ORIGINAL PAPER

Statistical bias correction for daily precipitation in regional climate models over Europe

C. Piani · J. O. Haerter · E. Coppola

Abstract We design, apply, and validate a methodology for correcting climate model output to produce internally consistent fields that have the same statistical intensity distribution as the observations. We refer to this as a statistical bias correction. Validation of the methodology is carried out using daily precipitation fields, defined over Europe, from the ENSEMBLES climate model dataset. The bias correction is calculated using data from 1961 to 1970, without distinguishing between seasons, and applied to seasonal data from 1991 to 2000. This choice of time periods is made to maximize the lag between calibration and validation within the ERA40 reanalysis period. Results show that the method performs unexpectedly well. Not only are the mean and other moments of the intensity distribution improved, as expected, but so are a drought and a heavy precipitation index, which depend on the autocorrelation spectra. Given that the corrections were derived without seasonal distinction and are based solely on intensity distributions, a statistical quantity oblivious of temporal correlations, it is encouraging to find that the improvements are present even when seasons and temporal statistics are considered. This encourages the application of this method to multi-decadal climate projections.

C. Piani (✉) · E. Coppola
Abdus Salam, International Center for Theoretical Physics,
Strada Costiera 11,
Trieste 34014, Italy
e-mail: cpiani@ictp.it

J. O. Haerter
Max Planck Institute for Meteorology,
Bundesstr. 53,
20146 Hamburg, Germany

1 Introduction

It is well known that general circulation model (GCM) precipitation output cannot be used to force hydrological or other impact models without some form of prior bias correction if realistic output is sought (Sharma et al. 2007; Hansen et al. 2006; Feddersen and Andersen 2005). The errors in GCM daily precipitation afflict the entire intensity spectrum: a low number of dry days, which are compensated by too much drizzle, a bias in the mean, and the inability to reproduce the observed high precipitation events (Boberg et al. 2007; Leander and Buishand 2007). It is customary for climate modelers to present future global or regional temperature or precipitation projections in terms of the relative changes in the statistics (Piani et al. 2007; Gutowski et al. 2007). For these projections to be translated into forcing fields for impact models, metadata with realistic statistics and which incorporate the projected statistical changes must be derived.

A realistic representation of precipitation fields in future climate projections from climate models is crucial for impact and vulnerability assessment (Semenov and Doblas-Reyes 2007; Schneider et al. 2007; Wood et al. 2004). Hence, crop modelers use bias correction techniques that correct all ranges of the intensity histogram (Biagorria et al. 2007). Often, this involves some form of transfer function derived from the observed and simulated cumulative distribution functions (cdfs) (for example in Ines and Hansen 2006). These methods are given a wide range of names in the literature: statistical downscaling, quintile mapping, and histogram equalizing, rank matching are among these. In this study, we refer to our method as a statistical bias correction. In applying a hindcast-derived correction to simulations of projected climate, one must assume that the correction still holds for the projected

climate, which is not a trivial assumption (Trenberth et al. 2003). This assumption is more palatable if the transfer function between raw and corrected GCM output is robust, which is the case if it depends on fewer parameters to be derived from the data. In this study, we develop a robust and practical statistical bias correction method, which we apply and validate using regional model output over Europe from the ENSEMBLES project. In Section 2, we describe the methodology, in Section 3 we present the results followed by discussion and conclusions in Section 4.

2 Methodology

The statistical bias correction method validated in this study is based on the initial assumption that both observed and simulated intensity distributions are well approximated by the gamma distribution:

$$\text{pdf}(x) = \frac{e^{\left(-\frac{x}{\theta}\right)} x^{(k-1)}}{\Gamma(k)\theta^k} \tag{1}$$

where x is normalized daily precipitation, where k and θ are the form and scaling parameter, respectively.

This is a generally accepted practice for observed daily precipitation where $k>1$ (for example: Wilks 1995; Katz 1999). In the case of simulated precipitation, there may be a case where $k=1$ (exponential distribution) or were the best fit is achieved with $k<1$. These are the cases with a high number of drizzle days and a rapid drop in occurrences for higher precipitation values. Of course, we cannot define the gamma distribution at $x=0$ when $k<1$ because it is unbounded there; hence, we chose to restrict the fit to $x>\varepsilon>0$ and add the number of dry days to the list of parameters in the correction method. As an example, let us assume that the pdf of simulated daily precipitation (excluding dry days) over a certain grid point is well fitted by the gamma distribution shown in Fig. 1a (solid line) with $k=1$ and $\theta=0.8$. Let us also assume that the dashed line in Fig. 1a fits the pdf of observed daily precipitation (excluding dry days) at the same grid point. This too is a gamma distribution but with $k=2$ and $\theta=0.7$. To construct a transfer function $y=f(x)$, where x and y are the simulated and corrected values of daily precipitation, respectively, and such that the distribution of y matches that of the observations, we proceed by plotting the cdfs for the simulated and observed variables, defined as:

$$\text{cdf}(x) = \int_0^x \frac{e^{\left(-\frac{x}{\theta}\right)} x'^{(k-1)}}{\Gamma(k)\theta^k} dx' + \text{cdf}(0) \tag{2}$$

where cdf(0) is the fraction of days with no precipitation and shown in Fig. 1b. The desired transfer function $y=f(x)$ obeys
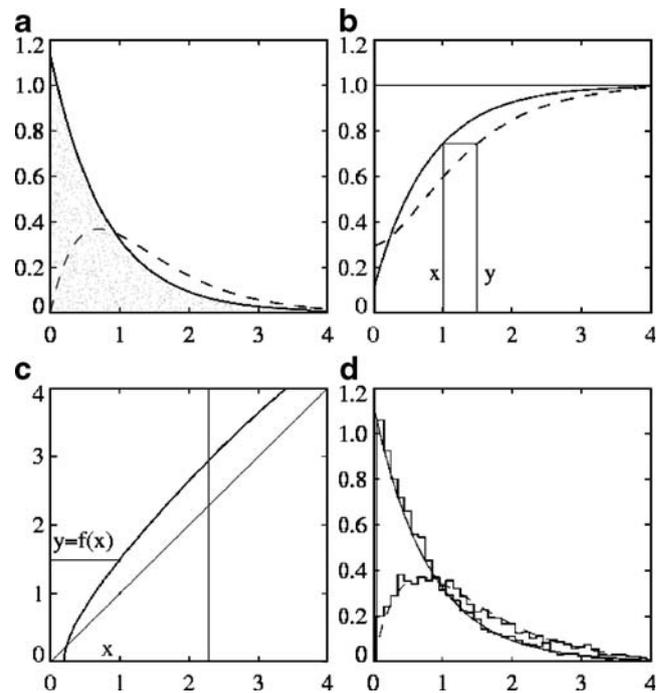
Fig. 1 Statistical correction applied to a synthetic dataset. **a** Synthetic pdf of simulated daily precipitation (*solid line*), synthetic pdf of observed daily precipitation (*dashed line*). **b** cdfs obtained by integrating the corresponding pdfs in **a**. **c** Transfer function obtained graphically from **b** by solving: cdf$_{obs}(y)$=cdf$_{sim}(x)$ (*thick solid line*). **d** Histogram of synthetic dataset given by x-coordinate of points evenly scatter under solid pdf in **a** superimposed onto the same pdf (*thin solid line*), histogram of transformed dataset f(x) superimposed onto dashed pdf from **a** (*thin dashed line*)

the equation: cdf$_{obs}$(f(x))=cdf$_{sim}$(x) and can be derived graphically as shown in Fig. 1b. The $y=$f(x) function itself is shown in Fig. 1c. The degree to which f(x) deviates from the $y=x$ line (also shown in Fig. 1c) is a measure of the difference between the observed and simulated pdfs.

For illustration purposes, the methodology was applied to a synthetic random data set. In Fig. 1a, the area below the solid pdf is populated by randomly distributed points with a constant surface density. Hence, the distribution of the x-coordinate of these points is well approximated by the pdf itself. We refer to this as our X' dataset. In Fig. 1d, we replot both the solid (simulated) and dashed (observed) pdfs. We then superimposed the histogram obtained from the X' dataset which, as expected, closely follows the simulated pdf. The X' dataset is then transformed according to the defined methodology; that is, we derive a new dataset given by Y'=f(X'). The histogram of Y' points is superimposed on Fig. 1d, and as anticipated, it follows the dashed (observed) pdf closely. We stress that this example does not constitute a validation of the correction methodology. It simply illustrates how the method works. A proper validation is carried out in the following section where the

transfer (or correction) function $y = f(x)$ is inferred using simulated and observed daily precipitation from a given time period and then applied to simulated data from a different time period and subsequently compared with observations.

The methodology described above was applied to the simulated daily precipitation data from the DMI regional model over Europe interpolated onto the CRU (Jones et al. 1999) 25-×25-km grid. The DMI regional climate model used for ENSEMBLES simulation is the HIRHAM model version 5. This model combines a limited-area high-resolution short-range weather forecasting module with a general circulation model (Christensen et al. 2008). The DMI model has undergone some improvements, mainly in the glacier parameterization, since these simulations were carried out (Christensen, personal communication). We stress that subsequent improvements to the model do not affect the results presented in this study, since we are addressing model error correction techniques. Observations were provided by the ENSEMBLES observational dataset (Haylock et al. 2008), which is defined on the CRU 25-×25-km grid as well. The transfer function $y = f(x)$ defined above was inferred using data from 1961 to 1970. Histograms were calculated for every grid point for both the DMI model and the observed daily precipitation. No subdivision in seasons was done at this point. The bin size was 0.4 mm/day, while the lower limit of the lowest bin was set at 0.01 mm/day. This was done to remove dry days from the statistics. For each grid point, the histograms of both observed and simulated daily precipitation were fitted with the two-parameter (k, θ) gamma distribution defined in Eq. 1. The fitting was done minimizing the square error (least squares), and a weighting equal to the value of precipitation itself was applied. The effects of using different weighting functions will be discussed in Section 4. Hence, for each grid point, we identified six parameters: k, θ and the number of dry days for both observed and simulated data. These six parameters allowed us to graphically derive the transfer function in the same way as described in Fig. 1c.

The transfer function was not obtained for all grid points. In some cases, missing observed data invalidated the automated procedure. This does not imply that the methodology failed or that it is not applicable. Rather, it indicates that reduced time segments or different bin sizes should be adopted specifically for those grid points. In this study, however, we chose to leave these grid points blank for clarity of evaluation. Clearly, no transfer function can be obtained for grid points where all observations are missing as in the case of sea points.

The transfer functions thus obtained were applied to the DMI model daily precipitation data from 1991 to 2000. The two decades furthest apart among those available to us were

chosen to validate this methodology to maximize the exposure of possible weaknesses. In the light of climate predictions, as they may be carried out by the same models in scenario computations, the use of these two decades also serves the purpose of testing how bias corrections obtained on data for the 'pre-climate shift' period (before 1975) perform in a 'post-climate shift' period (after 1977) (Trenberth et al. 2007). For the same reason, only one transfer function was derived for each grid point instead of separate ones for each season.

3 Results

In Fig. 2, we show the mean daily precipitation for the solstice seasons (DJF and JJA) for 1991 to 2000 inferred from the observed ENSEMBLES dataset and alongside that inferred from the corrected and raw DMI dataset. Again, we stress that the observed data shown in Fig. 2a, d played no role in inferring the data shown in Fig. 2b, e. This, however, is not a surprising result; indeed, it would be a poor bias correction if the mean of the corrected field looked nothing like the observations. In particular, we notice that, for both seasons, the biases over topography, which characterize most climate models, is all but removed (Norway, Alps and Pyrenees). This is also the case for northeastern Spain and the eastern Adriatic coast in winter. The variance of the corrected field (not shown) is also very similar to the observed. This too could have been anticipated, since mean and variance are statistical quantities that do not depend on the temporal statistics of the data (for example on the autocorrelation spectrum).

In Fig. 3, we show a meteorological drought index averaged over the same seasons and years as in Fig. 2. This drought index is solely based on precipitation (hence, the adjective meteorological) and consists of the average number of consecutive dry days (days with no precipitation) immediately preceding each day. We should point out that, for the purpose of calculating this index, a precipitation threshold was defined within the model below which a day was considered dry. This is because, without the threshold, the model index would be identically zero since simulated precipitation was always greater than zero. The threshold chosen was 0.01 mm/day for consistency with Section 2. The results are encouraging, given that such a drought index is highly dependent on the autocorrelation of the precipitation time series while the transfer function is not. The correction is particularly effective around the Mediterranean where it effectively reestablishes the observed gradients in the summer and the observed maxima over the Alps and southeastern Spain in the winter. There is a positive correction over northern Europe as well, but it is not as marked.
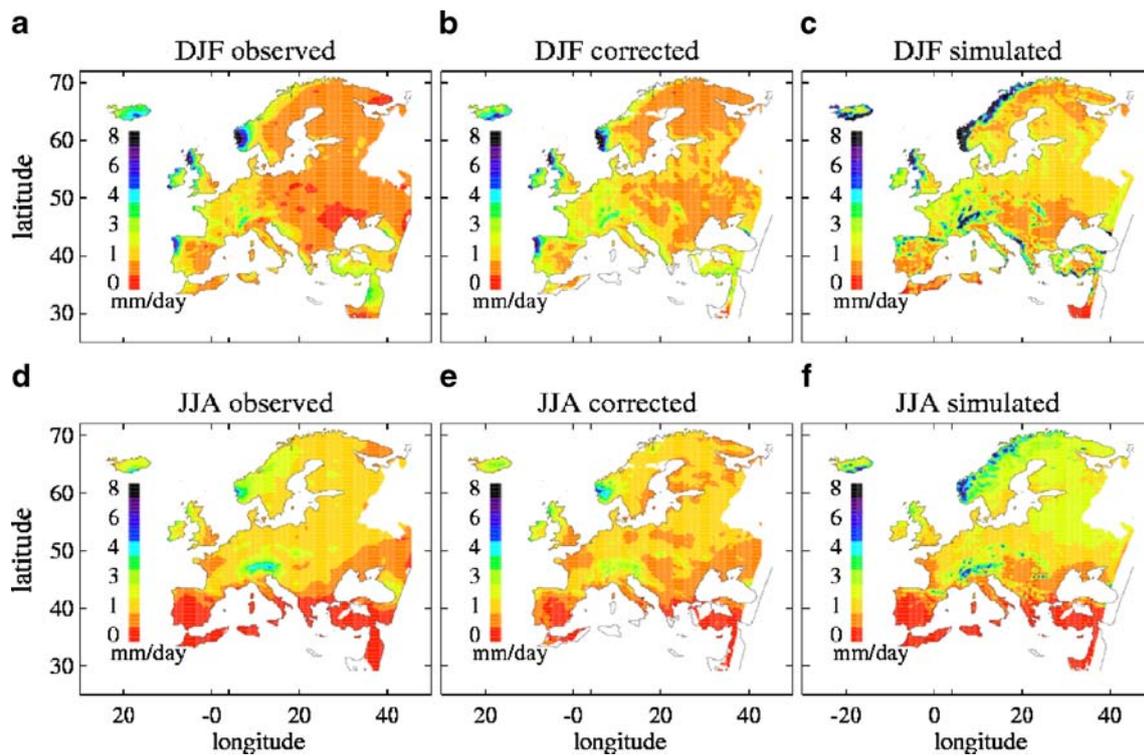
Fig. 2 Validation of methodology: seasonal mean daily precipitation. Application of bias correction, derived from simulated and observed data from 1961 to 1970, to model data from 1991 to 2000. **a** Mean observed daily precipitation for winter (DJF) 1991 to 2000, **b** same as **a** but for corrected simulated data, **c** same as **a** but for uncorrected simulated data. **d**–**f** Same as **a**–**c** but for summer (JJA)
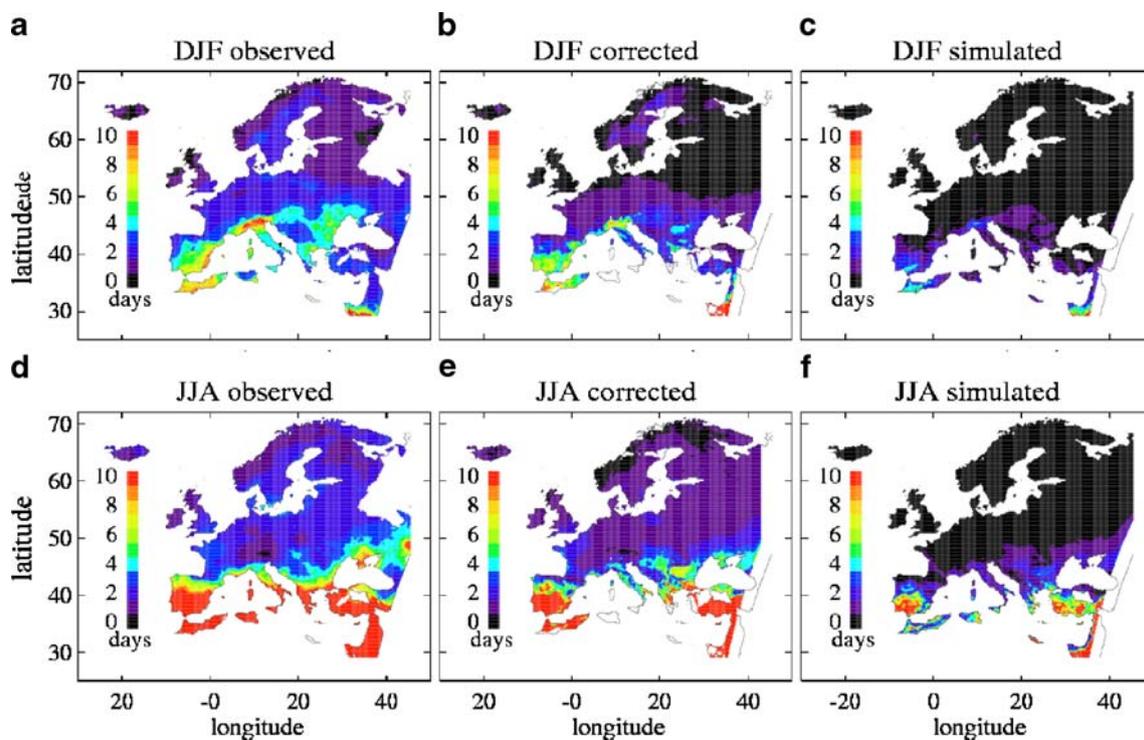
Fig. 3 Same as Fig. 2 but for seasonal mean drought index

In Fig. 4, we show a heavy precipitation index averaged over the same seasons and years as in Figs. 2 and 3. This index was calculated as follows. For each day, the number of immediately preceding consecutive days with heavy precipitation is considered. Heavy precipitation is defined differently for each grid point as twice the grid point decadal mean. Subsequently, daily precipitation is integrated over this range of days leading to the total precipitation for the entire "heavy precipitation event" preceding each day. As for the former, this index too is significantly affected by the temporal structure of the precipitation data. In this case, there is the added factor of the dimensionality of precipitation. We chose to validate with the actual measure of total precipitation, as opposed to the simple count of heavy precipitation days, because this would be the relevant quantity for hydrological purposes. The validation is positive though arguably not as good as in the case of the drought index. The bias over Norwegian topography is removed in both seasons. Central Europe is greatly improved in both seasons; most of the simulated local maxima are removed, though some of the unobserved maxima over the Transylvanian Alps remain. The winter distribution in Spain is also significantly improved.

4 Discussion and conclusions

It is important to note that the same corrected precipitation field was used to produce panels b and e in Figs. 2, 3, and 4.

This implies that the field can be used directly as input to hydrological models. This is crucially not the case when simple additive or multiplicative bias corrections are used. Arguably, the methodology presented in this study could be tailored a posteriori to give better results when applied to these particular model simulations; for example, histogram bin size and the fitting algorithm could be changed to minimize the validation error. Of course a posteriori optimizations would require further validation possibly involving simulations with other models. This is part of a planned work schedule including sensitivity studies to algorithm parameter settings. One experiment was done to recalculate the heavy precipitation index using a non-weighted least-square fitting algorithm for the pdfs. As expected, because this decreases the weight on the tail of the distributions were absolute errors translate into large relative errors, the corrected data showed little or no improvement in the high precipitation index. Results would also improve if seasons were corrected separately. With the present dataset, this may have made sense, but one of the aims of this study was to show that the method has potential even when the corrections are calculated without seasonal distinction. This will have to be the adopted procedure when limited observational datasets are available. A further experiment will be to evaluate the improvement of the hydrological model simulation with and without the bias-corrected precipitation field (Leander and Buishand 2007; van der Linden and Christensen 2003).
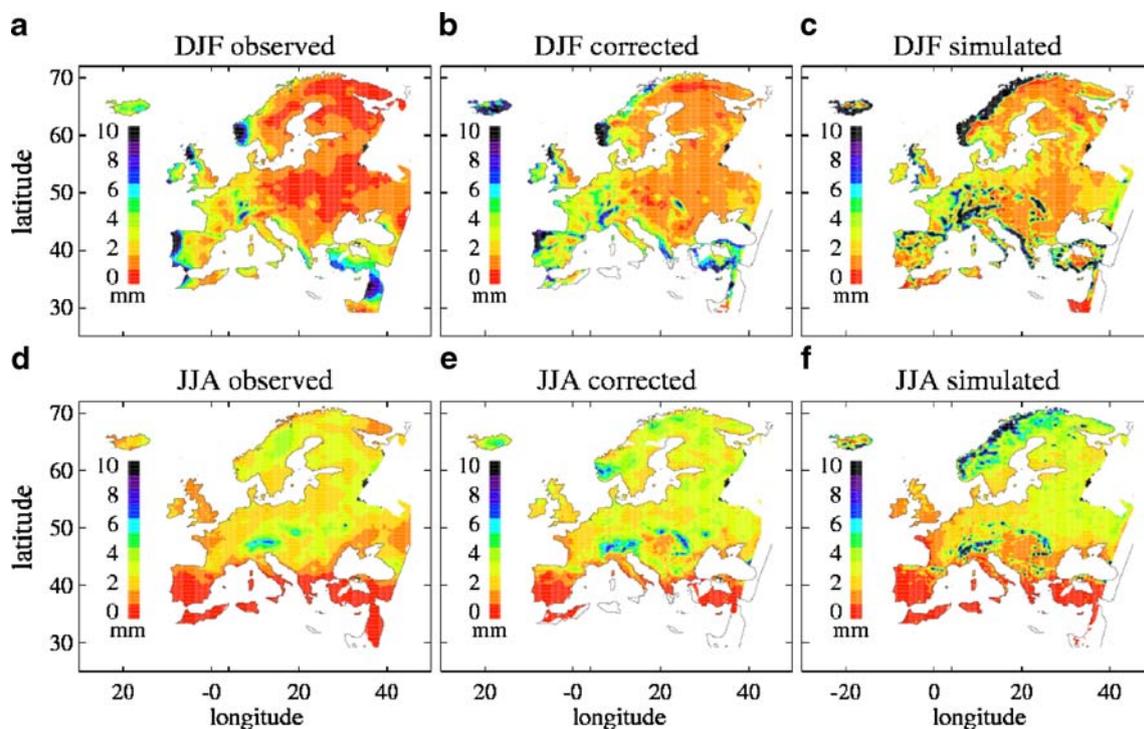
Fig. 4 Same as Fig. 2 but for seasonal mean heavy precipitation events