Hempel, S., Frieler, K., Warszawski, L., Schewe, J., Piontek, F. (2013): A trend-preserving bias correction - the ISI-MIP approach. - Earth System Dynamics, 4, 2, 219-236

DOI: 10.5194/esd-4-219-2013

Earth Syst. Dynam., 4, 219–236, 2013
www.earth-syst-dynam.net/4/219/2013/
doi:10.5194/esd-4-219-2013

Earth System Dynamics
Open Access

A trend-preserving bias correction – the ISI-MIP approach

S. Hempel, K. Frieler, L. Warszawski, J. Schewe, and F. Piontek

Potsdam Institute for Climate Impact Research, Potsdam, Germany

Correspondence to: S. Hempel (sabrina.hempel@pik-potsdam.de)

Received: 15 January 2013 – Published in Earth Syst. Dynam. Discuss.: 21 January 2013
Revised: 7 June 2013 – Accepted: 23 June 2013 – Published: 31 July 2013

Abstract. Statistical bias correction is commonly applied within climate impact modelling to correct climate model data for systematic deviations of the simulated historical data from observations. Methods are based on transfer functions generated to map the distribution of the simulated historical data to that of the observations. Those are subsequently applied to correct the future projections. Here, we present the bias correction method that was developed within ISI-MIP, the first Inter-Sectoral Impact Model Intercomparison Project. ISI-MIP is designed to synthesise impact projections in the agriculture, water, biome, health, and infrastructure sectors at different levels of global warming.

Bias-corrected climate data that are used as input for the impact simulations could be only provided over land areas. To ensure consistency with the global (land + ocean) temperature information the bias correction method has to preserve the warming signal. Here we present the applied method that preserves the absolute changes in monthly temperature, and relative changes in monthly values of precipitation and the other variables needed for ISI-MIP. The proposed methodology represents a modification of the transfer function approach applied in the Water Model Intercomparison Project (Water-MIP). Correction of the monthly mean is followed by correction of the daily variability about the monthly mean.

Besides the general idea and technical details of the ISI-MIP method, we show and discuss the potential and limitations of the applied bias correction. In particular, while the trend and the long-term mean are well represented, limitations with regards to the adjustment of the variability persist which may affect, e.g. small scale features or extremes.

1 Introduction

Climate simulations of historical periods often show systematic deviations from the observed climate resulting, for example, from imperfect model representations of the atmospheric physics, incorrect initialisation of the model or errors in the parameterisation chain (Ehret et al., 2012). These deviations must be treated carefully in the context of climate impact simulations, because the predicted impacts depend on the statistical properties of the climate input. While considering anomalies of impact projections with respect to a reference period might provide a way out in case of a linear dependence of impacts on climate input data, in many other cases this is not appropriate, e.g. when impacts are activated when certain absolute climatic thresholds are exceeded. Moreover, impact models (e.g. crop models, hydrological models, etc.) often require driving climate data that is statistically similar to the observational datasets with which they were calibrated.

Bias correction methods are designed to bridge the gap between the information that is provided by the climate modelling community and the climate data necessary for quantitative climate impact projections. Basic bias correction methods include an adjustment of the mean value by adding a temporally constant offset, or by applying an associated correction factor to the simulated data. This additive or multiplicative constant quantifies the average deviation between the simulated and the observed time series over the historical period. Since the constant is time independent such a method preserves the trend (in absolute terms for an additive approach and in relative terms for a multiplicative approach) whilst adjusting the mean value. However, it does not necessarily correct the variability of the data. Hence, in many cases differences in the variance or even higher moments of the simulated data are adjusted to the observations

Published by Copernicus Publications on behalf of the European Geosciences Union.

by parametric or non-parametric (empirical) quantile mapping (Boe et al., 2007; Piani et al., 2010; Themeßl et al., 2011). Non-parametric methods have been shown to be very successful in mapping simulations to the present climate owing to there flexibility (Gudmundsson et al., 2012). However, to correct values that fall outside the calibration range additional assumptions have to be made. While this is less critical as long as the correction is applied to historical simulations it might be an oversimplification when it comes to its application to future climate data, where a substantial portion of the event distribution may be shifted outside the calibration range. For this reason we prefer a parametric approach for our application.

Bias correction, however, must not be confused with a spatial downscaling. The correction of the misrepresented local variability is limited, since a disaggregation of the simulation data cannot be performed by a purely deterministic approach. If the resolution of the simulations and observations are considerably different high extremes are usually exaggerated while low events are overcorrected (Maraun, 2013). Moreover, whilst adequately representing the mean state of the observed period and the variability at a particular time scale, these bias correction methods may change the climate signal, or trend, arising from the climate simulations. The impact of bias correction on the climate signal is only rarely explicitly quantified and whether or not adjustment of the climate signal is advisable remains a topic of discussion (Ehret et al., 2012). In any case bias correction is tantamount to introducing a new level of uncertainty comparable in magnitude to the spread of the climate projections across the climate models or with regards to the emission pathways (Hagemann et al., 2011). The choice of an appropriate methodology depends strongly on the context. A review of state-of-the-art bias correction methods is given by Maraun et al. (2010).

Statistical bias correction of simulation data is broadly applicable to the climate impacts research (Robock et al., 1993; Berg et al., 2003; Ines and Hansen, 2006; Hagemann et al., 2011; Dosio and Paruolo, 2011), since it offers crucial advantages for impact modelling applications compared to using raw climate model output:

1. Statistical bias correction methods facilitate the comparison of observed and simulated impacts during the historical reference period and a continuous transition into the future. Without such an adjustment of the mean behaviour in the historical period, future impacts that depend on the exceedance of critical absolute thresholds of, for example, temperature (Rötter et al., 2011), cannot be accurately described. Studying the change in impacts starting from the reference level provided by a climate model would in general result in a mistiming of the threshold exceedance under global warming scenarios.

2. Many bias correction techniques include an implicit downscaling of the simulated data to the potentially higher resolution of the observational data. While a simple interpolation to the finer grid would not account for the increase in variability expected for the higher resolution data, an appropriate increase can be achieved by a bias correction method that adjusts the variance. In the general case, however, this adjustment will be limited since the temporal structure is still determined by the dynamics represented in the larger grid box and do not describe local phenomenons, e.g. small scale turbulence.

3. Bias correction also serves as a way to adjust the simulated climate data to the more detailed altitude-stratified information associated with observational data, so long as changes in mean and variability are resolved in the observational dataset.

On the other hand, there are several shortcomings of statistical bias correction:

1. Stationarity in the bias in the historical data with respect to the future data is assumed when applying the bias correction to future periods, which introduces additional uncertainty (Raisanen and Raty, 2012; Maraun, 2012).

2. The quality of the bias-corrected simulation data is limited by both the observational dataset and the representation of physical processes within the climate model.

3. Statistical bias correction (e.g. by adding the mean deviation from the observed data to the simulated one) often destroys the physical consistency of the different climate variables. For example, after the application of bias correction the temperature might be sub-zero, whereas rainfall is not converted into snowfall.

While the second issue can be tackled to a certain extent by testing the sensitivity to different sources of observational data, differentiation between statistical and phenomenological errors is not straightforward. With respect to the third issue, bivariate parametric quantile mapping was recently introduced by Piani and Haerter (2012) to provide consistency between temperature and precipitation corrections (not implemented in our present study). However, no multivariate approach exists that preserves the consistency between more than two variables.

In regional studies a way to overcome this third major deficit is to use dynamical downscaling in addition to statistical bias correction. In this approach, physical consistency is ensured by bias-correcting low resolution model data (e.g. sea surface temperatures) in order to provide correct boundaries. Subsequently this data is used to drive a higher resolution regional climate model (RCM) or a global circulation model (GCM) with locally enhanced resolution (Xu and Yang, 2012; Holland et al., 2010; Patricola and Cook, 2010; Cook and Vizy, 2008; Sato et al., 2007; Wu and Lynch, 2000). This does not necessarily solve the problem since the RCM also has a bias, for example, caused by inconsistencies between the physics of GCM and RCM, imperfect parameterisation or incorrect energy balance closure (Ehret et al., 2012). However, the two-step procedure is expected to reduce the deviation between high-resolution simulations and observations while ensuring physical consistency of different climate variables as provided by the high resolution model (Ehret et al., 2012).

The Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP) is designed to provide a consistent set of *global* impact simulations. Thus, within the ISI-MIP context a similar regional approach is not feasible as the involved impact models need climate input data that cover the entire global land area. The project relies on the relatively low-resolution GCM runs performed in the fifth phase of the Coupled Model Intercomparison Project (CMIP5, Taylor et al., 2012[1]). In that context, the described advantages of bias correction are essential to the project, which is intended to synthesise impact projections in multiple sectors at different levels of global warming. However, a complete bias correction directly from the simulations may not be advisable everywhere (Eden et al., 2012). There are regions where, e.g. the simulated precipitation is so wrong that a statistical bias correction with a transfer function may result in an even worse dataset as some extremes are very much amplified in order to adjust the parameters of the distribution. Different thresholds are incorporated in our bias correction algorithm to restrict the modifications in such cases.

Moreover, the algorithm which we propose is designed to preserve the long-term trend in the GCM data, and hence, for example, its climate sensitivity. For some applications it might be desirable to modify the trend of the GCM, however, this introduces a new level of uncertainty that we would like to avoid within the ISI-MIP context. For temperature (T), conservation of the absolute trend is essential in ensuring consistency between the projected global mean temperature change (land + ocean), based on the non-bias-corrected data, and the bias-corrected warming signal over land areas that is used as input by the impact models, since no bias correction is performed over oceans. As ISI-MIP intends to quantify the climate change impacts and the related uncertainty at different levels of global warming that aspect is particularly important in the project's context and relevant for decision makers wishing to better quantify possible consequences of specific temperature targets. In addition, modification of the (local and) global temperature trend would also modify the climate sensitivity, which would not be well founded based on the available 40 yr observational dataset. More generally, we choose to conserve the trend in other climate variables in order to ensure a transparent method, with some control over the GCM properties that are preserved. However, due to positivity constraints for some variables, we preserve the relative rather than absolute trend in those cases. For the same reason, other bias correction methods also applied multiplicative correction factors instead of additive constants to correct, for example, precipitation data (Ines and Hansen, 2006). A multiplicative correction of the monthly precipitation data (P), and an additive correction of the temperature data, conserve the hydrological sensitivity, i.e. the relative change in precipitation [%] with respect to absolute temperature changes [K] at each grid point.

In Sect. 2 we describe the climate model and observational datasets which are relevant for the ISI-MIP project. The details of the ISI-MIP bias correction are outlined in Sect. 3. We explain our methodology and describe the properties of the bias-corrected climate data exemplarily for the HadGEM2-ES GCM.

In Sect. 4 we demonstrate that the climate signal is preserved in comparison to the original method proposed by Piani et al. (2010) and discuss how well the statistical moments of the bias-corrected data match the observations during the reference period. In case of precipitation we compare the ISI-MIP dataset with an updated version (ISI-MIP extended) where we improved the adjustment of the variability of daily data about the monthly mean and corrected a bug in the code. This issue affects the variability of the daily data, but not the correction of the monthly means (cf. Sect. 4.3 for the results of the extended algorithm compared to the ISI-MIP precipitation).

2 Climate input data

The ISI-MIP dataset comprises bias-corrected daily data for the variables[2] listed in Table 1.

Simuations

We use data from five GCMs from the CMIP5 archive as input: HadGEM2-ES, IPSL-CM5A-LR, MIROC-ESM-CHEM, GFDL-ESM2M, and NorESM1-M. These five models were selected based on the availability of daily data for the required variables covering the period from 1 January 1950 to 31 December 2099 – historical and all Representative Concentration Pathway (RCP) scenarios (Moss et al., 2010) – in the CMIP5 archive at the beginning of the project.

The available climate model outputs are bi-linearly interpolated in space to a $0.5° \times 0.5°$ grid. The time series are linearly interpolated to the standard Gregorian calendar (365 days per year plus leap days) wherever necessary.

[1] cf. http://cmip-pcmdi.llnl.gov/cmip5/

[2] Surface pressure is derived from sea-level pressure, temperature and height assuming adiabatic conditions, since no daily data was available for surface pressure in the CMIP5 archive.

Table 1. Bias-corrected variables in the ISI-MIP dataset.

variable name	abbreviation	symbol
average temperature[1]	tas	T
minimum temperature[3]	tasmin	T_m
maximum temperature[3]	tasmax	–
total precipitation[2]	pr	P
snowfall[3]	prsn	S
shortwave radiation[2]	rsds	–
longwave radiation[2]	rlds	–
near-surface wind speed[2]	wind	W
near-surface eastward wind[3]	uas	K
near-surface northward wind[3]	vas	–
surface pressure[2]	ps	–

[1] (additive), [2] (multiplicative), and [3] (indirect) according to the applied bias correction approach. The last column refers to the symbols used in the algorithmic description.

Observations

We use the WATCH Forcing Data (WFD, Weedon et al., 2011) for the period from 1 January 1960 to 31 December 1999 (the reference period) as an observation-based reference dataset. It is a combination of the ERA-40 daily data, the 40 yr reanalysis of the European Centre for Medium-Range Weather Forecasts (ECMWF), and the Climate Research Unit TS2.1 dataset (CRU), that provides observed time series of month-by-month variations in the climate over the last century on a high resolution grid (0.5°). The ERA-40 dataset provides day-to-day variations but on a lower resolution grid (2.5°). Both datasets overlap for the 40 yr reference period.

The WFD are available on the 0.5° grid over land area points using the land-sea mask from the CRU, excluding Antarctica. It approximates the daily variability of different climate variables. A correction for the elevation differences between ERA-40 and CRU is included in the WFD. Additionally, the monthly mean for precipitation is corrected with the Global Precipitation Climatology Centre full dataset version 4 (GPCC) to account for the systematic underestimation of precipitation measurements in the WFD (cf. Hagemann et al., 2011). Thus, the WFD combines the daily statistics of ERA-40 with the monthly mean characteristics of CRU and GPCC datasets and represents a complete gridded observational dataset for bias correction of global climate data over land.

3 The trend-preserving bias correction method

In the following we describe our bias correction method, which preserves the long-term absolute (relative) trend of the simulated temperature (precipitation, pressure, radiation, wind) data. The method modifies the daily variability of the simulated data about their monthly means to match the observed daily variability. The monthly variability and mean are corrected only using a constant offset or multiplicative correction factor that corrects for long-term differences between the simulated and observed monthly mean data in the historical period. In this way the absolute or relative trend of the simulation data is preserved.

We present and discuss the properties of the bias-corrected temperature in the ISI-MIP dataset, and compare the results of two versions of the multiplicative algorithm exemplarily for precipitation: a basic version that was used to produce the ISI-MIP climate input (hereafter denoted ISI) and a corrected and extended version (hereafter denoted ISIe) that overcomes several limitations in adjusting the daily variability. We focus on the extended version, whilst noting crucial departures from the basic version (cf. Table 2 for comparison of the extended algorithm with the ISI-MIP dataset and the WATCH approach – on which the ISI-MIP method is based).

The correction of the daily variability is described by calendar-month and grid-cell-specific transfer functions that are applied to the daily simulated data. In what follows we select the April values for the grid cell corresponding to 55.75° N, 68.25° W (hereafter referred to as "example grid cell") for illustration of the method. Similar results can be obtained for other months and regions. We will not index the grid cell or the selected month for which the transfer function is created. Thus, let X_{ij}^{data} denote the April value for year i and day j at one particular grid cell of the simulated (data = GCM) or observational (data = WFD) time series, where $X = T$ for daily average temperature and P for precipitation. In addition, X_i^{data} describes the monthly mean at that grid cell. Residual data is denoted by $\Delta X_{ij}^{\mathrm{data}}$, while $\delta X_{ij}^{\mathrm{data}}$ refers to normalised data. Bias-corrected simulation data is denoted $\tilde{X}_{ij}^{\mathrm{GCM}}$ (daily) or $\tilde{X}_i^{\mathrm{GCM}}$ (monthly).

3.1 Correction of monthly mean data

The first step, is to adjust for the long-term differences between the simulated and observed monthly mean data during the historical period. The daily variability about the monthly mean remains unchanged at this stage.

3.1.1 Temperature: additive correction

For temperature we add to the entire time series a constant offset C that is equal to the average difference between the observations and the simulations during the 40 yr reference period,

$$C = \left(\sum_{i=1}^{m=40} T_i^{\mathrm{WFD}} - \sum_{i=1}^{m=40} T_i^{\mathrm{GCM}} \right) \Big/ 40 \,, \tag{1}$$

as is demonstrated in Fig. 1. The corrected temperature is then

$$\tilde{T}_{ij}^{\mathrm{GCM}} = C + T_{ij}^{\mathrm{GCM}}, \tag{2}$$

Table 2. Comparison of bias correction algorithm for precipitation (multiplicative approach), main algorithmic steps.

	WATCH	ISI-MIP FAST-TRACK	ISI-MIP extended
define dry months	WFD threshold	WFD threshold GCM threshold	WFD threshold GCM threshold
define dry days	WFD threshold GCM threshold	WFD threshold GCM threshold	WFD threshold GCM threshold
define outlier days	WFD and GCM outside 99 % (Gauss)	WFD and GCM outside 99 % (Gauss)	none
redistribute precipitation drizzle	no	no	uniformly over wet days (additive)
normalise daily values	no	using mean of all days of the month	using mean of wet days of the month
select values for fitting based on	rank ordered timeseries	rank ordered timeseries	rank ordered normalised timeseries
criteria for choice of fitting algorithm	predefined parameter thresholds or convergence of nonlinear fit	predefined parameter thresholds or convergence of nonlinear fit	convergence of nonlinear fit
hierarchy of possible transfer functions $g(x)$ cf. Fig. 6 red curve	1. linear fit 2. exponential fit initialized with linear fit 3. exponential fit initialized with linear fit fixed slope 4. only multiplicative monthly mean correction 5. only additive monthly mean correction	1. linear fit 2. exponential fit initialized with linear fit 3. exponential fit initialized with linear fit fixed slope 4. only multiplicative monthly mean correction 5. only additive monthly mean correction	1. exponential fit initialized with identity line 2. exponential fit initialized with linear fit 3. linear fit
fit function and application based on the same set of data	no	no	yes
preserve relative trend	no	yes	yes
adjust long-term mean	with transfer function (mixture of time scales)	with mean ratio c $(0 \leq c \leq 10)$	with mean ratio c $(0.1 \leq c \leq 10)$
adjust variability	with transfer function (mixture of time scales)	partially	with transfer function
truncation at upper bound	no	yes	yes

which preserves the absolute change in temperature in the simulations, i.e.

$$\tilde{T}_{ij}^{\mathrm{GCM}} - \tilde{T}_0^{\mathrm{GCM}} = T_{ij}^{\mathrm{GCM}} - T_0^{\mathrm{GCM}}, \tag{3}$$

where T_0^{GCM} and $\tilde{T}_0^{\mathrm{GCM}}$ are the uncorrected and corrected reference temperatures.

The method is the most basic temperature correction regularly applied in impact studies (e.g. called "unbiasing method" in Deque, 2007). It preserves the absolute trend, and the variability of the simulated data at all time scales.

3.1.2 Precipitation: multiplicative correction

Given the positivity constraints on precipitation data, a similar additive approach is not appropriate. Instead we correct the monthly mean precipitation values using a multiplicative factor, which is defined:

$$c = \sum_{i=1}^{m=40} P_i^{\mathrm{WFD}} \Bigg/ \sum_{i=1}^{m=40} P_i^{\mathrm{GCM}} \,. \tag{4}$$

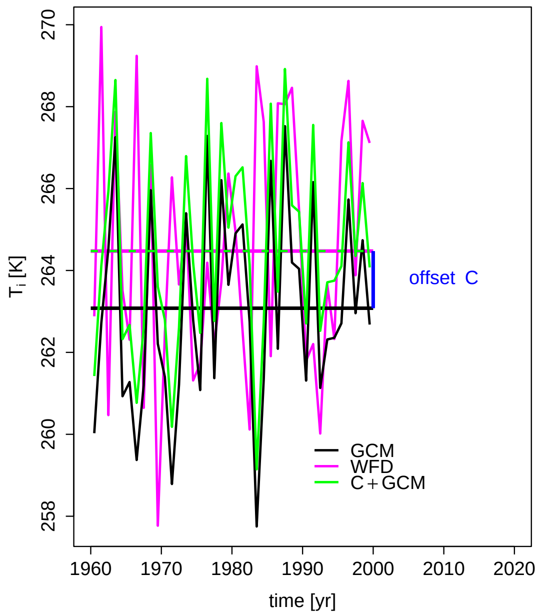

Fig. 1. April temperature means for the example grid cell during the reference period. The offset between observational and simulated data, C, in the reference period is illustrated, together with the shifted GCM data. The horizontal lines refer to the associated long-term means.

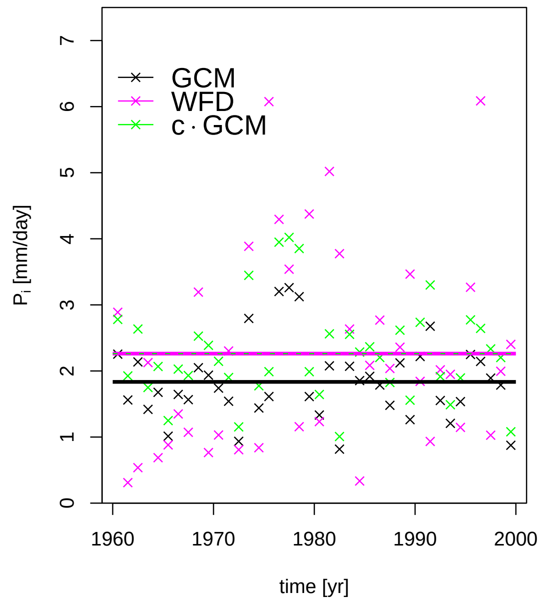

Fig. 2. April precipitation means for the example grid cell during the reference period. WFD (red), uncorrected (black) and scaled GCM (green) data are shown. The horizontal lines refer to the associated long-term means.

The precipitation time series is then

$$\tilde{P}_{ij}^{\mathrm{GCM}} = c \cdot P_{ij}^{\mathrm{GCM}}, \tag{5}$$

which maps the 40 yr mean of the GCM data to the observational one as demonstrated in Fig. 2. We impose an upper bound of 10 on c, in order to avoid unrealistically high precipitation values. This is justified by the fact that a very high c indicates a large discrepancy between the model and the observations. Possible reasons might be that the available time series is too short to well approximate the statistical properties or crucial physical processes are not included in the model. In those cases correcting the time series with the estimated values might lead to nonphysical values which we seek to avoid by truncation of c. In addition, in the extended version of the algorithm c is also truncated at the lower end following the same line of reasoning. This allows for the possibility that the model output in very dry regions can still get wetter in the future, since c cannot be zero over the reference period anymore.

The proposed multiplicative approach, modifies the simulated absolute precipitation change, but preserves the relative change in precipitation,

$$\frac{\tilde{P}_{ij}^{\mathrm{GCM}} - \tilde{P}_{0}^{\mathrm{GCM}}}{\tilde{P}_{0}^{\mathrm{GCM}}} = \frac{P_{ij}^{\mathrm{GCM}} - P_{0}^{\mathrm{GCM}}}{P_{0}^{\mathrm{GCM}}}, \tag{6}$$

where P_0^{GCM} and $\tilde{P}_0^{\mathrm{GCM}}$ are the uncorrected and corrected reference precipitation values.

3.2 Correction of daily variability

The second step, is to correct the daily variability of the simulated data to that of the observational dataset. This step is crucial for a proper representation of many impacts that depend on changes in both the mean and variability of the data: in this way, extreme weather events are better represented in the corrected data, although a careful analysis requires better understanding this important topic. Adjustment of daily variability also plays an important role when the climate data are interpolated to a finer grid before use by the impact model, which is often the case. Simple interpolation cannot account for the enhanced temporal variability that is expected at smaller spatial scales. Bias-correcting the variability of the interpolated data can alleviate this problem.

In the following section, we present a method to adjust the daily variability of the residual temperature

$$\Delta T_{ij}^{\mathrm{GCM}} = T_{ij}^{\mathrm{GCM}} - T_{i}^{\mathrm{GCM}}, \tag{7}$$

and the normalised precipitation data

$$\delta P_{ij}^{\mathrm{GCM}} = P_{ij}^{\mathrm{GCM}} \Big/ P_{i}^{\mathrm{GCM}}. \tag{8}$$

Note that this means that a specific correction value will refer to an anomaly instead of an absolute value. Thus, the same correction value for temperature is applied to different absolute temperatures and consequently to different weather situations, which may exhibit systematic differences in variability. Generally the variations of the temperatures are not expected to be very large as the correction is done on a calendar-month specific basis. Nevertheless, any dependence of the daily variances on the monthly mean values evident in the model or observed data are not perpetuated by the bias correction. In order to characterise this dependence, we have calculated the mean and the variance over each month in the reference period and estimated the coefficient of determination (R-squared) from the linear Pearson correlation between mean and variance at each grid point. For temperature the R-squared values are low for the entire global land area, indicating that the application of a common variance correction factor is consistent with both the observational and GCM data. For temperature, variance is based on the residual temperature, whilst for precipitation variances are calculated after normalising the daily data with the monthly mean. In the Supplement we show that for temperature and precipitation the R-squared values in this case are small at each grid point for both an exemplary GCM run and the WFD (cf. Fig. S1). This indicates that the residual and relative variance respectively over a month is almost independent from the related monthly mean. We therefore conclude that considering the anomalies instead of absolute values should not greatly impair the performance of the bias correction.

Moreover, in the case of precipitation, special care must be taken to account for low-precipitation (hereafter referred to as "dry") months. The correction of the daily variability comprises two steps: (1) correction of the frequency of dry days and (2) correction of the intensity of precipitation on rainy days. The proposed correction of the variability in daily data extends the method described by Piani et al. (2010) and applied in Water-MIP (Hagemann et al., 2011).

3.2.1 Temperature: linear regression

In order to correct the variability of the daily average temperature values to the observational data, we adjust the residual distribution of the GCM (cf. Eq. 7 for the definition of the residuals) to that of the WFD using a parametric quantile mapping. In general temperature values are considered to follow a normal distribution. This means the distribution is expected to be well described by only two moments (mean and standard deviation). For that reason a linear fit is considered an appropriate approximation in most cases and has thus been chosen to map the simulated to the observational temperature values.

Histograms of example time series from the WFD and GCM are shown in Fig. 3 for the April values at the example grid cell. We derive a transfer function

$$f\left(\Delta T^{\mathrm{GCM}}\right) = B \cdot \Delta T^{\mathrm{GCM}} \tag{9}$$

where B is the slope of a linear regression on the rank ordered WFD (ΔT^{WFD}) and GCM data (ΔT^{GCM}) for a given calendar-month over the 40 yr reference period (as plotted as black points in Fig. 4). An analogous procedure is described in Haerter et al. (2011), except that they allow for an additional offset, which we set here to zero, since the residual values have zero mean by definition.

3.2.2 Precipitation: nonlinear regression

In the case of precipitation we consider normalised values (cf. Eq. 8) to adjust the variability about the monthly mean, where both datasets should be described by the same distribution function. As in previous bias correction applications (e.g. in Water-MIP), we assume that the observational and simulated datasets are well approximated by a gamma distribution (excluding the days with zero precipitation). Following that assumption, we must correct the frequency and the intensity of precipitation separately, since the gamma distribution is not defined at zero. We perform a parametric quantile mapping with three parameters to adjust the intensity of precipitation, where a nonlinear fitting algorithm based on the gradient-expansion method adapted from Marquardt (1963) is used.

In dry months (zero mean or very small, i.e. in the range of measurement noise) a normalisation by the monthly mean is not possible. To solve this dry month problem we define threshold values for the monthly means,

$$\epsilon_{\mathrm{m}} = \max\left[P_k^{\mathrm{GCM}}\middle|\left(P_k^{\mathrm{WFD}} \leq 0.01,\ P_k^{\mathrm{GCM}} \leq 0\right)\right], \tag{10}$$

to classify the months into dry and wet, where the daily variability is only adjusted for the wet ones. The variables P_k^{WFD} and P_k^{GCM} represent the rank ordered sets of monthly precipitation values P_i^{data}. A similar procedure was described by Piani et al. (2010) for dry days.

Months with mean precipitation below 0.01 mm day^{-1} in the WFD (roughly 3.6 mm yr^{-1}, which approximates average precipitation in desert areas) are denoted as dry. Then we consider two cases: (i) if there are more dry months in the WFD than months with zero precipitation in the GCM, months are excluded in order of increasing monthly mean precipitation until the desired number (i.e. number of dry months in the WFD) is met, starting from the driest GCM month. (ii) If the number of months with zero precipitation in the GCM is larger than the number dry months in the WFD, only the months with zero precipitation in the GCM are classified as dry in the GCM. By applying Eq. (10) we ensure that the same number of months from the GCM and the WFD set are omitted. In both cases, the mean precipitation of the last month to be excluded in the GCM defines the threshold ϵ_{m} for the simulated monthly time series.

All daily data associated with a dry month (i.e. $P_{ij}^{\mathrm{GCM}}|(P_i^{\mathrm{GCM}} \leq \epsilon_{\mathrm{m}})$) are excluded from the estimation of the transfer function. The variability of the daily data belonging to these dry months is not modified. For the remaining wet months the bias correction proceeds in two stages: (i) increasing the frequency of dry days where needed.

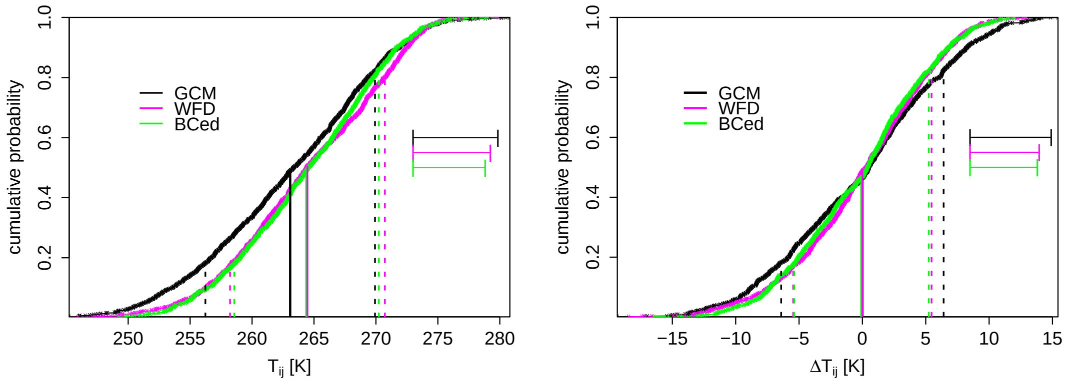

Fig. 3. Observational and simulated daily April temperature values during the reference period (left) and associated residual values (right) for example grid cell are shown as normalised cumulative sum. The vertical lines refer to related mean (solid) and mean ± standard deviation (dashed). Horizontal bars are for comparison of the standard deviation.

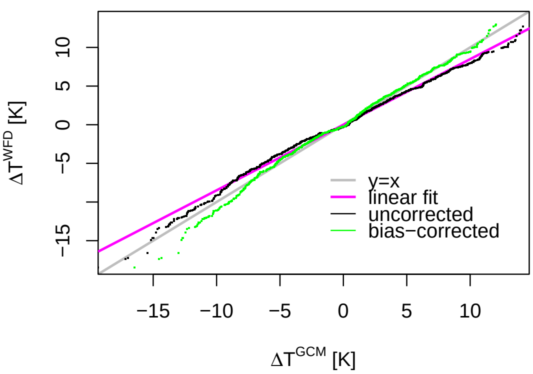

Fig. 4. Rank ordered residual observational versus simulated temperature values for all April days during the reference period for the example grid cell. The uncorrected GCM data are shown in black with the fitted regression curve overlaid (red). Statistically identical data would lie on the $y = x$ curve (grey). The bias-corrected GCM data are plotted in green.

(ii) Adjusting the precipitation intensity for wet days. We use a similar approach as proposed by Piani et al. (2010).

Correction of the frequency of dry days

Correction of the frequency of dry days is derived from the wet months of the reference period. In many cases there are artificially large amounts of drizzle in GCMs, i.e. days with low precipitation, while the observations suggest a larger number of dry days (i.e. zero precipitation). In order to correct for that discrepancy, we determine the number of observed dry days, N_{dry}, during the reference period by counting the occurrence of $P_{ij}^{\mathrm{WFD}} < 1\,\mathrm{mm\,day}^{-1}$ from the WFD daily data associated with wet months. The threshold value $1\,\mathrm{mm\,day}^{-1}$ was used already in earlier studies and is related to measurement noise. The same number of days (beginning with those having the lowest precipitation values) is set to zero in the GCM daily data and excluded from the dataset used to generate the transfer function for the correction of the intensity of precipitation.

In this way low-precipitation, or drizzle, in the GCM is truncated if the intensity of the precipitation is below a threshold

$$\begin{aligned}\epsilon_{\mathrm{d}} = 0.5 \cdot P_{ij}^{\mathrm{GCM}}\Big|\Big(P_i^{\mathrm{GCM}} > \epsilon_{\mathrm{m}},\ P_{ij}^{\mathrm{GCM}} \leq P_1^{\mathrm{GCM}}\left[N_{\mathrm{dry}}\right]\Big)\\ +0.5 \cdot P_{ij}^{\mathrm{GCM}}\Big|\Big(P_i^{\mathrm{GCM}} > \epsilon_{\mathrm{m}},\ P_{ij}^{\mathrm{GCM}} > P_1^{\mathrm{GCM}}\left[N_{\mathrm{dry}}\right]\Big). \quad (11)\end{aligned}$$

The variable P_1 represents the rank ordered simulated precipitation values in wet months, starting from the lower end.

Since precipitation values smaller than ϵ_{d} are set to zero, the frequency of dry days (i.e. those without measurable precipitation) can be increased in the model data. If there are more days with zero precipitation in the GCM than in the observational dataset N_{dry} is chosen equal to that number of days in order to calculate the threshold (cf. Eq. 11), i.e. no additional dry days are introduced in this case. Additional wet days are never introduced, since this could lead to crucial physical inconsistencies (e.g. rain without clouds).

Exclusion of drizzle days can modify the monthly means, which must be avoided if the long-term trend is to be preserved. An appropriate normalisation can ensure this. However, if identical normalisation for construction and application cannot be ensured in any case (as in the approach applied for the ISI-MIP dataset) this limits the capacity to adjust the daily variability, since multiplying the data with any factor different from one modifies the width of the probability distribution. Thus, in the extended approach, for each month we redistribute the amount of precipitation in dry days uniformly among the wet days. This is achieved by an additive constant m_i^{data} which is the total amount of precipitation from dry days (drizzle) divided by the number of wet days. It is calculated for each year and month separately. Redistribution of the precipitation leads to new values

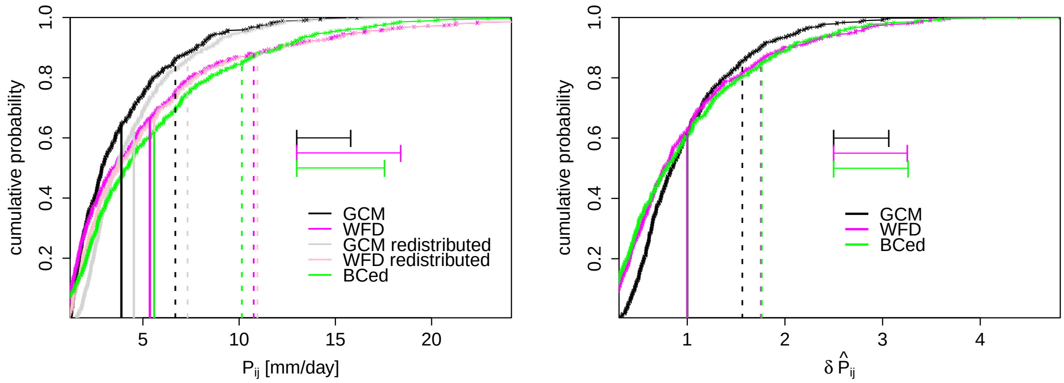

Fig. 5. Normalised cumulative sums of daily observational and simulated April precipitation values during the reference period for the example grid cell. Dry days and months are omitted. For all wet days of the reference period uncorrected values before and after the redistribution of drizzle (left panel), and normalised values (right panel) are shown. The vertical lines refer to related mean (solid) and mean + standard deviation (dashed). Horizontal bars are for comparison of the standard deviation.

$$\hat{P}_{ij}^{\text{data}} = \begin{cases} P_{ij}^{\text{data}} + m_i^{\text{data}} & \text{if wet} \\ 0 & \text{if dry} \end{cases}. \tag{12}$$

The mean over all wet days in a particular month, $\hat{P}_i^{\text{WFD}}$ and $\hat{P}_i^{\text{GCM}}$, is used for normalisation (cf. Fig. 5):

$$\delta \hat{P}_{ij}^{\text{data}} = \frac{\hat{P}_{ij}^{\text{data}}}{\hat{P}_i^{\text{data}}}. \tag{13}$$

Correction of the precipitation intensity of wet days

Correction of the precipitation intensity of wet days by fitting a transfer function is performed, if there are more than 80 wet days in the whole reference period (1960–1999) and the monthly mean is above 0.01 mm day^{-1}. The cut-off value 80 is motivated by sensitivity studies performed in WaterMIP.

In general a transfer function $g(\delta \hat{P}^{\text{GCM}})$ is derived using nonlinear regression on the rank ordered sets $\delta \hat{P}^{\text{WFD}}$ and $\delta \hat{P}^{\text{GCM}}$, which are the sets of normalised wet days in wet months over the 40 yr reference period (cf. Fig. 6). The lowest wet precipitation value in that period, $\delta \hat{P}_{\text{min}}^{\text{GCM}}$, is a parameter of the transfer function

$$g\left(\delta \hat{P}^{\text{GCM}}\right) = \left[a + b \cdot \left\{\delta \hat{P}^{\text{GCM}} - \delta \hat{P}_{\text{min}}^{\text{GCM}}\right\}\right] \times \left[1 - \exp\left\{-\frac{\delta \hat{P}^{\text{GCM}} - \delta \hat{P}_{\text{min}}^{\text{GCM}}}{\tau}\right\}\right]. \tag{14}$$

The offset a and slope b of the linear part of the function, as well as the decay constant τ of the exponential part must be fitted.

In the extended algorithm this nonlinear regression is preferentially applied. Only if the nonlinear fitting procedure (iteration according to gradient-expansion method) does not converge for two different sets of initial values, is a linear transfer function,

$$g\left(\delta \hat{P}^{\text{GCM}}\right) = \left[a + b \cdot \delta \hat{P}^{\text{GCM}}\right], \tag{15}$$

with offset a and slope b applied. For the ISI-MIP dataset we used a different set of selection rules for the transfer function (adopted from the Water-MIP procedure). However, for our normalised values these selection rules omitted the nonlinear fit in many cases. In addition, the frequency and variability of the precipitation were at multiple grid points not adjusted at all, because insufficient points were selected to be included in the fit due to the bug in the code or the fitted parameters were too extreme. Those issues have been solved in the extended version of the algorithm in order to improve the correction of daily variability. The resulting differences will be discussed in Sect. 4.3.

In both versions of the algorithm, where there are less than 80 wet days in the whole reference period (1960–1999), or the long-term monthly mean is below 0.01 mm day^{-1}, the daily variability of the precipitation is not adjusted due to a shortage of statistical information. In this case we consider a linear transfer function with zero offset $a = 0$ and unit slope $b = 1$ (cf. Eq. 15).

In the Supplements we provide month-specific maps which show the areas where the daily variability is not adjusted, where a linear and where a nonlinear transfer function is applied (cf. Fig. S2).

3.3 Application of the bias correction

In the following sections, we present how the values that were derived during the reference period are applied to bias-correct the simulation data in the past, present and future (application period 1950 to 2099). To adjust both the monthly mean and the daily variability of the data, we combine the two approaches described in Sects. 3.1 and 3.2.

3.3.1 Temperature

We calculate the residual daily average temperature values from the GCM for the whole application period in the same

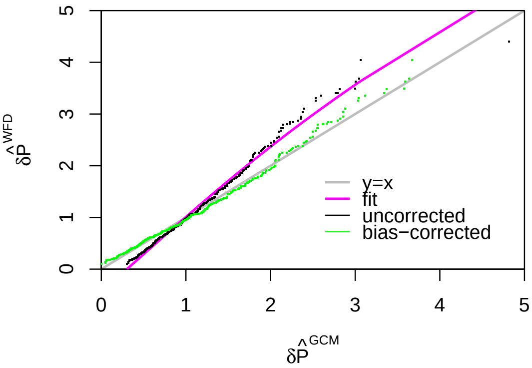

Fig. 6. Rank ordered normalised observational and simulated precipitation values of all wet April days during the reference period for the example grid cell (black). The associated regression curve (red) and the bias-corrected normalised data (green) are presented in addition. The identity line $x = y$ is shown in grey.

way as before for the reference period (cf. Eq. 7). The linear transfer function f (cf. Eq. 9) is then applied to adjust the daily variability. In order to avoid discontinuities at the transition between months, weighting factors for the previous (index m), present (index 0) and following month (index p)

$$d_m = 0.5 \cdot (|d| - d), \tag{16}$$

$$d_0 = 1 - |d|, \tag{17}$$

$$d_p = 0.5 \cdot (|d| + d) \tag{18}$$

are evaluated depending on the day of the month i_{day} and the number of days in that month n_{day}, with

$$d = \frac{i_{\mathrm{day}} - 1}{n_{\mathrm{day}} - 1} - 0.5. \tag{19}$$

Thus, for the first (second) half of the month the slope of the linear transfer function of the previous (following) month B_m (B_p) is taken into account. The weighted sum of the slopes

$$\overline{B} = d_m\, B_m + d_0\, B + d_p\, B_p \tag{20}$$

is then applied to the residual daily average temperature values, which leads to bias-corrected residual values

$$\Delta\tilde{T}_{ij}^{\mathrm{GCM}} = \overline{B} \cdot \Delta T_{ij}^{\mathrm{GCM}}. \tag{21}$$

Together with this equation the correction suggested in Eq. (2) can be extended to

$$\tilde{T}_{ij}^{\mathrm{GCM}} = C + T_i^{\mathrm{GCM}} + \Delta\tilde{T}_{ij}^{\mathrm{GCM}}. \tag{22}$$

This successfully preserves the long-term absolute temperature change in the simulations, whilst adjusting the daily variability about the monthly mean (if $\overline{B} = B$). The constant C arises from the monthly mean correction of temperature (Eq. 1) and assures the agreement between the long-term monthly means of the observed and the corrected simulated data. The monthly values of C are interpolated to daily ones, $\overline{C}$, using the same weighting approach as for the slope B (cf. Eq. 20), thus preventing jumps in the time series at the transitions between months. For $\overline{C} \approx C$ the trend is, except for very small deviations, preserved.

3.3.2 Precipitation

In the case of precipitation, similar interpolation (cf. Eq. 20) of the monthly correction factor, c (Eq. 4), to daily values is less appropriate since the derived value can vary strongly from month to month (because of the high variability at different time scales). The same applies to the parameters of the transfer function (a, b and τ in Eqs. 14 and 15). However, the continuity at the crossover between two months is not as problematic as for temperature. Therefore, we retain the individual monthly values for c, a, b and τ.

We use the thresholds ϵ_{m} and ϵ_{d} (Eqs. 10 and 11) defined previously for the reference period (cf. Sect. 3.2) in order to distinguish dry days and months from wet ones in the application period.

For all days in dry months, we apply only the multiplicative factor c for the long-term mean correction (cf. Eqs. 4 and 5).

In wet months the frequency of dry days is adjusted by setting all values below the dry day threshold ϵ_{d} to zero

$$\hat{P}_{ij}^{\mathrm{GCM}} = 0, \quad \text{if} \quad \left(P_{ij}^{\mathrm{GCM}} \leq \epsilon_{\mathrm{d}}\right) \quad \text{and} \quad \left(P_i^{\mathrm{GCM}} > \epsilon_{\mathrm{m}}\right). \tag{23}$$

Following the same line of reasoning as in Sect. 3.2, we redistribute the total precipitation from these dry days uniformly amongst the wet days of the month (cf. Eq. 12). The obtained precipitation values are normalised by the mean over the wet days (cf. Eq. 13), and the transfer function g (Eqs. 14 and 15) is applied to these normalised values. For application to the reference period (where the transfer function was derived) this procedure ensures that corrected precipitation values are not negative. However, this does not necessarily hold for all time periods, since the lowest precipitation value in the non-bias-corrected GCM data might be below $\delta\hat{P}_{\min}^{\mathrm{GCM}}$ (although those exceptions are rare). Thus, negative values arising from the correction process are set to zero. However, such a truncation modifies the monthly mean. In order to avoid this change in monthly mean precipitation, a correction factor is used to ensure that the mean of the corrected normalised wet days is unity in each month and year. In this way conservation of $\hat{P}_i^{\mathrm{GCM}}$ is ensured, i.e. the mean over the wet days of the month after the redistribution of the drizzle but before the normalisation. In addition, the variability adjustment is preserved. The correction factor can be applied since the new monthly mean is already close to unity by construction, and thus this multiplication does not significantly affect the width of the probability distribution. The latter could not be assumed for the redistribution of drizzle, therefore an additive approach was used in that case.

Finally, the correction in Eq. (5) can be extended to

$$\tilde{P}_{i,j}^{\mathrm{GCM}} = c \cdot \hat{P}_{i}^{\mathrm{GCM}} \cdot \delta \tilde{P}_{i,j}^{\mathrm{GCM}}. \tag{24}$$

With the addition of the dry day and dry month conditions, redistribution of drizzle, and normalisation of corrected values, Eq. (24) preserves the relative precipitation change in the simulations. The correction factor c is taken from the monthly mean correction of precipitation (cf. Eq. 4) and maps the long-term monthly mean of the simulated data to the observational one. Additionally, applying Eq. (24) adjusts the frequency of dry days and the variability about the mean.

An upper bound for precipitation (400 mm day^{-1}) was introduced to avoid single extremes blown up to nonphysically high precipitation values. This final truncation may slightly change the mean. However, this rare case is an accepted consequence.

3.4 Correction of other climate variables

Often there are also biases in other variables than temperature and precipitation, e.g. radiation or wind speed (Haddeland et al., 2012), most of which must not become negative. Within ISI-MIP we use a similar multiplicative approach as described for precipitation to adjust surface pressure, long- and shortwave radiation and wind speed (cf. Table 1). Modifications to the algorithm described earlier are made with regards to the selection of thresholds for pressure and radiation (ϵ_{d} was set to 0). Moreover, the final truncation of the bias-corrected values for pressure, radiation and wind speed plays no important role, since the threshold values were set very high (75 m s^{-1} for wind, 1420 W m^{-2} for short wave radiation, 1000 W m^{-2} for longwave radiation, and 1200 hPa for pressure, cf. precipitation 400 mm day^{-1}).

In addition, daily minimum (maximum) temperature correction is derived from the correction of daily average temperature. We calculate the mean distance to the average temperature value over the reference period for both observations and simulations:

$$\kappa = \frac{\sum_{i=1}^{m=40} \left(T_{\mathrm{m},ij}^{\mathrm{WFD}} - T_{ij}^{\mathrm{WFD}} \right)}{\sum_{i=1}^{m=40} \left(T_{\mathrm{m},ij}^{\mathrm{GCM}} - T_{ij}^{\mathrm{GCM}} \right)}, \tag{25}$$

where T_{m} refers to the daily minimum (maximum) temperature. In the application of the correction

$$\tilde{T}_{\mathrm{m},ij}^{\mathrm{GCM}} = \kappa \cdot \left(T_{\mathrm{m},ij}^{\mathrm{GCM}} - T_{ij}^{\mathrm{GCM}} \right) + \tilde{T}_{ij}^{\mathrm{GCM}} \tag{26}$$

the original distance to the daily average temperature is scaled with the factor κ and the result is added to the bias-corrected daily average temperature.

For snowfall the portion of snow (S_{ij}^{GCM}) from the total precipitation (P_{ij}^{GCM}) in the uncorrected model data is calculated at each grid cell. Application to the bias-corrected precipitation leads to bias-corrected snowfall data

$$\tilde{S}_{ij}^{\mathrm{GCM}} = \frac{S_{ij}^{\mathrm{GCM}}}{P_{ij}^{\mathrm{GCM}}} \cdot \tilde{P}_{ij}^{\mathrm{GCM}}. \tag{27}$$

The same procedure applies to the wind components, i.e.

$$\tilde{K}_{ij}^{\mathrm{GCM}} = \frac{K_{ij}^{\mathrm{GCM}}}{W_{ij}^{\mathrm{GCM}}} \cdot \tilde{W}_{ij}^{\mathrm{GCM}}, \tag{28}$$

where W_{ij}^{GCM} refers to the total wind speed (bias-corrected in the same way as precipitation according to Eq. 24) and K_{ij}^{GCM} represents the eastward (northward) wind component. The wind components are scaled in the same way as the total wind speed to obtain the bias-corrected components. The wind direction is preserved in that way.

4 Evaluation of the methodology

In this section we present the bias-corrected temperature and precipitation data of the HadGEM2-ES April climate.

We demonstrate that bias correction alters several statistical properties of the data in the desired fashion, but also discuss its limitations.

4.1 Sensitivity study

Due to the limited availability of observational data a validation of the applicability of the statistical bias correction setting which we applied earlier is in general not straightforward. However, strong evidence for applicability arises from the consideration of short sensitivity studies (Piani et al., 2010). Thus, in a first step, we study global maps of different statistical quantities of uncorrected and bias-corrected GCM data for a period 1980 to 1999. In this setup the parameters are derived only from a 20 yr reference period 1960 to 1979 in order to avoid an overlap of the "training" and the "validation" dataset.

As described earlier, the first step of our bias correction (cf. Sect. 3.1) adjusts the long-term monthly mean. Thus, this average is the first statistical quantity to be considered. Furthermore, the second step of the bias correction (cf. Sect. 3.2) modifies the width and in case of a nonlinear fit also higher moments of the distribution. Those parameters of the probability distribution are, for example, represented by the lower (50–10 %) and upper (90–50 %) inter-percentile range. Deviations between GCM and WFD in the three mentioned statistical quantities, as well as the improvement by our trend-preserving bias correction, are shown in Fig. 7 for (a) temperature and (b) precipitation.

The bias correction based on the reference period 1960 to 1979 yields a significant improvement of the matching of the long-term mean for the period 1980 to 1999 in most areas of the globe (cf. Fig. 7 upper row). There are only a few areas where the matching between the WFD and the bias-corrected

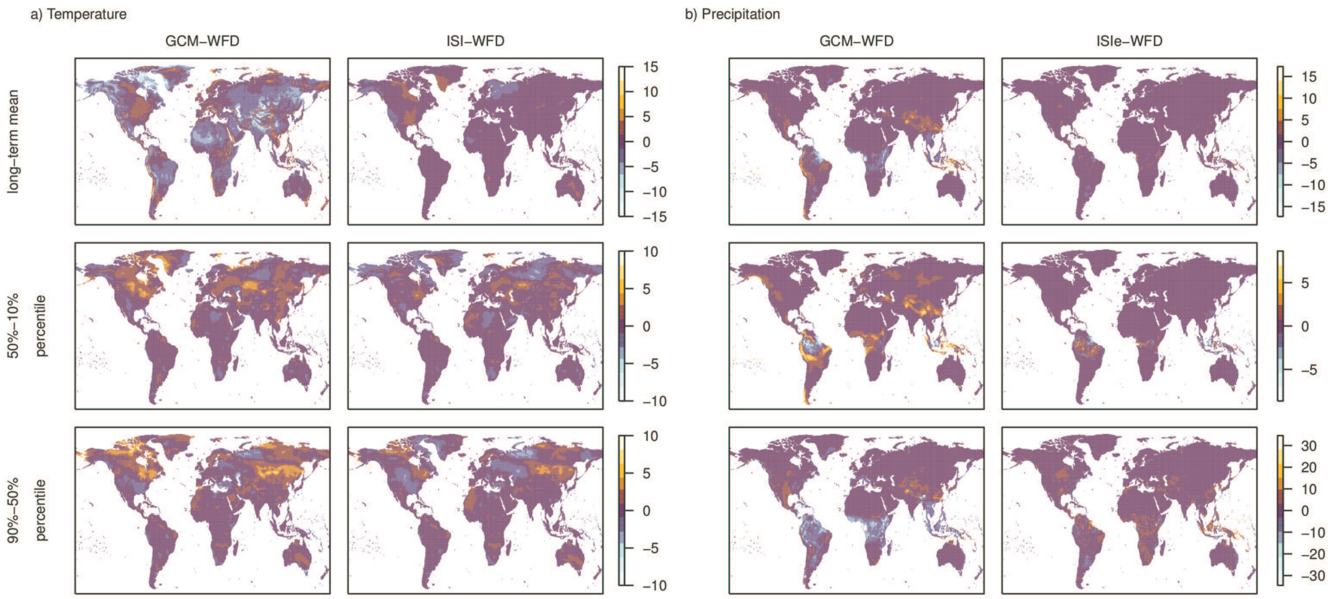

Fig. 7. The anomalies (i.e. simulation – observation) of statistical properties in raw (GCM) and bias-corrected (ISI and ISIe) model data are shown. The trend-preserving ISI-MIP methodology was applied to the period 1980 to 1999 for bias correction based on a 20 yr reference period (1960 to 1979). In case of precipitation we present the results of the extended algorithm. The long-term mean, lower inter-percentile range and upper inter-percentile range of the April daily **(a)** temperature and **(b)** precipitation from 1980 to 1999 are shown. The 50–10 % percentile refers to the lower inter-percentile range, while 90–50 % percentile denotes the upper inter-percentile range. Colours refer to **(a)** temperature values in K and **(b)** precipitation values in mm day^{-1}.

GCM data is slightly worse than for the uncorrected temperature and precipitation data. Deviations occur particularly for temperature, for example, in northern Scandinavia or southern US. They are, however, small compared to the maximum values of departure that occur around the globe when we consider the uncorrected data, and could be related to variability on a time scale that is not properly sampled within the 20-yr reference period. Moreover, we find a general improvement of the matching of the inter-percentile ranges, although the deviations that persist after bias correction are more extended than in the case of the long-term mean. This affects mainly North America and Asia for temperature and the equatorial region for precipitation.

Note that the 20-yr reference period in the sensitivity study is comparatively short, and thus the bias correction parameters estimated from this period are likely to be less robust than in the actual application where we used the full 40 yr reference period for training. Nevertheless, in most areas of the globe the bias correction results in a significant improvement of the matching between simulations and observations even with a 20 yr reference period.

Next we go back to our initial reference period which is 1960 to 1999 and was applied to obtain the ISI-MIP dataset. It samples variability on decadal and multi-decadal time scales and should thus result in more robust parameter estimates. We consider the distributions of simulations and observations over the whole reference period, since this is the time span where they are intended to match by construction.

While the mean climate signal (long-term trend) should be preserved by the ISI-MIP bias correction algorithm, different parameters of the probability distribution are modified. The latter was already illustrated in Figs. 3 and 5 for an example grid cell. Although the mapping of the probability distributions remains imperfect (see particularly left panels), it is significantly improved with the applied bias correction. In case of temperature (Fig. 3) the mean values show very good agreement, while the standard deviation is slightly underestimated in the bias-corrected data. For precipitation (Fig. 5) a substantial harmonisation of the standard deviations and the mean values was achieved.

In order to check if theses results are robust, next, we consider global maps of the different statistical quantities of the probability distributions which we already described at the beginning of this section (cf. Fig. 7). However, this time the reference period for the bias correction and the time span to construct the histogram are the same (1960 to 1999). Deviations between GCM and WFD in the three mentioned statistical quantities, as well as the improvement by our trend-preserving bias correction, are shown in Fig. 8 for (a) temperature and (b) precipitation.

In case of the long-term temperature mean, shown in the upper panels, we observe deviations of WFD and GCM data between approximately −15 and 15 K (i.e. a span of 30 K).

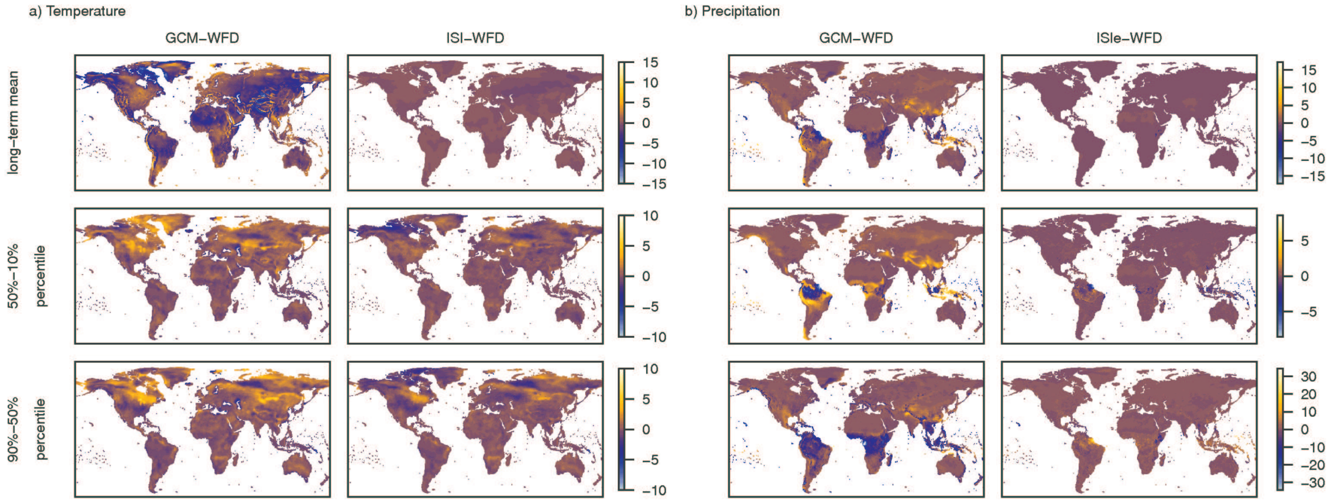

Fig. 8. The anomalies (simulation – observation) of statistical properties in raw (GCM) and bias-corrected (ISI and ISIe) model data are shown. The trend-preserving ISI-MIP methodology was applied to the period 1960 to 1999 for bias correction based on a 40 yr reference period (1960 to 1999). The illustration is analogous to Fig. 7. Colours refer to **(a)** temperature values in K and **(b)** precipitation values in mm day^{-1}.

This is much narrower than the span of the long-term temperature mean values themselves which range from approximately 236 to 308 K (i.e. a span of 72 K). Thus, the deviations are comparatively small. Discrepancies between WFD and GCM are mainly related to the coarse resolution of the model affecting the altitude information in some regions. These anomalies are significantly reduced by our bias correction, as illustrated in the upper right panel of Fig. 8a in comparison to the upper left one.

Moreover, the width (and skewness) of the distribution of the daily averaged temperature values shows good agreement between WFD and GCM data, as reflected in the inter-percentile ranges shown in the middle and lower panels in Fig. 8a. Here the departure between WFD and GCM spans 20 K. Larger deviations in the inter-percentile ranges occur mainly in the Northern Hemisphere (particularly in Cental Asia and North America). With the two-parameter quantile mapping applied to residual time series these differences between observation and model dataset are significantly reduced (right panels compared to left ones). The patterns, however, persist. A total matching cannot be achieved with the linear transfer function, since we do not adjust higher moments of the distribution. In addition, interpolation of the slope of the transfer function from monthly to daily values also prevents total matching.

In the case of precipitation (Fig. 8b) some regions in the Southern Hemisphere show larger deviations between the long-term mean and the inter-percentile ranges of the WFD and GCM data (particularly in Central Africa, South America, and Indonesia). In addition several regions in South-East Asia are affected. The extended trend-preserving bias correction algorithm reduces theses departures significantly (cf. Fig. 8b, right panels to left ones), although adjustment of the probability distributions is imperfect. In the bias-corrected dataset largest deviations from the WFD distribution persist in North Brazil and Indonesia with regards to the inter-percentile ranges (middle and lower right panel in Fig. 8b). The largest differences between WFD and bias-corrected GCM long-term mean precipitation occur in North Africa and China (Fig. 8b, upper right panel). While in North Africa the values obtained with the uncorrected dataset were already comparatively low, in China model and observations are substantially harmonised by the bias correction.

Moreover, the algorithm proposed here allows to significantly improve the dry days statistics in most regions of the globe (cf. maps of the number of dry days between 1980 and 1999 in the Supplements Fig. S3). Nevertheless, the adjustment is imperfect particularly in eastern North America, northern Asia, central Australia and in the Sahara, and it does not yield any improvement in some mountain regions (e.g. on the western side of the Andes). The latter is most likely related to the limited downscaling ability of the method.

We focus in our sensitivity study on the range between between the 10 % and the 90 % quantile. For this central range bias correction methods are expected to perform well, while the correction in the outer ranges of the distribution is typically worse, since there are less events (Maraun, 2013). In general bias correction methods tend to exaggerate extreme events, since the limited number of data points prohibits a robust analysis of the relationship between observations and simulations, potentially resulting in an overestimation of these events. In addition, the extreme events always cover the whole gridbox area, i.e. their spatial extent is typically too large. However, since we introduced an upper bound for

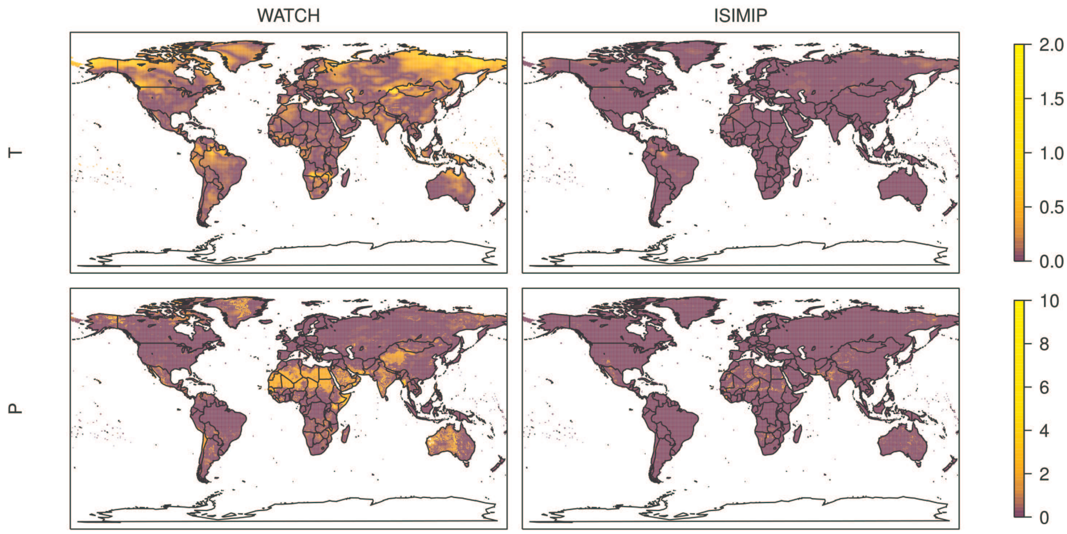

Fig. 9. Absolute differences between the trends in the interpolated GCM April data before and after bias correction. The absolute trend for temperature (in K) is estimated as $T_{2095-2099} - T_{1960-1964}$. In the case of the relative precipitation trend, we consider the logarithm $\log(P_{2095-2099}/P_{1960-1964})$. Similarly trends in the bias-corrected data are estimated from $\tilde{T}$ and $\tilde{P}$. Values of the absolute differences between the trends in bias-corrected and uncorrected data are truncated at the upper bound of the colourbar, i.e. yellow refers to the denoted or higher values. White areas belong to regions where no information about the trend is available. The results with the quantile mapping applied on the time series themselves (WATCH method) and on the residuals or normalised values (ISIMIP method) are shown for temperature and precipitation (cf. Table 2 for main algorithmic differences). We consider the end of the 21st century (mean 2095–2099) in comparison to the beginning of the reference period (mean 1960–1964) to define the trend.

the bias-corrected values in ISIMIP, the impact of this effect is not arbitrarily large. On the global scale the bias-corrected variables show good agreement with the observational data even in the tails of the distribution (cf. Supplement Fig. S4).

However, the influence of the overestimated spatial correlation on the impact simulations clearly depends on the degree to which impacts at one grid cell are influenced by impacts or meteorological events in the neighbouring grid cells that might be particularly relevant with regard to hydrological models. In Supplement Fig. S5, we provide maps of the bias-corrected and observed variances over the reference period (April precipitation aggregated to the original resolution of the HadGEM2-ES GCM) for illustration. Moreover, the ratio of both is given.

Furthermore, maps which evaluate the bias correction of the other ISI-MIP variables in terms of the long-term mean and inter-percentile ranges are provided in the Supplement. Patterns of discrepancies between observational and bias-corrected data can be identified in all statistical properties, however, their spatial distribution is not consistent among all variables (cf. Fig. S6). Changes in the inter-percentile ranges are small and rather localized for all four variables. In any case, bias correction yields a significant improvement in the long-term mean, particularly of pressure and radiation, as can be seen when comparing Figs. S6 and S7. Note that the colour bar for the long-term mean of pressure, long- and shortwave radiation differs between the two figures.

4.2 Trend: comparison with WATCH method

We illustrate that in contrast to a quantile mapping of the time series itself (as used e.g. in Water-MIP; Hagemann et al., 2011 or WATCH; Weedon et al., 2011), our approach preserves the long-term trend with respect to the monthly mean values either in absolute or relative terms (cf. Fig. 9).

The proposed additive approach does not modify the absolute trend in the temperature data compared to the interpolated GCM output (except for small deviations related to the interpolation of the transfer function, cf. Eq. 20). As an example, we consider monthly means over two 5 yr periods, one at the beginning (1960–1964) and one at the end (2095–2099) of the application period. The difference between those monthly mean values,

$$T_{2095-2099} - T_{1960-1964} = \tilde{T}_{2095-2099} - \tilde{T}_{1960-1964}, \quad (29)$$

is not affected if we apply the ISI-MIP algorithm for temperature correction described in Sect. 3, as shown in Fig. 9 (upper panels). This is in contrast to what is observed when applying the quantile mapping to the time series themselves (left panel, denoted as WATCH in Fig. 9). In the left panel significant changes in the temperature trend occur particularly in West Canada, Alaska, East Russia, North-West China, and North Brazil. With our proposed algorithm (right panel, denoted as ISIMIP in Fig. 9) a small region in North Brazil is most affected by the change in temperature trend.

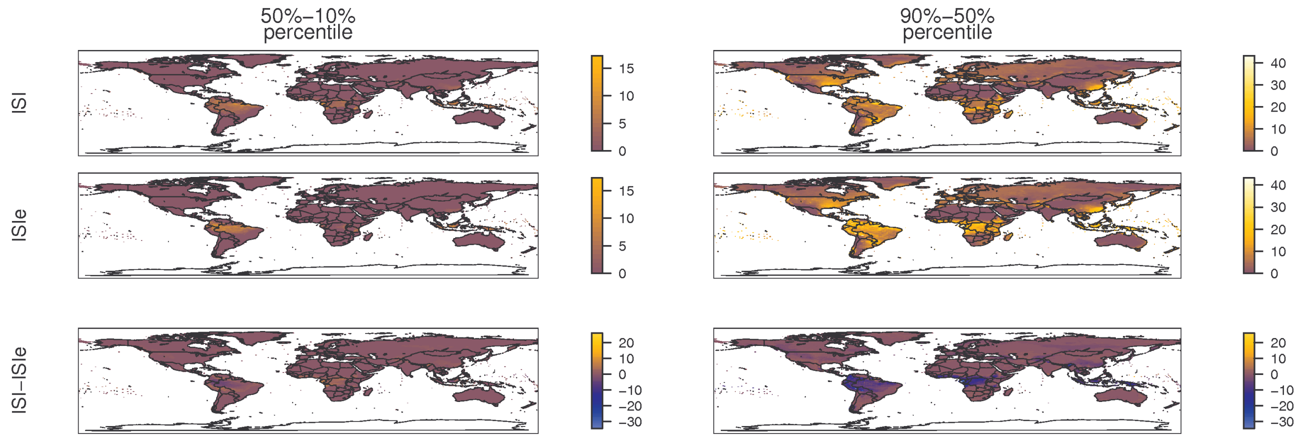

Fig. 10. The lower (50–10 %) and upper (90–50 %) inter-percentile range of the April daily precipitation from 1960 to 1999 are shown for the ISI-MIP dataset (ISI) and with the extended version of the algorithm (ISIe). In addition the differences of both version are shown. The inter-percentile ranges are analogues to Fig. 8. Colours refer to precipitation in mm day^{-1}. Difference values outside the range of the shown colourbars are white in order to increase visibility of the map (over land this affects only few small areas).

However, that shift is small compared to the changes observed with previous approaches.

The lower panels of Fig. 9 illustrate that the multiplicative approach preserves the relative trend of the precipitation in the same sense:

$$\frac{P_{2095-2099}}{P_{1960-1964}} = \frac{\tilde{P}_{2095-2099}}{\tilde{P}_{1960-1964}}. \tag{30}$$

This is valid for the precipitation that is bias-corrected with the extended version of the multiplicative algorithm (ISIe) as well as for the ISI-MIP climate input (ISI), since the modifications to the code affect only the variability of the the daily data, but not the correction of the monthly mean. White areas in Fig. 9 occur if no conclusions about the relative trend can be made. We show the absolute difference of the logarithms,

$$\log\frac{P_{2095-2099}}{P_{1960-1964}} - \log\frac{\tilde{P}_{2095-2099}}{\tilde{P}_{1960-1964}}, \tag{31}$$

in Fig. 9 in order to weight departures in increasing and decreasing precipitation amounts the same. To increase visibility a nonlinear colourscale which is truncated at 10 has been chosen. However, there are regions (particularly in the left panel of Fig. 9) where the change in trend exceeds that value (e.g. in the West Sahara).

We observe that significant changes in the relative trend in April precipitation occur mainly in regions which (in spring) are characterised by rather arid conditions. The quantile mapping applied to the time series themselves (denoted as WATCH in Fig. 9) results in extended areas of large changes in the trend. As shown in the lower left panel of Fig. 9, most affected regions are North Africa, Australia, India, West China, Namibia, Botswana, and few small regions in Chile, Argentina, Mexico, Southern US, Northern Canada, Greenland and East Russia. In contrast, the precipitation that was bias-corrected with the ISI-MIP approach (denoted as ISIMIP in Fig. 9) shows fewer and smaller changes in the trend. Modifications of the trend persist in North Africa, Australia, North-West India, Namibia, Botswana, Mexico, Southern US and East Russia, which is most likely related to numerical effects in arid regions.

4.3 ISI-MIP algorithm and its extension

The bias correction method for variables with positivity constraints that was applied to generate the ISI-MIP dataset, due to time constraints, suffers from some unresolved problems and a bug in the programme code. As a result, while the long-term mean is adjusted in the desired fashion, the variability of the variables (e.g. precipitation) is corrected only to a limited extent. Hence, compared to the results shown and discussed in the previous paragraph (cf. Fig. 8b), variability in the ISI-MIP dataset is typically closer to that in the GCM. This holds in particular for the upper inter-percentile range, whilst the lower inter-percentile range is slightly enlarged by introducing zero precipitation days.

In order to characterise the problems in the ISI-MIP dataset, in Fig. 10 we compare the inter-percentile ranges over the reference period (1960–1999) in the ISI-MIP precipitation dataset and the precipitation that is bias-corrected with the extended version of the algorithm. Since the same methodology is used for correction of the long-term mean in both datasets, the resulting long-term mean is the same by definition (identical to what is shown in the upper row in Fig. 8b).

In case of the lower inter-percentile range basically the northern part of South America, Congo, South-East China and Indonesia show significant deviations between the two bias-corrected datasets (cf. Fig. 10, lower left panel). In addition, we observe that the lower inter-percentile range is in

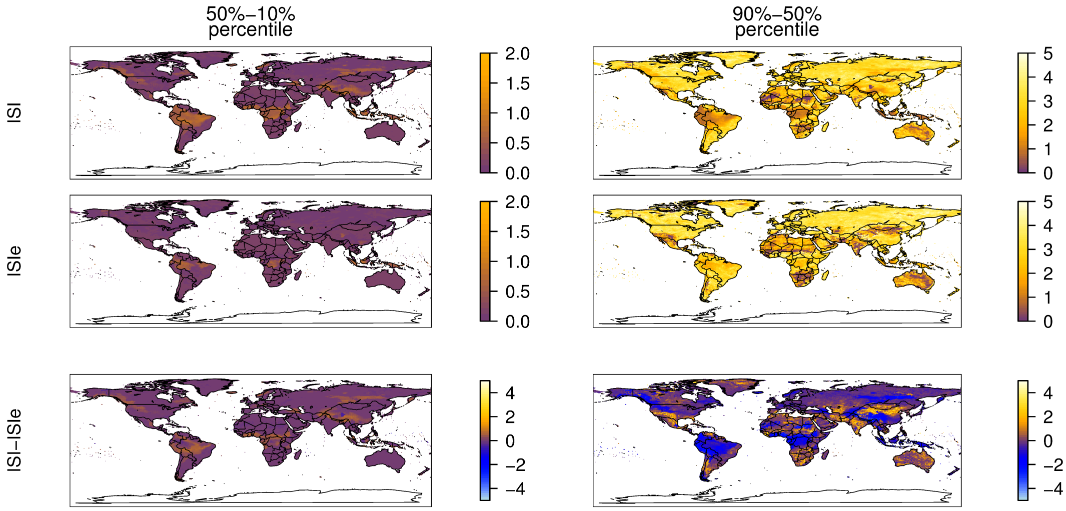

Fig. 11. The lower inter-percentile range (50–10 %) and upper inter-percentile range (90–50 %) of the normalised April daily precipitation from 2091 to 2099 are shown for the ISI-MIP dataset (ISI) and with the extended version of the algorithm (ISIe). In addition the differences of both version are shown. The inter-percentile ranges are analogues to Fig. 8.

general less affected than the upper one, both with regards to absolute values and to spatial extent (cf. Fig. 10, lower panels). This indicates that in the ISI-MIP dataset extreme high precipitation events in most areas are less likely than in the precipitation dataset that is bias-corrected with the extended algorithm. Particularly South-Eastern USA, Northern Brazil, South-Eastern China, and several countries in Central Africa are affected by this deviation between the datasets.

Furthermore, the limited correction of the variability in the ISI-MIP data set results in several places in a narrower distribution than the one obtained with the extended algorithm. This can be concluded from Fig. 10 if we sum up the values shown in the two lower panels. Most affected areas in that context are Eastern USA, Southern Greenland and South-Eastern China in the Northern Hemisphere, as well as South America, Central Africa and Indonesia in the Southern Hemisphere.

The level of agreement between the width of the WFD and GCM probability distributions (including daily and monthly variability) during the reference period dictates also the width of the distribution of bias-corrected values in the future. Thus, when performing the analysis shown in Fig. 10 for a period at the end of the 21st century (RCP 8.5) we obtain basically the same patterns, although absolute values are in general larger.

Since in both datasets the monthly variability is modified in the same way (by a constant multiplicative factor), the described differences must result from the correction of the daily variability. For a more detailed investigation and comparison of variability only on this short time scale, we consider the inter-percentile ranges of the normalised values for the end of the 21st century (RCP 8.5). We chose this scenario and time period, since deviations are expected to be most pronounced here. Results are shown in Fig. 11.

The time series are divided by the monthly mean (including *all* days of the month) in order to normalise them before the inter-percentile ranges are calculated. While the same normalisation was applied in the algorithm used to produce the ISI-MIP dataset, in the extended algorithm a different normalisation is applied (cf. Sect. 3.2.2). This means for the extended algorithm the normalisation applied before plotting the results does not coincide with the normalisation used during the corresponding bias correction process. Thus, we cannot expect to find agreement of the distributions of the two precipitation datasets across all locations. However, while particularly the patterns in South America and Central Africa reflect the results which we found for the unnormalised datasets, those in Central Asia and western North America did not occur before. This means consideration of the inter-percentile ranges of the normalised values in Fig. 11 reveals patterns of changes in the distributions, which are masked on the larger scale (cf. maps for the unnormalised values Fig. 10). A general statement on the deviation of the width of distributions of normalised values, as done for the unnormalised values in Fig. 10, is however not straightforward. This is because in many cases the discrepancy for lower and upper inter-percentile range is of opposite sign.

5 Conclusions and future work

We presented a novel, trend-preserving statistical bias correction approach, which adjusts the monthly mean and daily variability of simulated climate data to observations, whilst preserving the climate signal (long-term trend) much better than previous algorithms. The proposed bias correction method extends the approach by Piani et al. (2010) to conserve the trend. An additive approach preserving the absolute changes (for temperature) and a multiplicative one preserving the relative changes (for precipitation) were developed and described in detail. Quantile mapping was applied only to residual or normalised data. We demonstrated that our approach is capable of adjusting the probability distribution over the reference period, whilst widely preserving the long-term trend in the data. We showed and discussed that, although daily weighting of monthly correction factor (temperature algorithm) or truncation of extreme high values (precipitation algorithm) can affect the trend, even with those limitations the methodology proposed by ISI-MIP performs well in preserving the trend. This is essential for the project and not necessarily ensured by other methods (as shown for the method used within Water-MIP). In addition, our approach separates the bias correction at different time scales from each other.

In the case of temperature the proposed procedure is similar to the cascade bias correction method described by Haerter et al. (2011). The major difference is the bias correction on the largest time scale. We refrain from multiplicative correction in that case in order to preserve the trend, whilst Haerter et al. (2011) use a linear transfer function instead of an offset. However, the benefits of the cascading procedure persist. Thus, the method chosen by ISI-MIP avoids that improvement of the matching of probability distributions based on daily data leads to impairment of the one based on monthly data. A non-cascading bias correction on the other hand mixes the adjustment of short-term and long-term mean and variability and leads only to improvements on both scales if the fluctuations at the different time scales are aligned, as shown by Haerter et al. (2011).

Furthermore, the cascading approach allows to further extend the method to bias-correct GCM data at multiple time scales. For example, the same methodology as described in Sect. 3.2 could be applied to temperature by replacing the monthly mean by an annual mean and the daily data by monthly data. With that approach the annual mean temperature will be adjusted, while the trend based on the annual values is preserved. Moreover, the monthly variability, which was preserved in the ISI-MIP dataset, will be adjusted. In the next step, daily temperature variability will be corrected as described in Sect. 3.2.

With the bias correction method that we proposed in Sect. 3 a similar cascade bias correction can be assigned to precipitation and other variables with positivity constraints. This allows us to bias-correct the related GCM data at multiple time scales as well. Within the ISI-MIP approach described here, we corrected only the variability of the daily data about the monthly mean, while variability at other time scales was neglected. However, the bias in the weekly or monthly variability of precipitation, for example, affects the representation of droughts and floods. A bias correction of the variability at multiple time scales – e.g. the (relative) variability of the monthly data about the annual mean, i.e. the seasonal cycle – is in principle possible with the method proposed by ISI-MIP, but has not been applied so far. Such an extension will be crucial for future impact studies, even though bias correction can only be applied to processes that operate on time scales that are considerably shorter than the reference period.

Supplementary material related to this article is available online at: http://www.earth-syst-dynam.net/4/219/2013/esd-4-219-2013-supplement.pdf.

Acknowledgements. This work has been conducted under the framework of ISI-MIP. The ISI-MIP Fast Track project is funded by the German Federal Ministry of Education and Research (BMBF) with project funding reference number 01LS1201A. Responsibility for the content of this publication lies with the authors. We acknowledge close collaboration with the SURVIVE project funded by the German Federal Ministry for the Environment, Nature Conservation and Nuclear Safety (BMU) with project funding reference number 11_II_093_Global_A_SIDSandLDCs. K. Frieler was supported by SURVIVE. F. Piontek received founding from the European Union's Seventh Framework Programme (FP7/2007–2013) under grant agreement no. 266992. Furthermore with regards to data provision, we acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modelling groups (HadGEM2-ES, IPSL-CM5A-LR, MIROC-ESM-CHEM, GFDL-ESM2M, and NorESM1-M) for producing and making available their model output. For CMIP the US Department of Energy's Programme for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organisation for Earth System Science Portals.

Edited by: M. Floerke

References

Berg, A. A., Famiglietti, J. S., Walker, J. P., and Houser, P. R.: Impact of bias correction to reanalysis products on simulations of north american soil moisture and hydrological fluxes, J. Geophys. Res, 108, 4490, doi:10.1029/2002JD003334, 2003.

Boe, J., Terray, L., Habets, F., and Martin, E.: Statistical and dynamical downscaling of the Seine basin climate for hydro-meteorological studies, Int. J. Climatol., 27, 1643–1655, doi:10.1002/joc.1602, 2007.

Cook, K. and Vizy, E.: Effects of twenty-first-century climate change on the amazon rain forest, J. Climate, 21, 542–560, doi:10.1175/2007JCLI1838.1, 2008.

Deque, M.: Frequency of precipitation and temperature extremes over france in an anthropogenic scenario: Model results and statistical correction according to observed values, Global Planet. Change, 57, 16–26, doi:10.1016/j.gloplacha.2006.11.030, 2007.

Dosio, A. and Paruolo, P.: Bias correction of the ensembles high-resolution climate change projections for use by impact models: Evaluation on the present climate, J. Geophys. Res, 116, 1–22, doi:10.1029/2011JD015934, 2011.

Eden, J. M., Widmann, M., Grawe, D., and Rast, S.: Skill, correction and downscaling of GCM-simulated precipitation, J. Climate, 25, 3970–3984, doi:10.1175/JCLI-D-11-00254.1, 2012.

Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K., and Liebert, J.: HESS Opinions "Should we apply bias correction to global and regional climate model data?", Hydrol. Earth Syst. Sci., 16, 3391–3404, doi:10.5194/hess-16-3391-2012, 2012.

Gudmundsson, L., Bremnes, J. B., Haugen, J. E. and Engen Skaugen, T.: Technical Note: Downscaling RCM precipitation to the station scale using quantile mapping – a comparison of methods, Hydrol. Earth Syst. Sci., 16, 3383–3390, doi:10.5194/hess-16-3383-2012, 2012.

Haddeland, I., Heinke, J., Voß, F., Eisner, S., Chen, C., Hagemann, S., and Ludwig, F.: Effects of climate model radiation, humidity and wind estimates on hydrological simulations, Hydrol. Earth Syst. Sci., 16, 305–318, doi:10.5194/hess-16-305-2012, 2012.

Haerter, J. O., Hagemann, S., Moseley, C., and Piani, C.: Climate model bias correction and the role of timescales, Hydrol. Earth Syst. Sci., 15, 1065–1079, doi:10.5194/hess-15-1065-2011, 2011.

Hagemann, S., Chen, C., Haerter, J. O., Heinke, J., Gerten, D., and Piani, C.: Impact of a statistical bias correction on the projected hydrological changes obtained from three GCMs and two hydrology models, J. Hydrometeorol., 12, 556–578, doi:10.1175/2011JHM1336.1, 2011.

Holland, G., Done, J., Bruyere, C., Cooper, C., and Suzuki, A.: Model Investigations of the Effects of Climate Variability and Change on Future Gulf of Mexico Tropical Cyclone Activity, Offshore Technology Conference, Houston, Texas, USA, 2010.

Ines, A. V. and Hansen, J. W.: Biascorrection of daily GCM rainfall for crop simulation strudies, Agr. Forest Meteorol., 138, 44–53, doi:10.1016/j.agrformet.2006.03.009, 2006.

Marquardt, D. W.: An algorithm for least-squares estimation of nonlinear parameters, J. Soc. Ind. Appl. Math., 11, 431–441, doi:10.1137/0111030, 1963.

Maraun, D.: Nonstationarities of regional climate model biases in European seasonal mean temperature and precipitation sums, Geophys. Res. Lett., 39, 1–5, doi:10.1029/2012GL051210, 2012.

Maraun, D.: Bias Correction, Quantile Mapping, and Downscaling: Revisiting the Inflation Issue, J. Climate, 26, 2137–2143, doi:10.1175/JCLI-D-12-00821.1, 2013.

Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J.,Widmann, M., Brienen, S., Rust, H. W., Sauter, T., Themeßl, M., Venema, V. K. C., Chun, K. P., Goodess, C. M., Jones, R. G., Onof, C., Vrac, M., and Thile-Eich, I.: Precipitation downscaling under climate change. Recent developments to bridge the gap between dynamical models and the end user, Rev. Geophys., 48, doi:10.1029/2009RG000314, 2010.

Moss, R. H., Edmonds, J. A., Hibbard, K. A., Manning, M. R., Rose, S. K., van Vuuren, D. P., Carter, T. R., Emori, S., Kainuma, M., Kram, T., Meehl, G. A., Mitchell, J. F. B., Nakicenovic, N., Riahi, K., Smith, S. J., Stouffer, R. J., Thomson, A. M., Weyant, J. P., and Wilbanks, T. J.: The next generation of scenarios for climate change research and assessment, Nature, 463, 747–756, doi:10.1038/nature08823, 2010.

Patricola, C. M. and Cook, K. H.: Northern African climate at the end of the twenty-first century: an integrated application of regional and global climate models, Clim. Dynam., 35, 193–212, doi:10.1007/s00382-009-0623-7, 2010.

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

Piani, C., Weedon, G. P., Best, M., Gomes, S. M., Viterbo, P., Hagemann, S., and Haerter, J. O.: Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models, J. Hydrol., 395, 199–215, doi:10.1016/j.jhydrol.2010.10.024, 2010.

Raisanen, J. and Raty, O.: Projections of daily mean temperature variability in the future: cross-validation tests with ENSEMBLES regional climate simulations, Clim. Dynam., 1–16, doi:10.1007/s00382-012-1515-9, 2012.

Robock, A., Turco, R., Harwell, M., Ackerman, T., Andressen, R., Chang, H.-S., and Sivakumar, M.: Use of general circulation model output in the creation of climate change scenarios for impact analysis, Clim. Change, 23, 293–335, doi:10.1007/BF01091621, 1993.

Rötter, R. P., Carter, T. R., Olesen, J. E., and Porter, J. R.: Crop-climate models need an overhaul, Nat. Clim. Change, 1, 175–177, doi:10.1038/nclimate1152, 2011.

Sato, T., Kimura, F., and Kitoh, A.: Projection of global warming onto regional precipitation over mongolia using a regional climate model, J. Hydrol., 333, 144–154, doi:10.1016/j.jhydrol.2006.07.023, 2007.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, B. Am. Meteorol. Soc., 93, 485–498, doi:10.1175/BAMS-D-11-00094.1, 2012.

Themeßl, M. J., Gobiet, A., and Leuprecht, A.: Empirical-statistical downscaling and error correction of daily precipitation from regional climate models, Int. J. Climatol., 31, 1530–1544, doi:10.1002/joc.2168, 2011.

Weedon, G. P., Gomes, S., Viterbo, P., Shuttleworth, W. J., Blyth, E., Österle, H., Adam, J. C., Bellouin, N., Boucher, O., and Best, M.: 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, 823–848, doi:10.1175/2011JHM1369.1, 2011.

Wu, W. and Lynch, A.: Response of the seasonal carbon cycle in high latitudes to climate anomalies, J. Geophys. Res., 105, 22897–22908, doi:10.1029/2000JD900340, 2000.

Xu, Z. and Yang, Z.: An improved dynamical downscaling method with GCM bias corrections and its validation with 30 years of climate simulations, J. Climate, 25, 6271–6286, doi:10.1175/JCLI-D-12-00005.1, 2012.