Hydrol. Earth Syst. Sci. Discuss., 11, C4745–C4755, 2014
www.hydrol-earth-syst-sci-discuss.net/11/C4745/2014/

Hydrology and Earth System Sciences
Open Access
Discussions

Interactive comment on "Inter-comparison of statistical downscaling methods for projection of extreme precipitation in Europe" *by* M. A. Sunyer et al.

M. A. Sunyer et al.

masu@env.dtu.dk

Received and published: 30 October 2014

Anonymous Referee #1 Received and published: 10 July 2014 The authors apply different bias-correction/downscaling (BC/DS) methods to RCM simulated daily precipitation time series in 11 European catchments. They evaluate how the methods differ both with respect to the agreement with observations in the control period and with respect to future changes, in both cases with focus on extreme values of duration between 1 day and 1 month. Overall limited differences are found with weak dependence on e.g. location and duration. Precipitation BC/DS is generally a key activity in hydrological climate change impact studies and evaluation/comparison of methods is an

important activity. The experiment is very comprehensive, spanning a wide range of climate projections, methods and catchments. The outcome must be worth sharing in the scientific community, but I think substantial revision of this manuscript is needed before publication, mainly for the reasons discussed in the following.

1. The overall conclusion is essentially that there is no best method but we must use many, and on an as large an ensemble of RCM-projections as possible. In reality, however, it is extremely rare to have such resources but the impact study must be limited to one BC/DS method applied to a small RCM-ensemble (or even just one projection, it is not unusual). Thus I think that a key objective of this kind of study must be to provide advice and recommendations for real-world, limited-resource impact studies. If precipitation extremes are the key focus, what should we do? With all these results and all these (prominent) authors it must be possible to provide more useful knowledge, conclusions and advice than what is currently the case, only "highlight the need for considering an ensemble" is not of enough help. For example different methods are to various degree prone to different problems like (1) cannot be applied/create unrealistic extreme values under some circumstances, (2) increases the bias in precipitation extremes under some circumstances, (3) modifies the climate change signal with respect to changes in extreme precipitation under some circumstances, (4, 5: : :). Further different methods are more or less prone to deviate from the rest in other aspects. A systematic review of this kind of key issues would provide very useful information.

- We have now extended the discussion on the selection of downscaling methods. It is difficult to point to a specific best method as it depends on several factors and the results might be different depending on the application, but we have pointed out some issues that arise from the use of some methods (such as BCM, BCMV, CFM,. . .), which implies that they should not be selected for some specific applications. Throughout the paper we have also extended the discussion regarding the ability of the models to correct the RCM outputs as well as preserve the climate change signal from the RCMs (as recommended in some other comments below). This helps to provide information on which methods should and should not be selected and why. We think that it is important to highlight the need of using (when possible) an ensemble of methods to be able to assess the uncertainty in extreme precipitation projections. We have described the main characteristics that the methods included in the ensemble should cover (different underlying assumptions, different RCM outputs, preserving climate change signal, . . .).

2. As I understand it the authors have calibrated the methods on the full set of reference data available and then applied to the same set of data (as well as the future-period data). However, I think this type of study needs to also include some kind of cross-validation analysis for historical periods, i.e. calibrate on one period and verify for another. This may not be possible for all catchments but in many there are some 50 years of data so you could split equally, divide 30/20 or something else. I think this kind of analysis is crucial for assessing the uncertainty when applying the methods to future periods.

- We agree that the validation of these methods is crucial. However, the proper validation of statistical downscaling methods is difficult and should be carried out with care. As recommended in several studies (e.g. Refsgaard et al. 2014; Teutschbein and Seibert, 2013), it should be done using data that has different properties to be able to assess whether the downscaling methods can be used to project climate changes. It would be possible to carry out validation analyses with the data available, but if the observational data do not show a pronounced change in extremes, then the results of the validation analyses would be questionable. The validation of statistical downscaling methods is very relevant and needed but it would require an almost new study in itself. A range of recent studies (e.g. Räisänen and Räty 2013; Refsgaard et al. 2014; VALUE cost Action http://value-cost.eu/) focus exclusively on methods and techniques to properly validate statistical downscaling methods. We consider that further research continuing this study should focus on the validation of the methods used here. This discussion regarding the challenges in the validation of statistical downscaling methods has been added to the manuscript.

Räisänen, J. and Räty, O.: Projections of daily mean temperature variability in the future: cross-validation tests with ENSEMBLES regional climate simulations, Clim. Dyn., 41, 1553–1568, doi:10.1007/s00382-012-1515-9, 2013. Refsgaard, J. C., Madsen, H., Andréassian, V., Arnbjerg-Nielsen, K., Davidson, T. A., Drews, M., Hamilton, D. P., Jeppesen, E., Kjellström, E., Olesen, J. E., Sonnenborg, T. O., Trolle, D., Willems, P. and Christensen, J. H.: A framework for testing the ability of models to project climate change and its impacts, Clim. Change, 122, 271–282, doi:10.1007/s10584-013-0990-2, 2014. Teutschbein, C. and Seibert J.: Is bias correction of regional climate model (RCM) simulations possible for non-stationary conditions?, Hydrol. Earth Syst. Sci., 17, 5061–5077, doi: 10.5194/hess-17-5061-2013, 2013.

3. I find the presentation of results hard to follow. Results from different periods focusing on different aspects are mixed under the headings "all catchments" and "selected catchments". I think the paper would benefit from being structured more like the bullets 1-4 on l 13-25, p 6181.

- We agree that the headings used for the different sections are not very helpful for the reader to identify what is discussed in each section. The sections are organized as the 1-4 bullet points in page 6181. We have renamed the sections in order to link their name with the bullet points. For example, the section 4.1 called "Statistical Downscaling methods" has been renamed to "Comparison of the downscaled time series for the control and future periods" so it matches the comparison in bullet point 1. Similar changes have been applied to the other headings in the results section.

4. Related, I think the paper would benefit substantially from more distinct objectives. Now they are rather vaguely and incompletely formulated like "comparing BC/DS methods" or "assess the changes in extreme precipitation". It would be better to formulate some distinct hypotheses to investigate. Focus on apparent current knowledge gaps that this study can help filling.

- We agree that the objectives were not clearly stated. We have now described more clearly the key objectives of this study and how they relate with the current knowledge gaps. The key objectives of this study are: to identify if there are general similarities and/or dissimilarities between the statistical downscaling methods; to assess whether there are common trends regarding changes in extreme precipitation over Europe; to assess the main sources that lead to the variability in the results.

The Results section is largely a textual description of the tables and figures with little interpretation or explanation or speculation of the reasons behind the findings. Sometimes it is speculated, but in those cases they could often have gone back to the results and for verification. See further specific comments below. Specific comments: 5. 20, 6169: It sounds puzzling that bias correction improves the agreement with observations only "in most cases", clarify that is concerns the extremes.

- The fact that some of the bias correction methods do not lead to an improvement compared to the RCMs is now discussed in more detail. See discussion in comment 22.

6. 13-23, 6171: You need to indicate also what relevant knowledge that was found in these studies.

- The main conclusions regarding the influence of different sources of uncertainty from these studies are now discussed in the text. These are: Bürger et al., (2013) concluded that the main influences on the overall results for different extreme indices were the downscaling method followed by the climate model used. Sunyer et al. (2012) and Hanel et al. (2013) highlighted that the influence of the statistical downscaling method used on the variation of the results is more pronounced in the case of extreme events (extreme precipitation in the case of Sunyer et al. (2012) and droughts in the case of Hanel et al. (2013)). Wilby and Harris (2006) concluded that the in the case of low flows the main sources of variation are the statistical downscaling methods (SDM) and climate models used. Lawrence and Haddeland (2011) showed that in rainfall dominated catchments, the uncertainty arising from the hydrological parameters was more

significant than other sources. But in snow dominated catchments, climate scenarios and SDMs were the main source of uncertainty.

7. 24, 6171: Is not comparing BC/DS methods rather the main focus?

- Yes, the comparison of statistical downscaling methods is one of the main objectives of this study. In addition, this study also addresses the main sources of variability in the results as well as comparing the magnitude and direction of the changes in eleven European catchments. See the comment 4 regarding the need for more distinct objectives. The different objectives of the study have been stated more clearly in the manuscript.

8. 3-6, 6173: Was the gridded data not based on observations?, if so how were they derived?

- The gridded data was based on observations. The information regarding the methods used to obtain the gridded data is given in the references specified in Table 1. It has been clarified in the text that the gridded data is based on observations from station data.

9. 2.1, 2.2: What was the volume resolution in the observations?, 0.1 mm?, was the same cut-off used for the RCM data for consistency?

- The cut-off value in the observations varies depending on each catchment. The cut-off values were not applied to the RCMs as it is not relevant for the study of extreme precipitation. Nonetheless, several downscaling methods include a threshold of wet/dry days which addresses the issue of the cut-off values (in addition to accounting for changes in the frequency of wet/dry days in a future climate). The different cut-off values of the observations and the fact that they are not used in the RCMs is now discussed in the text.

10. 5, 6174: Rummukainen is misspelled and not in reference list, check carefully before submitting, should not be my task.

- Corrected.

11. 10-12, 6175: I think harmonisation would have been much better, why not used?

- This paper is a result of a coordinated effort where different research groups contributed with the different methods that they use. We agree that harmonisation would have been nicer but the methods were applied as previously used in other studies by the group. The reason for the approach used is now mentioned in the manuscript.

12. 22-23, 6176: What if there are more wet days in OBS than in RCM?

- This is not the case for any of the RCMs when the bias correction is applied to the full time series, without distinguishing months or seasons, as has been done in this application.

13. 19, 6177 – 5, 6178: As acknowledged it seems questionable to train on RCM-ERA and then apply on RCM-GCM without any correction. Can you provide some information on the additional error introduced?

- With respect to the downscaling, the XDS model is optimally trained given the available RCM data. With respect to the RCM data, however, better RCM realizations could be achieved by a second data assimilation with the RCM-ERA-40 runs (beyond the data assimilation which is already done for the ERA-40 reanalysis) in a similar way as done e.g. in the North American Regional Reanalysis (NARR) (Mesinger et al., 2010). The additional error introduced when applying XDS to RCM-GCM simulations cannot be quantified. For clarification we have added the following sentence to p.6177, l.26: "A second data assimilation with the RCM-ERA-40 runs (beyond the data assimilation which is already done for the ERA-40 reanalysis) would overcome this problem to some degree. However, such runs are not available for the RCMs accessible from the ENSEMBLES archive."

Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P. C., Ebisuzaki, W., Jović, D., Woollen, J., Rogers, E., Berbery, E. H., Ek, M. B., Fan, Y., Grumbine, R., Higgins,

W., Li, H., Lin, Y., Manikin, G., Parrish, D. and Shi, W.: North American Regional Reanalysis„ 2010.

14. 9-12, 6179: The intervals are very small (and different from BCQM). It is well known that there may be an enormous variation in the very highest quantiles (above 99 or so), I think it needs to be demonstrated that using 0.0005 works well, instead of smoothing out the fluctuations a bit with larger intervals.

- The probability intervals are estimated using linear interpolation from the empirical quantiles, so as in the case of CFQP and BCQM the fluctuations in the CFs are due to the fluctuations in the empirical quantiles. The possibility of large fluctuations in the changes estimated using empirical quantiles is a disadvantage of using empirical methods as used in this study, but empirical methods have the advantage that there is no need to fit a distribution to the data. In methods where distributions are fitted changes would be smoothed; however, with a risk to over-smooth the changes and introduce a bias in the estimated quantiles. It should also be noted that in most catchments the results of CFQM and CFQP are virtually the same, and CFQP uses different intervals than CFQM (but also estimated from the empirical quantiles). This discussion on the problem of fluctuations in the CF has been added to the text.

15. 15, 6185: What is a "threshold return level"?

- Corrected to only "thresholds".

16. 20-27, 6185: Assessing to which degree BC/CS modifies the CC signal is important and needs more attention. More analysis and interpretation is needed. What is the change increased after BC/DS? What is the reason for the regional changes? Dig in the results.

- We agree that it is important to assess whether the difference in the change projected for extreme precipitation using the downscaled time series differs from the change projected by the uncorrected RCMs. This has now been discussed in more detail using the overall results for all the catchments (results in Figure 3) and in detail for each downscaling method for three of the catchments (results in Figure 4). Overall the change in signal introduced by the downscaled series is not significant compared to the change projected by the RCMs, but some SDMs (CFM and in some cases XDS, BCM, and BCMV) tend to lead to different values than the ones obtained from the RCMs. These methods do not specifically correct or take into account the changes in extremes which might lead to differences in CC signal. On the other hand, the methods that are expected to better account for changes in extremes tend to show more similar results to the ones obtained for the RCMs, especially the BCQM method. It is now discussed in the paper (in the description of the methods and further emphasised in the discussion of results) how the different methods are suited for handling changes in extremes. There are not specific regional differences in the change of the CC signal from the downscaled series and the RCM series, except the fact that larger differences are obtained for the catchments in Turkey and Cyprus, as already shown some downscaling methods do not work well in these catchments.

17. 7, 6188: “most likely” – check in the data

- The large EPI value found for this method in winter and in the Mulde catchment is due to the influence of two very large extremes created in the correction of the RCMs for the future. These are two events of 60 and 55 mm/day. From the BCM and BCQM the largest values obtained are approximately 40mm/day and the largest value obtained for the control period is 27mm/day. This is now described in a more clear way in the manuscript. In addition, the fact that BCMV may lead to unrealistic results is discussed in the selection of downscaling methods in the summary and conclusions section.

18. 21, 6188: “might be similar” – check in the data

- We have checked that there are very few rainy days in the catchment in Turkey as in the case of the catchment in Cyprus, only approximately 20% of days are rainy days. This has been added to the text.

19. 26-27, 6188: Why expect a larger impact in TR in summer?

- Because in this catchment and in summer there are only very few rainy days. This implies that in some cases all the rainy days in a season are included in the selection of extreme events. Therefore, in this catchment the change in the number of wet days may have an effect on the changes in extreme precipitation. CFQM and CFQP differ in the way that CFQP accounts for the change in the number of wet days. This discussion has been added to the manuscript.

20. 4.2: I think it would be more natural to have 4.2 before 4.1, i.e. first an evaluation in reference period and then results from future period. - We prefer to keep 4.2 after 4.1. We consider that the use and comparison of the 8 statistical downscaling methods is the main part of this study. 4.2 considers only four of the methods, and is an additional analysis which is only possible for the bias correction methods.

21. 13, 6190: Not “extreme value index” but only “extreme values”, right?

- Corrected.

22. 5-15, 6191: Key paragraph which is far too compact. Only listing values is not very helpful; go much further, dig in the data, find out concretely why and under which circumstances the BC/DS methods decrease agreement with observations.

- We agree that this is an important point to highlight. In order to address it in more detail we have discussed the issue from the summary of the results for all the catchments (Figure 6) and then in detail for three of the catchments. The results obtained from the RCMs have been added to Figure 7 in order to make this discussion possible. In general, in the catchments where the agreement between the observations and RCMs is good the downscaling methods used in this study are not able to improve it. In general, the simple BCM method tends to fail in improving the uncorrected RCMs more often than the other methods, but this should be tested for each case as it depends on the catchment.

23. Table 3: Add the methods' abbreviations.

- The abbreviations of the downscaling methods have been added to Table 3.

24. Table 3: That seasonality is not taking into account in BCQM is an application issue, not a disadvantage of the method.

- This has been clarified in Table 3. In addition a note has been added to the table's caption to point out that some of the advantages and disadvantages are specific to the application, which is now indicated in all cases in Table 3.

25. Table 3: CFQP: ACF can be checked for all methods.

- This has been clarified in Table 3. See comment also reply to comment 24.

26. Fig 5: Add legend, referring to other figure is not sufficient.

- The legend has now been added to both Figure 5 and 8.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 11, 6167, 2014.