Earth Syst. Dynam. Discuss.,
https://doi.org/10.5194/esd-2017-81-RC2, 2017

Earth System Dynamics
Discussions

Interactive comment on "Bias correction of surface downwelling longwave and shortwave radiation for the EWEMBI dataset" *by* Stefan Lange

Anonymous Referee #2

Received and published: 2 November 2017

The author examines different methods for bias correction (BC in the following) of surface downwelling shortwave and longwave radiation (rsds and rlds). More specifically, he uses the Surface Radiation Budget (SRB, 1 degree resolution, 07/1983 - 12/2007) data set to adjust biases in the EartH2Observe (E2OBS, 0.5 degree resolution, 1979 - 2014) data set. The BC methods used differ in the time scale on which they operate (daily or monthly means), the way how the different spatial resolutions (1 degree versus 0.5 degree) are dealt with, and the details of the parametric quantile mapping (beta or normal distribution, use of physical constraints). Two best BC methods, one for longwave the other for shortwave, are identified based on cross-validation and variation considerations. Comparison with independent site-specific measurement data

from the Baseline Surface Radiation Network (BSRN) leads to more mixed results.

The topic of the study - performance of different BC methods for rsds and rlds - is clearly of interest and suitable for ESD. However, in its current form the study suffers from several shortcomings as detailed below. I therefore recommend major revisions.

General comments:

A first concern is the focus of the paper: is the focus the evaluation of different methods or the quantitatively correct bias correction rsds and rlds in an absolute sense? Overall, the paper seems to suggest the former (comparison of methods). However, the use of BSRN data as an independent quantitative check points to the later (quantitatively correct rsds and rlds in an absolute sense). If the latter is indeed part of the goal, more work has to go into ascertaining the quantitative correctness of the SRB data used for bias adjustment.

A second major point is the overall clarity of the manuscript. The methods used are complex, the figures shown are (too) packed with interesting information. However, explanations and descriptions come in often (very) long sentences, with lots of details, making it difficult to grasp the essentials. More focused and shorter sentences would help, as would some more information (possibly equations) on the parametric methods. The reason for specific choices (e.g. why comparing these methods, why using these metrics?) are not given. Conclusions read in wide parts more like an extensive summary.

Ideally, the statement that there are two best methods (one for rsds the other for rlds, and measured in terms of cross-validation) would be further embedded. Can these methods be used for bias correction of the entire E2OBS period without introducing artifacts? Could the methods be further improved? Are the other methods just slightly or clearly worse?

Specific comments:

p.3, l.27: Why use to different versions of SRB for rlds and rsds?

p.4, l.9: "If deviations of SRB from SRBQC data quantify methodological uncertainty inherent to SRB data then these findings justify the bias correction of E2OBS rlds and rsds using SRB data over land at least." Two points here. For rsds, one may argue on the same ground that wide parts of the oceans also need adjustment. More generally, you assume here that SRB is correct (at least more correct than E2OBS). How can you be sure? For example, how does SRB compare to CERES data? Or to global mean estimates of rsds and rlds? A number of papers, e.g. by Trenberth et al., give numbers for the latter. An alternative may be to focus only on the methods and not argue at all about the quality of the SRB data.

Figure 1: Which of the differences are statistically significant?

Table 1: How about the altitude dependence of short wave radiation? (See e.g. Marty, Philipona, Frohlich, Ohmura, Theor. Appl. Climatol. 2002)

p.6, l.6: What do you mean by bilinear interpolation from coarse (SRB) to fine (E2OBS) grid? Copying? Same question on p.11, l.18.

p.6, l.8: "For the BCvtp2 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." Please try to clarify. I think I understood much later, in Section 3.2.1, that you adjust the mean and variance of E2OBS data on the E2OBS grid with mean and variance of SRB data on the corresponding, coarser SRB gird. True?

p.6, l.14: "... of the underlying four E2OBS values." The two grids thus are such that four E2OBS cells correspond to one SRB cell? They are not shifted against each other?

p.6, l.16: It would be helpful if you added some information, possibly equations, on

transfer functions, target distributions, estimation of means and variances of beta functions etc. in an appendix, as these are absolutely central to your study. Currently, the reader has to know all this or has to check out the references. After all, you even devote an appendix to explaining Kolmogorov-Smirnov.

Figure 2d: Why are the colored lines so far away from the black and gray lines?

p.7, l.8: What do you mean by "The rsdt climatology at a given latitude is rescaled such that it sits just above the multi-year maximum..."? Why do that?

p.10, l.9: "... one possibility to define ..." What would other possibilities be?Why your choice?

p.10, Eq. 1: Where does the equation come from? Can you give a reference? The explanation following eq. 1 reads rather lengthy but not too clearly.

p.11, l.9: How often does this "99%" condition kick in?

p.11, l.16: How often does this "40%" condition kick in?

p.11, l.27: "Metrics used..." Why these? Why, for example, skewness? What do I learn from this measure? And why a Kolmogorov-Smirnov test? Why not a test that gives more weight to tails, e.g. Anderson-Darling? More generally, when do you say that your bias adjustment is good? When the adjusted E2OBS distribution is identical (mean, variance, skewness...) to the SRB distribution? Why then adjust at all and not just take the SRB data? Can you use your method to adjust E2OBS data beyond the time span where SRB data is available?

p.12, l.2: Does the remark about CVCC imply that your method cannot be used to correct E2OBS data outside the SRD period (1983-2007)?

p.12, l.11: "In the following, cross-validation results are only shown and discussed for the BCvtp0 and BCvtp1 methods, since results for the corresponding BCvtp1 and BCvtp2 are virtually identical." What do you mean? That the difference between

BCvtp0 and BCvtp1 is similar as between BCvtp1 and BCvtp2? And, consequently, BCvtp0 and BCvtp2 differ more?

p.12, l.17: "... overall performance ..." What do you mean by overall performance?

p.12, l.24: Why now looking at relative differences?

Figure 3: I guess a good bias correction in your metrics results in a white map. True? The color / hue coding may be better explained upon first use.

Figures 4 and 5: Why are the quantities shown of interest? And, again, what is good and what is bad? If white means "good", then none of the methods performs well here?

p.14, l.15: Why should bias adjustment on monthly timescales outperform daily bias adjustment with subsequent monthly averaging?

p.15, l.3: "Rather, the p-value distributions depicted in Fig. 6b,d suggest that if sampling errors are taken into account then the BCvdp1 methods correct the distributions of monthly mean values almost as well as the BCvmp1 methods." I do not see this point from the text and / or figure.

p.15, l.7: "For BCvdp1, this is linked to an insufficient adjustment of third-and higher-order moments..." Not sure what you mean. That you should use another parametric method that takes into account higher moments? At what point do you start to "overfit" if you do this?

p.15, l.11: "... correct the upper tail of the rlds and rsds distributions." Can you say this if you use Kolmogorov-Smirnov, which focuses on the center of the distribution?

Section 4.2: Comparison with BSRN data. Here you compare point data with area mean data. This comes with potentially quite some uncertainty. See e.g. papers by M.Z. Hakuba et al. 2013 / 2014 / 2016 or N.A.J. Schutgens et al. 2016. Part of your disagreement could have its roots there. More generally, you are looking here more into how good your SRB data is than how good your bias adjustment is. If this is of

interest, you should also consider other data, e.g. CERES or global mean estimates for rlds and rsds, e.g. by Trenberth et al. In its current form, the comparison with BSRN data is rather confusing than helping, I think.

Figures 7 and 8: What is the colored rectangle to the lower left in each panel?

p.18, l.1: "... and differences between standard deviation biases generated by BCdsdp0, BCsdp1 and BCsdp2 are in line with cross-validation results." What do you mean?

p.18, l.5: "... 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." So the BSRN comparison does not make sense?

p.18, l.8: I do not understand this paragraph.

p.19, l.1 to 14: I think much of what you are describing here has to do with the fact that you are comparing point measurements with area means. See the above mentioned papers by Hakuba, Schutgens, and references therein.

p.19, l.26: Why use a staggered grid?

Figure 10: The figure seems to suggest that variability is strongly enhanced (red areas) by the bias adjustment. True?

Appendix C: What is the take home message? Figure C2 seems to suggest that the window length is irrelevant. True?

Interactive comment on Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2017-81, 2017.