We now use any single station within the grid box as constituting the only available station for a given practical situation. Our hypothesis is that SABC yields better agreement with ρ_{obs} than the standard correction with matching time resolutions (SBC1). To exemplify the bias correction procedure, we use simple empirical quantile mapping by estimating a regularly spaced quantile distribution, following *Gudmundsson et al.* [2012]. Our approach is, however, generic and should also apply for more sophisticated bias correction techniques [e.g., *Piani et al.*, 2010b; *Mehrotra and Sharma*, 2012; *Rocheta et al.*, 2014]. As a benchmark, we first perform a correction with ρ_{obs} (SBC0). This represents the *ideal* correction obtainable when sufficient data are at hand. As expected, the resulting corrected model PDF is very close to ρ_{obs}, as seen in Figures 4a and 4b.

For the ML case, for each available grid box and for each possible choice of associated station, we now repeat the procedure using the hourly data—assuming that only this *single* station is available for the bias correction. For each combination of a grid box and station, this yields a histogram for the corrected data. SBC1 significantly shifts the original distributions to more extreme intensities (Figure 4a); affecting the overall PDF agreement and high percentiles. Repeating for SABC (gauge data coarsened to 3 h), markedly closer agreement with ρ_{obs} is reached. Besides the overall skill, also extremes and dry period fraction are consistently improved for each single station used as reference (Figures 4c–4e). Also, the average intensities are corrected well to that of the relevant reference station, which trivially produces a spread of results around the average of ρ_{obs} with only small differences between SBC0 and SABC.

An analogous comparison is performed for the GL case; now SBC1 employs temporal coarsening of model data to match the assumed daily resolution of the station, leading to generally too heavy intensities compared to ρ_{obs} (Figure 4b). Again, SABC overall yields substantial improvement for PDF, extremes and dry period fraction. Disaggregation back to hourly data (Figure 3f) retains the increased skill seen in the daily statistics but does not, e.g., affect other bias such as in the diurnal cycle.

We end with a comment on GL corrections: When multiple options are available for coarsening, one relevant criterion for optimal choice might be loss of data samples in the statistics. Reduction of sample size is greatest for intermediate options of spatial coarsening, but low for both the original spatial resolution and very high values of the spatial coarsening. We consider the latter only a theoretical option, as it may require substantial domain sizes and removes all spatial information, e.g., orographic effects (details on sample reduction: Appendix A).

5. Discussion and Conclusion

Statistical bias correction of precipitation has emerged as an indispensible tool at the intersection between climate and impact modeling—but hinges on the availability of adequate observational data. Especially in data sparse regions of the globe, successful bias correction is hampered by the lack of agreement of model precipitation output resolution and the resolution observed. We have shown that even with observational data from a single rain gauge station, considerable improvement can be achieved when aggregating the available data to yield maximal PDF agreement between model and observations. We suggest the use of this method, which we have called scale-adapted statistical bias correction (SABC), in situations where bias correction is required but only limited point measurements are available.

Crucially, SABC capitalizes on the ability of the model to capture the space-time dynamics of precipitation. The model is hence employed to associate adequate scales. In SABC, the variability of precipitation intensity—measured at a station—is compared to model data measured over a spatial domain but accordingly chosen, shorter time intervals. Our results point to the utility of outputting temporally highly resolved model data, even (or especially *if*) spatial resolution is low—the information on fluctuations, encoded in the higher temporal output rate, will then be preserved and can be used for SABC. In this way, the method can help remedy the inflation issue [*von Storch*, 1999; *Maraun*, 2013]. SABC can naturally not impact on inherent model shortcomings in simulating temporal variability [*Maurer and Pierce*, 2014]. Also, when atmospheric advection is not described adequately, the association of scales needed for SABC may itself be biased.

The methodology is straightforward and can be applied for any type of climate model data. It requires only to compute probability density functions for several choices of coarsened resolutions. This is simply done by data aggregation, e.g., by doubling or tripling of spatial and temporal scales. Pairs of matching resolutions then define appropriate gauge resolutions. Our results are encouraging in that quantitative improvement of intensity distributions is reached for the overall histogram, extreme precipitation as well as dry periods.

SABC should not be limited to precipitation correction but may also be relevant for other meteorological variables where the hypothesis of frozen turbulence applies, i.e., where advection of the quantity is the dominant cause of local fluctuations. While our method works well even in regions of moderate topographic variation (Southern Germany), stronger variation, e.g., mountainous regions, may introduce features that naturally require fine-scale knowledge of local climate, not captured by single gauges. Especially when gauge density is low, as is the case in vast parts of the globe, our method may allow for substantially improved bias correction at essentially no cost in terms of model output, data storage, or mathematical complexity.

Studying precipitation intensity at resolutions finer than that of convective systems has recently become of widespread interest [e.g., *Lenderink and Van Meijgaard*, 2008; *Berg et al.*, 2013; *Eggert et al.*, 2015]. Observationally, data describing such extremes is often limited to individual gauges: Measurements from gauges are usually considered more reliable than other sources of data, i.e., constitute the preferred source of information. Yet modeling of station scale characteristics will—for the time being—not be feasible. Our study speaks to an alternative solution, where low spatial resolution could again be compensated by high temporal resolution. A station with hourly temporal resolution could then be modeled by a regional climate model with approximately 12 km spatial and 5 min temporal output. Again, simply using the same temporal output rate as available from the station would lead to sizeable error in the comparison (compare Figures 2a and 2b).

As an alternative to recent stochastic methods [*Eden et al.*, 2014; *Wong et al.*, 2014], SABC could also be used for simple, direct, downscaling of model data to the point scale—possibly circumventing the need for a statistical model. In the example of the previous paragraph, each value of (12 km, 5 min) model output would then be a proxy for an hourly average for a point measurement. An analysis of such downscaling is, however, left to a future study.

Finally, it may be promising to combine SABC with two-dimensional bias correction [*Piani and Haerter*, 2012; *Mehrotra and Sharma*, 2015]. Given the large amount of data necessary to populate 2-D histograms of dynamical variables, for example, temperature and precipitation, gridded data sets with sufficiently long time series are hard to come by. Using SABC, climate impact modelers can access the information in station

data directly without prior gridding. This has the potential to make 2-D bias correction a standard procedure and will be the focus of future work.

Appendix A: Sample Size Reduction

We comment on the possible choices for gauge-limited corrections: When multiple choices are available for possible coarsening, one relevant criterion for optimal choice might be loss of data samples in the statistics. When coarsening, using equivalent resolutions along the relevant ridge, the reduction of sample size is $R(\tilde{n}) \equiv \tilde{n}^2 \cdot \tilde{m}(\tilde{n})$. Dropping the tilde for simplicity, and noting that $m = \delta_t/\delta_{t0}$, $\delta_t = \delta_t^* - r\delta_x$, and $n = \delta_x/\delta_{x0}$, we have

$$R(n) = \frac{\delta_t^*}{\delta_{t0}} n^2 - \frac{r}{r_0} n^3, \tag{A1}$$

where we have defined $r_0 \equiv \delta_{t0}/\delta_{x0}$. Equation (A1) states that R initially increases as a function of n but decays for large n. Noting that $n, m \geq 1$, we have the bounds

$$n_{\min} \equiv 1 \leq n \leq \frac{\delta_t^* - \delta_{t0}}{r\delta_{x0}} \equiv n_{\max}. \tag{A2}$$

Intermediate choices of n yield extreme reductions of data, for the case studied in Figure 4b, $R_{\text{extr}} > 100$ is possible. Generally, $n = 1$ may be a reasonable choice (in our example $R(1) = 19$) but in some cases also $n = n_{\max}$ should be considered, especially when high temporal resolution, i.e., small δ_{t0}, is available. In those cases, $R(n_{\max}) < R(1)$ is possible, allowing a larger sample size to be preserved. In the plot, n ranges from unity to $n_{\max}$. Extremal R_{extr} occurs at intermediate values of n and can substantially exceed $R(1)$ and $R(n_{\max})$. Dependencies on system parameters: $R(1) = \delta_t^*/\delta_{t0} - r/r_0$. The value of n where extremal R is reached: $n_{extr} = 2r_0\delta_t^*/3r\delta_{t0}$; extremal R: $R_{extr} = 4r_0^2\delta_t^{*3}/27r^2\delta_{t0}^3$. In practice, however, possible degradation of PDF agreement, even along the ridge, should be evaluated when using large n. Assessment of this question should be left for future work.

Acknowledgments

The authors acknowledge the radar and gauge data from the German Weather Service (DWD). J.O.H. acknowledges financial support by the Danish National Research Foundation through the Center for Models of Life. B.E. acknowledges support from Climate Service Center 2.0, C.M. acknowledges support from the project HD(CP)^2, funded by the German Federal Ministry of Education and Research. P.B. acknowledges support from SMHI.

The Editor thanks two anonymous reviewers for their assistance in evaluating this paper.

References

Berg, P., H. Feldmann, and H.-J. Panitz (2012a), Bias correction of high resolution regional climate model data, *J. Hydrol.*, *448*, 80–92.

Berg, P., S. Wagner, H. Kunstmann, and G. Schädler (2012b), High resolution regional climate model simulations for Germany: Part I. Validation, *Clim. Dyn.*, *40*, 401–414, doi:10.1007/s00382-012-1508-8.

Berg, P., C. Moseley, and J. O. Haerter (2013), Strong increase in convective precipitation in response to higher temperatures, *Nat. Geosci.*, *6*, 181–185.

Berg, P., T. Bosshard, and W. Yang (2015), Model consistent pseudo-observations of precipitation and their use for bias correcting regional climate models, *118–132*, *3*, doi:10.3390/cli3010118.

Dee, D. P., et al. (2011), The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Q. J. R. Meteorol. Soc.*, *137*, 553–597, doi:10.1002/qj.828.

Deidda, R. (2000), Rainfall downscaling in a space-time multifractal framework, *Water Resour. Res.*, *36*(7), 1779–1794.

Doms, G., and U. Schättler (2002), A description of the nonhydrostatic regional model LM. Part I: Dynamics and numerics, *COSMO Newsletter*, *2*, 225–235.

Eden, J. M., M. Widmann, D. Maraun, and M. Vrac (2014), Comparison of GCM- and RCM-simulated precipitation following stochastic postprocessing, *J. Geophys. Res. Atmos.*, *119*, 11,040–11,053, doi:10.1002/2014JD021732.

Eggert, B., P. Berg, J. Haerter, D. Jacob, and C. Moseley (2015), Temporal and spatial scaling impacts on extreme precipitation, *Atmos. Chem. Phys. Discuss.*, *15*(2), 2157–2196.

Fosser, G., S. Khodayar, and P. Berg (2014), Benefit of convection permitting climate model simulations in the representation of convective precipitation, *Clim. Dyn.*, *44*, 45–60, doi:10.1007/s00382-014-2242-1.

Gudmundsson, L., J. Bremnes, J. Haugen, and T. Engen-Skaugen (2012), Technical note: Downscaling RCM precipitation to the station scale using statistical transformations—A comparison of methods, *Hydrol. Earth Syst. Sci.*, *16*(9), 3383–3390.

Haerter, J., S. Hagemann, C. Moseley, and C. Piani (2011), Climate model bias correction and the role of timescales, *Hydrol. Earth Syst. Sci.*, *15*(3), 1065–1079.

Lenderink, G., and E. Van Meijgaard (2008), Increase in hourly precipitation extremes beyond expectations from temperature changes, *Nat. Geosci.*, *1*(8), 511–514.

Li, H., J. Sheffield, and E. F. Wood (2010), Bias correction of monthly precipitation and temperature fields from intergovernmental panel on climate change ar4 models using equidistant quantile matching, *J. Geophys. Res.*, *115*, D10101, doi:10.1029/2009JD012882.

Maraun, D. (2013), Bias correction, quantile mapping, and downscaling: Revisiting the inflation issue, *J. Clim.*, *26*(6), 2137–2143.

Maraun, D., et al. (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.

Maurer, E. P., and D. W. Pierce (2014), Bias correction can modify climate model simulated precipitation changes without adverse effect on the ensemble mean, *Hydrol. Earth Syst. Sci.*, *18*, 915–925.

Mehrotra, R., and A. Sharma (2012), An improved standardization procedure to remove systematic low frequency variability biases in GCM simulations, *Water Resour. Res.*, *48*, W12601, doi:10.1029/2012WR012446.

Mehrotra, R., and A. Sharma (2015), Correcting for systematic biases in multiple raw GCM variables across a range of timescales, *J. Hydrol.*, *520*, 214–223.

Perkins, S. E., A. J. Pitman, N. J. Holbrook, and J. McAneney (2007), Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions, *J. Clim.*, *20*(17), 4356–4376, doi:10.1175/JCLI4253.1.

Piani, C., and J. Haerter (2012), Two dimensional bias correction of temperature and precipitation copulas in climate models, *Geophys. Res. Lett.*, *39*, L20401, doi:10.1029/2012GL053839.

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

Piani, C., G. Weedon, M. Best, S. Gomes, P. Viterbo, S. Hagemann, and J. Haerter (2010b), Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models, *J. Hydrol.*, *395*(3), 199–215.

Rocheta, E., J. Evans, and A. Sharma (2014), Assessing atmospheric bias correction for dynamical consistency using potential vorticity, *Environ. Res. Lett.*, *9*(12), 124010.

Steiner, M., J. A. Smith, and R. Uijlenhoet (2004), A microphysical interpretation of radar reflectivity-rain rate relationships, *J. Atmos. Sci.*, *61*(1), 1114–1131.

Taylor, G (1938), The spectrum of turbulence, *Proc. R. Soc. A*, *164*, 476–490.

Teutschbein, C., and J. Seibert (2012), Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods, *J. Hydrol.*, *456*, 12–29.

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

Weedon, G., S. Gomes, P. Viterbo, W. J. Shuttleworth, E. Blyth, H. Österle, J. C. Adam, N. Bellouin, O. Boucher, and M. Best (2011), Creation of the watch forcing data and its use to assess global and regional reference crop evaporation over land during the twentieth century, *J. Hydrometeorol.*, *12*(5), 823–848.

Wong, G., D. Maraun, M. Vrac, M. Widmann, J. M. Eden, and T. Kent (2014), Stochastic model output statistics for bias correcting and downscaling precipitation including extremes, *J. Clim.*, *27*, 6940–6959.

Wood, A. W., L. R. Leung, V. Sridhar, and D. Lettenmaier (2004), Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs, *Clim. Change*, *62*(1–3), 189–216.