Review, STORM v.2 by Gaona et al.

This paper describes an updated version of the STOchastic Rainfall Model first introduced by Singer et al. (2018). The paper provides a detailed description of the package and the underlying concepts that are used for stochastic rainfall generation, data used to calibrate the model, and an evaluation of its performance in simulating rainfall in the Walnut Gulch Experimental Watershed (WGEW), which is located in Arizona.

I don't have comments at this stage concerning the stochastic modelling approach used, which simulates a total of seven variables, including a bivariate model of storm intensity and duration that is built by using a Gaussian copula to connect marginal distributions for intensity and duration.

I do, however, have quite a few other concerns:.

1. The paper should clearly state the intended applications of the model and also be clear about inappropriate applications. The model seems to have been developed to simulate convective rainfall events only. This may be suitable for a place like Arizona but would certainly not be appropriate for locations affected by frontal rain systems, atmospheric rivers, tropical and extratropical cyclones, and so on.
2. The authors should provide evidence that there is a suitable user audience for the package. The original Singer et al (2018) paper has been cited only 27 times (Google Scholar, 23 Nov 2023). There is only one paper amongst the 21 that are not self-citations that actually uses STORM 1.0; the remaining 20 papers only cite STORM 1.0 in passing as an example of one of several stochastic rainfall generators.
3. There is a claim that STORM 2.0 output would now be suitable for driving hydrologic models, which I think is grossly overstated given the limited amount of evaluation provided in the paper and the concerns that arise from that evaluation (more on that below).
4. The paper needs to be much better organized:
    a. The model should first be motivated scientifically, summarizing the statistical methods and concepts used, and providing readers with a clear indication of how the various bits fit together conceptually. A flow chart or similar tool for depicting the flow of information and how components are interconnected might be useful.
    b. The parameters that control model behaviour should be clearly detailed, with demonstrations provided of their effects on model performance and discussion of how the parameters are set, presumably based on the fitting the complex combination of statistical models to station data, such as that available for WGEW. The impact of parameter estimation uncertainty and how that depends on the quality and quantity of observational data for the watershed that is of interest to the user should also be discussed, together with consideration of the sensitivity of model behaviour to parameter misspecification and estimation error. Note that the abstract makes a claim that STORM 2.0 is a parsimonious

model. It's hard to know whether this claim is merited given the current (not very clear) presentation of the model.

c. I would suggest that this be followed by a brief user manual, with details relegated to an appendix and the github page for the model.
d. This could then be followed by an evaluation of STORM 2.0 performance. The evaluation strategy should be clearly laid out at the outset, including what aspects of performance you considered and how, and whether the evaluation was based on "out-of-sample" performance. it was only in the summary and conclusions that it became apparent that this was actually the case, with the small paragraph beginning at line 482, where it is explained that the test application had been calibrated using the data from the analog instruments that operated in the WGEW up to 2000, and that evaluation was subsequently performed by comparing simulated data against the data from the digital instruments that have been in operation since that time.
e. Recommendations should reiterate points about appropriate and inappropriate applications, both in terms of the types of events that the model is designed to simulate, and potential applications of the simulated rainfall.
f. The paper recognizes a limitation (sentence beginning at line 382 and text beginning at line 505) that would have serious consequences for many hydrologic modelling applications, but despite this evidence, it makes the broad claim that the output can be used to drive hydrologic models! Output may be suitable for some types of applications (e.g., in small, urban, drainage basins where intense rainfall events result in "flashy" streamflow responses), but it would certainly be inappropriate for others.

5. Concerning the evaluation that is performed:
    a. The authors seem to think that it is a virtue that the model can simulate rainfall events and wet season rainfall totals over a substantially wider range than observed, as is apparent in Figures 2 and 4. This may be reasonable given the very large datasets that can be generated from the model, but I think we need quite a bit more consideration of the physical plausibility of these extended ranges to treat this characteristic as a virtue.
    b. I am confused by the evidence in Figure 5, however, which seems to contradict Figure 4 by indicating that the range of measured seasonal rainfall is much wider than simulated seasonal rainfall. I've likely missed something important …
    c. For the results shown in Figure 5, there would be no reason to expect other than zero correlation since there is nothing from the observed climate record that would impose a specific time ordering on the model output. I don't think a demonstration is needed – a simple statement would suffice. The time labels are not needed, and indeed, would induce confusion amongst readers. Good performance would, presumably, correspond to a circular cloud of points in which the vertical spread of points is similar to the horizontal spread. In fact, a two dimensional scatter plot is not needed – simply plotting two frequency histograms on the same axis, smoothed in some way, would be sufficient.