

## Comments from Referee 2:

Citation: <https://doi.org/10.5194/nhess-2023-160-RC2>

This manuscript proposes a catchment-scale stochastic weather model to generate precipitation ensembles using hourly ERA5-Land precipitation data as input. While the authors state that the model is validated for the Rangitāiki-Tarawera catchment, New Zealand, it seems that the results are actually limited to only a few grid boxes. Several critical issues regarding stochastic modelling of precipitation and its validation remain unaddressed. I detail the major concerns below for the authors' consideration.

Major comments:

1. Methodology (Section 2 and Fig. 1): the generation of stochastic rainfall series only involves the selection of a location or point (longitude, latitude), and the random resampling and combination of the identified wet and dry blocks along the time dimension. For this procedure, two key issues need clarification in terms of space-time statistics: (a) for the point-scale temporal process, how can this resampling and combination operation maintain long-term climatology (from ERA5-Land)? e.g., the point-scale climatological mean, rainy days/hours, and other moments that characterize the time process;

We can assemble supplementary material based on Figures 2b, 7, and 9 in Kim and Onof (2020) suggested by the Reviewer that will address this point.

(b) from the space process perspective, how does this model account for the spatial correlation between the generated point-scale time series? This is particularly important for catchment-scale applications. The authors might want to check out these reference papers for a better understanding of the two aspects: Waymire, E. and Gupta, V. K., WRR, 1987; A. Burton et al., Environ. Model. Softw., 2008; D. Kim and C. Onof., JoH, 2020.

The method is proposed for use in multihazard cascade modelling at short spatial scales [needs to be emphasised in revision]. The key issue is the ability to rapidly and efficiently generate vast numbers of longitudinal series, using a complexity fit for purpose. Hence spatial variation is handled in a simplistic fashion whereby one location controls the temporal pattern of synthetic data, and data at all other locations are transplanted with it via means of the original timestamp. In solely hydrologically focussed settings, a more complex model such as Burton et al. (2008), or Papalexiou (2022) would be a more appropriate choice. This would be discussed in the revision, subject to space considerations as the Editor might permit.

2. Statistical validation (Sections 4 & 5): it's unclear how many grid boxes are used in the validation. For example, in Fig. 3 the authors should indicate whether the results are for a randomly selected grid, or average across all grids (11\*14 grids). Additionally, the authors should include figures to verify the spatial pattern derived from the model: e.g., to compare maps of the model-derived 40-year mean (75%, 95% quantiles, and rainy days/hours) with

the ERA5-Land reference statistics. Lastly, the absence of a split of training and validation datasets needs justified.

The results in Figure 3 are shown for a single randomly selected grid point, the checks, however, were done for all locations. We agree this should be indicated more clearly in the text.

We are happy to include the suggested figures on spatial pattern if the reviewer still thinks this is necessary after our response to 1(b) above. However, these would need to go into supplementary material due to figure number limitations.

To accomplish the objective in 1(b) above, long-term climate trends are not included in the model, hence a training / validation split was not considered. The methodology can be executed on any temporally contiguous subset of ERA5-land to reproduce the desired level of rainfall variation.

3. The Introduction section highlights that this model can “produce realistic precipitation patterns to improve the sampling of atmospheric properties and support robust hazard assessments”. I’d argue that for most hazard assessments, a model for real-time precipitation ensembles is typically required (see Samantha H. Hartke. WRR, 2022 for a review). Can this model serve that purpose? If not, the authors should clearly explain how information from this stochastic rainfall model can be ingested into hazard assessments.

The hazard space that we work within is predominantly multi-hazard, including volcanology, seismology, flooding, and geomorphology to name a few. For these, we are looking at long-term probabilities and consequent risks/impacts. Thus, our aim is to produce realistic (but not real / forecasted) rainfall data over a region of interest over long periods of time very quickly so that we can run our downstream models (e.g., tephra re-dispersal, lahar triggering and characteristics etc.) many times with varying but plausible rainfall datasets. This model is currently used within a multi-hazard assessment for the catchment highlighted in this manuscript, but this multi-hazard assessment is still in progress and thus not yet published in any peer-reviewed journal (<https://resiliencechallenge.nz/multihazard-risk-model-flooding-case-study-simulation/>).

Minor comments:

4. Abstract: please revise this part, to include the contribution and fundamentals of the stochastic model, the study period (1981-2020 ERA5-Land data), the validation catchment, and how the modeled information can be used in precipitation-related hazard assessments.

No problem.

5. Figure 1: the rectangle “COMBINE” is not clear – it would be better if described as “combination of time, longitude and latitude dimensions to build a 3-D rainfall cube”. In addition, please clarify the term “starting point” in Figure 1(b).

Happy to replace “COMBINE: with the reviewers extended description.

“Starting point” is the random selection of the point within the 11 x 14 grid that drives the model (i.e., the one that is used to split the time series into wet and dry blocks). Perhaps a different term would be clearer – “driver location” perhaps?

6. Lines 37-39: I cannot quite follow this sentence. Does this model only generate rainfall statistics for a specific day? And “ERA5 data” should be “ERA5-Land”.

Sentence removed as reviewer is correct, it is confusing and unnecessary. ERA5 data changed to ERA5-Land data.

7. Lines 46-47: I don’t understand why the model needs to specify the start date as 30th April. For example, what’s the difference between the 10-year data starting from 30th April and 10-year data starting on 31st July.

This is a downstream user detail – as noted elsewhere, the aim for this rainfall model is to support long-term probabilistic hazard analyses, if an eruption occurs during a dry season, it is less likely to trigger lahars (volcanic mudflows), than if it occurs during a wet season. A start-date also provides opportunities to test reproducibility of downstream-hazard model results.

8. Lines 67-69: What’s the rationale behind conducting 95 runs for 95th percentile bounds? I mean, the 95th percentile can be also estimated from 100, 1000, or 5000 runs. In those cases, the model derived 95th percentile might be more statistically robust due to the increased number of runs.

Subject to both reviewers’ responses, we are happy to run, e.g., 999 runs and pull the 95<sup>th</sup> percentiles from this. The reason we did not do this for initial submission was because the memory allocation of 95 \* 40 years of hourly data at 11 x 14 locations was close to the operating capacity of the then-standard desktop computer used by Whitehead, so we thought this would be easier to replicate.

9. Section 4: this section describes how to validate the model; a more appropriate title might be “Evaluation Method” or “Validation Method”. In addition, if “Sample 1” and “Sample 2” refer to two independent runs of the model at a specific gridbox, it would be better to use “Realization 1” and “Realization 2”.

Agree – happy to change section title to “Evaluation Method”, and “Sample” to “Realisation.”

10. Conclusions: I don’t think the presented work sufficiently supports these conclusions, as many points lack supporting results. For instance, the monthly and daily validation results (despite their flaws as noted above) cannot support the claim that the model can “generate realistic, long-term, hourly precipitation data”.

The original conclusions were written in favour of brevity and thus didn't include a lot of the qualifiers or context that otherwise may have been included.

What about something like the following? Added or altered text is underlined.

“The method and code provided through this brief communication can be used to rapidly generate multiple sets of realistic, long-term, hourly precipitation data over a spatial region. While the outputs do not have the nuances that come with more complex models, the efficient open-source code, written in an open-source language, and based on open-source data facilitate an easy to plug-in input for hazard simulations to support long-term, time and spatially varying, probabilistic risk assessments, uncertainty quantification, and multi-hazard models.”