Interactive comment on "Indirect downscaling of global circulation model data based on atmospheric circulation and temperature for projections of future precipitation in hourly resolution" *by* F. Beck and A. Bárdossy

F. Beck and A. Bárdossy

ferdinand.beck@iws.uni-stuttgart.de

Received and published: 14 October 2013

Thank you for your detailed and helpful comments. Based on your remarks and suggestion, we revisited the whole paper. The method section has been extended to make it clearer in the critical points (e.g. regarding the CP definition) that you mentioned. Most of all, we reformulated parts of the introduction and the conclusion, being less affirmative about the results of this study. In the new version we point out that this is a first attempt to test a new downscaling method and that further research should be

conducted including data from other scenarios and other GCMs.

In the following you find the responses to your questions one by one and how we incorporated them in the new version of the paper draft:

1. Why using NCEP data instead of the new ERA-interim d ataset extended back to 1987 that offer a better resolution. Is there a reason of preferring NCEP dataset?

The spatial resolution seems not to be crucial for the CP definition. The classification was also tested with NCAR sea level pressure of 5 x 5° resolution (https://climatedataguide.ucar.edu/guidance/ncar-sea-level-pressure), which lead to similar results, not only concerning the pressure maps of each CP, but also the precipitation response. To our experience, any gridded SLP data is suitable for the classification as long as the resolution is fine enough to represent the cyclonic and anti-cyclonic systems which are the main features of the CPs. Since the spatial resolution does not affect the CP definition very much, the choice for NCEP reanalysis was made for more practical reasons: The first reason is that NCEP offers a long data archive, which enables reliable comparisons with GCM data and trend estimations. The second reason is related to the stochastic optimization scheme that generates the CP definition: Any optimization has to test a certain (very small) portion of all possible combinations to choose the best one. In our case, the optimization is based on the grid points of the reanalysis and the associated five possible fuzzy pressure states. Therefore, every additional grid point induces 5 times more possible combinations. A twice as fine spatial resolution is equivalent to four times more grid points and therefore raises the number of possible combinations by 5 to the power of 4 times the original number of grid-points in the coarser grid. This means that the risk of missing the optimal classification increases with every refinement. (On a 2.5° x 2.5° grid, the optimization runs already for several days on a standard PC). Finally, the fact that the resolution of the reanalysis is close to the resolution of the GCM assures the transferability of the derived fuzzy rule set without significant interpolation errors. We are not sure whether these considerations are essential for the paper. We opted to include a short notice that the NCAR

SLP- Data set leads to very similar results.

2. CPs are extracted on the 1960-2003 period and the validation with rainfall is made on the 1991-2003 period. To my mind, results would be more robust if you compute the CP only on the 1991-2003 period. Otherwise, you might show that CPs are similar if you consider 1960-2003 and 1991-2003 (but It's not sure).

The rainfall data is not used for validation of the CPs. Instead, it is used in the set-up of the CP definition. This means that the CPs are based on MSLP and precipitation data from 1991 to 2003. The longer time interval of the NCEP/NCAR data set from 1958 to 2003 is only used for the analysis of the occurrence frequencies according to the derived CPs and not for any precipitation related validation. We rewrote and extended the method section concerning the CP definition hoping that this becomes clearer in the new version. It is the available precipitation data that limits the calibration period to the years from 1991 to 2003. We would have wished to use a longer time period, but this was not possible since the optimization that sets up the CP definition requires a homogenous set of precipitation data. This means that the number and location of the stations must be the same in every year and that the time series must be close to complete. If not, it would induce an implicit weighting between years (or seasons) with many available observations and the years with very few. In total we had time series from 292 rain gauges ranging from 1958 to 2003. We tested, if the precipitation response for the whole time series is the same as for 1991 to 2003 period. Overall, we found very similar characteristics. However, the results can only be compared qualitatively. The complete data sets exhibits an average of about 30% of missing values at number of available stations ranges from under 40 in the mid-90ties to more than 100 in the last years, which makes implicit weighting between the years and the seasons (missing values occur more frequently in summer months) very likely. Based on our data, we cannot judge if the deviations are due to changes in the CP's precipitation response or due to statistical effects of the inhomogeneous data set. Therefore, we did not include the results for the whole data period in our paper.

C5601

3. I don't really understand if you compute CPs on the whole year and then consider the summer and winter or if you compute CP separately for winter and for summer. This is very important because circulations patterns are seasonally different and it's usually better to compute CP separately for winter and for summer.

The CP definitions are the same for the whole year. The precipitation response, however, is calculated separately for September to April and the summer months from May to August, when convective events are most likely to occur. In the objective function O the chi-square values of the two seasons are combined as a sum. Some CPs (with westerly flow field) show similar precipitation during summer and winter months, some differ very fundamentally in their precipitation response in summer and winter, as it is explained shortly in the text. If there are CPs that can only occur during one season, as you suggest, we did not find them with our method. The optimization setting up the CP definitions, is not restricted regarding the occurrence frequency of the CPs during the two seasons. If the optimal classification required CPs that only occurred during summer or winter, it would generate them. This was not the case. Nevertheless, the frequencies of the CPs are different in summer and winter months, e. g. high pressure situations are more likely during summer. To avoid confusion regarding this point, we state more clearly in the new version of the method-section that the CP definition is constant.

4. Why using temperature for subdivision? Is there a physical reason? Could you explain why? Moreover, what is the added value of a subdivision with temperature? It would be nice to compare results with CP derived from SLP only and then show the added value of the subdivision with temperature.

According to the research stated in the introduction (e.g. by Trenberth and Lenderinck) temperature governs both potential evaporation and the moisture capacity of the atmosphere. Therefore, it has a strong influence on precipitation, especially on short aggregation intervals. Since GCM state in agreement that the atmospheric temperature will rise in the future, we suggest that it will most probably affect short temp precipitation

C5602

intensities. Surface temperature is used as a first approximation of the atmospheric temperature. We revisited the paper to improve the link between the introduction and the section dealing with the temperature subdivision. We hope that the physical justification for using surface temperatures becomes clearer in the new version. The added value of the temperature subdivision is in the difference between the overall CP precipitation response and the response in each temperature sub-class. The overall response is simply the average over all five temperature classes, since classification is based on quantiles. If the temperature subdivision would not add any value, the subclasses should all exhibit the same precipitation frequency and the same frequency of extreme precipitation. Thank you for the suggestion to compare the results with and without temperature subdivision. We added some new results, repeating the calculation of the expectations shown in Fig. 10, considering the CP of the day, but ignoring the temperature level. The comparison between the results with or without temperature reveals that the predicted increase in extreme precipitation is missed if only the change in CP sequence is considered. Overall, the use of surface temperature as classification criteria reveals a significant effect. It could be tested in subsequent studies if other temperature measures (e.g. reanalysis of higher atmospheric levels) could improve the result by quantifying effects that daily surface temperature is ignoring, e.g. latent heat flows.

5. Why considering 12 CPs, is there a physical or a statistical reason?

The number of CPs has to be chosen manually before the optimization. In this respect, the classification is subjective. In Bárdossy 2010 the optimal number of CPs was defined by repeating the classification with varying number of CPs. By different precipitation related performance measures, it was found that the optimal number of CPs is between 16 and 20, depending on the objective function. Nevertheless, we decided to reduce the number of classes here so that we still have enough values in each temperature subclass for valid statistical analysis. The criteria why we reduced to 12+1 CPs were that beyond twelve, the frequency of days that do not fit in any class becomes much higher. In the new version of this draft, this explanation is included as

a paragraph in the method section.

6. Using only one GCM and/or one run for future is not very robust. E.g. many other GCMs and runs are available (for free) in CMIP5. To my mind, it could be nice to use other GCMs (I know that it means a lot of extra work and computation but it could really improve the paper)

Since this is the first attempt with this new method, we restricted the study to one scenario from one GCM. On the long run, the method should indeed be tested with other scenarios and most important other GCMs. Unfortunately, for the moment we don't have the resources to conduct further research on this topic. Therefore, we have reformulated the paper, especially the conclusion to make it clearer that this is the first test of a new downscaling approach. One might see it as a case study which proved that the method leads to physically sound results. These results, however, have to be verified by subsequent research including other scenarios and data from other GCM.

7. Did you validate the spatial structure of the CP extracted from ECHAM? It would be interesting to show the spatial structure of the CPs computed with ECHAM in order to validate if the GCM is able to well reproduce the CPs you computed with NCEP.

The average anomaly pressure maps can hardly be distinguished, whether they are derived from the NCEP data set for the calibration period from 1991 to 2003, the complete NCEP time series since 1948 or the ECHAM5 data sets. This means that ECHAM5 is able to reproduce the same CPs and that the deviations between GCM and reanalysis only concern the frequencies of the CPs. A map of average pressure anomalies according to the ECHAM5 20th century pressure data has been added to illustrate this.

8. At the end, It's a bit strange because you show that trends is CPs frequency are not well simulated by ECHAM on the 1960-2003 period. If the model is not able to simulate observed trends, it's not possible to use the model to analyze and discuss future trends, right? Once again, using several models or an ensemble mean could make the result more robust and could also change the result. I don't know how to express it clearly

but I don't really understand what is the aim of analyzing and discussing the future if the model is not really able to simulate the present.

We considered your suggestion and reformulated the conclusion in a more careful way to point out the uncertainties in the results. Strictly speaking, any GCM is wrong, especially regarding precipitation. If we want to make a prognosis, we can only accept and communicate these errors. Nevertheless, we still think that the results are worth publishing because they agree with physical considerations and observed trends during the last decades. The new version of the paper draft ends with the following statement: "In perspective of the potential model errors and stochastic variability, this study can only be a first step. It proved that the proposed downscaling scheme is able to create physically sound results. These results, however, have to be verified by subsequent investigations including other scenarios than A1B and data from different GCM."