Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2017-81
Manuscript under review for journal Earth Syst. Dynam.
Discussion started: 25 September 2017


Earth System Dynamics
Discussions
Open Access
EGU

# Bias correction of surface downwelling longwave and shortwave radiation for the EWEMBI dataset

Stefan Lange[1]

[1]Potsdam Institute for Climate Impact Research, Telegraphenberg A 31, 14473 Potsdam, Germany

*Correspondence to:* Stefan Lange (slange@pik-potsdam.de)

**Abstract.** Many meteorological forcing datasets include bias-corrected surface downwelling longwave and shortwave radiation (rlds and rsds). Methods used for such bias corrections range from multi-year monthly mean value scaling to quantile mapping at the daily time scale. An additional downscaling is necessary if the data to be corrected have a higher spatial resolution than the observational data used to determine the biases. This was the case when EartH2Observe (E2OBS; Calton et al., 2016) rlds and rsds were bias-corrected using more coarsely resolved Surface Radiation Budget (SRB; Stackhouse Jr. et al., 2011) data for the production of the meteorological forcing dataset EWEMBI (Lange, 2016). This article systematically compares various parametric quantile mapping methods designed specifically for this purpose. The methods vary in the time scale at which they operate, in their way of accounting for physical upper radiation limits, and in their approach to bridging the spatial resolution gap between E2OBS and SRB. It is shown how temporal and spatial variability deflation related to bilinear interpolation and other deterministic downscaling approaches can be overcome by downscaling the target statistics of quantile mapping from the SRB to the E2OBS grid such that sub-SRB-grid scale spatial variability present in the original E2OBS data is retained. Cross-validations at the daily and monthly time scale reveal that it is worthwhile to take empirical estimates of physical upper limits into account when adjusting either radiation component and that, overall, bias correction at the daily time scale is more effective than bias correction at the monthly time scale if sampling errors are taken into account. A validation against independent ground observations from the Baseline Surface Radiation Network (BSRN; König-Langlo et al., 2013) suggests that the bias correction of E2OBS surface downwelling radiation using SRB data that was done for the production of EWEMBI had a positive and neutral overall effect on rlds and rsds, respectively. Using any of the other methods tested here would have given similar results as the biases relative to BSRN remaining after bias correction are dominated by the corresponding SRB data biases.



## 1 Introduction

High-quality observational datasets of surface downwelling radiation are of interest in many fields of climate science, including energy budget estimation (Kiehl and Trenberth, 1997; Trenberth et al., 2009; Wild et al., 2013) and climate model evaluation

Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2017-81
Manuscript under review for journal Earth Syst. Dynam.
Discussion started: 25 September 2017




(Garratt, 1994; Ma et al., 2014; Wild et al., 2015). As part of so-called climate or meteorological forcing datasets such as those generated within the Global Soil Wetness Project (GSWP; Zhao and Dirmeyer, 2003), at Princeton University (Sheffield et al., 2006) and within the WATer and global CHange project (WATCH; Weedon et al., 2011), the longwave and shortwave components of surface downwelling radiation (abbreviated as rlds and rsds or just longwave and shortwave radiation in the following) are used to, e.g., correct model biases in climate model output (Hempel et al., 2013; Iizumi et al., 2017; Cannon, 2017) and drive simulations of climate impacts (Müller Schmied et al., 2016; Veldkamp et al., 2017; Chang et al., 2017; Krysanova and Hattermann, 2017; Ito et al., 2017).

These meteorological forcing datasets are global, long-term meteorological reanalysis datasets such as those produced by the National Centers for Environmental Prediction-National Center for Atmospheric Research (NCEP-NCAR; Kalnay et al., 1996; Kistler et al., 2001) and the European Centre for Medium-Range Weather Forecasts (ECMWF; Uppala et al., 2005; Dee et al., 2011), refined by bias correction using global, gridded observational data. For the components of surface downwelling radiation, such a bias correction is often necessary as observations of these variables are not assimilated in the reanalyses, which makes them subject to modelling biases of, e.g., land-atmosphere interactions and cloud processes (Kalnay et al., 1996; Ruane et al., 2015).

Different approaches are adopted in order to carry out these bias corrections. Weedon et al. (2011, 2014) apply indirect corrections at the monthly time scale using near-surface air temperature observations for rlds and observations of atmospheric aerosol loadings and cloudiness for rsds. Sheffield et al. (2006) directly rescale rlds and rsds to match observed multi-year monthly mean values. Ruane et al. (2015) directly adjust distributions of daily mean rsds. The observational dataset commonly used for such direct adjustments of rlds and rsds is the Surface Radiation Budget (SRB) dataset assembled by the National Aeronautics and Space Administration (NASA) and the Global Energy and Water EXchanges project (GEWEX; Stackhouse Jr. et al., 2011).

Another meteorological forcing dataset, the EartH2Observe, WFDEI and ERA-Interim data Merged and Bias-corrected for ISIMIP (EWEMBI; Lange, 2016), was recently assembled to be used as the reference dataset for bias correction of global climate model output within the Inter-Sectoral Impact Model Intercomparison Project phase 2b (ISIMIP2b; Frieler et al., 2016). The surface downwelling longwave and shortwave radiation data included in EWEMBI are based on daily rlds and rsds from the climate forcing dataset compiled for the EartH2Observe project (E2OBS; Calton et al., 2016). In order to reduce deviations of E2OBS rlds and rsds statistics from the corresponding SRB estimates over tropical land (Dutra, 2015), for EWEMBI, the former were bias-adjusted to the latter at the daily time scale using two newly developed parametric quantile mapping methods.

These methods are conceptually similar to the Ruane et al. (2015) method, which fits beta distributions to reanalysed and observed daily mean rsds for every calendar month, thereby accounting for upper and lower physical limits of rsds using the multi-year monthly maximum value as the upper and zero as the lower limit of the distribution, and then uses quantile mapping to adjust the distributions. In contrast to Ruane et al. (2015), the methods developed to adjust E2OBS rlds and rsds for EWEMBI applies moving windows to estimate beta distribution parameters for every day of the year. This precludes discontinuities at the turn of the month (Rust et al., 2015; Gennaretti et al., 2015) and promises a better bias correction where the seasonality of radiation is very pronounced such as for rsds at high latitudes. Also, the new methods estimate the physical upper limits of rlds

Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2017-81
Manuscript under review for journal Earth Syst. Dynam.
Discussion started: 25 September 2017




and rsds differently, acknowledging that these limits are necessarily greater than or equal to the greatest value measured over any fixed period. Lastly, while Ruane et al. (2015) linearly interpolate SRB rsds from its natural horizontal resolution of 1.0° to the 0.5° reanalysis grid prior to bias correction, the new methods aggregate the E2OBS data from their original 0.5° grid to the 1.0° SRB grid, where the bias correction is done, and disaggregates these aggregated and bias-corrected data back to the E2OBS grid. Depending on the disaggregation method, this approach promises to generate bias-corrected data with more realistic temporal as well as spatial variability.

The new methods are comprehensively described and cross-validated in this article, and in order to assess the value added by the bias correction, the E2OBS and EWEMBI rlds and rsds are compared to independent ground observations from the Baseline Surface Radiation Network (BSRN; König-Langlo et al., 2013). Moreover, several modifications of the new methods are tested that differ in how they handle the spatial resolution gap between E2OBS and SRB, and how they account for the physical upper limits of rlds and rsds. Also tested are bias correction methods that operate at the monthly time scale as it is unclear a priori if bias correction of daily or monthly mean values yields better validation results at either time scale. The lessons learned from these analyses shall benefit bias corrections of surface downwelling radiation to be carried out in future generations of climate forcing datasets.

## 2 Data

### 2.1 E2OBS

The EartH2Observe (E2OBS; Dutra, 2015; Calton et al., 2016) daily mean rlds and rsds data bias-corrected for EWEMBI cover the whole globe on a regular $0.5° \times 0.5°$ latitude-longitude grid and span the 1979–2014 time period. Over the ocean, E2OBS rlds and rsds are identical to bilinearly interpolated ERA-Interim (ERAI; Dee et al., 2011) rlds and rsds. Over land, they are identical to WATCH Forcing Data methodology applied to ERA-Interim reanalysis data (WFDEI; Weedon et al., 2014) rlds and rsds. WFDEI rlds, in turn, is identical to bilinearly interpolated ERAI rlds, adjusted for elevation differences between the ERAI and Climatic Research Unit (CRU; Harris et al., 2013) grids. WFDEI rsds is identical to bilinearly interpolated ERAI rsds bias-corrected at the monthly time scale using CRU TS3.1/3.21 mean cloud cover and considering effects of interannual changes in atmospheric aerosol optical depths (Weedon et al., 2010, 2011, 2014).

### 2.2 SRB

The E2OBS data are bias-corrected using the NASA-GEWEX Surface Radiation Budget (SRB; Stackhouse Jr. et al., 2011) primary-algorithm estimates of daily mean rlds and rsds from SRB Release 3.1 and 3.0, respectively. These data cover the whole globe on a regular $1.0° \times 1.0°$ latitude-longitude grid and span the 07/1983–12/2007 time period. For bias correction and cross-validation, a 24-year subsample of these data is used which spans the 12/1983–11/2007 time period. Additional data from the adjacent months 11/1983 and 12/2007 are employed for computations of running mean values. The SRB estimates of rlds and rsds are based on satellite-derived cloud parameters and ozone fields, reanalysis meteorology and a few other ancillary

Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2017-81
Manuscript under review for journal Earth Syst. Dynam.
Discussion started: 25 September 2017


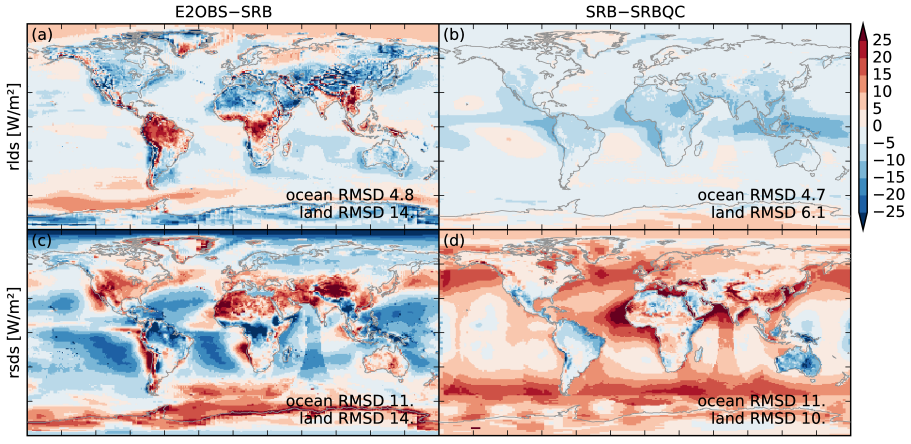


**Figure 1.** Deviations of E2OBS from SRB (**left**) and SRB from SRBQC (**right**) 12/1983–11/2007 mean longwave (**top**) and shortwave (**bottom**) radiation. Root-mean-square deviations (RMSDs) over all ocean and all land grid cells are given at the bottom of each panel.

datasets. Due to a lack of satellite coverage during most of the 07/1983–06/1998 time period over an area centred at 70°E, SRB data artefacts are present over the Indian Ocean (https://gewex-srb.larc.nasa.gov/common/php/SRB_known_issues.php; cf. Fig. 1).

Deviations of E2OBS from SRB long-term mean rlds and rsds are shown in Fig. 1, together with corresponding deviations of SRB from SRB Release 3.0 quality-check (SRBQC) products. The SRB and SRBQC products were produced with different algorithms (Stackhouse Jr. et al., 2011). Since the primary-algorithm products are more reliable than the quality-check products (Zhang et al., 2015; Stackhouse Jr. et al., 2011) the former were used for the bias correction of E2OBS rlds and rsds for EWEMBI. Over land, differences in long-term mean radiation between E2OBS and SRB are greater in magnitude than those between SRB and SRBQC. Over the ocean, the differences are of similar magnitude. If deviations of SRB from SRBQC data quantify methodological uncertainty inherent to the SRB data then these findings justify the bias correction of E2OBS rlds and rsds using SRB data over land at least.

### 2.3 BSRN

Ground observations of longwave downward and shortwave downward (global) radiation made at 54 stations of the Baseline Surface Radiation Network (BSRN; Table 1; König-Langlo et al., 2013) are used as independent validation data for rlds and rsds, respectively. BSRN measurements began at a few stations in 1992. The latest measurements included here are from 2014. Daily mean values of BSRN measurements, which are taken every minute or every few minutes, depending on the station, are computed in two steps. First, gaps no longer than 467/11 minutes in the original rlds/rsds time series are filled by linearly interpolation between values right before the beginning and after the end of a gap, as suggested by Schild (2016; for statistics of BSRN data gaps see Roesch et al., 2011). Daily mean values are then calculated for days that are fully covered by these gap-filled values. Prior to data comparison, the resulting BSRN data availability masks are applied to the original and bias-

Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2017-81
Manuscript under review for journal Earth Syst. Dynam.
Discussion started: 25 September 2017


Earth System Dynamics Discussions
Open Access
EGU

**Table 1.** Observations made at the following 54 BSRN stations are used in this study. In order to adjust rlds for elevation differences between BSRN stations and E2OBS-grid cells, prior to data comparison, BSRN rlds values are offset by the values listed in the rightmost column, based on the formula proposed by Stackhouse Jr. et al. (2011; see text).

| station | latitude [°N] | longitude [°E] | offset [$\mathrm{Wm^{-2}}$] | station | latitude [°N] | longitude [°E] | offset [$\mathrm{Wm^{-2}}$] |
|---|---|---|---|---|---|---|---|
| ALE | 82.451 | -62.508 | -6.580 | IZA | 28.500 | -16.300 | 57.316 |
| ASP | -23.798 | 133.888 | -4.256 | KWA | 8.720 | 167.731 | 0.252 |
| BAR | 71.323 | -156.607 | 0.000 | LAU | -45.045 | 169.689 | -7.420 |
| BER | 32.267 | -64.667 | -0.112 | LER | 60.140 | -1.185 | 1.232 |
| BIL | 36.605 | -97.515 | -0.560 | LIN | 52.210 | 14.122 | 1.764 |
| BON | 40.060 | -88.370 | -0.280 | MAN | -2.058 | 147.425 | -0.952 |
| BOU | 40.048 | -105.007 | -9.772 | MNM | 24.288 | 153.983 | 0.000 |
| BRB | -15.601 | -47.713 | -0.504 | NAU | -0.521 | 166.916 | -0.084 |
| CAB | 51.971 | 4.927 | -0.056 | NYA | 78.925 | 11.950 | -3.388 |
| CAM | 50.217 | -5.317 | 0.588 | PAL | 48.713 | 2.208 | 1.932 |
| CAR | 44.083 | 5.059 | -14.840 | PAY | 46.815 | 6.944 | -6.076 |
| CLH | 36.905 | -75.713 | 0.896 | PSU | 40.720 | -77.930 | 0.028 |
| CNR | 42.816 | -1.601 | -3.500 | PTR | -9.069 | -40.320 | 0.504 |
| COC | -12.193 | 96.835 | 0.140 | REG | 50.205 | -104.713 | 0.336 |
| DAA | -30.665 | 23.993 | 0.476 | SAP | 43.060 | 141.328 | -4.200 |
| DAR | -12.425 | 130.891 | 0.812 | SBO | 30.860 | 34.779 | -1.764 |
| DOM | -75.100 | 123.383 | 0.534 | SMS | -29.443 | -53.823 | 1.932 |
| DRA | 36.626 | -116.018 | -3.780 | SON | 47.054 | 12.958 | 39.424 |
| EUR | 79.980 | -85.930 | -5.740 | SOV | 24.910 | 46.410 | -3.640 |
| FLO | -27.533 | -48.517 | -9.324 | SPO | -89.983 | -24.799 | 0.290 |
| FPE | 48.310 | -105.100 | -1.204 | SXF | 43.730 | -96.620 | -0.056 |
| FUA | 33.582 | 130.375 | -1.092 | SYO | -69.005 | 39.589 | -14.012 |
| GCR | 34.255 | -89.873 | -0.112 | TAM | 22.790 | 5.529 | -1.120 |
| GOB | -23.561 | 15.041 | -4.788 | TAT | 36.058 | 140.126 | -0.924 |
| GVN | -70.650 | -8.250 | -0.097 | TIK | 71.586 | 128.919 | -1.484 |
| ILO | 8.533 | 4.566 | 2.492 | TOR | 58.254 | 26.462 | -0.028 |
| ISH | 24.337 | 124.163 | -0.504 | XIA | 39.754 | 116.962 | 0.280 |

corrected E2OBS time series from the respective E2OBS-grid cells. Additionally, BSRN rlds values are adjusted for elevation differences between BSRN stations and E2OBS-grid cells as proposed by Stackhouse Jr. et al. (2011). For elevations $z_{\mathrm{BSRN}}$ of BSRN stations and $z_{\mathrm{E2OBS}}$ of E2OBS-grid cells, BSRN rlds values are offset by $0.028\,(z_{\mathrm{BSRN}} - z_{\mathrm{E2OBS}})\,\mathrm{Wm^{-3}}$ (cf. Table 1).

Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2017-81
Manuscript under review for journal Earth Syst. Dynam.
Discussion started: 25 September 2017




## 3 Methods

The parametric quantile mapping methods introduced in the following are named according to the scheme $\mathrm{BC}vtpx$, where $v,t,p$ are used to distinguish between methods for longwave and shortwave radiation ($v=\mathrm{l},\mathrm{s}$) operating at the daily and monthly time scale ($t=\mathrm{d},\mathrm{m}$) using basic and advanced distribution types or parameter estimation techniques ($p=\mathrm{b},\mathrm{a}$). Index $x=0,1,2$ is used for variants of these methods that differ in how they handle the spatial resolution gap between SRB and E2OBS. The $\mathrm{BC}vtp0$ methods correct E2OBS data directly at the E2OBS grid using bilinearly interpolated SRB data. For bias correction with the $\mathrm{BC}vtp1$ methods, E2OBS data are spatially aggregated to the SRB grid, the aggregated data are then bias-corrected and the resulting data disaggregated back to the E2OBS grid. For the $\mathrm{BC}vtp2$ methods, the sub-SRB-grid scale spatial structure of the original E2OBS data is imposed upon spatially disaggregated SRB data prior to bias correction at the E2OBS grid. The bias correction of E2OBS rlds and rsds for EWEMBI was done with methods $\mathrm{BCl da1}$ and $\mathrm{BCs da1}$, respectively. Since the $\mathrm{BC}vtp0$ and $\mathrm{BC}vtp2$ methods are based on the $\mathrm{BC}vtp1$ methods, the latter are introduced first.

### 3.1 Bias correction at the SRB grid

For the $\mathrm{BC}vtp1$ methods, E2OBS rlds and rsds are aggregated to the SRB grid using a first-order conservative remapping scheme (Jones, 1999). This ensures that each aggregated value is the grid-cell area-weighted mean of the underlying four E2OBS values. In the following, the methods of bias correction of these aggregated values are described.

The $\mathrm{BC}vtp1$ methods use parametric transfer functions of the form $F_{vtp}^{\mathrm{SRB}^{-1}}(F_{vtp}^{\mathrm{E2OBS}}(\cdot))$, where $F_{vtp}^{\mathrm{E2OBS}}$ and $F_{vtp}^{\mathrm{SRB}}$ are climatological cumulative distribution functions (CDFs) of aggregated E2OBS and SRB data, respectively, estimated at daily temporal resolution for each SRB-grid cell individually (Fig. 2). In order to quantify the extent to which bias correction results benefit from explicitly accounting for physical radiation limits, the basic and advanced methods $\mathrm{BCl}t\mathrm{b1}$ and $\mathrm{BCl}t\mathrm{a1}$ for longwave radiation use normal and beta distributions, respectively. For shortwave radiation, the relevance of physical limits is less questionable, given that the lower limit of zero matters at least during polar night, and that the solar radiation incident upon land and ocean surfaces is limited by the solar radiation incident upon the top of the atmosphere (cf. Fig. 2). Therefore, all $\mathrm{BCs}tp1$ methods use beta distributions and the basic and advanced methods only differ in how they estimate the beta distribution parameters (cf. Fig. 2, Table 2).

#### 3.1.1 Bias correction at the daily time scale

The parameters of the climatological CDFs $F_{\mathrm{vd}p}^{\mathrm{E2OBS}}$ and $F_{\mathrm{vd}p}^{\mathrm{SRB}}$ are estimated based on empirical multi-year mean values, variances and maximum values of daily mean radiation from the 12/1983–11/2007 time period. Data from the whole period were used for the production of EWEMBI rlds and rsds. Data from some half of the period (cf. Sect. 4.1) are used for cross-validation in this study.

For shortwave radiation, the basic daily bias correction method is designed to resemble the method outlined by Ruane et al. (2015, Sect. 3.4). $\mathrm{BCsdb1}$ estimates mean values and variances of climatological beta distributions by 25-day running mean values of multi-year daily mean values and variances, respectively, and their upper bounds by 25-day running mean values of

Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2017-81
Manuscript under review for journal Earth Syst. Dynam.
Discussion started: 25 September 2017


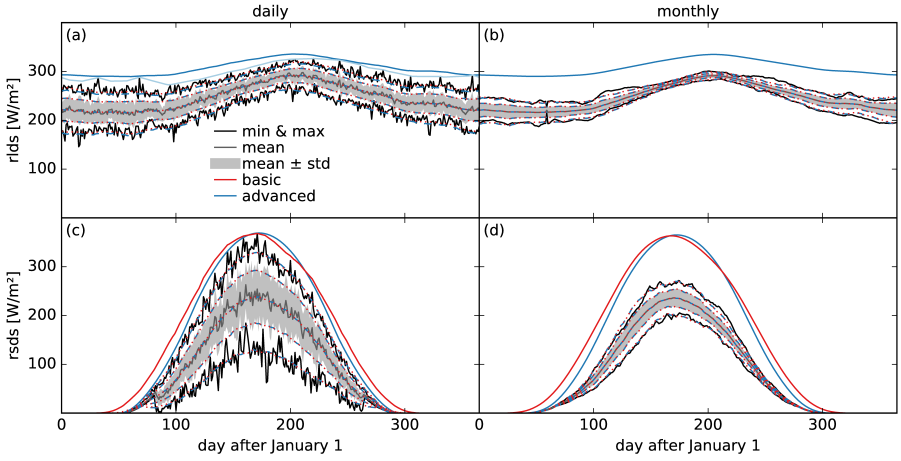


**Figure 2.** Parameters of climatological distributions of longwave (**top**) and shortwave (**bottom**) radiation are estimated based on empirical 24-year mean values (dark grey), standard deviations (light grey range around mean values) and minimum and maximum values (black) of daily mean (**left**) and 31-day running mean (**right**) radiation computed individually for every day of the year. The distribution parameters estimated for the basic (red) and advanced (blue) bias correction methods (cf. Table 2) include mean values and standard deviations (dotted red, dashed blue), and upper bounds (solid) where beta distributions are used. The light-blue line in (**a**) represents 25-day running mean values of 25-day running maximum values of 24-year maximum values of daily mean rlds, which are used to estimate the upper bounds of the climatological beta distributions used by the BCld$a1$ method (solid blue; see text). Please note that the lower bounds of all climatological beta distributions are set to zero. The lowermost and uppermost dotted red and dashed blue lines are the medians of sample minimum and maximum values of random samples of length 24 drawn from the estimated climatological distributions. This plot is based on SRB daily mean rlds and rsds data from 79.5°N, 12.5°E and the 12/1983–11/2007 time period.

25-day running maximum values of multi-year maximum values of daily mean rsds (solid red line in Fig. 2c). The idea behind this upper bound estimate is that 25-day running maximum values of multi-year maximum values of daily mean rsds resemble the multi-year monthly maximum values of daily mean rsds used by Ruane et al. (2015). Please note that using the same window length for the running maximum calculation and the additional smoothing ensures that the resulting upper bounds are always greater than or equal to the multi-year maximum values of daily mean rsds.

The BCsda$1$ method employs the climatology of daily mean shortwave insolation at the top of the atmosphere (rsdt; see Sect. A for how rsdt is calculated in this study) for the upper bound estimation. The rsdt climatology at a given latitude is rescaled such that it sits just above the multi-year maximum values of daily mean rsds on all days with rsdt $\geq 50\,\mathrm{Wm}^{-2}$. On a given day of the year, the maximum of this rescaled rsdt value and the empirical multi-year maximum daily mean rsds is then used as the upper bound of the beta distribution (solid blue line in Fig. 2c). The reason for handling days with rsdt below and above $50\,\mathrm{Wm}^{-2}$ separately is that during dusk and dawn of polar night, rsds can exceed rsdt due to diffuse radiation coming in from lower latitudes. Mean values and variances of the climatological beta distributions of the BCsda$1$ method are estimated by running mean values of multi-year daily mean values and variances, respectively. The window length used for these running

Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2017-81
Manuscript under review for journal Earth Syst. Dynam.
Discussion started: 25 September 2017


Earth System Dynamics Discussions Open Access EGU

**Table 2.** Distribution types and parameter estimation methods of bias correction methods BC$vtp$1 (cf. Fig. 2). Please note that the lower bounds of all climatological beta distributions are set to zero and that 24-year statistics are replaced by 12-year statistics for cross-validation.

| method | distribution type | mean value | variance | upper bound |
|---|---|---|---|---|
| BCldb1 | normal | rm25ym24[1] | rm25ys24[4] | — |
| BClda1 | beta | rm25ym24[1] | rm25ys24[4] | rm25rx25yx24-rm25ym24[7] |
| BClmb1 | normal | ym24rm31[2] | ys24rm31[5] | — |
| BClma1 | beta | ym24rm31[2] | ys24rm31[5] | rm31lda[8] |
| BCsdb1 | beta | rm25ym24[1] | rm25ys24[4] | rm25rx25yx24[9] |
| BCsda1 | beta | rm25*ym24[3] | rm25*ys24[6] | yx24-rsdt[10] |
| BCsmb1 | beta | ym24rm31[2] | ys24rm31[5] | rm31sdb[11] |
| BCsma1 | beta | ym24rm31[2] | ys24rm31[5] | rm31sda[12] |

[1] 25-day running mean value of 24-year daily mean values
[2] 24-year daily mean value of 31-day running mean values, with February 29 value replaced by average of February 28 and March 1 values
[3] 25-or-fewer-day running mean value of 24-year daily mean values (see text)
[4] 25-day running mean value of 24-year daily variances
[5] 24-year daily variance of 31-day running mean values, with February 29 value replaced by average of February 28 and March 1 values
[6] 25-or-fewer-day running mean value of 24-year daily variances (see text)
[7] affine transformation of mean value climatology of BClda1 that sits just above the 25-day running mean values of 25-day running maximum values of 24-year maximum values of daily mean rlds (see text)
[8] 31-day running mean value of upper bounds of BClda1 method
[9] 25-day running mean value of 25-day running maximum values of 24-year maximum values of daily mean rsds
[10] rescaled rsdt climatology that sits just above 24-year maximum values of daily mean rsds (see text)
[11] 31-day running mean value of upper bounds of BCsdb1 method
[12] 31-day running mean value of upper bounds of BCsda1 method

mean calculations is 25 days by default. On days that are fewer than 13 days away from the beginning or end of polar night (as defined by daily mean rsdt going to zero), the window length is shortened to $2n+1$, where $n$ is the number of days between the day in question and the beginning or end of polar night.

For longwave radiation, both the basic and the advanced daily bias correction methods use 25-day running mean values of multi-year daily mean values and variances to estimate climatological mean values and variances, respectively. The upper bounds used by BClda1 are not estimated by the often rather unsmooth 25-day running mean values of 25-day running maximum values of 24-year maximum values of daily mean rlds (rm25rx25yx24; solid light-blue line in Fig. 2a) but by a suitably shifted and rescaled mean value climatology: First, the mean value climatology curve is shifted and rescaled such that it best fits rm25rx25yx24 according to ordinary least squares. This fitted curve is then shifted once more such that the resulting upper bound climatology sits just above rm25rx25yx24 (solid blue line in Fig. 2a).

Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2017-81
Manuscript under review for journal Earth Syst. Dynam.
Discussion started: 25 September 2017




Since the choice of the window length used for all of the running mean and maximum calculations mentioned above is somewhat arbitrary, the window length dependence of the overall performances of the BC$v$da1 methods is investigated in Sect. C. Sensitivities are found to be very low for window lengths between 10 and 40 days.

#### 3.1.2 Bias correction at the monthly time scale

In order to mimic a bias correction at the monthly time scale as is was done by, e.g., Sheffield et al. (2006, Sect. 3.d.3), the BC$v$m$p$1 methods bias-correct 31-day running mean values and then rescale each daily value by the corrected-to-uncorrected ratio of the respective 31-day running mean value.

Mean values and variances of the climatological CDFs $F_{vmp}^{\mathrm{E2OBS}}$ and $F_{vmp}^{\mathrm{SRB}}$ of 31-day running mean values are simply estimated by 24-year (or 12-year for cross-validation) daily mean values and variances of 31-day running mean values, respectively, with February 29 values replaced by averages of February 28 and March 1 values. Upper bounds of beta distributions are estimated by 31-day running mean values of the upper bounds of the corresponding daily CDFs $F_{vdp}^{\mathrm{E2OBS}}$ and $F_{vdp}^{\mathrm{SRB}}$ (cf. Fig. 2, Table 2) as 31-day running mean values of multi-year maximum values of daily mean radiation are always greater than or equal to multi-year maximum values of 31-day running mean radiation.

#### 3.1.3 Disaggregation to the E2OBS grid

In principle, the disaggregation of aggregated and bias-corrected E2OBS data from the SRB to the E2OBS grid can be done in various ways. The simplest approach would arguably be a mere interpolation, which is disadvantageous since it ignores the sub-SRB-grid scale spatial variability present in the original E2OBS data. Probabilistic disaggregation methods, on the other hand, that are designed to retain that variability (cf. Sheffield et al., 2006, Sect. 3.b.1), are impractical if, as in the present case, the purpose of the disaggregation is the construction and publication of a dataset, because all variants of the dataset that can potentially be generated by a probabilistic algorithm are, as long as all conceivable constraints have been incorporated in the algorithm, equally plausible candidates for the one dataset to be published. Therefore, the following deterministic disaggregation approach was used for the construction of EWEMBI rlds and rsds and is adopted here for all BC$v$t$p$1 methods.

First, E2OBS-grid scale upper bounds of daily mean radiation are estimated by bilinearly interpolated maximum values of the climatological upper bounds of SRB all-sky and clear-sky radiation, which, in turn, are estimated using the BClda1 method for rlds and the BCsda1 methods for rsds (cf. Table 2 and blue lines in Fig. 2a,c). The clear-sky radiation data are included in order to prevent the E2OBS-grid scale upper bounds from being much lower than the real physical limits of daily mean radiation at that spatial scale, given that due to sub-SRB-grid scale spatial variability, upper radiation bounds at the E2OBS-grid scale may exceed those at the SRB-grid scale.

The original daily E2OBS data are then clamped between zero and these upper bounds, and the resulting values are rescaled day by day and SRB-grid cell by SRB-grid cell such that their SRB-grid scale aggregates match the respective bias-corrected values. More precisely, if the SRB-grid scale aggregate of the (clamped) original values from the four E2OBS-grid cells contained in one SRB-grid cell is greater than the bias-corrected value then the four values are all reduced by a common factor. Otherwise, the distances of the four values to their climatological upper bounds are reduced by a common factor.

Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2017-81
Manuscript under review for journal Earth Syst. Dynam.
Discussion started: 25 September 2017




### 3.2 Bias correction at the E2OBS grid

#### 3.2.1 The BC$vtp$2 methods

The disaggregation method introduced above corrects the original E2OBS values from the four E2OBS-grid cells contained in one SRB-grid cell as if they must all be too low/high if their area-weighted average is too low/high. This implicit assumption is questionable since it rules out the possibility that the area-weighted average is too low because one of the four values is much too low while the others are slightly too high, to give just one example. A statistical manifestation of this problem is illustrated and discussed in Sect. 4.3.

The assumption does not need to made if the bias correction is carried out directly at the E2OBS grid. With target distributions fixed at the SRB grid, one possibility to define target distributions at the E2OBS grid is to require the bias-corrected data to (i) have the SRB-grid scale target distributions and (ii) retain the sub-SRB-grid scale structure of the original data. For parametric bias correction methods such as those introduced above, this can be achieved via suitable definitions of the parameters of the E2OBS-grid scale target distributions. Here, for every BC$vtp1$ method, a corresponding BC$vtp2$ method is defined to operate at the same temporal scale and to use the same source (at the E2OBS grid) and target (at the SRB grid) distribution type and parameter estimation technique (cf. Table 2). E2OBS-grid scale target climatologies of mean values, variances and (where necessary) upper bounds are defined as follows.

The mean value estimates of the original E2OBS data are shifted by a common offset per SRB-grid cell and day of the year to obtain the E2OBS-grid scale target mean values. The offsets are chosen such that the E2OBS-grid scale target mean values aggregated to the SRB grid match the corresponding SRB mean value estimates. E2OBS data bias-corrected using these E2OBS-grid scale target mean values have SRB grid-scale aggregates that match the SRB grid-scale target mean values because (i) the aggregation is a linear operation and (ii) the mean value of a linear combination of random variables is equal to the same linear combination of the mean values of these random variables.

The variance estimates of the original E2OBS data are rescaled by a common factor $f_{ij}$ per day $i$ of the year and SRB-grid cell $j$ to obtain the E2OBS-grid scale target variances. For the derivation of the formula for $f_{ij}$ let $Y_{ijk}$ (and $X_{ijk}$) denote random variables representing bias-corrected (and original) E2OBS data from day $i$ of the year and E2OBS-grid cells $k = 1, 2, 3, 4$ contained in SRB-grid cell $j$. Then the estimated variance of the aggregate of $Y_{ijk}$ can be expanded to

$$\mathrm{Var}\left(\sum_{k=1}^{4} w_{jk} Y_{ijk}\right) = \sum_{k,l=1}^{4} w_{jk} w_{jl} \,\mathrm{Cor}(Y_{ijk}, Y_{ijl}) \sqrt{\mathrm{Var}(Y_{ijk})\mathrm{Var}(Y_{ijl})}, \tag{1}$$

where $w_{jk}$ is the area weight of E2OBS-grid cell $jk$ with $\sum_{k=1}^{4} w_{jk} = 1$ for all $j$, $\mathrm{Cor}(Y_{ijk}, Y_{ijl})$ is the estimated Pearson correlation between $Y_{ijk}$ and $Y_{ijl}$, and $\mathrm{Var}(Y_{ijk})$ is the estimated variance of $Y_{ijk}$. A bias correction would be deemed successful if the left-hand side of Eq. (1) was equal to the estimated variance of $Z_{ij}$, the SRB data from day $i$ of the year and grid cell $j$. On the right-hand side of Eq. (1), $f_{ij}\,\mathrm{Var}(X_{ijk})$ can be substituted for $\mathrm{Var}(Y_{ijk})$ by definition of the scaling factors, and $\mathrm{Cor}(Y_{ijk}, Y_{ijl})$ can be approximated by $\mathrm{Cor}(X_{ijk}, X_{ijl})$ since parametric quantile mapping preserves trends and therefore rank correlations and therefore approximately Pearson correlations. The variance scaling factors for method BC$vtp2$ are therefore

Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2017-81
Manuscript under review for journal Earth Syst. Dynam.
Discussion started: 25 September 2017




calculated based on

$$\mathrm{Var}Z_{ij} = f_{ij} \sum_{k,l=1}^{4} w_{jk} w_{jl} \,\mathrm{Cor}(X_{ijk}, X_{ijl}) \sqrt{\mathrm{Var}(X_{ijk})\mathrm{Var}(X_{ijl})}, \tag{2}$$

where the variances are estimated using the respective $\mathrm{BC}vtp1$ approach (Table 2), and the Pearson correlations are estimated as inversely Fisher-transformed 25-day running mean values of Fisher-transformed 24-year daily Pearson correlations of daily (for $\mathrm{BC}vdp2$) or 31-day running mean (for $\mathrm{BC}vmp2$) radiation data. The Fisher transformations are invoked here in order to approximately account for correlation value-dependent sampling error intervals (Fisher, 1915, 1921).

The E2OBS-grid scale target upper bounds are calculated in the same way as the E2OBS-grid scale target mean values. This way, the latter rarely exceed the former. Where they do, the latter are reduced to 99% of the former. This reduction is only necessary for some of the very low rsds values that occur under (near-)polar night conditions.

In order to obtain realistic E2OBS-grid scale target beta distributions, we further limit the E2OBS-grid scale target variances calculated using Eq. (2) to 40% of $\mu(b-\mu)$, where $\mu$ and $b$ are the E2OBS-grid scale target mean values and upper bounds, respectively. This limit is imposed because (i) the variance $\sigma^2$ of a random variable taking values from within the interval $[a,b]$ can generally not be greater than $(\mu-a)(b-\mu)$ if $\mu$ is the random variable's mean value, (ii) if that random variable is beta-distributed and $\sigma^2 > (\mu-a)(b-\mu)/2$ then the probability density function is U-shaped (Wilks, 1995), which is considered unrealistic for climatological distributions of rlds and rsds, and (iii) $\sigma^2/(\mu(b-\mu))$ has an empirical upper limit of about 40% in the original E2OBS radiation data.

#### 3.2.2 The $\mathrm{BC}vtp0$ methods

For the $\mathrm{BC}vtp0$ methods, daily SRB data are first bilinearly interpolated to the E2OBS grid. The E2OBS data are then bias-corrected directly at the E2OBS grid using the interpolated SRB data and transfer functions defined exactly as for the respective $\mathrm{BC}vtp1$ method.

## 4 Results

The bias correction methods introduced above are assessed in a threefold way. First, original and bias-corrected E2OBS data are compared to SRB data at the SRB-grid scale using a cross-validation approach. Secondly, they are compared to independent ground observations made at 54 BSRN stations. Thirdly, sub-SRB-grid scale spatial variability before and after bias correction are compared in order to measure the disaggregation performance of all methods.

Data comparisons are done at the daily and monthly time scale in order to identify strengths and weaknesses of bias correction methods operating at either of these time scales. Metrics used to quantify statistical dissimilarity between E2OBS and SRB or BSRN data include differences between multi-year mean values, standard deviations, skewness and maximum values, root-mean-square deviations (RMSDs) between time series, and $p$-values of two-sample Kolmogorov-Smirnov (KS) test statistics for empirical CDF comparisons (see Sect. B for details).

Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2017-81
Manuscript under review for journal Earth Syst. Dynam.
Discussion started: 25 September 2017




For the cross-validation against SRB data, 24 years worth of overlapping E2OBS and SRB data are divided into two 12-year samples of which the first is used to calibrate and the second to validate the method. Switanek et al. (2017) have shown that if climatological distributions differ substantially between calibration and validation samples of either the observed (here SRB) or modelled (here E2OBS) data (such differences are hereafter denoted as calibration-validation climate change signals or CVCCSs), then the remaining biases after quantile mapping trained on the calibration data sample and applied to the validation data sample are dominated by differences between observed and modelled CVCCSs. This implies that calibration and validation data samples should be made as statistically similar as possible if the cross-validation is to only measure the bias correction methods' imperfections. Hence, here, calibration and validation data samples are composed of data from every second and every other year or vice versa, respectively. The samples are accordingly labelled every1st and every2nd.

### 4.1 Cross-validation against SRB data

In the following, cross-validation results are only shown and discussed for the $\mathrm{BC}vtp0$ and $\mathrm{BC}vtp1$ methods since results for the corresponding $\mathrm{BC}vtp1$ and $\mathrm{BC}vtp2$ are virtually identical. In order to (i) measure how the use of spatially interpolated SRB data for bias correction impacts SRB-grid scale biases, and (ii) assess the value of the extra complications involved in the parameter estimations of the advanced compared to the basic bias correction methods, biases in multi-year daily mean values, standard deviations, skewness and maximum values remaining after bias correction with methods $\mathrm{BC}v\mathrm{da}0$, $\mathrm{BC}v\mathrm{da}1$ and $\mathrm{BC}v\mathrm{db}1$ are compared first. Then, bias correction methods operating at different temporal scales are compared with respect to their ability to adjust the interannual variability of monthly mean values. Lastly, the overall performance of all $\mathrm{BC}vtp1$ methods is assessed via CDF comparisons at both the daily and monthly time scale.

Maps of biases in multi-year mean values, standard deviations, skewness and maximum values of daily mean rlds and rsds remaining after bias correction at the daily time scale are depicted in Figs. 3 and 4. Remaining mean value biases for $\mathrm{BC}vdp1$ are small with medians over calendar months and validation data samples being within $\pm 1\,\mathrm{W m^{-2}}$ at most locations. At low/high latitudes, $\mathrm{BCsdb}1$ leaves smaller/larger mean value biases than $\mathrm{BCsda}1$. In comparison to $\mathrm{BC}v\mathrm{da}1$, $\mathrm{BC}v\mathrm{da}0$ leaves greater mean value biases in particular over coastal and mountainous regions, where spatial gradients are large.

Medians of relative standard deviation biases remaining after bias correction with $\mathrm{BC}vdp1$ are mostly within $\pm 4\%$. Underestimations by more then $4\%$ remain over large parts of subtropical Northern Hemisphere land. In most locations, $\mathrm{BCldb}1$ leaves smaller rlds standard deviation biases than $\mathrm{BClda}1$. Bias correction with $\mathrm{BC}v\mathrm{da}0$ yields systematically too low standard deviations in most locations, in particular for shortwave radiation. This is a result of the variance deflation the bilinear interpolation inflicts on the SRB data.

Large skewness biases with medians frequently exceeding $\pm 50\%$ remain after bias correction with any method. The median skewness of longwave radiation is mostly too low, in particular over the ocean and no matter if CDFs of beta or normal distributions are used in the transfer function. The median skewness of shortwave radiation is too low over most of the tropics and high-latitude oceans and too high over most land masses and subtropical oceans. The biases of third- and higher-order moments of the distribution of daily mean radiation would arguably be better adjusted by non-parametric quantile mapping methods.

Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2017-81
Manuscript under review for journal Earth Syst. Dynam.
Discussion started: 25 September 2017


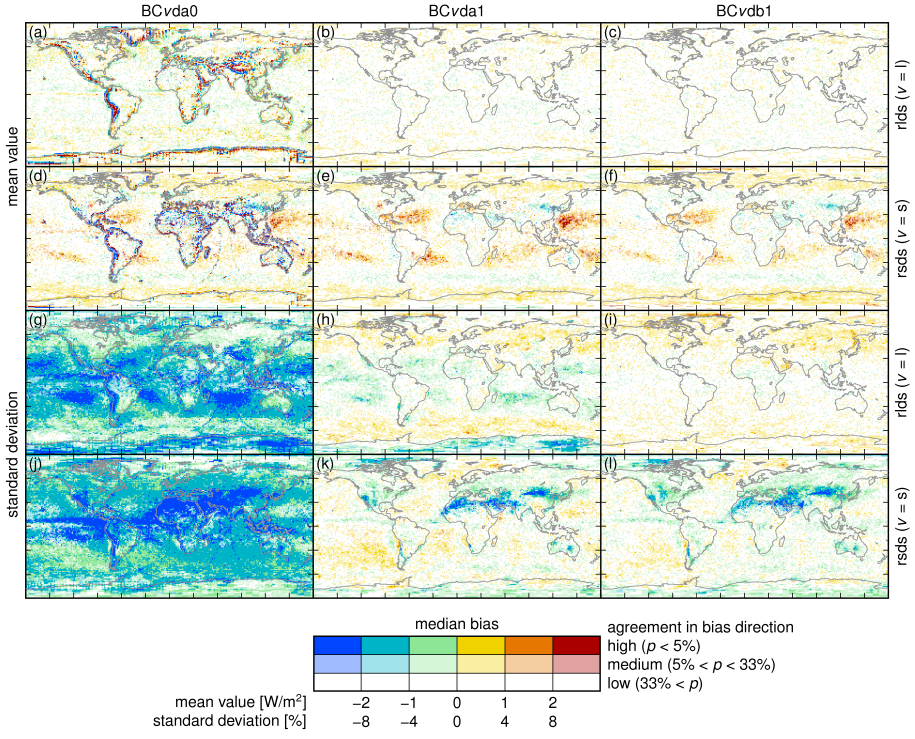


**Figure 3.** Biases relative to SRB in mean values (**a–f**) and standard deviations (**g–l**) of spatially aggregated (to the SRB grid) daily mean longwave (**a–c**, **g–i**) and shortwave (**d–f**, **j–l**) radiation after bias correction with methods BC$v$da0 (**left**), BC$v$da1 (**middle**) and BC$v$db1 (**right**). The biases are calculated individually for each calendar month (January to December) and calibration data sample (every1st, every2nd) pooling SRB and corrected E2OBS data from all years of the corresponding validation data sample (every2nd, every1st, respectively) and omitting shortwave radiation data from months with monthly mean rsdt less than 1 $\mathrm{Wm}^{-2}$ (cf. Sect. A and Fig. C1c). Depicted are median and agreement in direction of these individual biases, represented by hue and saturation of a grid cell's colour, respectively. Categories of agreement in bias direction are defined based on one-sided $p$-values obtained from modelling underestimations and overestimations for individual calendar months and validation data samples as possible outcomes of fifty-fifty Bernoulli trials.

Medians of remaining biases in 12-year maximum values are mostly within $\pm 10\,\mathrm{Wm}^{-2}$. Compared to BC$v$da1, these biases are shifted to more negative values for BC$v$da0. This is related to the negative standard deviation biases that remain after bias correction with BC$v$da0. For rlds, the use of beta instead of normal distributions clearly reduces the remaining maximum value biases. For rsds, the basic estimates of upper radiation bounds yield a slightly greater reduction of maximum value biases than the advanced ones.

Since multi-year mean values of monthly mean values are identical to multi-year mean values of the underlying daily values, bias correction at the monthly time scale adjusts multi-year mean values of daily rlds and rsds similarly well as bias correction at the daily time scale (cf. Fig. 3a–f). However, the BC$v$m$px$ methods leave larger and spatially less homogeneous biases of multi-year standard deviations and maximum values of daily mean radiation than the BC$v$d$px$ methods, with medians

Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2017-81
Manuscript under review for journal Earth Syst. Dynam.
Discussion started: 25 September 2017


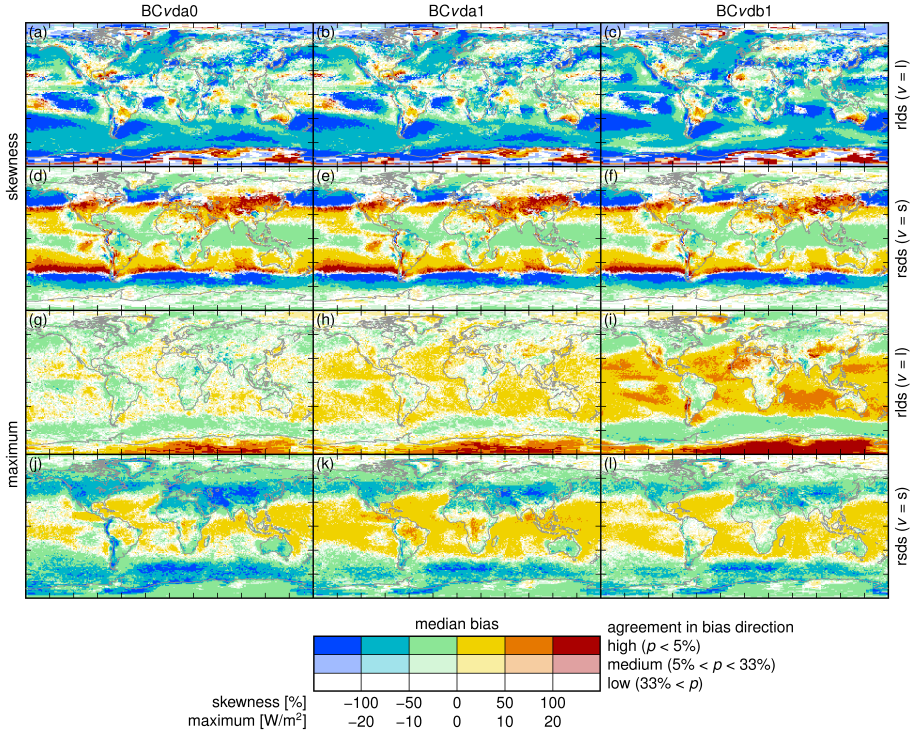


**Figure 4.** Same as Fig. 3 but for biases in skewness (**a**–**f**) and 12-year maximum values (**g**–**l**).

over calendar months and validation data samples being mostly within $\pm 20\%$ and $\pm 20\,\mathrm{W m^{-2}}$, respectively. In general, bias correction at the monthly time scale is expected to leave smaller biases at the monthly time scale than bias correction at the daily time scale. This is exemplified in Fig. 5, where median biases of interannual standard deviations of monthly mean rlds and rsds are shown to be mostly within/beyond $\pm 20\%$ after bias correction with $\mathrm{BC}v\mathrm{ma1}/\mathrm{BC}v\mathrm{da1}$.

The overall performance of the $\mathrm{BC}vtp1$ methods is examined next. As delineated above, it is measured by similarities of empirical CDFs of (spatially aggregated) E2OBS and SRB data before and after bias correction, quantified by $p$-values of two-sample KS test statistics (cf. Sect. B; greater $p$-values indicate stronger agreement of E2OBS and SRB distributions). For all radiation types, validation time scales, calendar months and $\mathrm{BC}vtp1$ methods, distributions of these $p$-values over calibration data sample, latitude and longitude are depicted as box-whisker plots in Fig. 6.

In all panels of Fig. 6, the $p = 10\%$ significance level is marked with a grey horizontal line. Any coincidence of a $p$-value distributions' 10th percentile with the 10% significance level suggests that that $p$-value distribution is in agreement with the null hypothesis of the KS test, which is that the compared samples were drawn from the same underlying distribution, and this indicates that the respective bias correction method did its job. Similarly, 10th percentiles of $p$-value distributions below/above the 10% significance level indicate undercorrections/overcorrections.

As expected by design, the $\mathrm{BC}vdp1$ methods outperform the $\mathrm{BC}vmp1$ methods at the daily time scale and the latter outperform the former at the monthly time scale. Yet performance gaps are much larger at the daily than at the monthly time

Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2017-81
Manuscript under review for journal Earth Syst. Dynam.
Discussion started: 25 September 2017


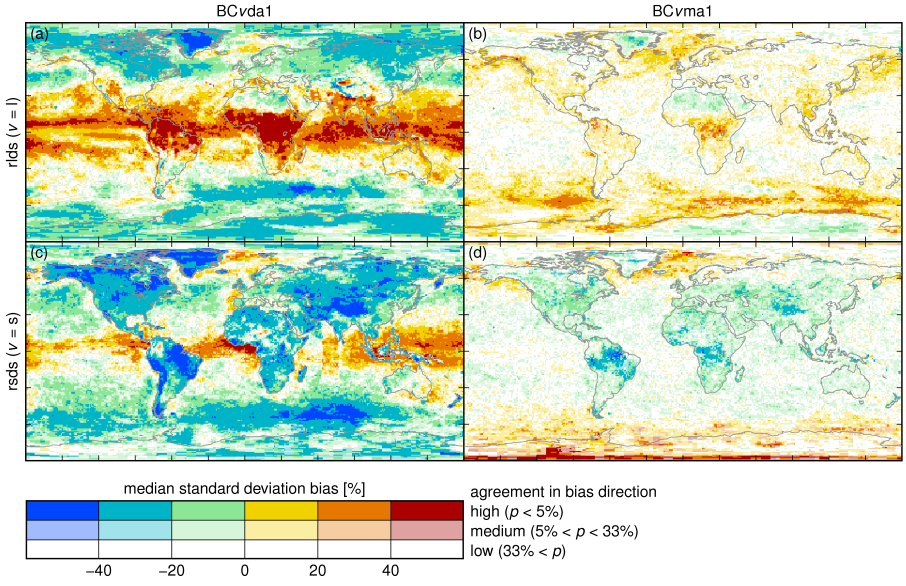


**Figure 5.** Same as Fig. 3 but for relative biases in interannual standard deviations of monthly mean radiation remaining after bias correction with methods BC$v$da1 (**left**) and BC$v$ma1 (**right**).

scale. The small performance gaps at the monthly time scale demonstrate that even though a direct bias correction of monthly mean values adjusts their distribution more precisely than a correction of the distribution of the underlying daily values (cf. Fig. 5), the statistical significance of this adjustment is low for sample sizes as small as in this cross-validation study. Rather, The $p$-value distributions depicted in Fig. 6b,d suggest that if sampling errors are taken into account then the BC$v$d$p$1 methods correct the distributions of monthly mean values almost as well as the BC$v$m$p$1 methods. In fact, both types of methods tend to overcorrect them.

In contrast, distributions of daily E2OBS data are undercorrected across the board. For BC$v$d$p$1, this is linked to an insufficient adjustment of third- and higher-order moments of the distributions of daily mean radiation (cf. Fig. 4). Throughout the year, BCldb1 performs slightly worse than BClda1 while BCsdb1 performs slightly better than BCsda1. The findings around Figs. 3 and 4 suggest that these differences in overall performance can be explained by how well the respective methods correct the upper tail of the rlds and rsds distributions. Finally, it is worth noting that rlds biases do not exhibit any pronounced seasonality whereas rsds biases are particularly large in the solstice months of June and December, both before and after bias correction with any method.

### 4.2 Validation against BSRN data

Rankings of original and bias-corrected E2OBS data according to their similarity in distribution to the corresponding BSRN data for daily and monthly mean longwave and shortwave radiation are depicted in Fig. 7. The distribution of rankings over BSRN stations and calendar months suggest that in most cases, bias correction of E2OBS using SRB data is beneficial either

Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2017-81
Manuscript under review for journal Earth Syst. Dynam.
Discussion started: 25 September 2017


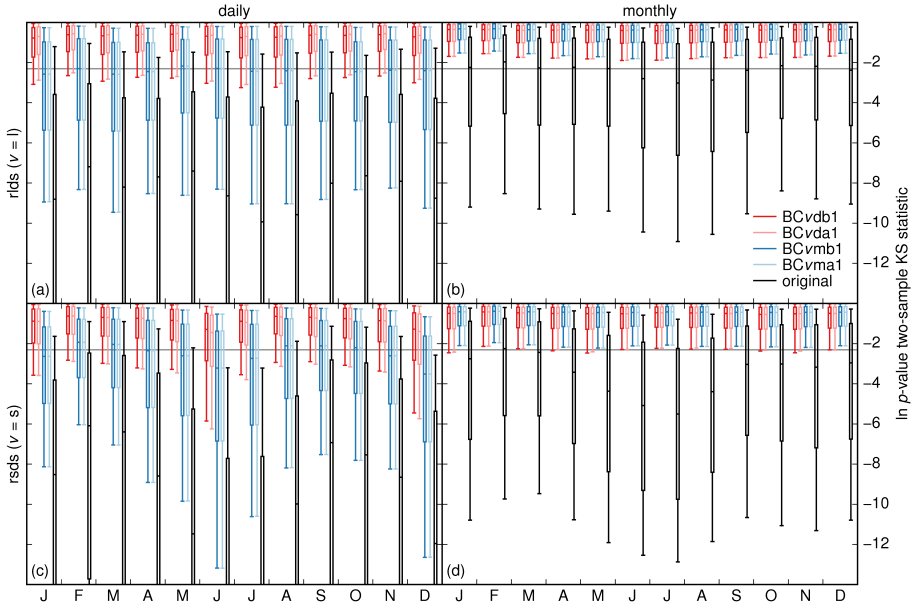


**Figure 6.** Overall performance of bias correction methods BC$vtp1$ for longwave (**top**) and shortwave (**bottom**) radiation at the daily (**left**) and monthly (**right**) time scale as quantified by $p$-values of two-sample Kolmogorov-Smirnov test statistics of the respective E2OBS and SRB data before (black) and after (colours) bias correction (cf. Sect. B; greater $p$-values indicate stronger agreement of E2OBS and SRB distributions). The $p$-values are determined individually for each grid cell, calendar month and calibration data sample, with all corresponding values pooled into one distribution and omitting shortwave radiation data from months with average rsdt less than 1 $\mathrm{Wm}^{-2}$. The horizontal lines of each box-whisker plot represent the 90th, 75th, 50th, 25th and 10th (from top to bottom) grid-cell area-weighted percentile of the natural logarithms of these $p$-values over calibration data sample (1sthalf, 2ndhalf), latitude and longitude. The grey horizontal line marks the $p = 10\%$ significance level. Compared to BC$vtp1$, $p$-values produced by BC$vtp0$ are slightly lower but qualitatively similar.

with any method or not at all, depending on whether the SRB or the original E2OBS data are less biased relative to the BSRN ground truth.

More often than with BCl$tp1$ or BCl$tp2$, the bias correction with BCl$tp0$ reduces rlds biases relative to BSRN. Smaller and opposite differences are found between the BCs$tpx$ methods. At the monthly time scale, the differences in rlds distribution similarity are mainly determined by long-term mean value biases (cf. Figs. 7b and 8b). This suggests that more often than not, the bilinear interpolation included in BC$vtp0$ yields more realistic long-term mean rlds values at the E2OBS grid than the disaggregation methods of BC$vtp1$ and BC$vtp2$. Presumably, this is due to an elevation correction implicitly carried out along with the interpolation.

At the daily time scale, standard deviation biases explain most of the method dependencies of distribution similarities between BSRN and bias-corrected E2OBS data (cf. Figs. 7a,c and 8a,c). For rlds, compared to BClm$px$, BCld$px$ leaves lower/higher standard deviation biases and yields higher/lower distribution similarities over the tropics/extratropics. For rsds,

Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2017-81
Manuscript under review for journal Earth Syst. Dynam.
Discussion started: 25 September 2017


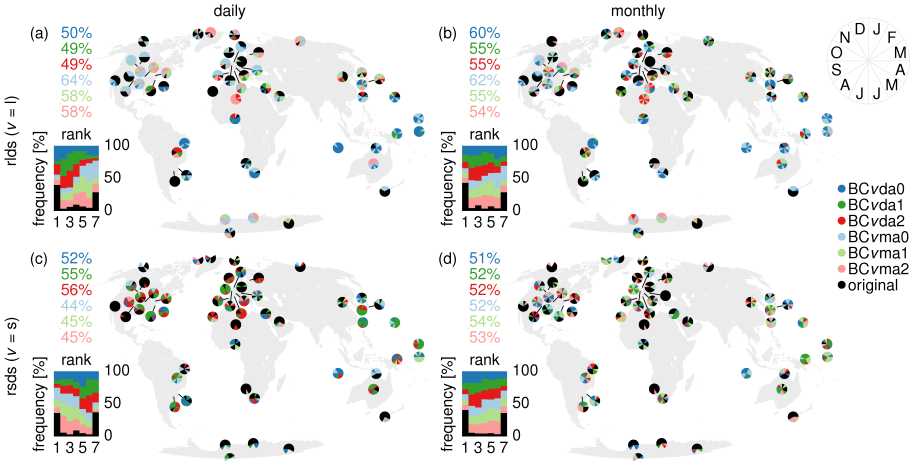


**Figure 7.** Rankings of original (black) and bias-corrected (colours) E2OBS data according to their similarity in distribution to the corresponding BSRN data from 54 stations (Table 1) and the 1992–2014 time period for daily (**left**) and monthly (**right**) mean longwave (**top**) and shortwave (**bottom**) radiation. Similarity in distribution is quantified by $p$-values of two-sample Kolmogorov-Smirnov test statistics (cf. Sect. B). Higher ranks indicate greater $p$-values and thus greater similarity in distribution. The map shows the highest ranking E2OBS dataset per station and calendar month (see legend). Rankings for rsds are not computed for months with average rsdt less than 1 $\mathrm{Wm}^{-2}$. Rank distributions over stations and months are shown in the inset at the lower left of each panel. The percentages at the upper left of each panel display in how many cases (stations, calendar months) E2OBS data bias-corrected with a certain method (colour) outrank the corresponding original E2OBS data. Very similar results are obtained for the corresponding basic bias correction methods, with outranking percentages deviating by $\pm 1\%$ at most.

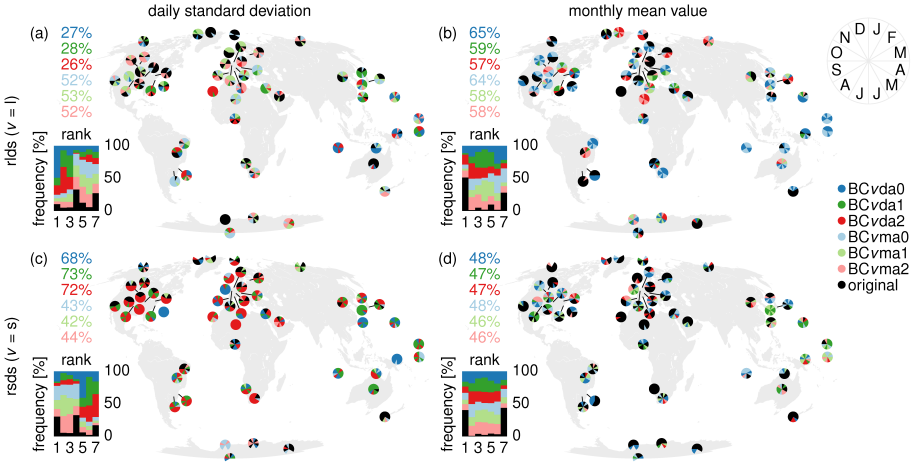


**Figure 8.** Same as Fig. 7 but for biases in multi-year daily standard deviations (**left**) and multi-year monthly mean values (**right**), with higher ranks corresponding to lower absolute values of these biases.

Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2017-81
Manuscript under review for journal Earth Syst. Dynam.
Discussion started: 25 September 2017


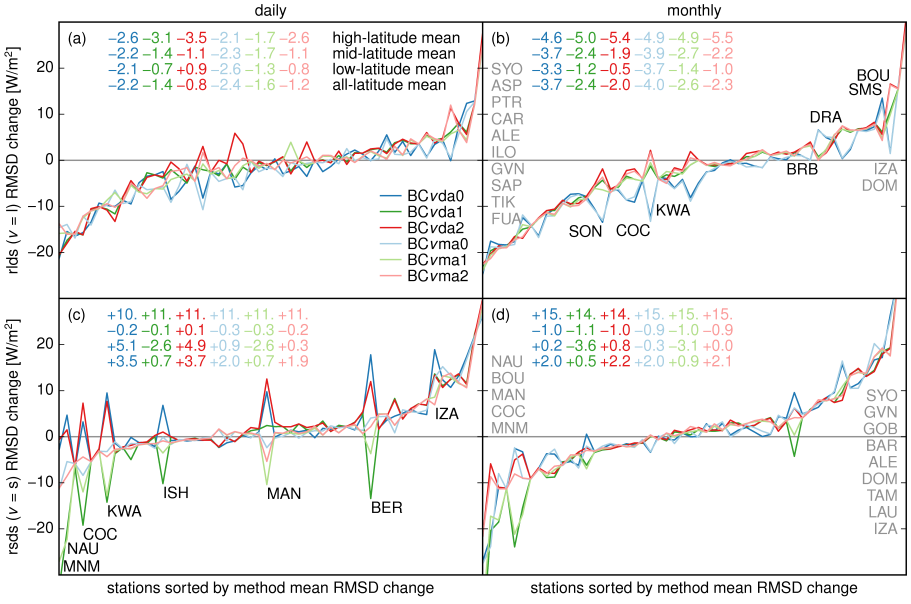


**Figure 9.** Changes in RMSDs of E2OBS from BSRN daily (**left**) and monthly (**right**) mean longwave (**top**) and shortwave (**bottom**) radiation at 54 stations (Table 1) after bias correction with methods BC$vtax$. At each station, RMSDs are computed over the whole time series of available BSRN data, only omitting shortwave radiation data from months with average rsdt less than 1 $\mathrm{Wm}^{-2}$. Listed in grey are the stations where bias correction on average decreases (left list; the more the lower) or increases (right list; the more the higher) RMSDs of monthly mean radiation by more than 10 $\mathrm{Wm}^{-2}$. Annotated in black are stations where RMSD changes (**b**) for BCl$t$a0 differ by more than 5 $\mathrm{Wm}^{-2}$ from those obtained with any other method, (**c**) spread a range larger than 10 $\mathrm{Wm}^{-2}$ over all bias correction methods. Numbers given in each panel are station mean RMSD changes per bias correction method (colour; $\mathrm{Wm}^{-2}$) over all high-latitude (beyond 66°S or 66°N; first row), all mid-latitude (66°S–33°S and 33°N–66°N; second row), all low-latitude (33°S–33°N; third row) and all (fourth row) stations. These numbers change by at most ±0.2 for the respective basic bias correction method.

BCsd$px$ leaves lower standard deviation biases and yields higher distribution similarities everywhere, and differences between standard deviation biases generated by BCsd$p$0, BCsd$p$1 and BCsd$p$2 are in line with cross-validation results (Fig. 3).

Bias correction-induced changes in RMSDs of E2OBS from BSRN time series of daily and monthly mean longwave and shortwave radiation are shown in Fig. 9. At both time scales and for both radiation types, the between-station variance of the RMSD changes is larger than their within-station variance, which again suggests that biases relative to BSRN after bias correction using SRB data depend more on the corresponding SRB data biases than on the method used for the bias correction.

However, there are two notable exceptions from this rule. First, more often than not, rlds RMSDs are systematically lower after bias correction with BCl$tp$0 than with BCl$tp$1 or BCl$tp$2. This is particularly well visible at the monthly time scale. As conjectured above, this might be the result of an elevation correction of long-term rlds mean values implicitly done by BCl$tp$0 along with its bilinear interpolation of SRB data to the E2OBS grid.

Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2017-81
Manuscript under review for journal Earth Syst. Dynam.
Discussion started: 25 September 2017


Earth System Dynamics Discussions

Secondly, at eight stations (BER, COC, ISH, IZA, KWA, MAN, MNM, NAU; cf. Table 1), daily rsds RMSDs after bias correction with different methods spread over a range wider than $10\,\mathrm{W m^{-2}}$ (in six cases even $20\,\mathrm{W m^{-2}}$). These stations are all located on islands that are smaller than one SRB-grid cell but large enough to be resolved by the original E2OBS data, which is to say that the E2OBS climatologies at the corresponding 0.5° grid cells stand out against those at the neighbouring 0.5° grid cells. Bias correction results at these stations can therefore be expected to depend on how a given method modifies the sub-SRB-grid scale spatial variability of the original E2OBS data. It turns out that, at all of these stations except IZA, daily rsds RMSDs are smaller after bias correction with $\mathrm{BC}stp1$ than with $\mathrm{BC}stp0$ and $\mathrm{BC}stp2$.

On average over the respective stations, rlds RMSDs are reduced by all bias correction methods at all latitudes. In contrast, bias correction results are more heterogeneous for rsds. At low latitudes (33°S–33°N), bias correction has a neutral average effect, at middle latitudes (66°S–33°S and 33°N–66°N) it reduces average rsds RMSDs, and at high latitudes (beyond 66°S or 66°N) it strongly increases them. Five out of the six stations with the greatest rsds RMSD increases are high-latitude stations (ALE, BAR, DOM, GVN, SYO; cf. Table 1). Stackhouse Jr. et al. (2011) believe that the difficult cloud characterisation over surfaces that are frequently covered by snow, ice or water is the primary reason for large SRB shortwave radiation biases at such polar sites.

### 4.3 Disaggregation and sub-SRB-grid scale spatial variability

As outlined in Sect. 3.2.1, the $\mathrm{BC}vtp1$ disaggregation method is based on the implicit assumption that the original E2OBS values of daily mean radiation onto the four E2OBS-grid cells contained in one SRB-grid cell must all be too low/high if their area-weighted average is too low/high. The four values are then all increased/decreased by the disaggregation method. In order to account for their upper/lower physical bounds, the increases/decreases are done by a common scaling factor applied to the distances to the respective upper/lower bounds. This leads to a reduction of the differences between the four values (necessarily if the four bounds are equal, in most cases if they are similar), i.e., to a deflation of sub-SRB-grid scale spatial variability.

In order to illustrate and measure the extent of this deflation, the RMSD of the four E2OBS values of daily mean radiation per SRB-grid cell from their area-weighted average is calculated for every time step both before and after bias correction with methods $\mathrm{BC}v\mathrm{da}1$. The median bias correction-induced deflation of these RMSDs over all calendar months is found to exceed 20% in some locations for both longwave and shortwave radiation (Fig. 10b,h).

In an analogous manner, RMSDs can be computed based on data from the four E2OBS-grid cells contained in each staggered SRB-grid cell, where the staggered SRB grid is a regular $1.0° \times 1.0°$ latitude-longitude grid shifted by 0.5° latitude and 0.5° longitude relative to the SRB grid, i.e., every staggered SRB-grid cell contains E2OBS-grid cells contained in four different SRB-grid cells. Bias correction with methods $\mathrm{BC}v\mathrm{da}1$ has a strong impact on these RMSDs (Fig. 10e,k), with increases/decreases found mostly over tropical and polar regions/middle latitudes. Most importantly, the RMSD change patterns at the staggered SRB grid are very different from those at the SRB grid. This is considered to be an artefact caused by the $\mathrm{BC}vtp1$ disaggregation method.

E2OBS data bias-corrected with $\mathrm{BC}v\mathrm{da}2$ do not suffer from the deflation of sub-SRB-grid scale spatial variability that results from bias correction with $\mathrm{BC}v\mathrm{da}1$ (Fig. 10c,i). Moreover, the RMSD change patterns at the SRB and the staggered

Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2017-81
Manuscript under review for journal Earth Syst. Dynam.
Discussion started: 25 September 2017


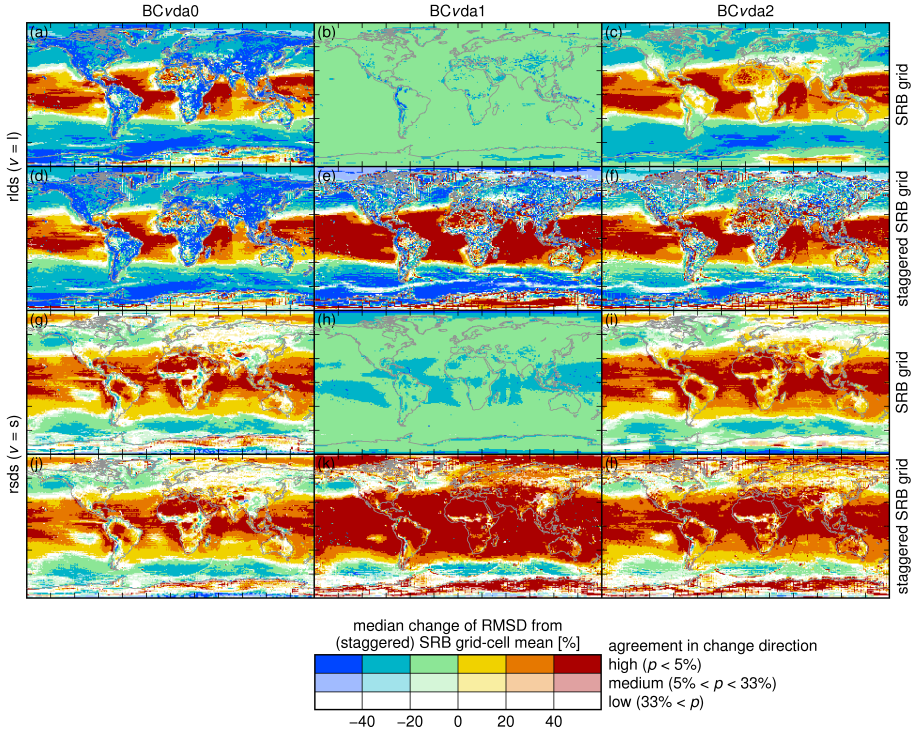


**Figure 10.** Relative change after bias correction with methods BC$v$da0 (**left**), BC$v$da1 (**middle**) and BC$v$da2 (**right**) of the RMSD of daily 0.5° longwave (**a–f**) and shortwave (**g–l**) radiation from the respective 1° grid-cell mean based on 1° grid cells of the SRB grid (**a–c**, **g–i**) and the staggered SRB grid (**d–f**, **j–l**; see text). For every 1° grid cell and calendar month, the RMSDs are calculated using original or bias-corrected E2OBS data from the four associated 0.5° grid cells and the 12/1983–11/2007 time period, omitting shortwave radiation data from months with average rsdt less than 1 $\mathrm{Wm}^{-2}$ (cf. Sect. A and Fig. C1c). Depicted are median and agreement in direction of individual RMSD changes after bias correction (cf. Fig. 3). Very similar results are obtained for the corresponding basic bias correction methods.

SRB grid are much more similar after bias-correction with BC$v$da2 than with BC$v$da1, except for rlds over land, where RMSDs are more noisy and smaller on average at the staggered SRB grid (Fig. 10f,l).

Bias correction with BC$vtp$0 yields virtually identical RMSD change patterns at the SRB and the staggered SRB grid. For rsds, these patterns are very similar to those obtained with BC$vtp$2. In contrast, rlds RMSDs over land are reduced much more by BC$vtp$0 than by BC$vtp$2. As a consequence, the land-sea contrast of rlds RMSD changes is much larger for BC$vtp$0 than for BC$vtp$2, in particular over the tropics. The strong deflation of rlds sub-SRB-grid scale spatial variability produced by BC$vtp$0 over land is considered to be another artefact caused by the bilinear interpolation of SRB data to the E2OBS grid. The magnitude of the deflation of more than 40% in most cases cannot be explained by the associated deflation of SRB-grid scale temporal variability, which in most cases does not exceed 8% (Fig. 3g,j). Presumably, it is mainly due to a deflation of the sub-SRB-grid scale spatial variability of long-term mean rlds values caused by the bilinear interpolation.

Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2017-81
Manuscript under review for journal Earth Syst. Dynam.
Discussion started: 25 September 2017




## 5 Conclusions

This article introduces various parametric quantile mapping methods for the bias correction of E2OBS surface downwelling longwave and shortwave radiation using SRB satellite estimates. Bias correction results are cross-validated as well as validated using independent BSRN ground observations. The quantile mapping methods differ in (i) the time scale at which they operate, (ii) if and how they take physical upper radiation bounds into account, and (iii) how they handle the spatial resolution gap between E2OBS and SRB.

As expected, cross-validation results suggest that statistics of daily mean radiation are mostly better corrected by methods operating at the daily time scale than by those operating at the monthly time scale, and vice versa for statistics of monthly mean radiation. However, compared to BSRN observations, daily mean longwave radiation is mostly better corrected by the methods operating at the monthly time scale because the methods operating at the daily time scale adulterate the day-to-day variability of the original data. Given the composition of the E2OBS data, this suggests that the day-to-day variability of ERA-Interim longwave radiation is mostly more realistic than the corresponding SRB estimates.

While the methods operating at the monthly time scale are best at adjusting the interannual variability of monthly mean values, our cross-validation results show that for calibration and validation sample sizes of only 12 years each, the methods operating at the daily time scale perform almost as well if sampling errors are taken into account. In that case, the methods operating at the monthly time scale are in fact overcorrecting. This result should be seen as an incentive to develop bias correction methods that take sampling errors into account.

Methods that do and that do not take upper radiation bounds into account during quantile mapping are applied to daily mean longwave radiation. It is found that multi-year monthly maximum values as well as the shape of the whole distribution is better adjusted by methods that respect the estimated upper bounds. For shortwave radiation, different approaches to estimating the upper bounds are tested. A simple method based on running maximum values is found to perform better in cross-validation than a more complicated one that uses rescaled daily mean insolation at the top of the atmosphere. Arguably, that is because upper physical bounds to downwelling shortwave radiation at the surface (rsds) are determined by downwelling shortwave radiation at the top of the atmosphere (rsdt) as well as by other factors such as atmospheric humidity. In fact, the highest rsds values usually occur under clear-sky conditions and they are the higher the drier the atmosphere. Atmospheric humidity, in turn, is limited by the water vapour holding capacity of the atmosphere, which is controlled by atmospheric temperature. Since the climatology of atmospheric temperature lags that of rsdt, the climatology of upper bounds to rsds can also be expected to deviate from any rescaled rsdt climatology.

The most simple approach tested here to bridging the spatial resolution gap between E2OBS and SRB data is to bilinearly interpolate the more coarsely resolved SRB data to the E2OBS grid and to use these interpolated data for a bias correction at the E2OBS grid. Methods operating this way are found to erroneously deflate both the temporal and the spatial variability of the original E2OBS data. On the other hand, relative to BSRN observations, the interpolation is found to more often than not benefit the multi-year monthly mean longwave radiation. This positive effect of the interpolation is interpreted as the result of an implicit elevation adjustment from the SRB to the E2OBS grid. This outcome encourages elevation adjustments preceding



future bias corrections of longwave radiation using, e.g., the Stackhouse Jr. et al. (2011) formula or the Cosgrove et al. (2003) method.

The second approach used here aggregates the original E2OBS data to the SRB grid, where the bias correction is done, and disaggregates these aggregated and bias-corrected data back to the E2OBS grid. The deterministic disaggregation method used for that purpose is found to deflate the sub-SRB-grid scale spatial variability of the original data. Yet it also has its merits, where sub-SRB-grid scale spatial gradients in radiation statistics are very large, such as over islands covering just one E2OBS grid cell. There, the aggregation-correction-disaggregation of rsds produces substantially lower root-mean-square deviations from daily mean BSRN values than the other approaches.

The third approach introduced here corrects biases at the E2OBS grid using target distribution parameters that are based on the more coarsely resolved SRB data as well as on sub-SRB-grid scale spatial variability present in the original E2OBS data to the end of adjusting SRB-grid scale biases while preserving the sub-SRB-grid scale spatial variability. The latter objective is achieved here by preserving sub-SRB-grid scale ratios between climatological standard deviations and offsets between climatological mean values and upper bounds. Potentially more suitable non-linear relationships might be tested in future studies. By design, the third approach precludes any of the variability deflations caused by the first and second approach.

The cross-validation reveal that substantial skewness biases remain after bias correction with any of the parametric quantile mapping methods introduced here, as these do not explicitly adjusted third- and higher-order moments. Better results might be obtained using non-parametric quantile mapping methods.

Deviations of bias-corrected E2OBS data from BSRN observations turn out to be dominated by the corresponding SRB data biases. This exemplifies that bias correction usually does not actually correct biases but merely adjusts them to those of another dataset, which is why some colleagues prefer the term *bias adjustment* over *bias correction*. In the example studied here, this is most painfully visible at polar BSRN stations, where in most cases bias correction using SRB estimates substantially increases E2OBS shortwave radiation biases relative to the BSRN ground truth. Yet apart from these cases, the validation against BSRN observations suggests that, overall, bias correction of E2OBS radiation using SRB data has a slightly positive effect on longwave radiation and a neutral effect on shortwave radiation.

For the EWEMBI dataset (Lange, 2016), E2OBS longwave radiation was adjusted to the SRB 3.1 primary-algorithm product using the BClda1 method, and E2OBS shortwave radiation was adjusted to the SRB 3.0 primary-algorithm product using the BCsda1 method. In that application, the full 24 years worth of SRB data and the same 24 years worth of E2OBS data aggregated to the SRB grid were used to derive the transfer function parameters. The present study identifies shortcomings of the BClda1 and BCsda1 methods and tests modifications of these methods as remedies. In terms of cross-validation results and variability deflation issues, the best methods tested here are BClda2 and BCsdb2, whereas biases relative to BSRN observations are most effectively reduced by BClmb0 and BCsdb1.

*Data availability.* The EWEMBI dataset is publicly available at https://doi.org/10.5880/pik.2016.004.

Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2017-81
Manuscript under review for journal Earth Syst. Dynam.
Discussion started: 25 September 2017


Earth System Dynamics Discussions
Open Access
EGU

## Appendix A: Daily mean insolation at the top of the atmosphere

Over the course of a year, the total solar irradiance, $S$, varies according to $S = S_0(1 + e\cos(\Theta))^2$, where $S_0 = 1360.8\,\mathrm{Wm}^{-2}$ is the solar constant (Kopp and Lean, 2011), $e = 0.0167086$ is the Earth's current orbital eccentricity and $\Theta$ is the angle to the Earth's position from its perihelion, as seen from the Sun. If the orbital angular velocity of the Earth is approximated to vary sinusoidally in time then the total solar irradiance on day $n$ after January 1 of the first year of a four-year cycle including one leap year is approximately given by

$$S = S_0\left(1 + e\cos\left(2\pi\frac{n-2}{365.25} + 2e\sin\left(2\pi\frac{n-2}{365.25}\right)\right)\right)^2, \tag{A1}$$

since $S$ is at its maximum when the Earth is at its perihelion, which on average occurs on January 3.

The daily mean insolation at the top of the atmosphere, rsdt, at some fixed geolocation depends on the location's latitude, $\phi$, and on the declination of the Sun, $\delta$, which varies over the course of a year. On day $n$ after January 1 of the first year of a four-year cycle including one leap year, the declination of the sun is approximately given by

$$\sin\delta = \cos\left(2\pi\frac{n+10}{365.25} + 2e\sin\left(2\pi\frac{n-2}{365.25}\right)\right)\sin\delta_{\min}, \tag{A2}$$

since $\delta$ is at its minimum value $\delta_{\min} = -23.4392811°$ at the December solstice, which on average occurs on December 22. Latitude and declination of the Sun determine the hour angle at sunrise, $h$, according to

$$\cos h = \min\{1, \max\{-1, -\tan\phi\tan\delta\}\}. \tag{A3}$$

The daily mean insolation at the top of the atmosphere at latitude $\phi$ on day $n$ is then given by

$$\mathrm{rsdt} = \frac{S}{\pi}(h\sin\phi\sin\delta + \sin h\cos\phi\cos\delta). \tag{A4}$$

For a given latitude, the rsdt climatology used to estimate the upper bounds of the climatological beta distribution of rsds in the $\mathrm{BCsda}x$ methods is derived using Eqs. (A1)–(A4) to compute rsdt over a four-year cycle including one leap year and then averaging calendar day values over the four cases of leap year occurrence in the four-year cycle.

## Appendix B: Two-sample Kolmogorov-Smirnov test

The overall effectivity of the bias correction methods introduced in this study is measured by similarities of empirical CDFs of SRB and E2OBS data before and after bias correction using the two-sample Kolmogorov-Smirnov (KS) test. Let $F_1$ be the empirical CDF of uncorrected or corrected daily or monthly mean longwave or shortwave E2OBS data for one particular grid cell, calendar month and validation data sample, with all corresponding values pooled into one distribution, and let $F_2$ be the empirical CDF of the corresponding SRB or BSRN data. Then the two-sample KS test statistic of these CDFs is given by $D = \sup_r |F_1(r) - F_2(r)|$.

Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2017-81
Manuscript under review for journal Earth Syst. Dynam.
Discussion started: 25 September 2017


Earth System Dynamics Discussions
Open Access
EGU

The null hypothesis of the KS test is that the two data samples whose empirical CDFs are compared have the same underlying distribution. According to Vetterling et al. (1992, Sect. 14.3), the probability $p$ of incorrectly rejecting this null hypothesis can be approximated by

$$p = 1 - F\left(\left[\sqrt{n} + 0.12 + 0.11/\sqrt{n}\right] D\right), \tag{B1}$$

where $F$ is the CDF of the Kolmogorov distribution, $n = n_1 n_2/(n_1 + n_2)$ is the effective sample size, and $n_1$ and $n_2$ are the sizes of the samples behind $F_1$ and $F_2$, respectively. This approximation of the true $p$-value is not only asymptotically accurate but already quite good for $n \geq 4$ (cf. von Mises, 1964; Vetterling et al., 1992).

In order to adjust these $p$-values for potential autocorrelations in the samples compared here, which are in fact time series, $n_1$ and $n_2$ in the formula for $n$ are replaced by $n_1(1-\rho_1)$ and $n_2(1-\rho_2)$, respectively, as proposed by Xu (2013), where the autoregression coefficients $\rho_1$ and $\rho_2$ of first-order autoregressive processes fitted to the time series are estimated by the respective sample autocorrelation at lag one.

## Appendix C: Window length for running mean and maximum calculations

The climatologies of mean values, variances, and upper bounds of daily mean radiation estimated by the $\mathrm{BC}v\mathrm{d}px$ methods are based on running mean values of empirical multi-year daily mean values, variances and running maximum values, respectively. A common window length of 25 days is used for these running mean and maximum calculations (cf. Table 2). An obvious question is how sensitive the bias correction results are to the choice of this window length.

The question is addressed here via variants of the $\mathrm{BC}v\mathrm{da1}$ methods that use uneven window lengths between 10 and 40 days for their running mean and maximum calculations and are otherwise identical to the $\mathrm{BC}v\mathrm{da1}$ method introduced in Sect. 3.1.1. The performance of these $\mathrm{BC}v\mathrm{da1}$ variants is then quantified by $p$-values of two-sample KS statistics of bias-corrected E2OBS data cross-validated against SRB data (cf. Sect. 4 and Sect. B). The window lengths that maximise these $p$-values vary considerably with location, calendar month and calibration data sample (Fig. C1). The reason for this high variability is illustrated in Fig. C2, where the overall performance of the $\mathrm{BC}v\mathrm{da1}$ variants, quantified by $p$-values of two-sample KS statistics aggregated over time (calendar months) and space (grid cells), is shown to only weakly depend on the chosen window length.

The optimal window length is thus highly uncertain. For longwave/shortwave radiation, the overall performance of the $\mathrm{BC}v\mathrm{da1}$ variants is slightly higher for window lengths from the upper/lower end of the investigated range (Fig. C2). For practical matters, one can apply the methods using any window length between 10 and 40 days and expect similarly well adjusted radiation biases. The choice of 25-day running windows made here for both longwave and shortwave radiation ensures a close-to-optimal performance of the $\mathrm{BC}v\mathrm{da1}$ methods for both variables.

*Competing interests.* The author declares that no competing interests are present.

Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2017-81
Manuscript under review for journal Earth Syst. Dynam.
Discussion started: 25 September 2017


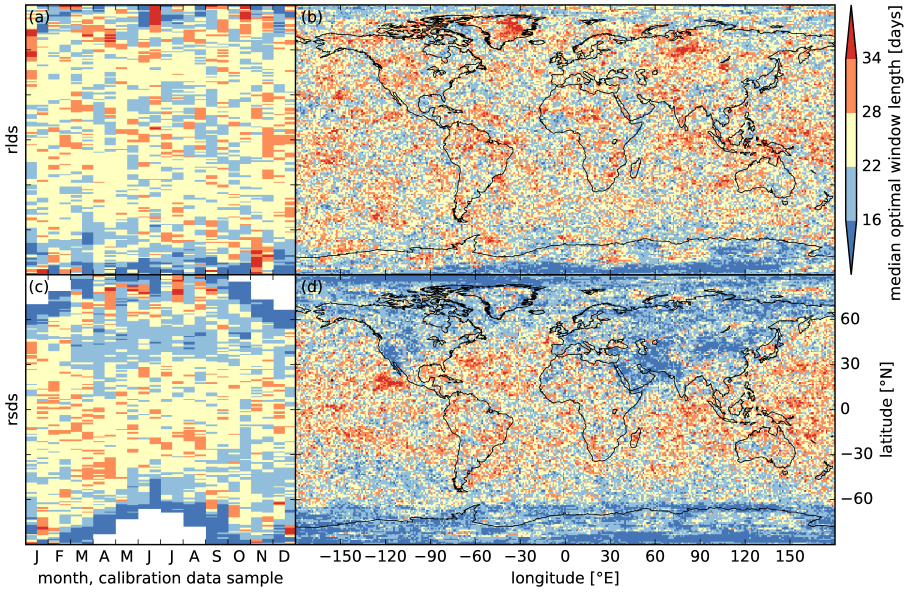


**Figure C1.** Optimal window length for running mean and maximum calculations that precede the estimation of parameters of the climatological distributions of longwave ($v = \mathrm{l}$; **top**) and shortwave ($v = \mathrm{s}$; **bottom**) radiation that are used for bias correction with $\mathrm{BC}v\mathrm{da1}$ (cf. Table 2). Window lengths are varied between 10 and 40 days. Optimal window lengths maximise the $p$-value of the two-sample KS statistic of bias-corrected E2OBS data cross-validated against SRB data (cf. Sect. 4 and Sect. B) and are determined individually for every grid cell, calendar month (with all corresponding values pooled into one distribution) and calibration data sample (every1st, every2nd). Zonal medians of optimal window lengths for each month and calibration data sample are shown in panels (**a**) and (**c**). Results are masked in (**c**) where and when the monthly mean rsdt (Eqs. (A1)–(A4)) is less than $1\,\mathrm{W m^{-2}}$. Panels (**b**) and (**d**) show medians of optimal window lengths over months and calibration data samples.

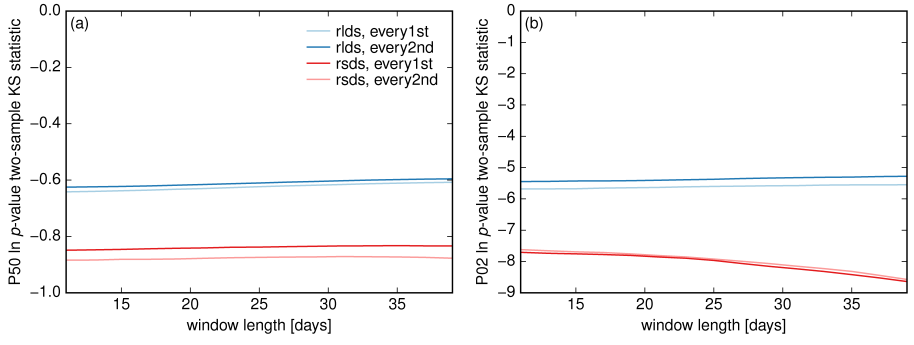


**Figure C2.** Dependence of two-sample KS statistic $p$-values on window length for different radiation types and calibration data samples (see text and Fig. C1). Plotted are the grid-cell area-weighted 50th (**left**) and 2nd (**right**) percentiles of the natural logarithms of the $p$-values over months, latitudes and longitudes.

Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2017-81
Manuscript under review for journal Earth Syst. Dynam.
Discussion started: 25 September 2017


*Acknowledgements.* The author is grateful to Katja Frieler, Jan Volkholz and Alex Cannon for various helpful discussions at different stages of this work, to Hannes Müller Schmied for suggesting the validation against BSRN data, to Gert König-Langlo and Amelie Driemel for their help during BSRN data acquisition and processing, to Paul W. Stackhouse Jr. for his guidance with SRB data products and the provision of SRB elevation data, and to Graham Weedon and Emanuel Dutra for their guidance during the initial stage of assembling the EWEMBI dataset. This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 641816 Coordinated Research in Earth Systems and Climate: Experiments, kNowledge, Dissemination and Outreach (CRESCENDO).



## References

Calton, B., Schellekens, J., and Martinez-de la Torre, A.: Water Resource Reanalysis v1: Data Access and Model Verification Results, https://doi.org/10.5281/zenodo.57760, 2016.

Cannon, A. J.: Multivariate quantile mapping bias correction: an N-dimensional probability density function transform for climate model simulations of multiple variables, Climate Dynamics, pp. 1–19, https://doi.org/10.1007/s00382-017-3580-6, 2017.

Chang, J., Ciais, P., Wang, X., Piao, S., Asrar, G., Betts, R., Chevallier, F., Dury, M., François, L., Frieler, K., Ros, A. G. C., Henrot, A.-J., Hickler, T., Ito, A., Morfopoulos, C., Munhoven, G., Nishina, K., Ostberg, S., Pan, S., Peng, S., Rafique, R., Reyer, C., Rödenbeck, C., Schaphoff, S., Steinkamp, J., Tian, H., Viovy, N., Yang, J., Zeng, N., and Zhao, F.: Benchmarking carbon fluxes of the ISIMIP2a biome models, Environmental Research Letters, 12, 045 002, https://doi.org/10.1088/1748-9326/aa63fa, 2017.

Cosgrove, B. A., Lohmann, D., Mitchell, K. E., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., Marshall, C., Sheffield, J., Duan, Q., Luo, L., Higgins, R. W., Pinker, R. T., Tarpley, J. D., and Meng, J.: Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project, Journal of Geophysical Research: Atmospheres, 108, https://doi.org/10.1029/2002JD003118, 8842, 2003.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

Dutra, E.: Report on the current state-of-the-art Water Resources Reanalysis, Earth2observe deliverable no. d.5.1, http://earth2observe.eu/files/Public%20Deliverables, 2015.

Fisher, R. A.: Frequency Distribution of the Values of the Correlation Coefficient in Samples from an Indefinitely Large Population, Biometrika, 10, 507–521, http://www.jstor.org/stable/2331838, 1915.

Fisher, R. A.: On the "probable error" of a coefficient of correlation deduced from a small sample, Metron, 1, 3–32, http://hdl.handle.net/2440/15169, 1921.

Frieler, K., Betts, R., Burke, E., Ciais, P., Denvil, S., Deryng, D., Ebi, K., Eddy, T., Emanuel, K., Elliott, J., Galbraith, E., Gosling, S. N., Halladay, K., Hattermann, F., Hickler, T., Hinkel, J., Huber, V., Jones, C., Krysanova, V., Lange, S., Lotze, H. K., Lotze-Campen, H., Mengel, M., Mouratiadou, I., Müller Schmied, H., Ostberg, S., Piontek, F., Popp, A., Reyer, C. P. O., Schewe, J., Stevanovic, M., Suzuki, T., Thonicke, K., Tian, H., Tittensor, D. P., Vautard, R., van Vliet, M., Warszawski, L., and Zhao, F.: Assessing the impacts of 1.5°C global warming – simulation protocol of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP2b), Geoscientific Model Development Discussions, 2016, 1–59, https://doi.org/10.5194/gmd-2016-229, 2016.

Garratt, J. R.: Incoming Shortwave Fluxes at the Surface—A Comparison of GCM Results with Observations, Journal of Climate, 7, 72–80, https://doi.org/10.1175/1520-0442(1994)007<0072:ISFATS>2.0.CO;2, 1994.

Gennaretti, F., Sangelantoni, L., and Grenier, P.: Toward daily climate scenarios for Canadian Arctic coastal zones with more realistic temperature-precipitation interdependence, Journal of Geophysical Research: Atmospheres, 120, 11,862–11,877, https://doi.org/10.1002/2015JD023890, 2015.

Harris, I., Jones, P. D., Osborn, T. J., and Lister, D. H.: Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset, International Journal of Climatology, https://doi.org/10.1002/joc.3711, 2013.

Hempel, S., Frieler, K., Warszawski, L., Schewe, J., and Piontek, F.: A trend-preserving bias correction – the ISI-MIP approach, Earth System Dynamics, 4, 219–236, https://doi.org/10.5194/esd-4-219-2013, 2013.

Iizumi, T., Takikawa, H., Hirabayashi, Y., Hanasaki, N., and Nishimori, M.: Contributions of different bias-correction methods and reference meteorological forcing data sets to uncertainty in projected temperature and precipitation extremes, Journal of Geophysical Research: Atmospheres, https://doi.org/10.1002/2017JD026613, 2017JD026613, 2017.

Ito, A., Nishina, K., Reyer, C. P. O., François, L., Henrot, A.-J., Munhoven, G., Jacquemin, I., Tian, H., Yang, J., Pan, S., Morfopoulos, C., Betts, R., Hickler, T., Steinkamp, J., Ostberg, S., Schaphoff, S., Ciais, P., Chang, J., Rafique, R., Zeng, N., and Zhao, F.: Photosynthetic productivity and its efficiencies in ISIMIP2a biome models: benchmarking for impact assessment studies, Environmental Research Letters, 12, 085 001, https://doi.org/10.1088/1748-9326/aa7a19, 2017.

Jones, P. W.: First- and Second-Order Conservative Remapping Schemes for Grids in Spherical Coordinates, Monthly Weather Review, 127, 2204–2210, https://doi.org/10.1175/1520-0493(1999)127<2204:FASOCR>2.0.CO;2, 1999.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-Year Reanalysis Project, Bulletin of the American Meteorological Society, 77, 437–471, https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2, 1996.

Kiehl, J. T. and Trenberth, K. E.: Earth's Annual Global Mean Energy Budget, Bulletin of the American Meteorological Society, 78, 197–208, https://doi.org/10.1175/1520-0477(1997)078<0197:EAGMEB>2.0.CO;2, 1997.

Kistler, R., Collins, W., Saha, S., White, G., Woollen, J., Kalnay, E., Chelliah, M., Ebisuzaki, W., Kanamitsu, M., Kousky, V., van den Dool, H., Jenne, R., and Fiorino, M.: The NCEP-NCAR 50-Year Reanalysis: Monthly Means CD-ROM and Documentation, Bulletin of the American Meteorological Society, 82, 247–267, https://doi.org/10.1175/1520-0477(2001)082<0247:TNNYRM>2.3.CO;2, 2001.

König-Langlo, G., Sieger, R., Schmithüsen, H., Bücker, A., Richter, F., and Dutton, E.: The Baseline Surface Radiation Network and its World Radiation Monitoring Centre at the Alfred Wegener Institute, WCRP Report 24/2013, GCOS 174, Geneva, Switzerland, http://www.wmo.int/pages/prog/gcos/Publications/gcos-174.pdf, 2013.

Kopp, G. and Lean, J. L.: A new, lower value of total solar irradiance: Evidence and climate significance, Geophysical Research Letters, 38, https://doi.org/10.1029/2010GL045777, l01706, 2011.

Krysanova, V. and Hattermann, F. F.: Intercomparison of climate change impacts in 12 large river basins: overview of methods and summary of results, Climatic Change, 141, 363–379, https://doi.org/10.1007/s10584-017-1919-y, 2017.

Lange, S.: EartH2Observe, WFDEI and ERA-Interim data Merged and Bias-corrected for ISIMIP (EWEMBI), https://doi.org/10.5880/pik.2016.004, 2016.

Ma, Q., Wang, K., and Wild, M.: Evaluations of atmospheric downward longwave radiation from 44 coupled general circulation models of CMIP5, Journal of Geophysical Research: Atmospheres, 119, 4486–4497, https://doi.org/10.1002/2013JD021427, 2013JD021427, 2014.

Müller Schmied, H., Müller, R., Sanchez-Lorenzo, A., Ahrens, B., and Wild, M.: Evaluation of Radiation Components in a Global Freshwater Model with Station-Based Observations, Water, 8, https://doi.org/10.3390/w8100450, 2016.

Roesch, A., Wild, M., Ohmura, A., Dutton, E. G., Long, C. N., and Zhang, T.: Assessment of BSRN radiation records for the computation of monthly means, Atmospheric Measurement Techniques, 4, 339–354, https://doi.org/10.5194/amt-4-339-2011, 2011.

Ruane, A. C., Goldberg, R., and Chryssanthacopoulos, J.: Climate forcing datasets for agricultural modeling: Merged products for gap-filling and historical climate series estimation, Agricultural and Forest Meteorology, 200, 233–248, https://doi.org/10.1016/j.agrformet.2014.09.016, 2015.

Rust, H. W., Kruschke, T., Dobler, A., Fischer, M., and Ulbrich, U.: Discontinuous Daily Temperatures in the WATCH Forcing Datasets, Journal of Hydrometeorology, 16, 465–472, https://doi.org/10.1175/JHM-D-14-0123.1, 2015.

Schild, P.: Macro-enabled Excel spreadsheet to calculate hourly-averages from BSRN .dat files, http://hdl.handle.net/10013/epic.48977.d002, 2016.

Sheffield, J., Goteti, G., and Wood, E. F.: Development of a 50-Year High-Resolution Global Dataset of Meteorological Forcings for Land Surface Modeling, Journal of Climate, 19, 3088–3111, https://doi.org/10.1175/JCLI3790.1, 2006.

Stackhouse Jr., P. W., Gupta, S. K., Cox, S. J., Mikovitz, C., Zhang, T., and Hinkelman, L. M.: The NASA/GEWEX surface radiation budget release 3.0: 24.5-year dataset, Gewex news, 21(1):10–12, http://www.gewex.org/resources/gewex-news/, 2011.

Switanek, M. B., Troch, P. A., Castro, C. L., Leuprecht, A., Chang, H.-I., Mukherjee, R., and Demaria, E. M. C.: Scaled distribution mapping: a bias correction method that preserves raw climate model projected changes, Hydrology and Earth System Sciences, 21, 2649–2666, https://doi.org/10.5194/hess-21-2649-2017, 2017.

Trenberth, K. E., Fasullo, J. T., and Kiehl, J.: Earth's Global Energy Budget, Bulletin of the American Meteorological Society, 90, 311–323, https://doi.org/10.1175/2008BAMS2634.1, 2009.

Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D. C., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., Berg, L. V. D., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, L., Janssen, P. A. E. M., Jenne, R., Mcnally, A. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J.: The ERA-40 re-analysis, Quarterly Journal of the Royal Meteorological Society, 131, 2961–3012, https://doi.org/10.1256/qj.04.176, 2005.

Veldkamp, T. I. E., Wada, Y., Aerts, J. C. J. H., Döll, P., Gosling, S. N., Liu, J., Masaki, Y., Oki, T., Ostberg, S., Pokhrel, Y., Satoh, Y., Kim, H., and Ward, P. J.: Water scarcity hotspots travel downstream due to human interventions in the 20th and 21st century, 8, 15 697–, https://doi.org/10.1038/ncomms15697, 2017.

Vetterling, W. T., Press, W. H., Teukolsky, S. A., and Flannery, B. P.: Numerical Recipes in C (The Art of Scientific Computing), Cambridge University Press, 2nd edn., 1992.

von Mises, R.: Mathematical Theory of Probability and Statistics, Academic Press, New York, 1964.

Weedon, G. P., Gomes, S., Viterbo, P., Österle, H., Adam, J. C., Bellouin, N., Boucher, O., and Best, M.: The WATCH forcing data 1958–2001: A meteorological forcing dataset for land surface and hydrological models, Technical report no. 22, http://www.eu-watch.org/publications/technical-reports, 2010.

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, Journal of Hydrometeorology, 12, 823–848, https://doi.org/10.1175/2011JHM1369.1, 2011.

Weedon, G. P., Balsamo, G., Bellouin, N., Gomes, S., Best, M. J., and Viterbo, P.: The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data, Water Resources Research, 50, 7505–7514, https://doi.org/10.1002/2014WR015638, 2014.

Wild, M., Folini, D., Schär, C., Loeb, N., Dutton, E. G., and König-Langlo, G.: The global energy balance from a surface perspective, Climate Dynamics, 40, 3107–3134, https://doi.org/10.1007/s00382-012-1569-8, 2013.

Wild, M., Folini, D., Hakuba, M. Z., Schär, C., Seneviratne, S. I., Kato, S., Rutan, D. A., Ammann, C., Wood, E. F., and König-Langlo, G.: The energy balance over land and oceans: an assessment based on direct observations and CMIP5 climate models, Climate Dynamics, 44, 3393–3429, https://doi.org/10.1007/s00382-014-2430-z, 2015.

Wilks, D. S.: Statistical Methods in the Atmospheric Sciences, Academic Press, San Diego, CA, 1995.

Xu, X.: Methods in Hypothesis Testing, Markov Chain Monte Carlo and Neuroimaging Data Analysis, Ph.D. thesis, Harvard University, http://nrs.harvard.edu/urn-3:HUL.InstRepos:11108711, 2013.

Zhang, T., Stackhouse, P. W., Gupta, S. K., Cox, S. J., and Mikovitz, J. C.: The validation of the GEWEX SRB surface long-wave flux data products using BSRN measurements, Journal of Quantitative Spectroscopy and Radiative Transfer, 150, 134–147, https://doi.org/10.1016/j.jqsrt.2014.07.013, topical issue on optical particle characterization and remote sensing of the atmosphere: Part I, 2015.

Zhao, M. and Dirmeyer, P. A.: Production and analysis of GSWP-2 near-surface meteorology data sets, vol. 159, Center for Ocean-Land-Atmosphere Studies Calverton, http://ww.w.monsoondata.org/gswp/gswp2data.pdf, 2003.