Review of Manuscript hess-2018-267

A Hybrid Stochastic Rainfall Model that Reproduces Rainfall Characteristics at Hourly through Yearly Time Scale

By Jeongha Park, Christian Onof and Dongkyun Kim

The authors here propose and test a composite method which appears able to generate synthetic rainfall time series for a wide range of aggregation time scales. The issue of producing reliable synthetic rainfall time series is undoubtedly a central problem in hydrology, and the extension to a wide range of scales attempted in this study has the potential to be a relevant contribution. Therefore, the scope of this study makes it suitable for consideration in Hydrology and Earth System Sciences. However, there are a number of issues that needs clarification and/or correction, which I detail below. In particular, my main comment concerns the validation of the method (comment 6 below). I would recommend the paper for publication only after these issues have been addressed by the authors.

Major Comments

1) The so-called Module 2 uses only one quantity (mean hourly rainfall, as obtained from the 1st module) to generate finer scale rainfall statistics (mean, variance, correlation, and dry fraction). Is not clear to me how the regression analysis for module 2 is carried out (Page 10, line 10). For example, the variance is computed from the mean assuming the two quantities are linearly related. How is the slope $\alpha_{[6]}$ computed? Using the entire dataset available for each station and each month? If this is the case, is not true that the proposed model solely use monthly information (as stated e.g., in Figure 4) to produce fine scale rainfall statistics, and this should be clarified. I find that section 3.2 is overall not very clear, and could be improved. You use 'functional relations' between many quantities, but a relation is only shown for the mean vs standard deviation relationship. Perhaps this linear relation could be introduced more generally (e.g., X = \alpha_i Y + \epsilon_i), and then state that X and Y can be the variables of interest, for example mean, variance, … ?

2) Also, it is not immediate to me how all these relations between rainfall statistics can be linearly related, especially rainfall mean and wet fraction. I think it would be helpful to show how these linear relations hold for all the stations in the study, not just a sample rain gauge. Is it possible they depend on season/rainfall regimes?

3) In general, figure captions should be improved and expanded throughout the manuscript, explaining more in depth what is in the figure. For example, the caption of figure 1 should state what the blue areas (cells) and shaded lines (total rainfall intensity I guess) are. Also, even if it is just a schematic figure, Figure 1 should have axes for time and rainfall depth. Figure 6 caption should specify that the results are for a single gauge, etc. The caption of figure 14 is probably swapped (blue and red lines – please check).

4) The authors present results for a particular station in some of the figures, and this station is not the same throughout the paper. I think it would be better to be consistent and present the results for the same station.

5) Page 15, line 17: it is stated that the Module 2 may fail in generating realistic fine scale rainfall statistics. The Authors should include a bit more explanation for this. How often does it happen? Given that Module 2 is based on linear relations, is it possible for some of this relations not to be linear in some cases, and cause these failures? (This may vary with precipitation types/rainfall regimes). This could be assessed checking if, in a cases of 'failure', any of the relations between rainfall statistics in Module 2 exhibit a divergence from linearity more marked than in other cases.

6) Figure 12 summarizes the performance of the rainfall generation method, comparing observed and simulated rainfall statistics. I think the method should be validated using an independent set of observations, not the same used for calibration (i.e., used for computing the regression coefficients). For example, the authors could divide the gauge time-series in two independent samples, using one of them for calibration and the second for validation. In the end, this is what we want to achieve with a synthetic rainfall generator: match as much as possible statistics of time series that we do not have available. This analysis could also resolve a second issue: Since the proposed model is more complex/has more parameters that the original MBLRP, using independent samples for validation would show to what extent the additional model complexity can improve method performance. Also for figures 9, 10 and 11 it would be helpful to show results obtained using independent samples.

7) In the MBLRP model, cell durations are represented by a double stochastic process which produces a long tail matching observations as stated at line 14. However, rainfall intensities are also known to be of sub-exponential nature (Wilson and Toumi 2005). While MBLRP uses an exponential distribution for rainfall intensities, the method proposed here appear to improve the performance in reproducing rainfall extremes. You state that this results on rainfall extremes (better performance compared to MBLRP) is 'surprising': However, this is likely due to the inter-annual variation of the parameters as you also discuss in section 5.1. On this effect, see Zorzetto et al (1016), where it is shown that the interannual variation of exponential-type rainfall distributions can, as is the case with your findings, can explain the emergence of fat-tailed extreme value distributions. The fact that your model generates time series with improved extreme value properties is very appealing, as this is a traditional shortcoming of Poisson-cells based rainfall models.

Minor Comments

Page 4, line 29 – Rain gauge with ID 85663, please correct

Page 5, line 2 – 'the discrepancy between their quartiles/ their range..'. Please also specify in the caption what the whiskers of the boxplots are (quartiles?)

The sentence at lines 27-28 is not clear, please clarify. The variance is not represented by the vertical axis in Figure 2, even though the boxplots do give an idea of the variability of the seasonal monthly rainfall distributions.

Figure 3: please state in the legend what the values in the map are (SARIMA parameters) and what the coloring is, even if it is already explained in the text.

Page 22, line 10: ‘It is important TO NOTE THAT..’ or something on this line.

Page 26, line 1: ‘and the subsequent EFFECTS ON human..’

Page 26, line 3; ‘time scaleS’

References

Wilson, P. S., and R. Toumi. "A fundamental probability distribution for heavy rainfall." *Geophysical Research Letters* 32.14 (2005).

Zorzetto, E., G. Botter, and M. Marani. "On the emergence of rainfall extremes from ordinary events." *Geophysical Research Letters* 43.15 (2016): 8076-8082.