Response to reviewer comments: HESS-2019-77

We thank the three anonymous reviewers and Peter Stucki for their constructive comments. In the following, we focus our reply to the major comments. Based on these comments, we conclude to rewrite the article. Later, we will take care of the minor comments, which then will be still important. Reviewer #1 pointed out that the title could be adjusted. However, we realized that not the title has to change but the content of the paper. Thus, a complete revision of the paper is necessary. We aim to use a different bias correction method and shorten/change the validation of the bias correction, and include corresponding literature. At this point we will also change the focus for possible applications in hydrology and what requirements are necessary for such purpose.
In the first version of the paper, we focused on downscaled ERA-Interim (and ERA-20C) simulations as an example. Now, we think that the second version of the paper would benefit a lot from the inclusion of a larger RCM dataset (ensemble of the MiKlip project, https://www.fona-miklip.de/). In total, we have over 10.000 simulated years, making it possible to do proper statistics, and which fits better to the chosen title "Towards the Development of a Pan-European Stochastic Precipitation Dataset".

RC2) Interactive comment on "Towards the Development of a Pan-European Stochastic Precipitation Dataset" by Lisa-Ann Kautz et al.

Peter Stucki (Referee)
peter.stucki@giub.unibe.ch

Received and published: 2 April 2019
Review of the manuscript Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-77

General comments

In this manuscript, the authors present the configuration and results from dynamical downscaling and subsequent bias correction for a number of heavy-precipitation events in five large river catchments in Central Europe since 1979. They make use of the ERA-Interim and ERA20C reanalyses for initializing the COSMO-CLM regional climate model with horizontal grid sizes of 25 km. Calibration and validation of the obtained new datasets are done using the gridded E-OBS daily precipitation dataset. With this approach, the authors claim to provide a proof of concept for producing a consistent pan-european precipitation dataset, which I think would be a relevant scientific contribution within the scope of this journal.

In general, the structure, language and style of the article are on a good level, and I have only minor comments on this. Also, I acknowledge the amount of conducted work, provided material and analyzed data. In contrast, I do have a number of major concerns about the concepts, datasets and methods used in the study. These need to be addressed before I can recommend accepting the manuscript for publication.

Major comments

My first concern is about the input data used and the downscaling resolution. In this study, the ERA-Interim dataset is downscaled from an 80-km grid to a 25-km grid. With the advent of the ERA5 at a 31-km grid, such a downscaling procedure results in hardly any gain in resolution, while it has substantial computational costs. The authors argue that a 25-km grid is adequate for driving hydrological models. Although I am not a specialist in this field, I would argue that convection-permitting model resolution is needed for many applications, and the authors also mention this in the manuscript (P16 L15). I can clearly see that research must be based on data availability at the time of the analyses. Nevertheless, I assume that the development of ERA5 was already known when this study was conceptualized, and also that future applications would rather switch to using the better resolved dataset. In addition, the study downscales from the long-term ERA-20C reanalyses for comparisons of performance. However, the current study does not exploit the long-term character of the reanalyses; early events that would be covered by ERA-20C only are not included. To conclude, I am reluctant to insist on redoing the experiments using the currently available datasets because of the massive additional work load, although it seems the way to go to achieve a solid contribution in the sense of a proof of concept for future hydrological applications.

We agree with reviewer that the added value of downscaled ERA-Interim data (with 25 km resolution) is not optimal, given that ERA-5 is now available. In the first version of the paper, we have chosen ERA-Interim-CCLM as an example (together with ERA-20C) and missed to make clear that in truth, over 10.000 simulated years are available on the 25km resolution (resulting from a cooperation with the MiKlip project). Note that ERA-5 was only a trial data set at the beginning of this project.
We prefer to validate the selected bias correction method in flooding cases after 1950 as for the second half of the century we can use high-quality observational data (HYRAS). In the new paper version, we will show long-term time series of our ensemble dataset and compare them to the uncorrected dataset and to observations. Furthermore, as our purpose is not a high-quality reproduction of historical events but stochastics, we think it is not necessary to include case studies from the beginning of the century.
Regarding the model resolution, the purpose of the stochastic event set is to evaluate flood losses for large floods on big rivers in Europe using rainfall-runoff modelling. The primary concern for such analysis is the total extent and total precipitation over large area and resolution of 25 km^2 is sufficient. Nevertheless, we plan to use further downscaling steps, eventually with a combination of statistical and dynamical downscaling. This will be next phase of the project and is not reflected in this paper.

My second concern is about the application of bias correction. In the first place, I have precipitation dataset. I take from the experimental setup that after downscaling from 'consistent' reanalyses (I agree with the authors, P2 L6), the bias correction based on E-OBS actually re-introduces inhomogeneities, because the E-OBS dataset is only quality checked, but by far not homogeneous (see also P16 L20). In addition, I see that a number of studies conduct calibration and validation of bias correction using the same data. However, in a clean experiment, the training dataset must be independent from the test dataset. Therefore, I suggest dividing the available data into these sets, or apply cross-validation or similar techniques. Moreover, there are also disadvantages of applying bias correction, especially EQM, which should be addressed in a proof of concept. These are (i) the assumption of climatological stability of the derived transfer function (ii) the potential physical distortion of the dataset by statistical correction including (iii) potential physical disconnection from other variables like temperature, for instance, and (iv) the problem of how to deal with unobserved extreme values when applying EQM to events outside the calibration period. I appreciate that the study tests a range of options for bias correction before selecting EQM, such as recommended by Maraun et al. (2010) and others. In turn, this means that finding EQM the best-performing method is not a novel contribution of the study. The authors discuss potential improvements, and there are even more options, e.g. applying a combination of change factor and quantile mapping, using detrended or region-aware quantile mapping, applying multivariate bias-correction or bias-correction based on synoptic weather situations. Such a comparison might be a valuable and novel step.

We agree with the reviewer that an independent test dataset should be used to validate the bias correction method. In the reworked paper, we will provide a case study for the river Rhine, in which HYRAS data are used as reference. In addition, we will discuss the advantages and disadvantages of the chosen bias correction method in the method section also under hydrological perspectives. We will partly use the results of this first paper version and treat ERA-Interim and ERA-20C as training data. The test dataset would be the novel large RCM ensemble of the MiKlip project.

My third major concern is about uncertainties along the downscaling – bias-correction – validation path. On a number of occasions, I miss information about potential or quantified uncertainties along the process of downscaling, bias correction and validation. An important feature of current observational and reanalyses products like more recent versions of E-OBS, ERA5 and CERA-20C (which cover the same periods as the ones used in the manuscript) is that they provide an ensemble to assess uncertainties or sensitivities. I do not insist on a thorough probabilistic analysis here, but I tend to think that such ensemble information should be exploited in the current study in one way or the other. Furthermore, no information about uncertainties in the configuration of the regional model are given, just a statement that it is 'suitable'.

We agree with the reviewer that the study would benefit from an ensemble approach. Thus, in the revised version, we want to include ensemble data (from the MiKlip project) and provide ensemble characteristics (e.g., ensemble mean and spread, standard deviation). In addition, we will add some information on the suitability of the regional climate model (COSMO-CLM) and name additional references.

For example:

Bellprat, O., S. Kotlarski, D. Lüthi, and C. Schär, 2012: Exploring Perturbed Physics Ensembles in a Regional Climate Model. J. Climate, 25, 4582–4599, https://doi.org/10.1175/JCLI-D-11-00275.1

Kotlarski, S., K. Keuler, O.B. Christensen, A. Colette, M. Déqué, A. Gobiet, K. Goergen, D. Jacob, D. Lüthi, E. van Meijgaard, G. Nikulin, C. Schär, C. Teichmann, R. Vautard, K. Warrach-Sagi, V. Wulfmeyer (2014): Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble. Geosci. Model Dev. Discuss., 7, 217-293, doi:10.5194/gmdd-7-217-2014

My fourth concern is about the achieved enhancements and the chosen metrics and variables. I agree with my fellow reviewer that the absolute differences between corrected model and observation-based dataset are still enormous at places. This contrasts the argument of the authors that the study provides a proof of concept for hydrological modeling. Then, the added value often comes from the second or third decimal place, or even slight worsening occurs regarding some metrics (e.g. LS and LOCI in Table 1). This leaves me uneasy, and I wonder if such differences are still (statistically) significant, can be called enhancement and worth all the involved efforts for practical applications. At one point, the authors mention that such small differences rather deserve the adjective 'similar' (P7 L14), or that missing the dynamics is a much larger problem (e.g. Figure S1). Furthermore, the distinctions between the 'captured' categories are comprehensible, but involve a lot of subjective judgement. For better comparability, I suggest using standard ETCCDI precipitation indices like annual precipitation maxima in terms of Rx5day, R20mm, R95pTOT or R99pTOT. For instance, I wonder how many and which heavy-precipitation events are missed or captured in terms of three- or five-day rainfall totals. Finally, please check if non-parametric measures should replace parametric measures (Pearson correlation, standard deviations, also involved in Taylor diagram) for precipitation.

We thank the reviewer for this comment. In the reworked paper version, we will show that the selected bias correction has clear advantages in comparison to other bias correction methods by including output from the hydrological model. In addition, we will include some indices of the standard ETCCDI framework to show the added value of the ensemble approach in comparison to a deterministic run.