December 10th, 2023.

**Reply to Anonymous Referee #2**.

Dear Referee,

We thank you for your comments. Please find below a detailed reply, from paragraph 1. onwards, to each one of them.

0. 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 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.

   **R/.** We slightly disagree with the reviewer. Most (if not all) of the Stochastic Rainfall Generators (SRG) are built (and calibrated) on rainfall data. Such data certainly reflects the type of rainfall it does come from. Therefore, a bottom-up approach like STORM (and most of all SRG) can be applied to any (small) catchment with rainstorm records, so an appropriate parameterization can be carried out, which ultimately will only enhance the capabilities and performance of any SRG model. Numerical Weather Prediction (NWP) are top-down approaches in which rainfall types can be modeled more explicitly (see e.g., Lavers et al., 2020; Haupt et al., 2017). From a catchment perspective, the performance of STORM with regard to 'global' statistics does not depend on the type of rainfall falling over the catchment. If anything, tropical and extratropical cyclones must be one of the rainfall types STORM would most likely excel at given, for instance, their circular patterns/shape and radial decay (from their cores) in rainfall intensity (see e.g., Rios Gaona et al., 2018); and STORM's capability to model such features. With regard to Atmospheric Rivers (AR), that is a phenomenon that mainly develops over oceans (see e.g., Gimeno et al., 2014), so no gauges there. Nevertheless, their landfalling (and consequently their type of rainfall) is highly influenced by the orography; which is something (already acknowledged by the Reviewer) our model is capable of doing.

   Following (some of) the Reviewer's suggestion, we make more explicit the applicability of STORM by:

   → adding the sentence "Nevertheless, nothing precludes the application of STORM to other (small) catchments in any climatic zone, as long as some detailed rainstorm records exist for the related area/catchment." to the new paragraph inserted/appended at the end of sub-section 2.1 (i.e., Walnut Gulch Experimental Watershed), and before sub-section 2.2 (i.e., Total Seasonal Rainfall [TOTALP]), i.e., "*The richness and careful curation (for more than half a century) of this dataset, especially with regard to high density of rain gauges and detailed and lengthy rainstorm records, was the main reason our model was designed and built with a focus on this particular catchment. <u>Nevertheless, nothing precludes the application of STORM to other (small) catchments in any climatic zone, as long as some detailed rainstorm records exist for the related area/catchment.</u> The effect of the number and extension of rainstorm data on the performance of STORM is beyond the scope of the present work. Given the set up of our model, it is expected that the richer the (rain-gauge) records the more robust the parameterization is, and therefore the better the performance of STORM will be.*" [Please see our reply to item (3) in **Reply to Anonymous Referee #1**].

   → rephrasing the first sentence in section 4 (i.e., Summary and Conclusions) "Built upon STORM 1.0, STORM[1] is an improved Stochastic Rainfall generator focused on gauged watersheds." as "Built upon

STORM 1.0, STORM[1] is an improved Stochastic Rainfall Generator applicable to (small) gauged watersheds (with detailed rainstorm records) in any climatic zone."

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.

   **R/.** As the reviewer will understand, the number of times a paper has been cited is not a reflection of the quality of the work or its potential utility today or in the future. We see this as a baseless comment. Furthermore, self-citations are often applicable when a scientist is building on their own work. Again, not controversial or wrong. We have had numerous queries from users of the model, but that does not always translate into new publications that cite the paper. This is especially the case if users sit outside academia and/or do not publish in peer review literature. Therefore, we feel the reviewer has misunderstood the point of developing STORM v.2. We are making a more appealing and versatile rainfall simulation model, hoping that this new version will appeal to a broader audience of potential users.

   With the two paragraphs inserted/appended at the end of section 1 (i.e., Introduction), and before section 2 (i.e., Data and Methods) [please see our reply to item (2) in **Reply to Anonymous Referee #1**], we definitely showcase the audience for our model (and also show how our model compares with other SRG).

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).

   **R/.** We disagree with the Reviewer's opinion here. The evaluation of our model showcases its improvement, and the new capabilities of its upgrade. STORM's output is now a gridded product, which can easily be applied to, and integrated with, distributed hydrologic models. In fact we are currently using it for various hydrological modelling studies, and plan to apply it in the context of operational seasonal hydrological forecasting in the Horn of Africa. Please see our replies to comments 4. and 5. (–in this document–).

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.

      **R/.** The first paragraph from the section 1 (i.e., Introduction) provides such a motivation, e.g., *". . . In this paper, we introduce STORM v.2 and highlight the novel aspects of the model that warrant a new version number. We made several changes to the model that make it more user-friendly and enhance its capability for simulating rainfall in a manner that supports computation of the water balance over gauged watersheds under historical climate or under various user-defined scenarios of climate change. . . . "*. We also consider that the statistical methods and concepts are well summarized and presented. Sub-sections 2.3 through 2.8 explain in detail how the key variables/parameters of STORM are modelled, i.e., RADIUS, BETPAR, MAXINT, AVGDUR, DOYEAR, DATIME, and the Scaling Factors. Not only is this explanation sequential (some parameters might be pre-requisite from others) but also congruently builds up on advanced concepts such as copulas and circular statistics. In our opinion, the pseudo-code presented as "Algorithm 2" not only is succinct and summarizes and clarifies the logic (and flow) behind STORM components, but is also equivalent to the 'flow chart' the Reviewer suggests.

   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

[1] https://github.com/feliperiosg/STORM2

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.

**R/.** The parameters controlling STORM are clearly detailed and explained (please see reply to comment 4.a. –in this document–). Nevertheless, we now expand on how the 'fitting' is carried out in STORM. To that end, the second paragraph in sub-subsection 2.9.1 (i.e., Pre-Processing Module) will be split into two:

→ the following paragraph will be appended/inserted after the sentence "... starting date - DOYEAR, and starting time - DATETIME."

The best-fitted PDFs are generated through Python's library *fitter* (Cokelaer et al., 2023). For a given variable/parameter, STORM's pre-processing module passes to *fitter* post-processed data along with several families of probability distributions that might be adequate for its fitting. At its core, *fitter* uses *scipy's* object `fit` (from the `stats` module; see Sec. 2.6) "to extract the parameters of that distribution that best fit the data". This is done via either the Maximum Likelihood Estimation method or the Method of Moments (Virtanen et al., 2020). Because several probability distributions are passed to *fitter* (distributions which users can modify according to their needs), the latter finds the best-fitted parameters for such distributions, and computes several assessment metrics: Error Sum of Squares (SSE), AIC (Akaike's information criterion), and BIC (Bayesian information criterion; see Appendix A). The pre-processing module selects the fitted PDF with the lowest BIC (this assessment metric can be modified too by the user). The impact of parameter misspecification and estimation error and/or uncertainty on the performance of STORM is also beyond the scope of the present work. Tools like *fitter* are practical implementations that ultimately reduce to a minimum these sort of potential impacts in STORM's performance/outcomes.

→ and the sentence "The statistical distribution parameters are exported to a CSV..." will be rephrased as "After the PDF fitting and selection is done, the PDF best-fitted parameters are then exported to a CSV...".

Please note that we now clearly state that the 'consideration of the sensitivity of model behaviour to parameter misspecification and estimation error' was beyond the scope of our present work. Likewise, 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' was also beyond the scope of our present work. This too is now clearly stated as "The effect of the number and extension of rainstorm data on the performance of STORM is beyond the scope of the present work. Given the set up of our model, it is expected that the richer the (rain-gauge) records the more robust the parameterization is, and therefore the better the performance of STORM will be.". These two sentences were added to the new paragraph inserted/appended at the end of sub-section 2.1 (i.e., Walnut Gulch Experimental Watershed), and before sub-section 2.2 (i.e., Total Seasonal Rainfall [TOTALP]) [please see our reply to comment 1. –in this document–].

With regard to the 'parsimonious' term, the sentence in the abstract "To fill this gap, we present the second version of our STOchastic Rainfall Model (STORM), an open-source, parsimonious and user-friendly modeling framework for simulating climatic expression as rainfall fields over a basin." will be rephrased as "To fill this gap, we present the second version of our STOchastic Rainfall Model (STORM), an open-source and user-friendly modeling framework for simulating climatic expression as rainfall fields over a basin."

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.

**R/.** STORM is a small and simple framework and does not require the creation of a user manual or another appendix. Its configuration and/or running (i.e., that 'brief user manual') is already very detailed in STORM's repository, i.e., README.md file (main page) at https://github.com/feliperiosg/STORM2. The code is also well documented to support straightforward user adoption.

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.

   **R/.** We did carry out an evaluation of STORM's performance; please see sub-section 3.1 (i.e., Evaluation of STORM) and 3.2 (i.e., Testing Climate Drivers). This evaluation was described earlier in the manuscript too, e.g. first paragraph of sub-section 3.1, and last sentence of sub-section 2.1 (i.e, Walnut Gulch Experimental Watershed). Nevertheless, and to make this description more fully early in our manuscript, the sentence "We parameterize STORM using $37$ years of analog data (i.e., from $1963$ to account at least for $80$ gauges per year); and we validate the performance of STORM over the $23$ years of digital/automatic data (see Sec. 3.1)." will be rephrased as the following (new) paragraph:

   We parameterize STORM using 37 years of analog data (i.e., from 1963 through 1999). Even though there are gauge records for the WGEW from 1953, we use them starting from 1963 to account at least for 80 gauges per year. This analog network amounts to 118 gauges sparsely deployed over the whole WGEW. We carried out simulations of 30 runs, each run having 23 simulated years (i.e., 690 simulation-years in total, per simulation), in order to evaluate the performance of STORM on features such as: seasonal total rainfall (over a small catchment), number of storms generated, and modelled climate impacts in rainfall intensity. The output of this evaluation exercise(s) were compared against 23 years of storm data from the aforementioned digital/automatic network, i.e., the one from 2000 onwards (see Sec. 3.1).

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.

   **R/.** This has been addressed already in the reply to comment 1. (–in this document–).

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.

   **R/.** We disagree with the Reviewer that the consequences for our model in not being able to simulate teleconnections are 'serious'. Figure 4 of our manuscript shows that STORM2 is able to reproduce the average seasonal precipitation of the catchment (including some of its intra-seasonal variability). At the end of the first paragraph of section 1 (i.e., Introduction), the original manuscript made clear that our model is applicable to "any small basin with available storm rainfall data". This is now more evident throughout the whole revised manuscript (please see replies to comments 1. and 4.d. –in this document–).

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.

      **R/.** Figure 2 shows that STORM2 is capable of reproducing wider ranges (either in storm intensity and duration) than what was measured. Not only is this feasible/possible in reality but also a numerical consequence in fitting a PDF over a data series (i.e., fitting along the distribution tail(s)). Nevertheless, Figure 4 also shows that, at least for this exercise, ensembles of simulated seasonal rainfall were never higher than ensembles from actual seasonal measurements (seasonal averages are well reproduced though). Furthermore, STORM offers the capability to control the sampling range of such variables (e.g., `MAX_I`, `MIN_DUR`, `MAX_DUR` parameters/variables).

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 . . .

   **R/.** The Reviewer may have missed the fourth and last lines (from top to bottom) of Figure 4's caption. They respectively read "Panel **b** - Percentile time series for the $90^{\text{th}}$-percentile of all time series. . . "; and "Supplemental Fig. B4 shows percentile time series for the $100^{\text{th}}$-percentile." $90^{\text{th}}$-percentile means that the bands presented in Figure 4 are the 'central' 90% of all possible values (either simulated or measured) along the Y-axis. All such possible values are represented by the bands shown in Supplemental Figure B4 (i.e., $100^{\text{th}}$-percentile series). Therefore, the minimum seasonal rainfall values one can read along the X-axis, in Figure 5, are $\sim$50 mm; which is about the lower boundary of the blue band presented in Supplemental Figure B4. The same goes for the upper threshold, which is deliberately not presented in the latter figure. Hence, Figure 5 does not contradict Figure 4.

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.

   **R/.** A 'low' or 'very low' correlation is not 'zero correlation'. Please see our reply to item (4) in **Reply to Anonymous Referee #1**, in which we have already addressed a similar comment to Figure 5.

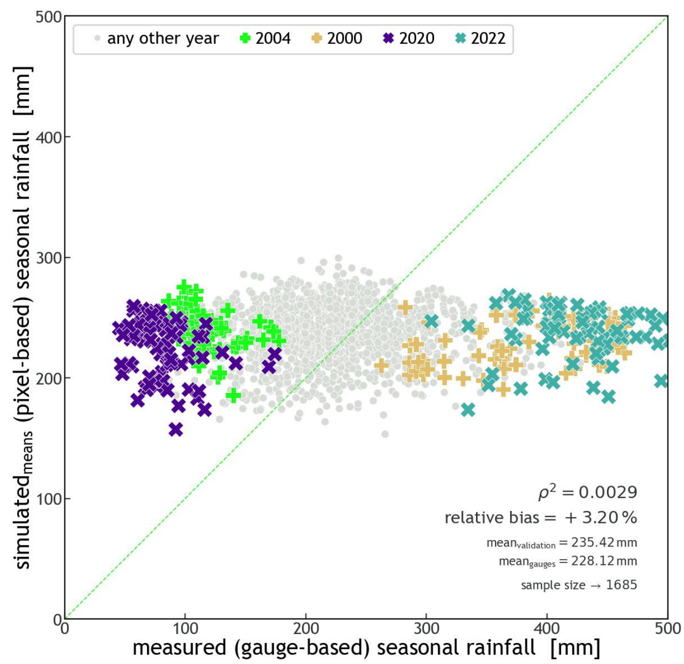


Figure 1: Scatter plot of simulated (means) seasonal rainfall against measured seasonal rainfall. Each marker represents a pixel/station for which the seasonal totals of $30$ simulations were averaged (y-axis), and the actual seasonal total recorded for that location (x-axis). The x-markers indicate the wettest ($2022$) and driest seasons ($2020$), from $2000$ through $2022$. Within the plot, it is indicated the coefficient of determination ($\rho^2$, which is the square of the coefficient of correlation); the medians of the datasets; the relative bias between them; and the size of the sample (an average of $73.3$ gauges per year). The green line indicates a $1:1$ line.

We are keeping Figure 5 in the manuscript, as it serves as a counterpoint to Figure 8, which is a visual representation of the effects of applying a positive scaling factor _SC to the variable TOTALP. Nevertheless, we have improved the readability of Figure 5 by removing most of the time labels, highlighting only those two wettest and driest seasons. [Figure 1 –in this document– is the improved Figure 5 (in the manuscript). Figure 8 (not presented in this document) was also improved to the same color scheme.]

A histogram is a histogram. A histogram 'smoothed in some way' is 'like a PDF'. A histogram is not a PDF. We're not sure what the reviewer means by 'smoothed histograms'. Furthermore, a scatter plot is the simplest way to contrast the different outcomes of the same variable.

Sincerely yours,

Manuel F. Rios Gaona

## REFERENCES


Cokelaer, T., Kravchenko, A., lahdjirayhan, msat59, Varma, A., L, B., Stringari, C. E., Brueffer, C., Broda, E., Pruesse, E., Singaravelan, K., Li, Z., mark padgham, and negodfre: cokelaer/fitter: v1.6.0, https://doi.org/10.5281/zenodo.8226571, 2023.

Gimeno, L., Nieto, R., Vázquez, M., and Lavers, D.: Atmospheric rivers: a mini-review, Front. Earth Sci., 2, https://doi.org/10.3389/feart.2014.00002, 2014.

Haupt, S. E., Jiménez, P. A., Lee, J. A., and Kosović, B.: 1 - Principles of meteorology and numerical weather prediction, in: Renewable Energy Forecasting, edited by Kariniotakis, G., Woodhead Publishing Series in Energy, pp. 3–28, Woodhead Publishing, https://doi.org/10.1016/B978-0-08-100504-0.00001-9, 2017.

Lavers, D. A., Ralph, F. M., Richardson, D. S., and Pappenberger, F.: Improved forecasts of atmospheric rivers through systematic reconnaissance, better modelling, and insights on conversion of rain to flooding, Commun. Earth Environ., 1, 39, https://doi.org/10.1038/s43247-020-00042-1, 2020.

Rios Gaona, M. F., Villarini, G., Zhang, W., and Vecchi, G. A.: The added value of IMERG in characterizing rainfall in tropical cyclones, Atmos. Res., 209, 95–102, https://doi.org/10.1016/j.atmosres.2018.03.008, 2018.

Singer, M. B., Michaelides, K., and Hobley, D. E. J.: STORM 1.0: a simple, flexible, and parsimonious stochastic rainfall generator for simulating climate and climate change, Geosci Model Dev, 11, 3713–3726, https://doi.org/10.5194/gmd-11-3713-2018, 2018.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors: SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, Nat. Methods, 17, 261–272, https://doi.org/10.1038/s41592-019-0686-2, 2020.