December 10^{th}, 2023.

Reply to Anonymous Referee #1.

Dear Referee,

We thank you for your valuable and constructive comments which have undoubtedly improved the quality of our work. Please find below a detailed reply, from paragraph (1) onwards, to each one of them.

(0) Sochastic rainfall generators are important tools for assessing the potential impact of climate change on hydrologic regimes across the world. They are especially valuable when bottom-up or scenario discovery approaches are used, i.e. those where the feasible input is explored. Gaona et al. present here the second version of their STORM tool in this context. The paper is suitable and interesting, but I have a few constructive comments that I believe require addressing. I list them briefly below.

(1) The title of the manuscript is confusing. First, the title mentions a decision-support tool and not a stochastic rainfall model. To me, the title does not fit the content of the paper. A decision support tool help stakeholders during their decision-making process. I see no such tool described in this manuscript. STORM could be part of a decision support tool, but by itself I do not see how it fits this characterization. In fact, the term 'decision' is not mentioned anywhere in the document apart from title and abstract. Second, the authors introduce a stochastic rainfall model, which is great. Why do you not state this in the title? Please change the title to be more suitable to the content of the manuscript. Second, the term gauged watershed is confusing given that it is not clear whether gauging refers to rainfall or streamflow.

R/. This is a fair comment. We felt this model could serve as a decision-support tool for a wide range of applications. However, since it is not explicitly discussed in this article, we propose changing the title to: <u>STORM v.2: A simple, stochastic rainfall model for exploring the impacts of climate and climate change at, and near the land surface in gauged watersheds</u>.

(2) The introduction section makes essentially not reference to any other stochastic rainfall model than the one discussed here. While I do not expect a full review, I believe it is paramount to understand what other tools exist and how STORM v2.0 compares to it.

R/. The two paragraphs below, in which we now acknowledge other existing tools and how our model compares against them, will be inserted/appended at the end of section 1 (i.e., Introduction), and before section 2 (i.e., Data and Methods).

STORM is our contribution to the wealth of Stochastic Rainfall Generators (SRGs) currently available. A state-of-the-art review on SRGs lies beyond the scope of this work. Nevertheless, here we briefly acknowledge some of the vast work carried out in this field during the last two decades. Stochastorm, developed by Wilcox et al. (2021), is a high spatio-temporal ("on the order of kilometers and minutes") SRG for convective storms. It is event-based (like STORM), built upon a meta-Gaussian framework, and also able to generate rainfall fields. Vu et al. (2018) evaluated the performance of five Stochastic Weather Generators (SWG), namely: CLIGEN, ClimGen, LARS-WG, RainSim and WeatherMan. Nevertheless, Vu et al. (2018) focused their analyses, for the aforementioned five models, on characteristics of rainfall such as occurrence, intensity, and wet/dry spells, over three climatic regions around the globe. Unlike STORM, which only focuses on rainfall, SWG (e.g., Papalexiou et al., 2021; Peleg et al., 2017) are frameworks build to also model climatic variables other than rainfall, e.g. temperature, wind velocity/direction. Like STORM, another open-source framework is that of Benoit et al. (2018), which generates conditional or unconditional high-resolution (100 m; 2 min) continuous rain fields over small areas. Kleiber et al. (2012) developed a Gaussian-processes framework that allows the generation of gridded rainfall fields (accounting for uncertainties in their estimates too). Similar to STORM, their model captures local and domain aggregated rainfall from daily to seasonal to annually scales. Other open-source SRG frameworks, both based on Neyman-Scott process, are STORAGE (STOchastic RAinfall GEnerator; De Luca and Petroselli, 2021) that generates long and high-resolution (1 or 5 min) time-series; and NEOPRENE (Neyman-Scott Process Rainfall Emulator; Diez-Sierra et al., 2023) that simulates rainfall from hourly to daily scale. Like STORM, NEOPRENE is also coded in Python. Other SRG based on Poisson process are: RainSim V3

(Burton et al., 2008), designed for catchments up to 5000 km^2 and time scales from hourly to yearly; and Let-It-Rain (Kim et al., 2017) that generates stochastic 1-h rainfall time series. AME (Advanced Meteorology Explorer) is a SRG for high spatial-resolutions at daily temporal scales. Developed by Dawkins et al. (2022), AME is based on a Hidden Markov Model (HMM), able not only to capture observed rainfall but also meteorological drought behaviour. STREAP (Space-Time Realizations of Areal Precipitation; Paschalis et al., 2013) is also a high-resolution (10 min to daily in time; few tens-of-km in space) SRG. Both Dawkins et al. (2022) and Paschalis et al. (2013) provide a plethora of references to alternative stochastic frameworks in rainfall simulation. The latter also provides a detailed review on advances on stochastic spatio-temporal rainfall modelling, grouped into for categories, namely, point process theory, random fields theory, multifractal processes, and "multisite temporal simulation framework".

Our model differs from the aforementioned ones in several features. STORM is an open-source framework (accessible via it repositories), and build upon a free/open-source software platform (i.e., Python). This implies that anybody, from any computational platform, can modify and improve STORM accordingly to their needs. Some of the SRGs here cited are not open-source; and those that are freely available (e.g., De Luca and Petroselli, 2021; Benoit et al., 2018; Kim et al., 2017) were built upon commercial applications (i.e., not open-source software). In our opinion, this is a key factor that plays against their widespread usability and/or deployability. Most of the SRG parameterize the 'inter-arrival' time (i.e., time between storms) via a PDF. STORM, via its circular framework, indirectly, and rather straightforwardly, characterizes this parameter by instead modelling storm starting dates and times (see Sec. 2.7). Through a couple of scaling factors (see Sec. 2.8), STORM is able to simulate rainfall fields from potential climate change scenarios. This is a modular functionality not present in many of the SRG.

(3) The authors use Walnut Gulch as the experimental case study. From the text, I see no mention that STORM v2.0 has been developed specifically for arid regions. So why do the authors test their model only on this very particular case study? The authors should justify this selection. If STORM v2.0 is not specific to arid catchments, then I would expect to see more examples in other climatic domains. I would also expect to see some justification regarding how all new elements of the model can be tested in this watershed, e.g. the orographic effects. Also, hardly any watersheds are as densely gauged for rainfall as Walnut Gulch is. How does gauge density matter for the results obtained by STORM v2.0?

R/. Strictly speaking, STORM was developed for any (small) catchment regardless of its climatic domain (i.e., arid or otherwise). We built STORM upon the richness of the dataset curated by the WGEW (Walnut Gulch), which afforded opportunities to explore details of different rainfall characteristics over a long time series in a setting where there is high spatial and temporal variability in rainfall. Testing our model on any other catchment, although very feasible even with limited gauge data, will require another readily available dataset similar (in form –not extension–) to that of the WGEW (to enable an adequate parameterization). We fully welcome other researchers to apply this model in other catchments. Hence, we let out of the scope of our current work not only the testing of our model on another catchment but also the influence of the amount of data (gauge density) on its performance. All the above is stated in the following paragraph, which will be 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]).

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.

With regard to testing of new elements of the model (e.g. orography, like the Reviewer suggests); we indeed performed such analyses. Nevertheless, given its limited influence within the Walnut Gulch catchment, we decided not to include these results in our first submitted version of the manuscript. However, given the suggestion of the Reviewer, we will now insert/append the following text (which makes reference to the results of STORM when accounting for orography) to the last paragraph sub-section 2.8 (i.e., Scaling

Factors & Orographic Stratification). [The figure accompanying the text (i.e., Fig. 1 in this letter) will also be included as supplemental figure in Appendix B.]

Supplemental Fig. 1 shows the capability of STORM to account for orography, in this case, the simulation of different sets of storms at three elevation bands, e.g., up to 1350, between 1350 and 1500, and above 1500 m.a.s.l. We found no significant differences in simulated or measured storms among any of these three bands, which can be attributed to the low orographic gradient within the WGEW.

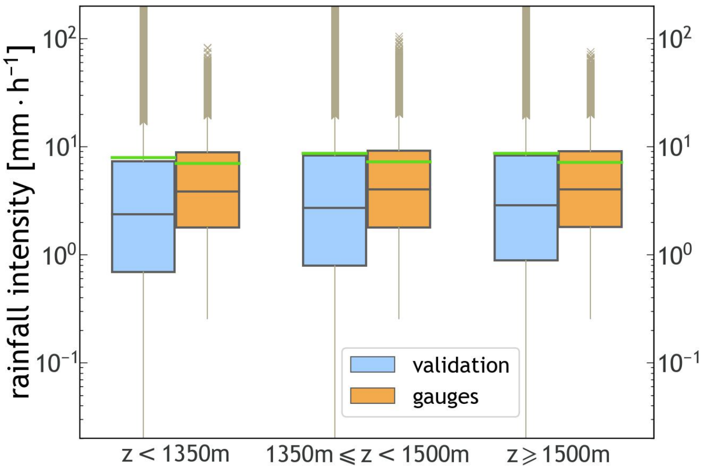

Figure 1: Distribution of storm station/pixel-based intensities by elevation bands (i.e., orography). Blue is for the validation dataset, whereas orange is for gauge data. The green lines represent the mean of the distributions. Please note the logarithmic scale of the y-axis. The x-axis, from left to right, indicate three elevation bands for which analyses were carried out, i.e, up to 1350 m.a.s.l., between 1350 and 1500 m.a.s.l., and above 1500 m.a.s.l. Given the low gradient in relative orography within the WGEW, rainstorm variability with regard to orography is not a stark characteristic in this case.

(4) I am a bit confused by the poor performances (e.g. Fig. 5). Why is there no correlation between estimated and observed seasonal rainfalls (Fig. 5)? Do other stochastic rainfall models have the same problem? I miss a discussion of the ability of STORM v2.0 in comparison to what other models of the same type can do. Also, the authors state that "The scatter plot presented in Fig. 5 clearly shows STORM's innability to depict extreme stormy seasons, either wetter or drier than those in the historical distribution" - is this not a requirement for a tool like this one?

R/. There certainly is correlation between estimated/simulated and observed seasonal rainfall, e.g. Figure 4, and Figure 6 (especially panel b). However, it should be noted here that stochastic models are not designed to obtain a deterministic fit between observed and simulated data because the stochastic simulations will always contain far more variability (based on multiple realizations) than what occurred historically, and this is an inherent strength of a stochastic approach. Furthermore, we would not expect the stochastic simulations to reproduce a historical time series of rainfall, since the stochastic procedure contains no element of real time (each season/year is computed independently unless a trend in rainfall intensity or totals is defined by the user). The poor correlation presented in Figure 5 indicates the relationship between historical observations over 23 years and all simulated runs (each one comprising 30 simulations, already averaged per station). It is always challenging to pin down model efficacy to a one-number metric for this sort of exercise. In fact, it does a disservice to our model to present such a poor performance metric. Nevertheless, we chose to include Figure 5 as its relative bias is equivalent to that of Figure 4 (obtained in a different way); and as a comparative counterpoint to Figure 8. Figure 5 ultimately reflects the (good) intrinsic performance of STORM's design, i.e, reaching the median (or any other set up metric). If we disaggregate the evaluation of Figure 5 by year, e.g., we compute coefficients of determination for markers of the same color/year, we would end up with ρ^2s ranging from 0.0006 to 0.3519 (bear in mind that ρ^2 is the square of the coefficient of correlation); thus indicating years with much better performance than others (see Fig. 2 in this letter). From Fig. 2 we can see that most of the evaluated years perform well (when accounting for seasonal rainfall averaged over the WGEW). But we must emphasize again here that the model is not designed to reproduce accurate time series of rainfall compared to the historical data but to produce multiple plausible rainstorms while remaining faithful to the 'global' statistics (meaning of the catchment-wide rainfall stats). For that purpose, one could use a deterministic model.

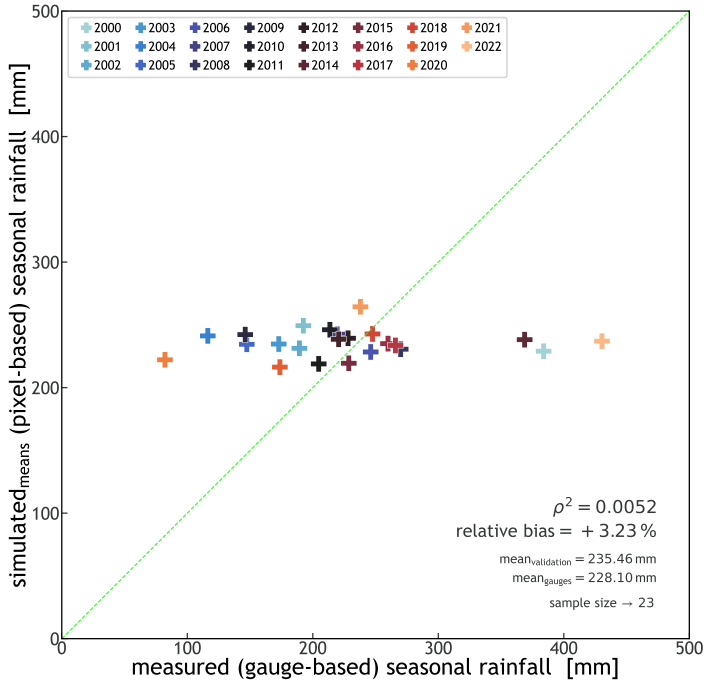

Figure 2: Scatter plot of simulated (mean) seasonal rainfall against measured seasonal rainfall. Each marker/cross represents the mean of all seasonal totals of 30 simulations of all stations, and the actual seasonal total recorded for that location (x-axis). The color scale varies over the 23 simulated years (from 2000 through 2022). Within the plot, it is indicated the coefficient of determination (ρ^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. The green line indicates a $1:1$ line.

Unlike Vu et al. (2018), the comparison of STORM's performance against other SRGs is beyond the aim of this work. Still, the poor performance on overall statistics is not uncommon in this type of model/framework. For instance, (Kim et al., 2017) showed in their work also a poor performance in the auto-correlation between observed and simulated rainfall series (please see their Fig. 10, third column). There, and very similar to the behaviour displayed in our Figures 5 and 8, simulated values tend to be grouped around a central value, whereas observed data is widely dispersed along its axis. (Kim et al., 2017) also indicate that a model reproducing mean rainfall (and variance) much better than auto-correlation (such as is the case with STORM) is more robust for practical applications, given that the basic watershed response is more sensitive to the former than to the latter (i.e., auto-correlation and probability of zero rainfall).

The paragraph below, which highlights the aforementioned point/discussion, will be inserted/appended at the end of the third paragraph of sub-section 3.1 (i.e., Evaluation of STORM).

Having a low (auto)-correlation is not considered a bad feature in STORM or other stochastic models. As suggested and demonstrated by e.g., (Kim et al., 2017), a model performing better in metrics such as mean rainfall (and variance) rather than in auto-correlation (and probability of zero rainfall) is more suitable in practical applications because watershed response variables like runoff volume and peak flow are much more sensitive to the former than to the latter. We emphasize again here that STORM is not designed to reproduce accurate time series of rainfall (compared to the historical data) but to produce multiple plausible rainstorms while remaining faithful to catchment-wide rainfall "global" statistics.

Finally, STORM does have 'switches' to simulate wetter and drier conditions, namely, PTOT_SC and/or STORMINESS_SF. This is stated as *"The modelling of teleconnection phenomena/patterns in STORM was beyond the scope of this work; but it could be represented empirically by altering the rainfall total PDF (TOTALP) by the scaling factors for particular seasons."* at the end of the paragraph which the Reviewer makes reference to when citing *"The scatter plot presented in Fig. 5..."*. To make this more readable and accessible, the sentence "The scatter plot presented in Fig. 5 clearly shows STORM's inability to depict extreme stormy seasons, either wetter or drier than those in the historical distribution (i.e., a very low coefficient of determination ($\rho^2 = 0.0028$))." will be rephrased as "The modelling of teleconnection phenomena/patterns in STORM was beyond the scope of this work. Nevertheless, such patterns could be empirically represented for particular seasons by altering either the rainfall total PDF (TOTALP) or maximum storm intensity (MAXINT) respectively via the scaling factors PTOT_{-} or STORMINESS_{-} (see Sec. 2.8). For a control run, i.e., one without climate drivers (see Sec. 3.2), the scatter plot in Fig. 5 shows STORM's innability to depict extreme stormy seasons either wetter or drier than those in the historical distribution (i.e., a very low coefficient of determination; $\rho^2 = 0.0028$).". [Please see our reply to item 5.c. in **Reply to Anonymous Referee #2**, in which we present an update on Figure 5.]

(5) The figure captions would benefit from more detail so that the reader does not have to read the manuscript text before understanding the figures. E.g. Figure 7 shows a particular experiment where increasing deviation is the goal. This is not clear from the caption. So please include some statement regarding the point of the figure in the caption to make it easier for the reader.

R/. The current caption of Figure 7 will be replaced by paragraph below.

Yearly boxplots for the validation (blue), and gauge (orange) datasets. The boxplots represent the distribution of storm intensities of three stations, i.e., RG012, RG042, and RG072 inside the WGEW (respectively top, middle, and bottom row). In all rows, the green line within each boxplot represents the mean of the distribution. Please note the logarithmic scale of the y-axis (i.e., rainfall intensity). Supplemental Fig. B8 shows the (sparse) location of the aforementioned gauges. 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).

(6) The authors conclude that "STORM's current weakness is its innability to account for other local hydrometeorological patterns, and global teleconnections that may contribute to intra- and inter-seasonal rainfall variability" - However, there is not concluding discussion how this might be rectified in the future and how this influences the application of the software. Is this not a problem?

R/. We do not see this as a pressing problem of the actual version of STORM. Besides, its identification and modelling from only gauge data (which if the core premise of STORM) is not so straightforward, if at all possible. Currently, we are focusing on upscaling STORM from the catchment to the regional scale (this was stated in the "Constraints and Recommendations" section).

The sentence *"Nevertheless, STORM's flexibility allows to roughly model the eventual impact of such teleconnections via the scaling factors $PTOT_{-}$ or $STORMINESS_{-}$ (see Sec. 2.8)."* will be added at the end of the paragraph the Reviewer refers to (i.e., last paragraph of section 5; Constraints and Recommendations).

Sincerely yours,

Manuel F. Rios Gaona

REFERENCES

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.

Burton, A., Kilsby, C., Fowler, H., Cowpertwait, P., and O'Connell, P.: RainSim: A spatial-temporal stochastic rainfall modelling system, Environ. Modell. Softw., 23, 1356–1369, https://doi.org/10.1016/j.envsoft.2008.04.003, 2008.

Dawkins, L. C., Osborne, J. M., Economou, T., Darch, G. J., and Stoner, O. R.: The Advanced Meteorology Explorer: a novel stochastic, gridded daily rainfall generator, J. Hydrol., 607, 127 478, https://doi.org/10.1016/j.jhydrol.2022.127478, 2022.

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.

Diez-Sierra, J., Navas, S., and del Jesus, M.: NEOPRENE v1.0.1: a Python library for generating spatial rainfall based on the Neyman–Scott process, Geosci. Model Dev., 16, 5035–5048, https://doi.org/10.5194/gmd-16-5035-2023, 2023.

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.

Kleiber, W., Katz, R. W., and Rajagopalan, B.: Daily spatiotemporal precipitation simulation using latent and transformed Gaussian processes, Water Resour. Res., 48, https://doi.org/10.1029/2011WR011105, 2012.

Papalexiou, S. M., Serinaldi, F., and Porcu, E.: Advancing Space-Time Simulation of Random Fields: From Storms to Cyclones and Beyond, Water Resour. Res., 57, e2020WR029 466, https://doi.org/10.1029/2020WR029466, 2021.

Paschalis, A., Molnar, P., Fatichi, S., and Burlando, P.: A stochastic model for high-resolution space-time precipitation simulation, Water Resour. Res., 49, 8400–8417, https://doi.org/10.1002/2013WR014437, 2013.

Peleg, N., Fatichi, S., Paschalis, A., Molnar, P., and Burlando, P.: An advanced stochastic weather generator for simulating 2-D high-resolution climate variables, J. Adv. Model. Earth Syst., 9, 1595–1627, https://doi.org/10.1002/2016MS000854, 2017.

Vu, T. M., Mishra, A. K., Konapala, G., and Liu, D.: Evaluation of multiple stochastic rainfall generators in diverse climatic regions, Stoch. Env. Res. Risk A., 32, 1337–1353, https://doi.org/10.1007/s00477-017-1458-0, 2018.

Wilcox, C., Aly, C., Vischel, T., Panthou, G., Blanchet, J., Quantin, G., and Lebel, T.: Stochastorm: A Stochastic Rainfall Simulator for Convective Storms, J. Hydrometeorol., 22, 387–404, https://doi.org/10.1175/JHM-D-20-0017.1, 2021.