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".

RC3) Interactive comment on "Towards the Development of a Pan-European Stochastic Precipitation Dataset" by Lisa-Ann Kautz et al.

Anonymous Referee #3

Received and published: 3 April 2019

The authors try to construct and validate a European long-term precipitation data set, by dynamically downscaling reanalysis data, and bias-correcting them to match the E-OBS dataset. I appreciate the work, and the serious efforts undertaken, but I have some serious concerns about the general setup of the study. Therefore, I regret that I cannot recommend this study for publication in its present form.

General comments:

(I) I had some serious problem understanding the basic setup of the study: The E-OBS dataset is taken as "ground truth" for the validation – and for the training of the bias correction scheme.

(1) You start with reanalysis data – which are known to have severe deficiencies in representing precipitation: "The Climate Data Guide: Atmospheric Reanalysis: Overview & Comparison Tables." (Dee et al., 2018): https://climatedataguide.ucar.edu/climate-data/atmospheric-reanalysis-overview-comparison-tables: "Diagnostic variables relating to the hydrological cycle, such as precipitation and evaporation, should be used with extreme caution."

We thank the reviewer for this comment. We are aware of the fact that diagnostic precipitation from ERA-Interim might contain errors, particularly when the precipitation has a convective character and thus cannot be resolved by the rather coarse grid spacing of ERA-Interim. This leads in fact to large discrepancies between ERA-I and observational datasets e.g. in the tropics, where most of the precipitation is convective. Nevertheless, precipitation is not directly downscaled by the regional model. The idea behind regional modelling is to provide information about the atmospheric flow and moisture (u,v wind components, specific humidity) at the boundaries of the model domain. The precipitation simulated by CCLM then depends on the applied physical parametrisation schemes for

large-scale and convective precipitation in the CCLM. Thus, the resulting precipitation CCLM precipitation is independent from and not affected by the diagnostic precipitation of ERA-Interim.
In general, (extreme) precipitation is one of the variables that exhibits a large added value when applying a RCM with higher resolution compared to coarse and global scale datasets (e.g. Flato, G. et al. 2013. Evaluation of climate models. In Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. T.F. Stocker, D. Qin, G.-K. Plattner, et al., Eds.: 126. Cambridge, UK and New York, NY: Cambridge University Press.)

(2) The reanalysis data are then downscaled to 0.25° x 0.25° employing an RCM – with its own deficiencies in representing precipitation – resulting in huge errors, even in mean annual precipitation (Figure 3). Those data cannot be reasonably used without performing bias correction – which requires reference data.

See comment above: the ERA-Interim precipitation is not considered for downscaling by CCLM as it is only a diagnostic variable. Precipitation in CCLM is generated based on the applied parametrizations and the atmospheric conditions (flow and moisture availability) provided by ERA-I. Of course, the simulated precipitation has its uncertainties emerging from the parametrization schemes. Therefore, bias correction is based on independent reference data is a common technique in hydrological studies.

(3) The bias correction is performed using gridded E-OBS data (0.25° x 0.25°) – which results indeed in some reduction of the bias (with respect to E-OBS). Given all that I cannot imagine that the results would be better than the E-OBS data – and throughout the paper I could not find convincing arguments for that (If you have some – please provide, e.g. by comparing with the HYRAS data). So why not using the E-OBS data to begin with – at least for the time period, where they are available? One possible application of this approach could be the extension of the dataset into time periods, where no E-OBS data are available (prior to 1950). For this purpose you would, however, need to employ ERA-20C – with even coarser resolution than ERA-Interim – and probably with even larger systematic errors (Table 1).

We thank the reviewer for that comment. We think that in the first version of the paper, we did not emphasize sufficiently that we wanted to describe how to generate a huge centennial precipitation dataset that can be used to make proper statistics. In the reworked version of the paper, we will include an ensemble dataset consisting of over 10.000 simulated years. This will show a clear advantage in comparison to the E-OBS dataset.

(II) To prove the concept you would need to show a Figure like Fig. 3 for ERA20C-CLLM – which should also include a presentation of relative errors. I fear that they might be intimidating in parts of the domain, e.g. in Tuscany. You should also provide at table like Table 2 for ERA20C-CLLM. You assume that the scaling factors for the EQM did not change over time, and present Fig. 4 to prove. This is not strictly true – if the centers of the squares are meant to represent the data: Almost all the crosses are clearly above the squares – and this a logarithmic plot.

The table (like Table 2) for ERA20C-CCLM was provided in the supplemental material. As we want to revise our bias correction method and the validation, Fig. 3 as well as Tab. 2 will be removed from the paper or at least strongly revised. In the reworked paper version, we will provide a similar figure as Fig. 4 including also global data and HYRAS data and improve its description. We will also change Fig. 4 from absolute occurrence to probability. We agree with the reviewer, the y-axis is logarithmic and so differences in order of 1.000 seem to be large. But this is absolute occurrence in the whole domain and time period with millions of data values going in. In terms of probability these difference are of order of 10e-5 or smaller. Changing Fig. 4 will clarify this point.

(III) Some of the metrics seem not to be well chosen. One of your major concerns should be the bias – which is not necessarily related to the correlation coefficient (e.g. Table 1). For example: You would still find a perfect correlation, if all data in one dataset would have a relative error of 50 %. After employing the bias correction, you should show how the bias has reduced. Fig. 3 already indicates that this is in fact the case, but it would be helpful to include some quantitative information (Table). The RMSE (Fig. 2., Table 1) is not a very good measure for that. At first sight, the RMSE reduction looks disappointing – but the bias reduction does not necessarily reduce the bias anyway. RMSE should come with a unit – could this be mm/day? ~3 mm/day seems to be too much, even when looking at Fig. 3, but 3 mm/year is not enough.

To emphasize the aim of the study (cf. title of the article) and to have more space to show the novelty of the study (over 10.000 simulated years), we decided to shorten the bias correction validation by skipping the tables and the RMSE calculation. We have chosen to present the skill score of Taylor (2001), as it has some advantages in comparison to other scores (as e.g. the RMSE). For example, the skill score of Taylor (2001) increases as the modeled variance approaches the observed variance and it increases monotonically with increasing correlation.

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, J. Geophys. Res., 106, 7183–7192, https://doi.org/10.1029/2000JD900719, 2001.

(IV) I do not understand the results for the Vistula catchment (Poland). The bias correction changes a positive bias – over a large area – into a negative bias of similar magnitude (Fig. 3c), how can this happen? You argued with data gaps in that area, but within the framework of your study E-OBS is the "truth". Why is there just a minimal bias reduction (Fig. 7c, d)? Figure 8 shows accumulated precipitation from June 19 until June 25, 2009. In most of the catchment area, there is no precipitation at all, according to the E-OBS data (Fig. 8a). Are you sure that his is the tight plot? It is hard to believe that the average over the cachtment area could correspond to the values shown in Fig. 7 (c,d).

In the empirical quantile mapping approach, the empirical rain distribution (separated by quantiles) of the dataset (which should be corrected) is adjusted to the distribution of the reference dataset (here E-OBS). In addition, monthly distributions are used. Thus, for individual events (on a daily or weekly scale), there can be still differences between the reference data and the bias-corrected data. The discrepancy between Fig. 7d (black line) and 8a results from 1) different time ranges, 2) the routine to calculate the spatial mean. In Fig. 8a, there is a small area of strong precipitation within the Vistula river catchment. In the surrounding areas (within the Vistula catchment), there are missing values which were not included in the calculation of the spatial mean, resulting in a relatively high value for the spatial mean in Fig. 7d.
In the revised paper, we will use a different bias correction method and we will go more into observational gaps in the E-OBS dataset that is especially relevant for the Vistula river catchment.

(V) The language needs to be improved – and careful copy-editing will be required. Some examples are listed below, but this list is not exhaustive (and I am not a native speaker either). (Linguistic) agreement (I hope that this is the right term) is a recurring problem:
heavy precipitation ... are
reanalysis (singular) ... provide (plural)

reanalysis ... are (here you could use “reanalyses”)
The added value ... are discussed
bias correction ... do

“Data” is a plural word, therefore “data ... are” instead of “data ... is”

Some words have a different meaning than you think. E.g., check your usage of “towards”

We thank the reviewer for this comment. We will take care of these linguistic matters in the next paper version.