

## ***Interactive comment on “The representation of location by regional climate models in complex terrain” by D. Maraun and M. Widmann***

**D. Maraun and M. Widmann**

dmarau@geomar.de

Received and published: 8 June 2015

We acknowledge the very helpful comments by all three reviewers. Please find detailed comments and suggestions for revisions regarding all major reviewer comments. In general we agree with most minor comments made by the reviewers and will address them accordingly.

### **Comments by Reviewer 1**

*The main result I got from looking at Figure 1 and Figure 3 and 4 is that in many cases in winter (DJF) local representativeness is high and that non-local bias correction can (only) improve the correlation in some small parts over Europe (Figure 3, center). I agree that this problem is generally linked to leeward mountain areas in winter*

C2010

*and summer, but not all leeward areas are affected (e.g. Scandes). This is good news! This means that bias correction methods can be applied to many parts in Europe without the necessity for non-local correction. For sure, in summer this is different due to the internal climate variability and climate model limitations. Still, the improvement by non-local correlation is clear but spatially limited. Hence, I would prefer a less “dramatic” argumentation in the discussion section on the need for testing local representativeness. Even more as climate change signals are not systematically deteriorated, but due to model limitations and internal climate variability.*

We agree with the reviewer that indeed for much over Europe, RCM simulations are not systematically displaced and may be further bias corrected (at least when driven with reanalyses). However, our main intention was to point towards the critical cases in mountain regions. In the revised version, we will therefore highlight the overall good performance of the chosen RCM, and point out the need to test for location representativeness in complex terrain.

*First, looking at the climate chart for “Domodossola” one sees that DJF is not a very relevant season for this location as it is mainly influenced by southerly flows in spring and autumn, as well as convective rainfall in summer. The observed correlation to the northwards grid box is hence more an exception than the typical situation. (Not to speak about the similarity in the correlation to the dry inner-alpine and very humid grid cells in Switzerland – indication the wrong representation of topography and weather within the climate model). Maybe observations in the Aosta valley (approx. one or two grid boxes away from the author’s choice) might be more appropriate.*

We appreciate the reviewers comment. In fact, the reviewer suggests to investigate other seasons as well. Based on the results for all seasons, we will select a suitable

example.

*Second, the longitudinal cross section irritates me as it crosses three different precipitation regions. Why not have a latitudinal cross section if you intend to investigate the lee effect – at least in addition. Hence, I suggest to use a slightly different grid box that more clearly expresses leeward effects and to extend the analysis to all seasons. The latter would also allow discussing the need for temporally changing representativeness.*

Depending on the grid box we will select as new case study region, we will adjust the cross section to best illustrate the lee effect. And, indeed, it is valuable to look at the transition seasons as well: both to assess the strength of synoptic forcing as well as potentially season dependent representativeness (as is likely, given the example discussed by the reviewer).

## **Comments by Reviewer 2**

*Effective climate model resolution and spatial averaging of climate model output: In the context of the presented manuscript, the issue of the effective resolution of climate models (e.g., Grasso BAMS 2000) is very relevant and should at least briefly be discussed. There's an ongoing debate in the climate model community whether climate model output for variables such as precipitation should really be analyzed at the grid cell level or if a spatial averaging is required prior to the analysis. Doing so could potentially improve the identified deficiencies in the local representativeness of climate model output. Furthermore, some approaches exist that indeed spatially average RCM output before applying it to impact models (e.g., Bosshard et al. HESS 2011). In the light of the presented results, would such approaches be favourable?*

C2012

We acknowledge the point made by the reviewer about effective climate model resolution, and will include a discussion in the revised version. Yet, the spatial averaging of neighboring grid boxes as suggested by the reviewer is typically done to improve the signal-to-noise ratio. Some authors even pool neighboring grid box values to increase the signal to noise ratio (e.g., Kendon et al., J. Climate, 2008), i.e., they interpret the data at the grid box scale. Note also that temporal averaging, similar to spatial averaging increases the signal-to-noise ratio as well. We consider (temporal) seasonal averages, which should be well interpretable at the grid-box scale. Anyway, the effect we observe would only be smoothed out by spatial averaging, it would not be removed. In fact, one might argue that averaging across a major mountain ridge might cause additional problems as it combines two different climatic regions into one average.

*Measure for location representativeness: Only one method for quantifying the location representativeness (temporal correlation of seasonal precipitation sums) is presented. However, other concepts could be thought of but are not discussed. One measure might for instance be the representation of the seasonal or daily precipitation PDF after correcting for a mean model bias.*

In the manuscript we argue explicitly that comparisons of (marginal) pdfs are NOT suitable to identify location representativeness problems. We argue that a wrong (marginal) PDF does not imply that a grid box is not representative. In contrast, we even argue that a wrong PDF is not a problem as it can in principle be corrected by bias correction, if the grid is representative. A low representativeness is caused by a mismatch between the observed and simulated local weather and the large scale flow (which is typically correctly simulated by the RCM). The easiest way to test for such a mismatch is by comparing the observed and simulated local weather sequence,

C2013

and thus by correlations. Note that an RMSE would not be a useful measure, as it would penalise errors in PDFs. One might, however (we will discuss that below), think of conditioning the precipitation fields on weather types, and consider the mismatch between the observed and simulated fields as an alternative measure of local representativeness.

*The cross-validation setup remains unclear to some extent, this section needs to be clarified. Apparently, the entire period is divided into four sub-periods, the most representative grid cell is determined for each combination of three periods and this time series is written into the remaining fourth validation sub-period. There's some danger that the resulting time series used for validation will consist of time series from different contributing grid cells, if the identified most representative grid cell changes from one 3x10/11 year block to another block. In my opinion this would be a shortcoming of the entire method as the spatial representativeness pattern is not stable in time. The authors should better clarify this point. Also, it would actually be nice to assess this temporal stability, i.e.: To what extent does the most representative RCM grid cell for a given observational grid cell depend on the analyzed time period?*

This is exactly the idea behind our cross validation. To eliminate artificial skill, one would select the most representative grid box based on a calibration period, and then assess the actual representativeness of that grid box by calculating the correlation in a validation period. This is what we do, only in a 3-fold cross validation. For the approach to be meaningful, the most representative grid-box has to be chosen individually for each fold. This does not necessarily imply that the (true) most representative grid box changes in time - the best estimate might in fact vary in time because of sampling errors. Consider, e.g., figure 2. Two neighboring grid boxes close to Lausanne show very similar correlations with the central observational grid box. In the cross validation,

C2014

there is the possibility that for some fold one is randomly chosen, for another fold the other. In the revision, we will explain the cross validation setup more clearly.

*Deficiencies of the gridded observations: It is well known that the EOBS dataset has deficiencies due to a rather coarse underlying station network. This is also true for the region of the European Alps. The spatial representativeness pattern shown in Figure 2 does therefore not necessarily result from a climate model problem in representing spatial precipitation variability, but could also result from deficiencies in EOBS.*

In principle, the reviewer is right, and we will discuss the point in the paper. We believe, however, that this effect is very unlikely, given the systematic pattern along the main ridge of the Alps. If the reviewer were right, this would imply that across the whole Alps, just to the South of the major ridge no observations would be available, but all along the northern side.

*Looking at Figure 2 (right panel) I'd argue that temporal trends in seasonal precipitation sums are very low and to a large extent masked by interannual and decadal climate variability. This is supported by the fact that the pattern shown in Figure is rather noisy and doesn't actually present systematic results. In my opinion this questions the entire analysis of trends in the present manuscript.*

Please note that we are not discussing the causes of the observed trends! Of course the time series contain a large fraction of internal climate variability. But for both RCM and observations this noise is synchronised. What we observe is that the long term trend fitted to these noisy time series (be it a forced signal or just random)

C2015

is improved during winter for most regions where a distant grid box improves the location representativeness at least for winter: green regions very much outway brown regions in figure 4, left panel. Thus, a non-local bias correction not only improves the interannual variability, but also long term trends. For summer, again, long term trends seem to be governed by RCM internal variability, which is not in sync with observed internal variability. Note the difference between winter and summer here: in winter, the internal variability is of synoptic scale and therefore synchronous between observation and simulation; in summer it is of small to mesoscale and therefore not synchronous. We will add a further discussion to the paper.

*In case the authors have a good reason to keep it in, it would at least be interesting to know if the most representativeness climate model grid cell for a given observational grid cell (determined according to the correlation measure) would also correspond to the grid cell with the best agreement of the seasonal precipitation trend (which I'd doubt very much).*

It is true that we only show that trends are improved, but we do not show whether they might be further improved by choosing a different grid box in the 11x11 field we consider. We argue, however, that any such result would not be very robust, and therefore not defensible: as trends are only one number, the probability is high that a distant grid box just by chance has a very similar trend.

*In other words: If a most representative RCM grid cell is identified in a reanalysis-driven setup, this might probably not be the most representative grid cell if the RCM is driven by some GCM. Furthermore, the identification of representativeness using the concept of temporal correlation is not possible in a GCM-driven setup. What would be*

C2016

*the solution to this? The authors should at least discuss this point.*

The reviewer raises a very serious and valid point. Representativeness should first of all depend on the major flow direction - as a systematic dislocation is caused by an interaction of the main flow with the wrongly represented topography, one can assume that for the same flow, one would get the same non-local representativeness pattern. A direct transfer of the most representative grid box based on our correlation measure to GCMs implies that the distribution (and pattern) of flow types is basically the same in the observations and the GCM. The reviewer is right that this assumption is questionable. However, one could in principle circumvent this problem: one might define flow types in the observations, determine the most representative grid box for each flow type, and then transfer the result to the same GCM simulated flow types (i.e., one has to define identical flow types in GCM and observations, e.g., by common EOFs). Such an extension of the method, however, would require extensive further analyses, and therefore beyond the scope of this paper. Nevertheless, we will add this discussion to the paper.

*One way to get out here might be the application of a different concept of defining spatial representativeness that could also be applied to GCM-driven setups (e.g. based on the marginal distribution).*

Again, we are convinced that a comparison of marginal distributions would not help (see above).

*Figure 3 shows the non-local representativeness as opposed to the local representa-*

C2017

tiveness in Figure 1. While correlations are generally higher in Figure 3, some regions can be identified where non-local representativeness seems to be LOWER than local representativeness (e.g. some parts of Finland have a larger DJF correlation in Figure 1 than in Figure 3). This means, that the application of the non-local concept actually deteriorates the representativeness over these areas and does not maximize it. Is that right? And if so, why?

The reason is our cross validation. For local representativeness, no cross validation is possible. For some regions, if the improvement is only marginal (or zero), sampling variability will - in the cross validation setup - cause lower correlations for non-local representativeness than for local representativeness. We will add a short explanation to avoid confusing the reader.

### Comments by reviewer 3

*My main criticism is related to the measure of representativeness. The authors use in this study the Pearson correlation coefficient. This coefficient is a measure of linear dependence. It is very sensitive to deviations from a linear relationship (Wilks, 2011). The authors do not mention these limitations in the concept and data section and also do not discuss alternative measures. Alternative measures would be measures of rank correlation such as Spearman rank correlation or Kendall tau, which provide a more robust estimate of dependence. This point is of particular importance and needs to be carefully addressed because state-of-the-art bias correction methods such as quantile mapping do not rely on the assumption of linearity, but instead on the assumption that ranks are comparable.*

As we consider seasonal total precipitation, our conclusions are robust and basically

C2018

independent on the correlation measure. In the (relevant) winter season, the central limit theorem kicks in, and deviations from Gaussianity are minor. Even more, linearity should be of minor importance, as long as the relationship is monotonous. Moreover, the generally very high values, often close to one, indicate that deviations from linearity are minor. Anyway, we will compare results for Spearman and Pearson correlations and, based on the outcome, may consider using Spearman correlations.

*The method is intended for bias correction of RCM data (p. 3020, l.1ff). The final goal of bias correction is to be applied during future periods. For this reason, stationarity of the relationships exploited for the bias correction needs to be assumed. The authors do not address this point. I suggest to discuss why the same non-local grid cell should be representative during future periods in a revision of the manuscript.*

Similar to the argument made by reviewer 2, this is a very important point. We would again argue that location representativeness would depend on the flow direction. To make the method as robust as possible, one could define flow types in the observations, determine the most representative grid box for each flow type, and then transfer the result to the same GCM simulated flow types (i.e., one has to define identical flow types in GCM and observations, e.g., by common EOFs). Such an extended version of our method is beyond the scope of this paper, but we will add a discussion.

---

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 12, 3011, 2015.

C2019