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 modeling 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; (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.

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.

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.

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.

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).

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”.

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.

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.

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”.

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”.

References:

[1] Burton, A., Kilsby, C. G., Fowler, H. J., Cowpertwait, P. S. P., & O’Connell, P. E. (2008). RainSim: A spatial–temporal stochastic rainfall modelling system. Environmental Modelling & Software, 23(12), 1356–1369.

[2] Hartke, S. H., Wright, D. B., Li, Z., Maggioni, V., Kirschbaum, D. B., & Khan, S. (2022). Ensemble Representation of Satellite Precipitation Uncertainty Using a Nonstationary, Anisotropic Autocorrelation Model. Water Resources Research, 58(8).

[3] Kim, D., & Onof, C. (2020). A stochastic rainfall model that can reproduce important rainfall properties across the timescales from several minutes to a decade. Journal of Hydrology, 589.

[4] Waymire, E., & Gupta, V. K. (1981). The mathematical structure of rainfall representations: 1. A review of the stochastic rainfall models. Water Resources Research, 17(5), 1261–1272.