# 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 #1

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

This paper assess whether bias correction techniques can be tested for nonstationarity and if it is either possible or not to assess this. Although the paper is well-written well presented, the presentation of the methodology is somewhat unclear. The methods are tested on a few very small catchments in Sweden, so the generality of the results is questionable. The question posed is relevant, but the conclusions are not enough supported by the results. I would also recommend that the reference list is updated. I therefore recommend that the paper undergoes a major review and is resubmitted to the journal.

RESPONSE

*We thank Referee #1 for the useful comments. We will try to clarify the methodology in the revised manuscript. We do not fully agree with the notion of 'very small catchments, and would actually argue that this meso-scale catchment size (150-300 km2) is often a relevant catchment size for (hydrological) impact modelers; it is also the catchment size, which approximately corresponds to the resolution of today's RCMs. We will revise the conclusions and update the reference list.*

### 2. Major comments

1. The authors perform a differential split-sample validation test (DSST) to assess the effect of bias correction methods. This was done by calibrating the methods on the coldest (driest) years for temperature (precipitation) and validating over the warmest (wettest) years over the period 1961-1990. The periods were selected from the observed data series. However, it is not clear to me how the wet/dry and cold/warm periods were selected from the RMCs. If the RCMs would have been driven by ERA40 (as in the Christensen 2008 paper) one can assume that a wet year in observations also would correspond to a wet year in the RCM-driven precipitation. However, since the RCMs in this paper are all driven by GCMs under scenarios this is not the case anymore. It is not true that the specific climate of a GCM projection will correspond to the same climate in observations, not even over a time period such as 30 years. A specific run of a GCM/RCM might have a cold and wet bias over the modelled period due to long-term climatic modes in the model. A bias correction using differential split-sample validation might then correct for biases that are caused by long-term climatic trends rather than model biases. This could have been avoided if the authors would have chosen the RCMs driven by ERA40, which are also available from the ENSEMBLES project, and I suggest that these are added to the study to see whether the conclusions are still valid.

RESPONSE

*The periods from the RCMs were selected in the same way as for observations. This was explained on page 12772, ll 3-13:*

> *>> Since our available 30-yr period was not long enough to show a considerable trend in precipitation or temperature data, we chose the two required segments 5 as follows: given that climate projections indicate an increase in future*

*precipitation and temperature for Northern Europe (IPCC, 2007), we compiled the two periods by sorting the years according to their amount of precipitation and temperature, respectively (Fig. 3). For the precipitation-bias correction assessment, we included the 15 driest years in the first subset ("calibration years") and the 15 wettest years in the second 10 subset ("validation years"). For the temperature-bias correction evaluation, we used the 15 coldest years as "calibration years" and the 15 warmest as "validation years". This procedure was done to all 11 RCM-simulated times series and the observed times series. <<*

*We will clarify the methodology in the revised manuscript to avoid any confusion.*

2. The authors claim that the split-sample test on different climate situations is novel, but it has been applied in earlier studies and can hardly be called novel (for a full discussion on stationarity, see Maraun et al 2010). Also, to fully use the method the opposite calibration-validation should be tested (warm-cold and wet-dry) to see whether the method is non-stationary. I suggest that full cross-validation to evaluate the performance of the bias-correction techniques. This should also be done over seasons or months, since a wet/dry year does not necessarily tell you anything on the distributions of rainfall. A few large events could turn an otherwise dry year into a normal or wet year. Furthermore, a validation over a control climate does not necessarily guarantee a valid method in a future climate. For a longer discussion on stationarity and validation methods, see Maraun et al. 2010.

RESPONSE

*We claim that the DIFFERENTIAL split-sample test has – to our best knowledge - not been used for testing the transferability of bias correction methods before. The test itself is not novel, as it was proposed by* Klemeš *[1986] almost 30 years ago. In which earlier studies has it been applied to evaluate bias-correction methods?*

*We agree with the referee that a full cross-validation is more appropriate and will try to corporate this in the revised manuscript.*

3. Stationarity is a problem when it comes to downscaling, but it is more related to statistical downscaling methods than dynamical. In a statistical downscaling method this assumption is fundamental. However, if RCMs are bias-corrected, this is more likely not so crucial, since the underlying precipitation derives from a dynamical model. On the other hand, also GCMs/RCMs have large inherent assumptions of stationarity incorporated in the parameterisation, especially in the land surface component.

RESPONSE

*We agree with the referee in that the stationarity is a problem when it comes to statistical downscaling. However, this is exactly the reason why it is a problem when it comes to bias correction of RCMs as most bias correction methods are by their nature a form of statistical downscaling. Therefore, stationarity is a serious problem of bias correction.*

4. In one of the methods, variance scaling is used (Chen et al, 2011). However, von Storch (2000) showed how variance inflation is built on the wrong assumption that the variance of the predictor, and that it the inflated variable will have larger squared errors than the original.

RESPONSE

*The referee points out a relevant drawback of variance scaling that we will add to the list of advantages/disadvantages in Table 3.*

5. The conclusion of the study is that the proposed methodology can evaluate the transferability of the methods to other climates, but I do not think that the results are strong enough to draw such conclusion. Firstly *(see RESPONSE A)*, the areas are very small and limited to Norhtern Europe. Secondly *(see RESPONSE B)*, a method that is applied to current conditions can only be valid under current conditions, even if it can separate different weather situations during that period. If a method is to be valid under future conditions it has to be tested under those conditions as well. One approach would be to test the methods in a "pseudoreality", as in Maraun 2012.

*We still believe that a differential split-sample test can give us an idea of how transferable bias correction methods are to future changed climate conditions.*

RESPONSE A

*We are not claiming to proof that the methods are transferable to completely different climate and land-use conditions (e.g., tropics). We are still talking about the same catchments (in our case Northern Europe) under changing climate conditions. Furthermore, we explicitly point out (on page 12772, ll 3-13) that*

>> These findings remain to be confirmed for other catchments and other geographic regions… <<

*Thus, we do not agree with the first part of this referee comment.*

RESPONSE B

*As we cannot foresee the future, it is in principle not possible to test how bias correction methods perform under future climate conditions (not even with the test performed by* Maraun *[2012]). This was pointed out several times in the manuscript. The applied differential split-sample test is simply an approach to get an idea how bias correction methods perform under conditions different from those that they were calibrated to. With help of the differential split-sample test, we created two periods, one representing current and one future climate conditions. The method used by* Maraun *[2012] is certainly another valid approach and will be discussed in the revised manuscript.*

**3. Minor comments**

1. P12766, L26-27. The authors mention the use of simpler bias-correction methods, but these are quite outdated, and methods using quantile matching or distributions to correct precipitation are more common now. See for example Maraun et al 2010, Themessl et al 2012 and Eden et al 2012

RESPONSE

*We agree with the referee in that the simpler bias correction methods are rather outdated, but only if one considers the climate modeling community. There are still many ongoing studies that use these rather simple methods. Especially impact modelers (e.g., modeling certain hydrological aspects, water chemistry, soil properties, etc.) still apply these methods because they are easy to understand and simple to apply. Impact modelers are often facing long and complex modeling chains and simply cannot handle rather advanced methods. Thus, the purpose of this paper was to test the most common methods of different complexity (also including distribution mapping) that can still be handled by impact modelers.*

2. P12769, L4. You use 11 RCMs from the ENSEMBLES project, but there are a lot more available, why not use them all? Then you can also see how much bias originates from the GCM rather than the RCM. But as mentioned above, the ERA40-driven RCMs would be needed to be added to the analysis.

RESPONSE

*We used all the RCMs (with 25 km resolution and scenario A1B) that were publicly available from the ENSEMBLE RT3 data archive and that had the required temperature and precipitation data for the time period available at the time of downloading (end of 2011).*

3. P12769. The first paragraph of section 2.3 could be moved to the introduction, since it is more of descriptive character.

RESPONSE

*We will consider this for the revised manuscript.*

4. P12770. L15-L24. The last paragraph could be moved to the discussion or the introduction, since it is rather arguing the drawbacks/benefits of bias correction.

RESPONSE

*We will consider this for the revised manuscript.*

5. P12770. L15. In the paper of Ehret et al 2012, they set out to discuss whether bias correction should be applied or not, since they claim that it hides the biases in climate models. It is a discussion paper, and therefore they are entitled to their opinion. However, In my opinion they are attacking bad science rather than actual bias correction, and they are depicting a too black and white picture of the science. If all the methods and results are clearly stated in a paper, then everything is transparent and the effects of bias correction are visible. Then their main argument fails. The paper also has errors in their interpretation of previous research. I am not saying that you should not use the reference, I am merely expressing my opinion that the particular paper is rather weak, which can also be seen in the interactive comments.

RESPONSE

*We can see the reviewers point, but still think the paper by Ehret et al. makes some useful comments. At this point we do not want to start a discussion about the quality of an already published paper.*

6. P12771. L22-23. I do not agree with this point. There are numerous bias papers who uses split sample approach and also DSST to evaluate their results, and I suggest that the authors do a more thorough literature review here.

RESPONSE

*The referee mentioned this before, but unfortunately no references are given. We are not aware of any study using DSST to test bias correction methods.*

7. P12772, L16-19. I would argue that these differences are most likely within the natural variability of current climate, and not a sign of climate change. Climate changes is expressed as changes in the mean climate over a long period, but the changes described have more to do with natural variability.

RESPONSE

*The referee points out a valid issue that was also raised by Referee #2. Thus a more detailed reply can be found in our response to Referee #2.*

8. P12774. Why would you get larger relative mean errors and variability in the validation period for precipitation and temperature? Could this be related to the setup of the methods?

RESPONSE

During calibration, the parameters are optimized to fit the observations for this period. It is, thus, rather common that the fit is not as good during validation with an independent data set. Therefore, the error and inter-RCM variability can increase. We do not see a problem here.

9. P12777, L1-3. I do not agree with this conclusion. There is very little difference between wet/dry years and cold/warm years over the observational time period, and I would hardly call this to push the bias correction to its limits. A full cross-validation over over regions with different climates would answer this question better.

RESPONSE

*The term was used, because the analysis focuses much on extremes that might not occur as extreme during "the real climate change". There are certainly other ways to push the bias corrections to their limits (or rather to "other" limits) and we will mention this in the discussion of the revised paper*

**References**


- Eden, J., M. Widmann, D. Grawe, and S. Rast, 2012: Skill, correction, and downscaling of GCM-simulated precipitation. J. Climate, 25, 3970-3984.
- Maraun, D. F. Wetterhall, A. M. Ireson, R. E. Chandler, E. J. Kendon, M. Widmann, S. Brienen, H.W. Rust, T. Sauter, M. Themessl, V. K. C. Venema, K. P. Chun, C. M. Goodess, R. G. Jones, C. Onof, M. Vrac, I. Thiele-Eich, Precipitation downscaling under climate change. Recent developments to bridge the gap between dynamical models and the end user, Reviews of Geophysics, 48 doi:10.1029/2009RG000314, 2010.
- Maraun: Nonstationarities of regional climate model biases in European seasonal mean temperature and precipitation sums. Geophys. Res. Lett. 39, L06706, 2012
- von Storch, H. (1999), On the use of “inflation” in statistical downscaling, J. Clim., 12(12), 3505–3506.
- Themessl MJ, Gobiet A, Heinrich G. 2012. Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal. Climatic Change 112: 449–468.


RESPONSE

*The mentioned references will be added to the revised manuscript.*