HESSD

Interactive
comment

Interactive comment on "Cross-validation of bias-corrected climate simulations is misleading" *by* Douglas Maraun and Martin Widmann

Douglas Maraun and Martin Widmann

douglas.maraun@uni-graz.at

Received and published: 11 July 2018

We would like to thank the reviewer for the thoughtful and constructive comments.

Regarding the possible deletion of section 4, we strongly support the reasoning of reviewer 1 to keep the example as is. The reason is twofold:

1. we believe that the analytical derivation might be helpful for some readers, whereas others may prefer a strong illustrative example. This holds in particular as - to our experience - the role of internal variability in climate research is still often underestimated. The analytical derivation might then be dismissed as being purely academic reasoning without practical relevance.

2. choosing the Maraun et al. (2017) example would not suffice. In fact, the reason for

[Printer-friendly version](#)

[Discussion paper](#)

this short article was that - during the review process of the Maraun et al. (2017) paper we realised that the situation was even worse than laid out in that paper. There, the key starting point for the discussion was that cross validation may not be able to identify a nonsense bias correction. This is the false negative case in the manuscript at hand. Here we show additionally that there is another problematic case: that a sensible bias correction may be rejected (false positive). This case is not contained in the Maraun et al. (2017) example. Of course, it makes sense to furthermore show the true positive and true negative cases as well. We also do not believe that the exaggerated examples are limited in applicability/transferability to more realistic cases: the true negative case is a realistic case where a well performing climate model is successfully bias corrected - here, the case is not at all exaggerated. Similarly the false positive case, where the bias correction is sensible, but the residual bias does not vanish, is far from exaggerated: this is exactly the case we would like to highlight with this paper.

The other two cases are chosen to display wrong applications of bias correction in a convincing case. To avoid any discussion about the sensibility of bias correction in one or the other situation, we decided to take examples where anybody would agree. In fact, the false negative case - the correction does not make sense, but it is not rejected - is of a very similar character as the example chosen in Maraun et al. (2017). One may actually argue that the correction of temperature against precipitation from two different regions in that paper is even more exaggerated than the example here (where the same variables are chosen, but different locations). In a real application, of course, the problem will not be as obvious as constructed in our examples. Here, the user of BC has to carefully assess whether the bias correction makes sense at all (see discussion below). This discussion, however, is not the main focus of our manuscript.

To summarise, we suggest to keep the example as is. But we agree that a brief discussion on the transferability of our examples might help.

Regarding the suggestion of a general discussion of the contexts in which cross validation may make sense: we believe this would go well beyond the scope of our

manuscript. As indicated, this even depends on the way cross validation is carried out (in the "statisticians way" or in the "atmospheric scientists way"), with several subtleties. A thorough discussion could easily distract the reader from our main point.

We do agree, however, that it would be useful to add a discussion about the role of the length of the observational record in comparison with the periodicities and amplitude of the relevant mode(s) of variability. We will therefore add a brief discussion accordingly and will try - as suggested by the reviewer - to pull the different points together to highlight the overall issue in a concise manner.

Regarding the question how BC can be evaluated in the context of extrapolation: the reviewer is of course right that this is in general an open question. We have had a discussion of this issue in the Maraun et al. (2017) paper, where we highlight the fact that BC has to be accompanied by a thorough evaluation of non-corrected features (in particular temporal and spatial), by a process-based evaluation of the underlying climate model (in terms of location biases, relevant feedbacks etc.), by an assessment of potential artefacts, and by reasoning about representativeness and trend modifications. We will consider to add a summary of this discussion in the conclusions with a reference to the other paper.