Authors response to comment on hess-2022-21

Anonymous Referee #3

Author comment on **A gridded multi-site precipitation generator for complex terrain: An evaluation in the Austrian Alps** by Hetal Dabhi et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2022-21-RC2, 2022

This paper discusses the results obtained using a space-time generator for daily rainfall in the Austrian Alps. This is an interesting topic and the paper is generally well written. The proposed model is very close to the one proposed by Kleiber et al. (2012) and the methodological contribution of the paper is rather limited. However, the authors have done an impressive work to validate the model on a complex data set. I think that this may be of interest for the readers of the journal Hydrology and Earth System Sciences. I suggest that the authors take into account the following comments before submitting a new version of their paper.

We are thankful to Reviewer-3 for the time and constructive feedback on our manuscript. In the following we provide our response comment by comment.

1) Line 3. "...as well as future climat ". Indeed, this would be great, but this is not discussed later in the paper. I think that it should be removed from the abstract or discussed in the paper.

   Response: The future climate is mentioned in the opening of the article as to let the reader know the fact about the requirement of high-resolution data for current as well as future climate. In our view, the 'future climate' here simply refers to a potential application of a precipitation generator. Later in the abstract it is made very clear that 'this study aims at evaluating…' (which is not possible for future climate). In the revised version we will replace 'as well as' by 'or' to make this even more obvious.

2) Line 5 . "... propose an extension...". Please detail, contributions should be clear when reading the abstract.

   Response: This will be considered and necessary changes will be made in the revised manuscript.

3) Line 38 "...typically of 1 km for spatial and daily for temporal scale". Could you precise an application where such scales are involved? I am not a specialist in hydrology, but I have the feeling that if you consider such a high spatial resolution, then the temporal resolution should also be finer?

   Response: This is indeed a very valuable comment. We think, however, to have made exactly this assertion a couple of lines later (L41. '…in the spatial scale of 100 m and hourly for temporal scale.'). In the paper we propose an approach to generate 100 m (spatial) but daily (temporal) data — which might seem to be inconsistent, but is due to data availability (there are very few hourly data sets over climate time scales available for training). Methods exist to 're-sample' daily variables to finer time scales (e.g. Delle Monache,et al. (2013)). We will add this information to the revised manuscript where the data is introduced.

4) Line 90. "Such WGs are of limited use if the observed gridded data are not available which is often the case". One option would be to "interpolate" the data on a grid before fitting the Wgs. Could you comment on the benefit of the proposed approach versus interpolation? It may be possible to obtain high quality gridded data using an interpolation method which merges all the available sources of information (meteorological stations, but also radar, models and so on), whereas the inclusion of such information in the proposed Wgs does not seem straightforward?

   Response: Indeed, it is an option to first produce gridded (interpolated) data and then fit the gridded WG. This however, is a formidable task of its own, which is not just a little extension to our gridded precipitation generator. This essentially is another study (and for once we think 'being out of scope for the present study' is an appropriate wording). Therefore, if the gridded data does not yet exist, it is not an option. For Austria INCA data (the nowcasting facility of ZAMG, the Austrian Weather Service) at 1 km spatial resolution are available (which would be already good) — but only starts from 1999, unfortunately.

5) Line 100 "However, Kleiber et al. (2012) tested the model only for the multi-site precipitation generation, i.e. at locations with observation and not for the generated gridded data of precipitation." Does it make an important difference? If yes, please detail.

   Response: Indeed, it makes a big difference. Since Kleiber et al. (2012) proposed the model that can generate gridded data of precipitation, but they tested their model only for the stations with observations, i.e. they tested only the **multi-site** model and not the **multi-site gridded** model. This means that the model was used to generate correlated data at stations with historical observations only and the interpolation step for producing gridded data was not carried out. Since many applications for impact studies require gridded data as input, before producing the gridded data for such applications, one must evaluate the model for its ability to reproduce gridded fields of precipitation.

   Another important point is, since the model can provide data at locations without historical observations, one can obtain the historical time-series of daily precipitation at those locations. In order to use the model for this purpose, it is necessary to assess the model performance for the gridded fields.

   We will add a sentence or two to clarify this in the revised manuscript.

General comment on Section 2. I find that the statistical methodology is not described precisely enough. The reader sometimes has to guess how the model is defined, and I do not think that there are enough details for someone who would be interested in reproducing the results. This is especially true in Section 2.2, but I think that more details should also be given in Section 2.1.

Response: We thank the reviewer for pointing this out. We will provide more details and make it more precise in the revised manuscript.

Please also

- explain how the model is fitted to the data, eventually provide the codes,

Response: This will be added in the revised manuscript.

- give the number of parameters involved in the model,

Response: Gaussian process is considered a non-parametric method. In a parametric model, the number of parameters stay fixed with respect to the size of the data and it is easy to report the number of parameters. This is not the case with non-parametric methods where the number of parameters grows with the number of data points. We will include this in the revised manuscript.

- comment on the computational time to fit and simulate the model.

Response: This will be included in the revised manuscript.

6) Line 164 "The Gaussian process itself provides a spatial interpolation method 'kriging' so that the model parameters βO associated with each covariate, which are estimated at observation locations, can be interpolated to any location of interest." Not clear for me, please reformulate.

Response: We thank the reviewer for pointing this out: we will reformulate this in the revised manuscript.

7) Line 230 "To reduce uncertainty and add more robustness to the observations, we increase the sample size of the observed data by considering a 7-days window centred at the day of interest." This sentence is mysterious for me, please reformulate.

Response: We will reformulate it in the revised manuscript.

8) Table 1. Is this table useful?

Response: We agree with the reviewer that Table 1 might not be useful if the stations were in regular terrain. In the present case, however, when the stations are in so different orographic settings, we think it might be helpful to keep the table as it is. In the revised manuscript we will therefore keep the table as it is in the original manuscript. Also, we have presented the results using the names of the stations, so it is easier for the reader to follow, if the details on the stations are given.

9) Line 267. "We select the covariates using both AIC and BIC …". If you consider AIC/BIC, then the model was fitted using Maximum Likelihood? Or only part of it?

Response: Yes, this is correct, a maximum likelihood approach was used for fitting the model. This will be mentioned in the revised manuscript.

10) General comment on Section 3.3. The authors have done an impressive work for validating their model. However, I have the feeling that the two following aspects are important but not discussed and thus should be further considered:

- Cross-validation. If I understand correctly, no cross-validation is performed although it is usually done when validating spatial Wgs. Cross-validation would consist in removing some stations when fitting the model and, then check if the model is able to generate realistic precipitations at these stations by comparing simulation and 'true' data. It may give confidence in the ability of the model to generate precipitation at locations where no data are available.

Response: We thank the reviewer for this comment. We will perform the hold-out cross validation and update the results in the revised version of the manuscript.

- Spatial dependance. The spatial dependence structure, which is an important aspect for many hydrological applications, is discussed only in the discussion (Line 585-610) and supplementary material. This should be discussed in Section 4.

Response: Indeed, spatial dependence of the hydrological aspects is only discussed in Section 5. This is due to an attempt not to mix the presentation of results (in a technical sense) and their discussion. We acknowledge that many readers might expect that discussion already in Section 4. In the revised manuscript, we will therefore add an opening sentence to Section 4.2 in which this expectation is directed to the 'Discussion' section.

11) Line 292. It is not clear for me why the authors use tolerance intervals instead of confidence intervals. Confidence (or fluctuation) intervals have the advantage of being widely used when validating Wgs and easily understood by most readers.

Response: Confidence intervals are sensitive to sample size. As the sample size increases the confidence interval narrows down, while tolerance intervals do not face such issue. This will be clarified in the revised manuscript.

12) Line 300. "To quantify the model performance…". Is it useful to have all these criteria? Do they bring complementary information? It takes space in the paper (with tables and plots) but it is barely discussed in the paper.

Response: Different evaluation metrics have different purposes. Showing one or two metrics only would not be sufficient to give the clear picture of the model performance to the reader. For example, RMSE is a commonly used error metric, however, showing only RMSE does not tell anything about the bias and for that reason we selected two other error metrics where MBE shows the positive or negative bias along with the absolute error MAE. Similarly, correlation shows the association between the two variables but R-squared shows the variation. In the revised manuscript, we will add a brief description about performance metrics.

13) Line 395. The Kolmogorov-Smirnov test and the Wilkoxon-Mann-Whitney test are valid for continuous distributions, whereas rain gauge measurements are usually discrete (e.g. every 0.2 mm for tipping-buckets). Could you comment on the validity of the tests in such situation?

Response: The reviewer is indeed right that the two tests are valid for continuous data. We respectfully disagree, however, with the assessment that the (low) resolution of a rain gauge measurement makes it discrete (otherwise, every technical process imposing a resolution in any measurement would result in a discrete measurement). In fact, 'discrete data' are usually considered to be countable — but the fact that a tipping bucket 'counts' tipping points doesn't make this the principle of the measurement. The measurement principle is indeed given by a physical law (the weight of a droplet with some assumed density overcoming the known resistance of the 'tipping structure') which technically yields a relatively coarse resolution.

14) Line 496. "It is evident that by allowing the elevation as a covariate in the kriging interpolation for prediction at each grid point, the amount of precipitation is considerably improved". Is it so obvious? Maybe there are some improvements, but I don't have the feeling that they are 'considerable'!

Response: In our view, it can be called 'considerable' except for the months October to December. There is indeed a considerable improvement in the amount during summer months. We will take this into account in the revised manuscript.

15) General comment on Section 5 and 6. There are repetitions in Section 5 and 6. I suggest that you concatenate both Sections and try to make it shorter.

Response: Indeed, separating the two sections gives rise to a certain degree of repetition. It does, on the other hand, also clearly have advantages which we wouldn't want to miss. In the revised manuscript we will try to minimize the repetitions so as to shorten it at a little bit.

References:

1) *Delle Monache L, Eckel FA, Rife DL, Nagarajan B, Searight K (2013). Probabilistic Weather Prediction with an Analog Ensemble. Monthly Weather Review, 141(10), 3498{3516. doi:10.1175/MWR-D-12*-00281.1.