April 10th, 2024.

Dear Referees,

Here we provide a response to all and each of your comments.

Report #1 **Anonymous Referee #1**

The authors responded in detail to my comments. There is still a bit of confusion left on my part, and I also think that some aspects could be presented better. See my comments below (numbering relates to previous review).

(1) Good that decision making is out of the title. I am still a bit unclear why it is used as key motivation, but the authors seem convinced.

R/. STORM is at its core a decision-support tool, as it provides flexible simulation of various scenarios of climate and their translation into rainfall. These can be used to explore analyses of the associated impacts of such changes, which is a fundamental component of decision support.

(2) Good that authors now acknowledge the work by others more.

(3a) OK. So the model has never been tested anywhere but in Walnut Gulch which is dominated by convective storm events. Why is it then presented as a generic rainfall model? Are rainfall properties not significantly different in other places? E.g. there is the Brue catchment with about 50 rain gauges in the UK [1]. If the model has so far only been tested in semi-arid regions, should this not be mentioned more prominently? Given that STORM 1 was introduced 6 years ago, it seems ample time to test it on a second catchment in a different climate zone.

R/. STORM is indeed presented as a generic rainfall model, because it is capable to model seasonal precipitation over any given catchment as long as the statistical distribution for five parameters are well know, i.e., total seasonal rainfall (TOTALP), maximum storm radius/extent (RADIUS), average intensity (for that RADIUS), average duration (AVGDUR), and storm start date (DOYEAR) (storm start time (DATIME) is optional; please see Section 2 of the manuscript). Yes, rainfall properties might significantly differ/vary around the globe and that is precisely why STORM must be parameterized for each new location to obtain reasonable results. The local variation in rainfall characteristics is captured via historical and spatially resolved rainfall records, from which the aforementioned statistical distributions are later modelled. Hence, nothing precludes the application of STORM to other (small) catchments, as long as there are detailed rainstorm records (extensive records will contribute to robust statistics). All of the above was clearly stated in the last paragraph of Section 2.1.

We looked into the reference [1] suggested by the Referee, and ended up in its official repository. Without focusing too much in the accessibility of such a repository to actually retrieve storm data (see also, the first remark of Referee #3 in this letter); its main (and rather crucial) drawback is the extent/duration of that dataset, i.e., 4 years of rainfall data (i.e., from 1993 to 1997). Thus, for example, if a user wants to characterize the seasonal rainfall (TOTALP) from that dataset, they would have a hard time fitting an adequate statistical distribution to just four values (i.e., one seasonal average per year). It would not be robust. The Walnut Gulch Experimental Watershed dataset does not have such limitations. For starters, if offers a quite accessible and readily available repository to download detailed storm data. The repository includes seven decades of continuous records, with a gross yearly-average of 80 stations (spread across $148\,\mathrm{km}^2$). Hence, there is high confidence in modelling mean-seasonal precipitation over a dataset of 70 values.

Yes, STORM 1 was introduced 6 years ago; but STORM 2 relies more heavily in robust datasets than its predecessor and offers several new advances that make it more effective and usable. Our premise here is not in testing it over different regions but to develop a robust methodology that generates reliable statistics from storm data. Nevertheless, we are aware of several instances where our model has been used/applied to other datasets in various regions, even if publications were not generated from the work.

This gives us confidence that STORM (as it currently is) is flexible enough to model seasonal rainfall in different climatic regions. Therefore, there is no need for it to be advertised as a model exclusively for semi-arid regions.

(3b) I also have a question regarding alternative models. The authors state in the new last paragraph of the introduction section that no other model exists that is both free and open source. This is not my area of study, but I find that hard to believe.

R/. We partially disagree with the Referee on this point. We did point out that there are other Stochastic Rainfall Generators (SRGs) which are open-source and build upon a free/open-source, e.g., "Like STORM, NEOPRENE is also coded in Python.". Nevertheless, to clarify that there indeed exist SRG that are both free and open source, the sentence *"Some of the SRG here cited are not open-source; and those that are freely available [e.g., 3, 2, 6] were built upo commercial applications (i.e., not open-source software)."* was rephrased as "Some of the SRG here cited are not open-source; and some of the freely available [e.g., 3, 2, 6] were built upon commercial applications (i.e., not open-source software)."

(4) I am afraid that I am still a bit confused. I am unclear what the point of Figure 5 is then. So there is not correlation here bause no temporal information is used. So why do you not include a statement in the cation that makes clear why there is no correlation? Currently this plot is confusing because its relevance is unclear. The reader must go through the paper text to understand its meaning. Why not say so in the caption?

R/. We slightly disagree with the Referee. Figure 5 is, in our opinion, easy to understand and/or read. Y-axis for simulations, and X-axis for measurements. The average of (all) the simulations roughly matches the average of (all) the measurements (which it is the premise STORM is built upon). All this information is condensed in the caption of Figure 5, and its discussion is presented in Section 3.1. (We believe that figure captions should only include the description of the figure and/or what it presents; its discussion and/or analyses is presented within the respective section).

There is temporal information appended to STORM's simulations (e.g. DOYEAR, DATIME –see item (3a) in this letter–). The correlation between simulations and measurements is very low because there is no modelling of inter-annual variability and/or teleconnections in STORM (that may cause an extreme dry or wet season), nonetheless. Still, and following the Referee's suggestion, we added at the end of Figure 5's caption this sentence: "The very low coefficient of determination ($\rho^2 = 0.0028$) indicates STORM's innability to explicitly simulate extreme rainfall seasons either wetter or drier than those present in the historical records.". This was also already paraphrased/presented in the third paragraph of Section 3.1.

Report #2 **Referee #3: Idrissou, Mouhamed**

the authors need to :

– make it clear to the reader that like any other stocastic model, STORM relies on heavily available data for it to perform well. there are very few catchments in the world that are as well equiped as the catchment used for this study;

R/. This was already done in the last paragraph of Section 2.1, which reads: *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. 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. 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.*

Still, to accommodate the Referee's suggestion, we inserted into the aforementioned paragraph this sentence: "Such data richness might not be the norm in other catchments around the globe.".

– delete in the manuscript the statement that says that STORM can well perfom in "any climate zone" because there is no evidence in this manuscript to support the statement;

  R/. We respectfully disagree with the Referee on this point. There is not yet evidence of the performance (whether good or bad) of STORM beyond the Walnut Gulch catchment. Hopefully, after the publication of this manuscript other groups/users will be encouraged to test the performance of STORM in several, and diverse, climatic zones. We feel that this Referee (and Referee #1) may have misinterpreted how the model works. It is not parameterized once and then applied to other locations. It must be parameterized to each catchment to which it is applied. That way, we can ensure that it will faithfully represent the rainstorm records of the new place. In STORM v2, we have given potential users a more friendly set of tools for doing this parameterization in new catchments.

– Provide more recent references when describing the study area;

  R/. Describing the study area is done in the first two paragraphs of Section 2.1. There we provided a total of 10 (un-repeated) references (12 if the official websites/repositories we make reference to are also taken into account). We consider this to be an adequate and thorough number of references.

– the authors might consider using climate scenario applied in IPCC reports instead of using simple scalar alteration of climate parameters that are not supported by any evidence in the study area

  R/. The model is not set up to explicitly simulate any specific scenario, but rather gives the user the control to simulate the impacts of any climate change scenario, including IPCC scenarios. The climate scenarios suggested in the IPCC reports can in fact be modelled by STORM via its scaling factors PTOT_ or STORMINESS_. From the Summary for Policymakers report [8], one can see (e.g., pag. 19, item B.3.1) that *"The average annual global land precipitation is projected to increase by... $1-13\%$ under the very high GHG emissions scenario (SSP5-8.5)..."*. STORM can indeed model such $1-13\%$ increase via its progressive scalar PTOT_SF, for instance. As a matter of fact, Supplemental Figure B7 (panel **a**, in the submitted draft) showcases a simulation exercise where storm rainfall intensity was decreased 3.5% every year (during a 23-year validation).

Implemented changes to the comments found in the file: *gmd-2023-98-referee-report-2.pdf*.

– page 5, line 132: "$1956-2005$" [$\sim$20 years ago!!!]

  R/. If we compute the average annual precipitation over the Walnut Gulch catchment from the dataset we have, we obtain a value of $288.9\,\mathrm{mm}$[1] for the 90% confidence-interval of annual values (means from the gauge network) between 1955 through 2022. When we repeat this average for the period referred by [4], we obtain a value of $282.0\,\mathrm{mm}$[2]. Such discrepancies are expected as there is now practical way for us to know how (and where from) exactly these authors arrived at such values, in spite of our estimates being very close. Therefore, our preference is to refer to values that come from specific literature directly related to the WGEW, instead of providing our own estimates.

  The sentence (in Section 2.1) *"The climate is semi-arid with low average annual rainfall of $\sim 312\,mm$ for the period $1956-2005$ [4]."* was updated to "The climate is semi-arid with low average annual rainfall of $\sim 312\,\mathrm{mm}$ for the period $1956-2005$ [4] (or $350\,\mathrm{mm}$ [5], for an undisclosed period)."

– page 5, lines 134-135: "[9, 10, 11]" [too old reference find recent reference or compute them your self!]

  R/. We found no inconvenience in referring the reader to 'old' but seminal work related to the WGEW, especially when we have also presented 'current' and valuable works. Please see our reply to your previous remark, with regard to computing the mean by ourselves. (plus, the works the Referee refers here bear no explicit relation the mean presented by Goodrich et al. [4]).

– page 5, line 139: "one" [check the structure of the sentence]

[1] with a standard deviation of $51.84\,\mathrm{mm}$
[2] with a standard deviation of $49.50\,\mathrm{mm}$

R/. *"one of the most highly instrumented semiarid experimental watersheds in the world"* is a sentence quoted *verbatim* from Moran et al. [7]. Therefore, there really is nothing to check out here... or we just did not understand what the Referee is referring to.

- page 6, lines 154-155: "nothing precludes the application of STORM to other (small) catchments in any climatic zone," [that need to be supported by evidence!!!]

  R/. Please see our reply to your first two remarks/suggestions, and/or to Anonymous Referee #1's item (3a) in this letter.

- Figure 1 (caption), line 3: "$11.37\,\mathrm{mm}\cdot\mathrm{h}^{-1}$" [Please, state the year or years if it is an average]

  R/. The year is right there in the figure (panel **a**; $2000.08.22\ 15{:}59{:}47$). Nevertheless, this year is purely incidental, i.e., a by-product of the simulation (which is clearly indicated at the beginning of the sentence *"One large simulated storm. . . "*). Therefore, we don't feel compelled to explicitly tie any simulation to a specific year as this would be truly misleading, and beyond the scope and focus of our stochastic simulation model.

- Figure 1 (caption), lines 3-4: "Cumulative seasonal distribution of 116 storms for the wet season, i.e., from June through October." [Please, state the year or years if it is an average]

  R/. The same argument of our previous reply holds here. "116 storms" was updated to "112 simulated storms" in Figure 1's caption, nonetheless. (116 was for a different case presented in the first submitted draft.)

- page 8, line 204: "mm" [mm/h?]

  R/. The sentence *"Figure 1 (top panel) shows a simulated storm with a steep β of $\sim 0.18\,km^{-1}$, and $I_{max} = 18.77\,mm$."* was changed into "Figure 1 (panel **a**) shows a simulated storm with a steep β of $\sim 0.22\,\mathrm{km}^{-1}$, and $I_{max} = 11.4\,\mathrm{mm}\cdot\mathrm{h}^{-1}$."

  Also (at the end of Section 2.3's first paragraph), the sentence *"Figure 1 (top panel) shows a simulated storm with a radius of $\sim 11\,km$."* was updated to "Figure 1 (panel **a**) shows a simulated storm with a radius of $\sim 7\,\mathrm{km}$.".

- page 21, line 432: "$5\,\%$ larger" [compared to what ?]

  R/. To "the median for the gauge data ($217.4\,\mathrm{mm}$)". It is right there in the sentence (i.e., $217.4 \times 1.05 = 228.3$).

- page 25, line 492: "One where TOTALP is increased by a fixed scalar throughout the whole period," [This seems too simple to me and it does not have any support from climate physics of the area therfore, un realistic. It might be much better the current scenarios applied in IPCC reports]

  R/. Yes, indeed, using the PTOT_SC scalar might not be adequate to simulate the IPCC scenarios. Nevertheless, STORM also offers the _SF scalars (manuscript's Section 2.8), which certainly are adequate for modelling the IPCC scenarios. Please see our reply to your last-'main' remark.

- page 27, line 538: "any climatic zone" [I would be careful with this statement because the manuscript does not demontrate it. I did not any evidence in this manuscript supporting the statement]

  R/. Please see our reply to your first two remarks/suggestions, and/or to Anonymous Referee #1's item (3a) in this letter.

- there are other several places/lines highlighted throughout the aforementioned document (i.e., *gmd-2023-98-referee-report-2.pdf*) but without any comment(s) whatsoever.

Report #3 **Anonymous Referee #4**

- As a model description paper, I found this paper is extremely difficult to follow. You should assume that people are likely to use your model/tool to do some real-world assessment, your paper should be restructured to make sure an average person with proper knowledge can use the model directly. Unfortunately, I don't think this is the case even though I have sufficient knowledge and modeling experience in hydroclimate modeling.

R/. We disagree with the unnecessarily harsh criticism of Anonymous Referee #4. We don't feel this work is difficult to follow, and proof of that is the positive and very respectful comments we have received so far from three Referees (two of whom we have addressed in this letter). We are very confident that our paper is well structured, and any user, even from those outside the academy, will have no outstanding issues running our model (see for instance, our reply to the second remark of Referee #3; and to comment (3a) of Anonymous Referee #1 –in this letter–). We urge Anonymous Referee #1 to have a look at STORM v2's repository, which offers an even simpler guide on how easy and straightforward it is to run STORM.

Given that the majority of the Referees have agreed that our work is valuable, we will mainly address the comments of Anonymous Referee #4 that fall in line with this majority view.

- The title is extremely confusing! Such confusion is also reflected in the length and difficult-to-follow main text. First of all, I don't your model is SIMPLE at all based on your description in the main text, so you should remove "simple" from your title. Second, you say it is a "rainfall model" -¿ what is the input and output of your model? Is rainfall the output of your model? Or are you taking rainfall data from weather gauges as inputs, and then calculate/estimate/summarize some characteristics about historical rainfall events? This should be clearly described. Third, "stochastic" is usually used to address uncertainties, what kinds of uncertainties is your model able to address? How do you reflect the uncertainties associated with climate change? Lastly, "the impacts of climate and climate change at and near the land surface in gauged watersheds" -¿ if you want to assess the impacts of climate change, you should use climate scenarios as inputs, but I don't think this is the case here. Using a simply scaling factor for total precipitation is not acceptable at all for extreme rainfall analysis due to the complexity and nonlinear nature of precipitation events. Future climate scenarios from acceptable sources should be used for testing. What do you mean by "at and near the land surface"? If you are talking about temperature, this is okay, but here you are referring to rainfall. Weather stations are set up at 1.5m or 2m above the ground to capture the rainfall, but the rainfall received on the ground is essentially the same as the amount received at a height of 1.5 or 2m. The authors emphasize "gauged watersheds" -¿ the model is only applicable to gauged watersheds? What about ungauged ones? For gauged watersheds, we have all the observational data to conduct different kinds of hydrological analyses, what are the advantages of your model? Certainly, the title needs to be rewritten, but mostly important, all the comments raised here are linked to the main text. I don't think the quality of the paper is good enough for a publication at this moment. Substantial revisions are still required.

  R/. Pertinent changes to the title were already addressed in reply to to item (3) in **Reply to Anonymous Referee #1** (first iteration). The model is really simple (just five parameters; please see reply to comment (3a) of Anonymous Referee #1 –in this letter–). Rainfall is the output of our model (this was made clear throughout the entire Section 1, and we also we presented its output in the manuscript's Figure 1, panel **a**; and in STORM's repository). In our case, STORM uses a stochastic approach to address the uncertainties of rainfall by randomising key rainfall variables, captured within historical records and then modelled via probability density functions (PDFs). STORM's focus is on modelling plausible sequences and spatial distributions of rainfall rather than explicitly simulating climate change. Nevertheless, the model has the ability to take into consideration potential future climatic impacts (please see our reply to Referee #3's last-'main' remark). Our use of 'at and near the land surface' refers to the impact of rainfall on hydrology, rather than where rainfall is measured. It is not recommended to run our model in ungauged catchments, unless one can be confident that the pdfs used for simulation are directly transferable to a new location. As any other model requiring any sort of input (you might call them empirical-stochastic), STORM needs gauge data to construct the PDFs so it can model rainfall (all this is clearly stated throughout the entire manuscript; see also reply to first remark of Referee #3, and item (3a) of Anonymous Referee #1).

- The authors claim that the STORM v.2 is an upgrade of the previous version, then you should include a detailed table or plot to compare the differences between the previous version and the current version. Your description in this paper should be mainly focused on the new functions/capabilities of the STORM v.2, as well as the testing and validation of these new functions.

R/. The current version of the manuscript contains all what the Referee here suggests. Right from the first paragraph of Section 1, we state the differences between this version and the previous one, e.g. *"a) treats rainstorm intensity and duration as joint variables in a copula framework, rather than as independent variables, which overcomes a shortcoming in the previous version of the model;. . . d) accounts for storm start date-times from the perspective of circular/directional statistics, which supports more realistic diurnal and seasonal timing of rainfall;. . . "* Then, we go and explain in detail these updates respectively in Section 2.6 and 2.7 (for instance).

- Figure 5 clearly shows a poor performance of the model. The scattered dots should be at least aligning with the diagonal line to indicate an acceptable performance. Here the simulated rainfalls (vertical axis) tend to present a horizontal line, which indicates that the simulations are NOT related to the measured values. In other words, no matter what values you have for the measured rainfalls (between 0 and 500mm), the simulated values are always fluctuating around 200mm. Similarly, Figure 8 also demonstrates the same issue with the model.

  R/. We disagree with the Referee. Yes, The simulation exercise(s) presented in Figure 5, show (overall) a poor performance (especially when recorded values were extremely wet and/or dry). Nevertheless, Figure 5 now also presents 'gray' markers/years indicating STORM simulations considered good/acceptable. Similarly Figure 8 is a clear example in which using our climatic drivers can improve the performance of simulations, which previously (extreme cases in Figure 5) were poor.

- Figure 7: as you move forward from 2000 to 2022, the differences between validation and gauges are getting bigger -¿ also an indication of the poor performance of your model? Why the tails of boxplots for RG012 are significantly longer?

  R/. It is clearly explained in Figure 7's caption. Please read the lines *"This plot is equivalent to Fig. 6b, except that here we force the sampled maximum storm intensity (MAXINT) to be $3.5\,\%$ lower than the previous year. Thus, and after 23 years of simulation, the (mean) decrease in maximum storm intensity is $77\,\%$ (less)."*. Tails of all simulated boxplots are 'significantly' longer because not only *"...STORM allows for plausible storm intensities larger and smaller than those (ever) recorded by the gauge network (see sup. Fig. B6a where..."*, which we state at the end of our manuscript's Section 3.1; but also *"the logarithmic scale of the y-axis"* (as we duly noted in Figure 7's caption).

Sincerely yours,

Manuel F. Rios Gaona

REFERENCES

[1] Bell, V. A. and Moore, R. J.: The sensitivity of catchment runoff models to rainfall data at different spatial scales, Hydrol. Earth Syst. Sc., 4, 653–667, https://doi.org/10.5194/hess-4-653-2000, 2000.

[2] Benoit, L., Allard, D., and Mariethoz, G.: Stochastic Rainfall Modeling at Sub-kilometer Scale, Water Resour. Res., 54, 4108–4130, https://doi.org/10.1029/2018WR022817, 2018.

[3] De Luca, D. L. and Petroselli, A.: STORAGE (STOchastic RAinfall GEnerator): A User-Friendly Software for Generating Long and High-Resolution Rainfall Time Series, Hydrology, 8, https://doi.org/10.3390/hydrology8020076, 2021.

[4] Goodrich, D. C., Keefer, T. O., Unkrich, C. L., Nichols, M. H., Osborn, H. B., Stone, J. J., and Smith, J. R.: Long-term precipitation database, Walnut Gulch Experimental Watershed, Arizona, United States, Water Resour Res, 44, https://doi.org/10.1029/2006WR005782, 2008.

[5] Keefer, T. O. and coauthors: Southwest Watershed Research Center and Walnut Gulch Experimental Watershed, Tech. Rep. SWRC Publ. Reference No. 1588, Southwest Watershed Research Center, 2000 East Allen Road, Tucson, AZ 85719, URL http://www.tucson.ars.ag.gov/unit/publications/PDFfiles/1588.pdf, 2007.

[6] Kim, D., Cho, H., Onof, C., and Choi, M.: Let-It-Rain: a web application for stochastic point rainfall generation at ungaged basins and its applicability in runoff and flood modeling, Stoch. Env. Res. Risk A., 31, 1023–1043, https://doi.org/10.1007/s00477-016-1234-6, 2017.

[7] Moran, M. S., Holifield Collins, C. D., Goodrich, D. C., Qi, J., Shannon, D. T., and Olsson, A.: Long-term remote sensing database, Walnut Gulch Experimental Watershed, Arizona, United States, Water Resour Res, 44, https://doi.org/10.1029/2006WR005689, 2008.

[8] on Climate Change (IPCC), I. P.: Summary for Policymakers, pp. 3–32, Cambridge University Press, 2023.

[9] Osborn, H. B.: Timing and duration of high rainfall rates in the southwestern United States, Water Resour Res, 19, 1036–1042, https://doi.org/10.1029/WR019i004p01036, 1983.

[10] Osborn, H. B. and Lane, L.: Precipitation-runoff relations for very small semiarid rangeland watersheds, Water Resources Res., 5, 419–425, https://doi.org/10.1029/WR005i002p00419, 1969.

[11] Renard, K. G. and Keppel, R. V.: Hydrographs of Ephemeral Streams in the Southwest, J. Hydr. Eng. Div-asce., 92, 33–52, https://doi.org/10.1061/JYCEAJ.0001419, 1966.