Earth Syst. Dynam. Discuss.,
https://doi.org/10.5194/esd-2017-81-AC2, 2017

Earth System Dynamics Discussions

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

S. Lange

slange@pik-potsdam.de

Received and published: 26 December 2017

Responses by the author (in italics) to comments (not in italics) by anonymous referee #2

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.

Many thanks to referee #2 for her comprehensive criticism of the validation agains independent surface observations. After carefully consulting the concerns presented and literature provided by the referee I have decided to completely remove this part of the manuscript. Indeed, the validation was a secondary goal of the paper, which clearly benefits from focusing on its main goal, which is the evaluation of the different quantile mapping (QM) methods.

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.

Since referee #1 also pointed to too packed figures, I will reduce their information content to some extent in the revised manuscript and provide more comprehensive explanations, see my responses to your specific comments. I will consult a native English speaker to help shorten sentences where needful. Reasons for choices of methods and metrics will be better motived, see my further responses below. The conclusions section will be distilled to the essentials.

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?

The methods can definitely be used for bias correction of the entire E2OBS period, see my response to your specific comment below. Clearly benefitial would be a simultaneous bias correction at both the daily and the monthly time scale. Also, non-parametric QM using upper radiation limits as estimated by my best methods might yield better cross-validation results. These potential improvements will be discussed in the revised manuscript. Qualitative differences between methods will be discussed in more detail.

Specific comments

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

These are the latest available versions of the SRB dataset. The version numbers differ between rlds and rsds. This will be explained in the revised manuscript.

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.

In the revised manuscript I will 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?

This figure will be removed from the manuscript, in line with focusing on the methods.

Table 1: How about the altitude dependence of short wave radiation? (See e.g. Marty,

Philipona, Frohlich, Ohmura, Theor. Appl. Climatol. 2002)

Also this table will be removed from the manuscript (and along with it the question of how shortwave radiation changes with altitude), in line with focusing on the methods.

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.

I will change "bilinearly interpolated" to "spatially bilinearly interpolated" in both cases. I think this is a standard term, which does not need further explanation.

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?

I will change this sentence to "the BCvtp2 methods adjust mean values and variances at the E2OBS grid such that mean values and variances of spatial aggregates to the SRB grid match the corresponding SRB estimates while the sub-SRB-grid scale spatial structure of mean values and variances present in the original E2OBS data is retained."

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?

Correct. I will add the sentence "Every SRB grid cell contains exactly four E2OBS grid cells." to the data description section.

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.

Thank you for pointing this out. I will add such an appendix to the revised manuscript.

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

Because my estimates of the upper bounds of monthly mean radiation are calculated based on the upper bounds to the corresponding daily mean radiation. The resulting upper bounds are typically much larger than observed maximum monthly mean radiation because 31 consecutive days of daily mean radiation at its physical upper limit are very unlikely to occur in reality. I will add such an explanation to Sect. 3.1.2.

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?

To answer your questions, I will rewrite the beginning of this paragraph as follows: "The BCsda1 method employs the climatology of daily mean shortwave insolation at the top of the atmosphere (rsdt; see Appendix 1 for how rsdt is calculated in this study) for the upper bound estimation. This is motivated by rsds being limited by rsdt in most locations and seasons, which suggests that the annual cycle of the upper bound of daily mean rsds has a similar shape as the climatology of daily mean rsdt. Therefore, method BCsda1 uses a rescaled daily mean rsdt climatology as the upper bound climatology of daily mean rsds (solid blue line in Fig. 2c). The rescaling is done with the smallest possible factor which guarantees that the resulting upper bounds are greater than or equal to the multi-year maximum values of daily mean rsds on all days of the year with rsdt $\geq$ 50 W m-2. An extension of this guarantee to days of the year with lower rsdt would inflate the rescaling factor because during dusk and dawn of polar night, rsds can exceed rsdt due to diffuse radiation coming in from lower latitudes. On days of the year with rsdt $<$ 50 W m-2, the maximum of the rescaled rsdt and the empirical multi-year maximum daily mean rsds is used as the upper rsds bound."

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

choice?

Another possibility would be to follow the BCvtp0 approach, i.e. to use interpolated data. The motivation of my choice is that it solves the problem illustrated and discussed in Sect. 4.3. I will rephrase the sentence as follows: "With target distributions fixed at the SRB grid, target distributions at the E2OBS grid can be defined such that the bias-corrected data have the SRB-grid scale target distributions and the sub-SRB-grid scale structure of the original E2OBS data."

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.

This does not need any reference. It is the standard formula for the variance of a linear combination of random variables. I will however insert one intermediate step using covariances in the equation to make its derivation easier to understand. I will ask a native English speaker to improve the explanation following Eq. 1.

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

For longwave (shortwave) radiation, this "99%" condition kicks in over four (about 11% of all) grid cells and there on 15% (about 5%) of all days of the year. I will add this information to the revised manuscript version.

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

The "40%" condition is never met for longwave radiation whereas for shortwave radiation it kicks in over 14% of all E2OBS grid cells and there on 2% of all days of the year. I will add this information to the revised manuscript version.

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?

The skewness is included because it is the first distribution moment which is not explicitly adjusted by my parametric QM methods. It is included here to illustrate this conceptual imperfection of my methods. I will include this motivation in the revised manuscript. You are right about the KS test and the relatively low weight it gives to tails. In the revised manuscript version, I will include Kuiper's test as one that (like the suggested AD test) gives the same weight to CDF differences at all quantiles. Qualitatively, however, the Kuiper's test results are the same as those of the KS test. You are right that I (and, as far as I know, everybody else who cross-validates bias correction methods) consider a bias adjustment good if the adjusted distributions are identical to the target distributions. I will include this definition of (overall) performance in the revised manuscript. In the ISIMIP framework, there are two reasons for doing the bias adjustment of E2OBS to SRB data and not just using the SRB data directly: It (i) promises a higher intervariable consistency (e.g. consistency of temperature and longwave radiation) in the EWEMBI dataset and (ii) prodces radiation data that cover a longer time span. Applying the methods to E2OBS data beyond the time span where SRB data are available is fine since the 1979–2013 period is in fact not much larger than the 1983–2007 period, so that the former is expected to be sufficiently well represented by the latter.

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

No, it does not, see my response to your previous comment.

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?

No, I mean that cross-validation results for the BCvtp1 and BCvtp2 methods are virtually identical. I will remove the word "corresponding" as I suspect that the confusion will be gone once this word is gone. I will also consult a native English speaker about this.

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

This will be answered in the introduction part of the results section of the revised manuscript, see my response to your comment on p.11, l.27.

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

Why not? For standard deviations, I think this makes more sense than to look at absolute 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.

Not true. White means low agreement in bias direction (positive or negative bias) over months and validation data samples. I will consult a native English speaker to see how the explanation can be improved.

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?

The cross-validation of multi-year maximum values shall reveal if it is worthwhile and if so, then how to explicitly adjust upper radiation bounds. I will include this sentence in the introduction part of the results section of the revised manuscript. For why skewness is of interest, see my response to your comment on p.11, l.27. As to the significance of Figure 5, see my answer to your next question. In terms of what white means, see my answer to your previous comment. The methods are clearly not perfect but I also did not expect that. It does not make sense to make an absolute statement such as "this shows that the method performs well." The only sensible question is if one method performs better than another. Figures 4 and 5 quantify the magnitude of biases of *selected statistics that remain after bias correction with different methods.*

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

Because of what is shown in Figure 5. I will revise the explanation of Figure 5 earlier in the text as follows in order to answer your question: "At the monthly time scale, lower biases are expected to remain after bias correction at the monthly than at the daily time scale. Most importantly, the interannual variability of monthly mean radiation is explicitly adjusted by the BCvmpx methods whereas it is not by the BCvdpx methods as the latter pool daily mean radiation values from all years before the adjustment and are therefore oblivious to variability at the interannual time scale. As an example, in Fig. 5, median biases of interannual standard deviations of monthly mean rlds and rsds are shown to be mostly within/beyond +-20% after bias correction with BCvma1/BCvda1."

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.

I will remove this sentence from the revised manuscript.

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?

I mean that my parametric methods explicitly adjust mean values and variances. Higher-order moments are only implicitly (and therefore most likely not perfectly) adjusted through the distribution fitting. In fact, with my methods you cannot overfit in your sense because both the normal and the beta (provided its bounds have been fixed) distribution only have two parameters, which are fixed once two moments have

been fixed. Therefore, they cannot adjust more than two moments explicitly. An alternative would be to use non-parametric QM methods. I think that all of this will become clearer thanks to the appendix about QM and downscaling that will be appended to the revised manuscript.

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?

Kuiper's test gives the same result. I will adjust the statement accordingly.

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.

I agree (see above). I will remove this section from the manuscript.

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

These figures will be removed from the revised manuscript.

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?

Irrelevant now that this part will be removed from the revised manuscript.

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?

Irrelevant now that this part will be removed from the revised manuscript.

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

Irrelevant now that this part will be removed from the revised manuscript.

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.

Maybe. Will be removed from the revised manuscript.

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

Smaller differences between RMSDs of adjusted E2OBS data from SRB-grid cell and staggered SRB-grid cell mean values are considered to indicate a better bridging of the E2OBS-to-SRB spatial scale gap. Ideally, there would be no such difference and it would therefore be impossible to tell from this analysis if the target distributions of the bias correction were defined on the SRB or staggered SRB grid. I will include this explanation in the revised manuscript.

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

True.

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

True.

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