Point-by-point response to Reviewer #2

J. J. Gómez-Navarro on behalf of all co-authors

May 15, 2018

Anonymous Reviewer #2:

This paper presents a bias correction method for regional climate simulations over the Alps at very high resolution. A observational database for the region is used for the validation, and ERAinterim and GCM-CESM forcing fields are used to WRF modelling work. To my opinion, it shows enough aspects to novelty and adequate analysis and understanding of the obtained results. I suggest it to be considered for publication, once the questions and requested item can be properly answered or at least taken into account in some way

We thank the reviewer for the time devoted to carefully read the manuscript, the positive vision expressed about it, and the constructive comments that will certainly improve the final version. We have tried to answer point by point all his/her comments below.

1. Missing references. It is always the case that not all the relevant references are included when a work is presented. Here I find some that I consider that are essential to be included, not only for the introductory aspects, but also for the methods and results description. Let me indicate them to the authors for them to be considered a properly used throughout the text

The new version of the manuscripts includes many more references, including most of those suggested by both reviewers, and that clearly allow to better contextualize this piece of work in the existing literature.

2. Apart from the pure bibliography missing items, there are some aspects that could be more deeply described by the authors. One of them should be to compare the proposed bias correction method with other similar ones, if there are some, to see more clearly differences and similarities with others already proposed. I am sure the quantile mapping procedures have been used before, if one goes to those references. Therefore, I recommend the ongoing work by Nikulin and others in the frame of EuroCORDEX activities, named BCIP. Take a look at this abstract at EGU2015: Nikulin, G., Bosshard, T., Yang, W., Bärring, L., Wilcke, R., Vrac, M., ... & Fernández, J. (2015, April). Bias Correction Intercomparison Project (BCIP): an introduction and the first results. In EGU General Assembly Conference Abstracts (Vol. 17). In a more general sense, perhaps a mention to this recommendation by CORDEX community could be made. take a look at http://cordex.org/data-access/bias-adjusted-rcm-data/, and from there, to a IPCC work focused on this topic: See Breakout Group 3bis: Bias Correction (pp. 21-23) in IPCC, 2015: Workshop Report of the Intergovernmental Panel on Climate Change Workshop on Regional Climate Projections and their Use in Impacts and Risk Analysis Studies [Stocker, T.F., D. Qin, G.-K. Plattner, and M. Tignor (eds.)]. IPCC Working Group I Technical Support Unit, University of Bern, Bern, Switzerland, pp. 171. (https://www.ipcc.ch/pdf/supporting-material/RPW_WorkshopReport.pdf). I can imagine that authors do not want to go too far on this aspect, but I think that some more comments, to have this work inside the wider context, should be made. Even a mention to some developed software for this kind of analysis could be included, such as Bedia, J., Iturbide, M., Herrera, S., Manzanas, R., & Gutiérrez, J. (2017). downscaleR: an R Package for Bias Correction and Statistical Downscaling. R Package Version 2.0-3.

We have included a discussion of the issues raised by the reviewer about on-going debate about the use of bias correction, including the mentioned references in the new version of the manuscript.

3. I am not sure if the authors have a comment about the fact that this bias correction method has been applied to a region with a very deep orography, and to precipitation field. Which could be the potential to apply it to other regions with smoother orography, and/or to other variables?

We have added a new paragraph in the conclusions to discuss how this method can be exported to other regions/variables. We reproduce it below:

We note that the rationale of the developed methodology is to divide a large domain into smaller subregions according to the behaviour of the target variable. We have applied it here to daily precipitation in Switzerland for being a variable strongly affected by complex orographical details that lead to strong horizontal gradients. With more generality, spatial regionalisation is an efficient method to break down complexity in areas and variables whose behaviour strongly varies through the domain. Still, the bias correction applied separately to subregions can be in principle adapted to other cases with simpler topography, or other variables with lower horizontal gradients. The only practical difference is that in this case the regionalisation will naturally lead to a lower number of subregions which are necessary to obtains clusters with coherent features.

Specific comments

1. It has been indirectly mentioned on the general comments section, but here I want to comment if explicitly: I miss a mention to the EuroCORDEX/MedCORDEX activities, that have used plenty of simulations at high resolutions (0.11) over Europe, and several studies with not a single RCM as here, but an ensemble of them, that have analyzed, also forced with ERAinterim fields, how precipitation is described. I do not mean a full comparison with other RCMs, but at least some mention and comparison with them, to see more clearly if WRF-RCM is similar to the state-of-the-art RCMs modelling alpine precipitation for current climate conditions.

We have added plenty of references and explicit mentions to EURO-CORDEX and MED-CORDEX activities in the newer version.

2. And also related to this point, I miss some comparison of your figure 5, for example, with figure 2 of Torma et al., 2015 or Fantini et al., 2016, figure 5, not only for RCMs, but also for observational datasets, I am not sure if they are totally consistent. Or for your figure 6 and 7, and their corresponding figures.

This is an important pitfall that has been corrected in the new version. We have enlarged the discussion of the results in Section 4.2 with explicit mention to the ones in the corresponding figures in the two sources pointed out by the reviewer. However, we have not included a discussion comparing the results about daily PDFs in Fig. 7. The reason is that the daily PDFs in those references include all seasons, and they are built to emphasise the different results across spatial resolutions. Therefore they are somewhat different figures, and it is difficult to stablish a fair and meaningful comparison.

3. I have a concern about the domain of study chosen here. On figure 1, D4 subdomain seems to be the one used for the analysis, but then figures with the political borders of Switzerland seem to be used. This relatively artificial borders could add some non-physical or modelling aspects to the analysis, and specially when obtaining the subregions from the clustering procedure. Which is the opinion of the authors about this aspect?.

This confusion between the simulated and analysed domains (the Alpine region vs. Switzerland) has also been pointed out by reviewer #1. We believe we were not clear enough in the former version in the description of the dataset and the methodology. Therefore we have clarified this in the newer version of the manuscript. The reason for using Switzerland, i.e. a political boundary, instead of a natural one, is that the observational product we used is limited to this domain. This imposes a unavoidable bottleneck of the validation. Certainly there are observations beyond the borders of Switzerland, but we believe that they do not contain the high density, even distribution, and quality-tested of the observations blended to create the gridded product developed by Meteoswiss.

4. Another point I would like to hear from the authors is about the very high resolution used for the WRF D4 domain (page 5, line 14): 2km. Which one is the real advantage here of using such resolution compared with the even-very-high 6km one?. It seems that no much mention or usefulness is made by the authors to this resolution, by far much larger than the mentioned 0.11 "high resolution" EuroCORDEX standard values these days. It is also a tricky aspect, since the comparison and bias correction method is made against the roughly 20km observational dataset information, and so some statements are made through the text related to this resolution differences. A more complete study should perhaps include at least some other resolution from the WRF model to a better understanding of the resolution topic?.

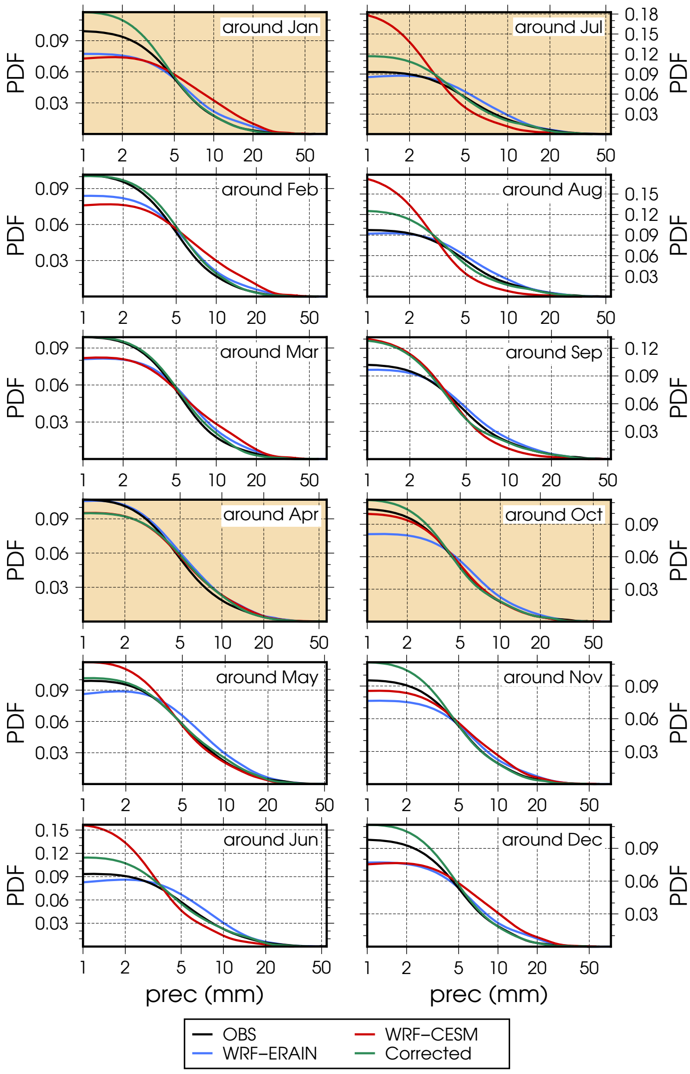
We have tried to motivate the added value of the high resolution. In particular, the fact that we can avoid the use of parametrisations of convective processes, and we provide references that back the added value of such simulations. The difference between 2 km and 6 km can be substantial. For instance, in response to reviewer #1 we have discussed the effective resolution. If we use the factor 3-4 mentioned in some references (e.g. Pielke Sr, 2013), 6 km of spatial resolution would have an effective resolution clearly above 10 km, which can be argued that it is not sufficient to account for all convective processes. This can be hardly put in doubt with a resolution of 2 km. More precisely, (Gómez-Navarro et al., 2015) investigated the particular issue of the skill as a function of spatial resolution, and found that there is indeed a large gain in switching from 6 km to 2 km. Unfortunately the latter study is based on the performance of surface wind, not precipitation. With this context, it seems reasonable to carry out a study of the added value of the model performance as a function of spatial resolution, using the precipitation produced within D3 or even D2 of this simulation. Unfortunately such an analysis can not be carried out with the present simulation. The reason is that this run was carried out with all domains nested two-way, as described in section 2.4. This implies that the precipitation as simulated by each coarser domain is replaced by the one within the innermost domain in the overlap region, i.e. the precipitation recorded for D3 inside the region span by D4 is actually a spatial smoothed version of the latter. Therefore it is does not correspond to the actual precipitation as resolved by a 6 km configuration, but an improved version that accounts for phenomena explicitly resolved within D4. This effectively precludes the use of this data for the fair evaluation of the model performance as a function of spatial resolution suggested by the reviewer. At this stage, a proper evaluation of this issue would require re-running great part of the simulations, which would involve a prohibitive computational cost.

5. I understand that the forcing GCM is always an open question, but the usage of just one instead of, at least, a couple of them, does not limit a little bit the representativity of the GCM-forced RCM analysis?

Certainly. It is always better to target at an ensemble, as such an approach allows to better characterise GCM-specific biases. This is indeed what we aim to some extent with the inclusion of the simulation driven by ERA-Interim in the analysis. However, computational cost is a bottleneck in this study. Carrying out a single realisation with a single GCM costed thousand of hours in one of the most powerful supercomputer available, CSCS. It is completely unaffordable for us the repetition of this simulation driven by alternative GCMs to produce a proper ensemble. We hope that this limitation is overcome in future studies, but unfortunately we are currently limited by this.
Still, we have added a paragraph in section 2.5 to discuss this issue.

6. The result shown in pages 10-11 that related intermediate seasons with cancellation artifacts sounds reasonable, but perhaps a more specific analysis could be made, with moving seasons, to see if more clear picture of that can be obtained. Because on the other hand, this result could be found non-intuitive, as one can think that precisely those transition seasons are more difficult to be properly captured. Which are the thoughts of the authors about it?.

The figure below shows the result of the calculation suggested by the reviewer. It shows PDFs of daily precipitation within "moving seasons". There are 12 panels, each one obtained considering as the window the given month, the previous and the former. The coloured panels highlight the standard seasons shown in Fig. 7 in the manuscript. The compensation of errors in intermediate seasons becomes apparent in WRF-CESM, as this simulation shows opposite biases in the previous and following seasons. We have briefly discussed this results in the manuscript, although we believe that the inclusion of the figure is not necessary.



7. Page 11, line 22. The bias corrected result over the frequency distribution that changes from underestimation to overestimation in winter looks a little bit peculiar. Could this result be a little bit further explained?

This issues has been raised by reviewer #1. As discussed by Themeßl et al. (2011), this effect occurs when models tend to underestimate the dry-day frequency (which is a rather infrequent feature of some RCMs, as most of them exhibit the oposite behaviour, i.e. drizzling-effect), as all these days become mapped onto a precipitation day, leading to a wet bias. This could be further corrected using frequency adaptation techniques, although we have not considered such techniques here. A brief discussion of this aspect has been included in the manuscript.

Technical corrections

1. When describing the experimental design (page 5, line 25) I do not understand those 6-day chunks and 12h spinup periods. I thought that a whole year or even two or more where needed for the soil moisture to be adapted. Could this aspect be explained a little bit more? I understand that more details can be found in Gomez-Navarro et al., 2015, but perhaps here it is too little what is said. It is the same about D1-D2-D3-D4 subdomains and nesting aspects.

We carry out the simulation in so-called reforecast mode. This approach is not new, but a well-settled methodology to conduct RCM simulations that splits the simulations into small tranches. As explained in the cited reference (Gómez-Navarro et al. 2015), *"The method consists of splitting a long simulation into shorter simulation periods of 1 to a few days, running each period separately and finally merging them. This method effectively minimises the impact of the boundaries, transforming the problem into a mostly initial-value problem. The reforecast method is regularly applied (Jiménez and Dudhia, 2012; García-Díez et al., 2013; Menendez et al., 2014, among others), and the increased skill of this method compared to continuous runs has been reported (Lo et al., 2008)."*. In a nutshell, splitting the simulation allows to bind the RCM to the driving dataset, and it can

be regarded as a form of nudging. Further, this strategy has computational advantages: several simulations can be run simultaneously, which naturally leads to the parallelization of the problem.
Of course there are drawbacks. As pointed out by the reviewer, the short spinup period does not allow the soil moisture to reach an actual equilibrium with the atmosphere, which in opinion of the authors reduces the land-atmosphere coupling (still, this coupling does not disappears, as it is borrowed from both the soil and atmosphere initial conditions used to run both submodels within each tranche). Although we believe this can bias certain applications of the simulations, for instance in the study of severe droughts and certain type of floodings, this approach does not impose a fundamental problem in general, as the successful validation of the simulations carried out in this same study demonstrates.
Regarding the domains, we clearly state their setup in Section 2.4 and even show them explicitly in Fig. 1: "Horizontally, we use four two-way nested domains with grid sizes of 54, 18, 6 and 2 km, respectively (top map in Fig. 1)"
We have added more details and a brief exposition of these arguments in the Section 2.5 in new version of the manuscript.

2. Close to this point, I do not also understand why nudging is applied to ERAinterim forced simulation, but not to the ESM one.

This is not arbitrary, but there is a rationale behind this choice. We developed it in the first version of the manuscript:

> The rationale behind this choice is that avoiding nudging gives the model more freedom to develop a more precise representation of the physical processes at regional scales (due to the higher resolution), and thus is potentially able to better correct systematic biases of the ESM, which, e.g., simulate a too strong zonal circulation (Bracegirdle et al., 2013)."

References

Gómez-Navarro, J. J., Raible, C. C., and Dierer, S.: Sensitivity of the WRF model to PBL parametrisations and nesting techniques: evaluation of wind storms over complex terrain, Geoscientific Model Development, 8, 3349–3363, https://doi.org/10.5194/gmd-8-3349-2015, 2015.

Pielke Sr, R. A.: Mesoscale meteorological modeling, vol. 98, Academic press, 2013.

Themeßl, M. J., Gobiet, A., and Leuprecht, A.: Empirical-statistical downscaling and error correction of daily precipitation from regional climate models, International Journal of Climatology, 31, 1530–1544, https://doi.org/10.1002/joc.2168, 2011.