

**Review of the manuscript, HESSD 9, 12765-12795, 2012,
Is bias correction of Regional Climate Model (RCM) simulations possible for non-stationary conditions?,**

by C. Teutschbein and J. Seibert.

Summary

The authors present a study which analysis and compares the performance of six widely used bias-correction methods under changing climate conditions. Their analysis setup is based on a differential split sample test. From the 30 years in the reference period, they identify the 15 coldest/15 driest and the 15 warmest/15 wettest years, with the former being used for calibrating the bias-correction models and the latter for validating them. The results show that all bias-correction methods show better validation statistics than the raw climate model data, and that distribution mapping outperforms the other methods.

General comments

The paper's topic is highly relevant for current impact research. It is to my knowledge one of the most comprehensive bias-correction methods evaluation studies for temperature and precipitation, combined. The chosen cross-validation framework is sound and any effort to promote cross-validation in bias-correction should be supported as this has often been neglected in previous studies. The authors have achieved to compress a large amount of data into a few meaningful graphics which generally are of high quality.

One of the main messages of the paper is that simple methods do not perform well. However, when looking at the figures, I found the simple methods to perform surprisingly well. Although I generally agree with the criticism of the simple methods, I think the results do not allow to draw the rather general conclusion that the simple methods should not be used any further (See also my major comment below). In fact, depending on the statistics one is interested in, one might come up with a different choice of a bias-correction method.

Based on the high relevance and quality of the paper, I suggest acceptance of the manuscript after some changes have been made.

I am very much looking forward to the revised article.

Major comments

1) Blending of climate change signal with natural variability

In the study, two climatically different periods are the basis for the cross-validation of the bias-correction and for answering the question whether bias-correction is possible for non-stationary climate conditions. The differences between the two periods are to a large extent caused by natural variability, and not by a greenhouse-gas forcing. As there might be differences in how non-stationary biases either due to greenhouse-gas forcing or due to natural variability evolve, and it is the former we are primarily interested in, the authors should clearly state this limitation of the study (which is also a limitation of any other similar study). For example, for biases due to natural variability, non-stationarity due to model-representation of the ocean dynamics do not matter so much, but they might become more important once we turn to time-scales relevant for greenhouse gas forcing.

2) General conclusion about using simple versus more advanced methods

Based on the results of the study, the authors question the use of simple bias-correction methods. However, when reading the paper and looking at the results, I got a more differentiated picture which is consistent with the section 3.2 in the paper, particularly lines 9-11 on page 12775. The simple methods perform surprisingly well in many statistics, but distribution mapping seems to combine all the strengths of the simple methods. The authors should soften their last statement in

the abstract in a way that is consistent with the discussion of the results in section 3.2, and that also accounts for the fact that the results are only valid for the investigated catchments and might look different in other climates. For example, in case of a dry bias in the RCMs which is often the case in Southern Europe during summer, the distribution mapping will have problems to correctly allocate additional precipitation days in the time series, and possibly leading to a wet-bias (e.g. Themessl et al. 2012). In dry climates with only a few precipitation days, there might be not enough data points for the distribution mapping to work (e.g. unstable parameter estimation).

Detailed comments

Page 12769, lines 17-19: „Therefore, climate variables simulated by individual RCMs do often not agree with observed time series (Fig. 2), which poses a problem for using simulations of a single RCM as input data for hydrological impact studies.“

I do not understand this reasoning. Even if RCMs would agree perfectly with the observations, they potentially disagree on the climate change signal and an ensemble approach is still necessary. Please clarify.

Page 12770, lines 11-14: The authors state that more details about the methods can be found in Teutschbein and Seibert (2012) and two other papers. If the methods have been applied in precisely the same way as in Teutschbein and Seibert (2012), please also state that. If not, give some more information about the method application. I'm particularly interested in the sub-annual periods chosen for the parameter estimation. Based on Teutschbein and Seibert (2012), I assumed it is based on monthly values, but from the text, it is not totally clear.

Page 12772, description of the DSST:

The DSST is based on the assumption that the years are interchangeable, i.e. that it does not matter whether the 15 wettest/warmest years in the observations correspond with the ones in the climate model. This should be clearly stated somewhere in the description of the DSST.

Page 12772, line 19 and thereafter:

Related to the major comment 1), the authors should discuss the blending of differences between the two periods based on natural variability and greenhouse gas induced differences.

Page 12772, discussion of Fig. 4:

I would like the authors to discuss in more detail the Figure 4, which is in my point of view a very central one. 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.

Page 12773-12774, Section 3.1:

As distribution mapping is the only method that is able to reproduce the observed annual temperature distribution in the validation period (Figure 6), it might be interesting to see an additional analysis of the distribution mapping. When studying the results, the question arose "Why can distribution mapping correct the substantially underestimated interannual variability of

temperature shown in Fig. 4?" I suggest to insert an additional plot that shows the distribution mapping correction function for temperature of the basin #2. Based on such a figure, one could discuss how the distribution mapping is able to alter the change signal of the GCM-RCMs in a non-linear way. I guess such an analysis would be illustrative for many readers, but I leave it to the authors to decide whether or not such an analysis is beneficial for the paper.

Conclusions: The authors should mention the limitation of the DSST, namely that it uses non-stationarity introduced by natural variability rather than non-stationarity introduced by a greenhouse-gas induced climate change to test the performance of the bias-correction methods (see major comment 1)

Fig. 5: It would be very interesting to see the results for the catchment #3 as it shows more pronounced non-stationarity in the bias (see Fig. 4). If this non-stationarity leads to different results compared to Fig. 5, please include two Figures for precipitation, one representative for the 4 catchments #1,#2,#4 and #5, and one for the outlier catchment #3. Otherwise, please state that the results are similar for all catchments, including #3.

Technical comments

Fig. 5: I suggest to stick to the name Brusaån as in the Fig. 1 and Tab.1 not to confuse the reader.

References

Teutschbein, C. and Seibert, J.: 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, 2012

Themessl, M. J., Gobiet, A., Heinrich, G.: Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal, *Climatic Change*, 112, 449-468, 2012