Comment on hess-2022-21
Anonymous Referee #2


The paper entitled 'A gridded multi-site precipitation generator for complex terrain: An evaluation in the Austrian Alps' by Hetal Dabhi and co-authors describes an extension of the daily stochastic spatio-temporal precipitation generator of Kleiber et al. (2012) to mountainous areas with complex topography, and illustrates the proposed framework in the Austrian Alps based on a network of 29 meteorological stations.

The paper is well written and structured, and the topic is relevant for the journal Hydrology and Earth System Sciences. The selected case study seems appropriate to test the proposed precipitation generator in an area with complex topography.

Despite the above qualities, the manuscript also contains major shortcomings, which, in my opinion, must be corrected before the paper can be considered for publication:

(1) It is claimed that the paper provides an extension of the model of Kleiber et al. (2012) to areas with complex topography, but the proposed extensions (Kriging with external drift (KED) of model parameters and altitude dependent covariance function) are not very convincing. According to assessment results, these extensions do not offer significant improvements compared to the original model despite the addition of a lot of model complexity. In my opinion, the authors should therefore test the original model in their case study, and propose extensions only if the added complexity generates clear improvements in simulation results. If it is not the case, I would recommend to stick to the simplest possible model, and introduce the current paper as a case study testing the performance of the original model in the presence of complex topography, which I believe is already an interesting contribution.
(2) Almost no information is provided about the spatial interpolation of model parameters (using different versions of Kriging, in particular KED), which is however a critical step of the model set-up. I think that it is inevitable to give more details about the Kriging step, including: (i) mention if a nugget term is used, and if yes what is the nugget contribution to the total variance, (ii) mention which variogram model is used, which method is used to fit or infer variograms, and maybe show some examples of adjusted variograms, (iii) prove by some data analysis than model parameters are (linearly) dependent of altitude to justify the use of KED, and finally (iv) display maps of kriged model parameters.

(3) A cross-validation is missing to evaluate the performance of the interpolation of model parameters by Kriging, and to assess the density of stations required for model calibration.

(4) The proposed model fails to reproduce precipitation at one of the three stations selected for illustration (i.e. Prutz), and according to Fig 9 also performs poorly for Dresdner Hütte, Kühtai, Nauders and Sankt Leonhard 2 (i.e. 17% of stations in total), and the reasons why the model fails at these locations are not investigated in enough details. The only explanation given for Prutz is that this station exhibits 'precipitation characteristics quite distinct from those of neighbouring (or more distant) stations', but nothing is said about why a model calibrated at station locations is unable to reproduce precipitation at the exact same locations. For me it is very probable that the Kriging of model parameters introduces errors when clustered stations exhibit distinct statistics (which is the case for Prutz, that forms a cluster with Ried in Oberinntal and Fendels) - and possibly also where sharp gradients of precipitation occur (Nauders or St Leonhard im Pitztal 2) as well as at the edges of the study area (Marienberg, St Martin or Kühtai) - but one cannot see it in the present manuscript because the Kriging step is completely overlooked.

(5) Regarding the problem at Prutz station, I would be interested in seeing (in supplementary material for instance) the raw time series for Prutz, Ried in Oberinntal, Fendels and Ladis to be sure that the modeling problem does not simply originate from instrumental errors at Prutz station... the fact that this station has so much differences with Ried in Oberinntal which is located 1km apart, at the exact same altitude, and with similar neighboring topography is very surprising to me.

In addition to these major concerns, I also have some minor comments detailed hereafter. Line and figure numbers refer to: https://hess.copernicus.org/preprints/hess-2022-21/hess-2022-21.pdf.

L 121: Threshold to define a wet day: I suppose that the value of 0.1mm corresponds to the lowest resolution of the rain gauges used in the case study. But I was not able to find this information. Please mention the resolution of the rain gauges in section 3.1.

L 125-128 and Eq4: Making the covariance of the occurrence altitude dependent seems a
relatively arbitrary choice, and leads to a complex (and hardly tractable) model. This must be supported by some preliminary data analysis, showing that such altitude dependent covariance of precipitation occurrence actually exists in your study dataset. In addition, I wonder if including station altitude as a covariate in the vector $X_0$ would not be a more convenient modeling choice. This may lead to a simpler model (also easier to 'validate' using the AIC/BIC model selection procedure introduced L260).

Eq 5 (L156): It is not clear to me how this equation derives from Eq 2 and Eq 4 considered at a single site, and using previous day's occurrence as a covariate. Could you give more details (maybe in supplementary or a reference)?

L166: Using KED with altitude as drift to interpolate regression parameters means that these parameters are all (linearly) correlated with altitude. This should be shown by a data analysis. In addition this leads to a complex model, and I wonder (as in my comment about Eq 5) if it would not be easier and equally effective to include altitude in $X_0$, and test if altitude is a relevant covariate.

Eq 6 and paragraph L179-188: the mean function of the latent process $\mu_{\{A\}}$, and the parameters of the Gamma distribution are space and time dependent, and in addition are interpolated by KED. This is a lot of parameters! You should quantify and acknowledge the complexity of your model. I'm not sure that so much model complexity (and degree of freedom) is necessary, but I'm ready to be convinced by a careful data analysis showing that all these dependencies are indeed present in your dataset. If i'm not mistaken, in Kleiber et al (2012), the mean function of the latent process $W_a$ used to model precipitation amount is fixed to zero (and not regressed on covariates with regression parameters additionally interpolated by KED), which makes the model of precipitation amount way simpler. Such addition of complexity must be supported by data.

Figure 1: I think station 22 (Pitztaler Gletscher) should be in red instead of station 24 (Obergurgl).

L218-220: Could this extreme value be an outlier? Prutz is surrounded by very nearby stations (few kilometer apart) in all directions, and none of them measure more than 35mm this day (compared to 156mm in Prutz). I agree that summer convective rains can be very localized, but I'm still surprised by this observation, and I think this requires more investigation. And this also rises concerns about the quality of data at Prutz.

L223-224: different precipitation features at St Leohard 2: more details are needed to ensure that this station operates properly (same comment for Prutz).

L229-232: I do not understand this paragraph. How this 7-days window increases the
amount of data? And how simulation will add robustness to the observations? This is very unclear. Maybe because I do not understand what you name data.

Eq 7 and Eq 8: A lot of temporal covariates are tested, but only one covariate linked to atmospheric circulation. Why such unbalance? I agree that NAOI can influence precipitation in the Austrian Alps, but it is definitely not the only covariate one can think about. In my opinion you should test other climate covariates, or justify why you think NAOI is enough.

L288-291: More information about the APGD dataset is required to allow the reader to understand the main features of this reference dataset. In addition, it should be mentioned somewhere that APGD is not a perfect reference. Finally, the use of a 5km spatial resolution reference does not make it possible to assess the fine scale patterns generated by your 1km resolution model. Hence, all fine scale patterns seen in Fig 11 and Fig 14 may only be artifacts of using KED driven by altitude. This must be mentioned somewhere in the paper, or Fig 11 and Fig 14 should be aggregated at 5km resolution to avoid over-interpretation of the results.

Fig 5 and Fig 6: Frequency -> Frequency (%).

Fig 9: Very useful figure. It could be improved by: (1) using the same range of values for abscissa and ordinates, and for all stations. (2) Add station Id in addition to station names, and maybe order stations according to their Id (as in Table 1). (3) Mention in caption which quantiles are used (percentiles I presume).

Section 4.2: It would also be interesting to display q-q plots of areal daily precipitation amount.

L420: The influence of topography on precipitation occurrence (and also amount) may be an artifact of the model. I don't say that it is the case, but just that you do not prove it in this paper.

Section 4.3: Results in this section prove that the anisotropic model is not relevant in the present case. They also show very few improvement when using KED instead of OK. When considering how much model complexity and arbitrary hypotheses about orographic precipitation enhancement are added with KED I wonder if a direct application of the model of Kleiber et al, (2012), with maybe altitude as a covariate (to be selected by BIC/AIC) would not be a better option.

L532-533: It would be interesting to compare the original and the extended model in a
schematic, including the number of parameters involved.

L678-827: many typos in the doi of the references.