Review of the manuscript HESS-2019-77

Towards the Development of a Pan-European Stochastic Precipitation Dataset

by Lisa-Ann Kautz, Florian Ehmele, Patrick Ludwig, Hilke S. Lentink, Fanni D. Kelemen, Martin Kadlec, and Joaquim G. Pinto

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."
(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.
(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).

(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.

(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.

(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).

(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"

(Some) specific comments:

(Page 1, Line 7) You focus on five river catchments in **central** Europe, but later the Rhine is located in **Western** Europe and the Vistula in **Eastern** Europe. I know that there are numerous definitions, but it would be good to stay consistent throughout the paper.

(P 1, L 23) It is true that extreme events are often not adequately represented in short time series. But if you use those time series for the training of the quantile mapping – couldn't it be happening that you would then unwittingly remove some the "true" extreme events in your modeled (downscale) data?

(P 2, L 28) "case sensitive" has a totally different meaning: (https://www.merriam-webster.com/dictionary/case-sensitive)

(P 3, L 21) Isn't the E-OBS resolution 0.25° ? (instead of 0.22°).

(P 6 L 25) "RPI": "I" will most likely be understood as capital "i".

(P 9 L 4) "improvement of at least −33%": What is the meaning of the "minus"?

(P 9 L 5) "pinrted"

(P 9 L 9) I would understand as "east-west gradient in the accuracy" as an **increase** in accuracy (better agreement) from east to west.

(P 11 L 3) "up to 70% of missing data at certain locations" How many stations are affected?

(P 14 L 2) "We speculate .." please check, if this indeed the case.

(P 15 L 21) "spring foods"

(P 16 L 21) "However, the suitability of the E-OBS dataset as reference data was shown by the high correlation coefficients between EOBS and HYRAS daily precipitation sums .."

No, not really – at least not in this paper. You could show it here – are provide a reference.