Interactive comment on “Is bias correction of Regional Climate Model (RCM) simulations possible for non-stationary conditions?” *by C. Teutschbein and J. Seibert*

Anonymous Referee #2

Received and published: 10 January 2013

Review of “Is bias correction of Regional Climate Model (RCM) simulations possible for non-stationary conditions?” by C. Teutschbein and J. Seibert.

1. Summary

The authors propose to use a split sample approach to test how bias correction methods perform under non-stationary conditions. The research question is definitely relevant, and only little has been published on this issue. Also, the manuscript contains some interesting aspects. Nevertheless, I am pretty sure that the authors cannot learn from their study what they intend to learn. Namely, they do not learn anything about the performance of bias correction methods under nonstationary conditions.

RESPONSE

We thank Referee #2 for the useful comments. We will consider these when revising the manuscript. As we discuss below, we still argue that one can learn about the performance of bias correction methods under non-stationary conditions, but we also agree that we need to be more careful in some formulations/conclusions. Please see our response below.

2. Main point

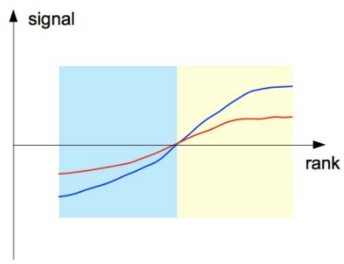
The main shortcoming of the study is that the authors do not clearly define what a bias is, and what nonstationarity of a bias means. In particular they do not distinguish between forced signals and internal climate variability. This imprecision has severe consequences for the design and interpretability of their study. What is a bias, and what is a bias nonstationarity? A bias is by definition the systematic difference between the observed mean climate and the simulated mean climate, i.e., the difference between the expected observed signal and the expected simulated signal. This implies, and this is the important point, that a bias is by definition unaffected by realisations of internal climate variability. For a good estimate of a bias, one would need many realisations of both the model and observations, or at least a long series in a stationary climate. Consquently, nonstationarities of biases cannot be caused by internal climate variability, but only by changes in the mean climate itself, i.e., by changes in external forcings (for a discussion, see Maraun, Geophys Res Lett, 2012).

RESPONSE

We would like to thank the referee for highlighting this issue and for the thorough explanations. We agree that we did not discuss this sufficiently detailed in the manuscript.

We agree that we missed to clearly define what we mean by bias and this apparently has led to some misunderstanding. We will include our definition in the revised manuscript; this definition is broader than the one used by the reviewer and includes all types of systematic errors, i.e. not only the mean as first moment but also other moments and statistical properties. We do not agree with the referee that a bias is only “the difference between mean observed and mean simulated climate”. We refer to Ehret et al. *[2012]:*

>> In the context of climate change impact studies, the definition of bias is not as strict: it varies with the scope of the studies and is often used in a general sense for addressing any deviation of interest (e.g. with respect to the mean, variance, covariance, length of dry spells, etc.) of the model from the corresponding "true" value. Typically, biases are calculated for precipitation or temperature on continental, river basin or model grid scale for annual, seasonal, or monthly aggregations.<<



Consider the following example (see figure): assume a GCM/RCM simulation in a stationary climate (i.e., no nonstationarities!) that correctly simulates the mean climate (i.e., no bias of the mean!), but with a biased representation of internal climate variability. In fact, assume that the internal variability is only half as strong as in the real world. The authors would take the observed data (e.g., temperature), sort them (blue line in the figure) and split the sample into the lower half for calibration (light blue), and the other half for validation (light yellow). The same procedure is carried out with the regional climate model (RCM) simulation. Because of the under-represented internal climate variability, the amplitude of the simulated signal is only half of the observed data (red line). Now in the calibration period, the RCM signal is higher than the observed signal, i.e., the calibration would estimate a positive bias in the mean climate. In the verification period, however, the RCM signal is lower than the observed signal. Consequently, the authors would detect a negative bias in the mean climate, i.e., a bias nonstationarity. Yet by construction, there is no mean bias at all, and also no nonstationarity! The discrepancy instead arises from a wrong representation of internal variability, i.e., a bias in the second moment of the distribution of the considered variable. To summarise: the chosen approach is not able to assess bias nonstationarities, basically because it cannot discriminate between changes in forcings (which would cause bias nonstationarities) and internal climate variability (which does not cause bias nonstationarities).

RESPONSE

Thanks for the illustration of the problem. As the referee mentioned above, we did not clearly specify the definition of bias, which "has severe consequences for the design and interpretability of the study". The manuscript will be revised accordingly.

In addition, we would like to add a comment of referee #3 at this point:

>> Figure 4 shows to me that the climate models underestimate the interannual variability in the mean annual temperature, which leads to a lower increase in the mean temperature between the two 15-year periods than in the observations. This figure strongly supports the study setup as it proves that the bias in the RCMs change from the calibration to the validation period (no matter whether it comes from natural variability or from greenhouse gas forcing). If the bias was stationary, the change signal in the RCMs should be just the same as in the observations, at least for mean precipitation and temperature over a 15 year period. Based on the results in Figure 4, I would expect the bias-correction methods to show the worst performance in catchments where the change signal in the RCMs differs most from the one in the observations, i.e. in all catchments for temperature and in Vattholmaan for precipitation.<<

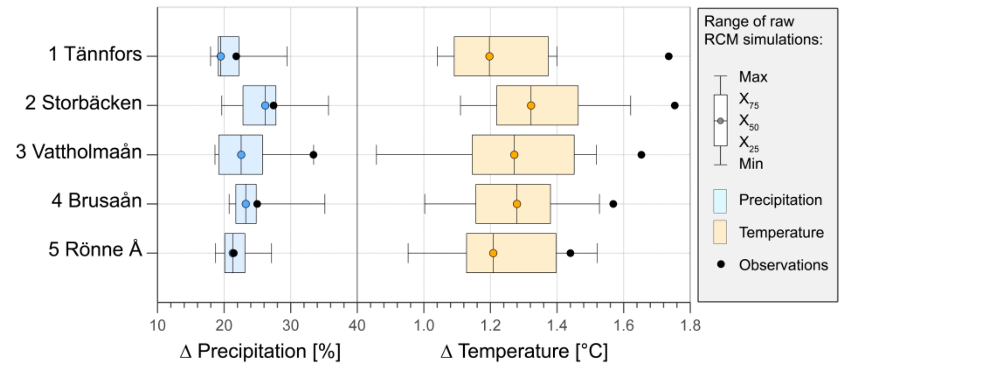


Figure 4: Differences between mean values of designed calibration and validation period for precipitation (left) and temperature (right) shown for raw RCM simulations (colored boxes) and observations (black circles).

We would also like to add that no test is perfect. The motivation of our study was that we experienced that for many impact modelers (who are using RCM data as input to other models) the source/cause of biases is not relevant, but the performance of bias correction methods is and we wanted to find a way to evaluate this performance. Thus, our experimental setup of the differential split-sample test is simply one possible approach to address the issue of biases and non-stationarity, but also biases in interannual variability. This will be further stressed in the revised manuscript.

3. Further points

1. The statement that we are not able to check whether the stationarity assumption is actually true or not (p 12771, l 2) is not correct. In fact, in a pseudo reality, one can at least test whether the assumption is wrong. Two recent studies have addressed these questions, Maraun, Geophys Res Lett, 2012; and Räisänen and Räty, Clim Dynam, 2012. These studies should be cited.

RESPONSE

Thanks for pointing this out. We will included the mentioned references in the revised manuscript and further discuss this issue.

2. Apart from the discussion above, the authors should also clearly discuss whether they address bias nonstationarities, or apparent bias nonstationarities caused by insufficient correction methods (for a discussion, see Maraun, Geophys. Res. Lett., 2012).

RESPONSE

Thanks for referring to the discussion by Maraun [2012]. So far we didn't distinguish between sensitivity related bias changes (SBC), variability related apparent bias changes (VABC) and mixture related apparent bias changes (MABC). We will analyze and discuss this issue further and include the mentioned reference in the revised manuscript.

3. For precipitation, the climate change signal in this short period is very weak only. Especially here, the fundamental flaw discussed above will come into play, as one only looks at internal variability, not at any forced trends.

RESPONSE

We don't agree in that the climate change signal is weak for precipitation. We mentioned in the manuscript (p. 12772, ll. 16-18):

> *>> In this study, the differences between designed calibration and validation period were within a range of 18–36 % for precipitation (Figure 4, left). <<*

This percentage change of 18-36 % is even higher than the values projected by IPCC [2007] for Northern Europe.

4. Please clearly explain how the bias correction methods have been applied! Did you apply them to the 4 seasons separately? Months? The whole year?

RESPONSE

We are sorry that we did not mention this in the current manuscript. The methods have been applied in the same way as in Teutschbein and Seibert *[2012], i.e. on a monthly basis. This will be clarified in the revised manuscript.*

5. The term „linear scaling“ does not make sense for temperature (table 3). Scaling is by definition multiplicative, as it keeps the proportions of the scaled object. See Widmann et al., J Climate, 2003, and http://en.wiktionary.org/wiki/scale#Verb or any relevant mathematical textbook. Also the term „statistical downscaling“ for quantile mapping (table 3) makes no sense. Statistical downscaling is the generic term for all

statistical methods attempting to bridge the gap between large scales and local scales. Please avoid using misleading and wrong terminology!

RESPONSE

Thanks for pointing this out. We will change the term to "linear transformation" in the revised manuscript.

6. Please clarify whether the example in Fig. 3 is artificial or not (what does it represent?). Also it makes no sense to add year from 1963 onwards, as the x-axis shows ranks, not years.

RESPONSE

As mentioned in the manuscript, Figure 3 shows an "exemplary procedure" and is of artificial nature based on random numbers. We believe that the years on the x-axis help to make the figure better understandable and removing them would not help the reader. However, it is possible to add the ranks on the x-axis in the revised manuscript.

7. Sometimes the authors have too many self-citations. For instance, l 12769: "As Teutschbein and Seibert (2010) concluded, multi- model approaches (i.e. ensembles) have two advantages:" This conclusion has not been drawn by Teutschbein and Seibert for the first time. Here, the standard literature should be cited. Also regarding downscaling methods, the classical papers introducing PP and MOS (i.e., bias correction, e.g., Klein, BAMS, 1974) and relevant reviews (e.g., Maraun et al., Rev. Geophys., 2010) should be cited.

RESPONSE

We will add further references in the revised manuscript.

8. The statement „In this study, the differences between designed calibration and validation period were within a range of 18–36 % for precipitation (Fig. 4, left) and 0.86–1.75 C for temperature (Fig. 4, right). These values represent a reasonable climate change signal that is likely to occur within this century (IPCC, 2007)." (p 12772, l 16) is obviously misleading, as the likely temperature increase is much higher.

RESPONSE

We assume the referee refers to the expected temperature change until the end of the 21st century. We are just saying "within" this century, which can be at any point, e.g., within the next 30 or 50 years. For example, the 11 uncorrected RCMs (A1B) used in this study project a temperature increase of 1.6-2°C from 1961-1990 to 2021-2050. This was not mentioned in the manuscript, but can be included in the revised version for clarification.

9. The term „ensure" (p 12776, l 6) is overly optimistic and rather naive.

RESPONSE

This will be changed in the revised manuscript.

REFERENCES

Ehret, U., E. Zehe, V. Wulfmeyer, K. Warrach-Sagi, and J. Liebert (2012), HESS Opinions "Should we apply bias correction to global and regional climate model data?", Hydrol Earth Syst Sci, 16(9), 3391–3404, doi:10.5194/hess-16-3391-2012.

IPCC (2007), Climate Change 2007: The Physical Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, and H. L. Miller, Cambridge University Press, Cambridge, UK, and New York, USA.

Maraun, D. (2012), Nonstationarities of regional climate model biases in European seasonal mean temperature and precipitation sums, Geophys Res Lett, 39(6), L06706, doi:10.1029/2012GL051210.

Teutschbein, C., and J. Seibert (2012), Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods, J Hydrol, 456–457, 12–29, doi:10.1016/j.jhydrol.2012.05.052.