Reviewer 2:

Summary:

This study introduces EGGS-WG, a new open-source stochastic weather generator based on the ERA5-Land dataset. The model can simulate weather over gridded land regions worldwide, extending previous approaches that were mainly designed for individual sites. It also provides hourly temperature simulations and includes dew point temperature and relative humidity. Validation across five regions with different climates shows that the model generally reproduces regional averages and spatial patterns well. However, it has clear limitations, including systematic biases in extreme precipitation and weaker performance in representing tropical weather dynamics in New Guinea.

Overall, the study is innovative, useful, and clearly presented. I recommend minor revisions to address the following comments:

Thank you for the positive comments. We have tried to address your comments as thoroughly as possible. In conjunction with comments from the other reviewer, we have also substantially improved the discussion around existing limitations.

Introduction section:

1. Line 80: “The full code base for EGGS-WG are available on a….” - Please use “is” instead of “Are.”

We have updated the line as suggested

2. Line 70: The Introduction states, “if our model fails to accurately handle extremes,” suggesting that this is only a possibility. However, the Abstract and Results already show that the model does not accurately simulate the most extreme precipitation values and tends to overestimate them in most regions. Therefore, the conditional wording is misleading. This limitation should be presented as a known finding that the study aims to address, rather than as a hypothetical Issue.

Wording has been adjusted in the introduction to make it clear this issue was identified. We have also altered some of the conclusions to reflect this point.

3. The Introduction presents the model’s global land coverage as a major advantage, stating that the parameters are “always solvable.” However, this claim should be qualified because the model’s performance is not consistent across all regions. The Results section reports significant biases in the New Guinea test region, including reduced variability and poor representation of spatial dynamics. These findings are important limitations for a model presented as a global tool and should be acknowledged in the Introduction.

Our wording here has been changed to reflect our intention that this sentence was specifically about being able to calculate parameters for the precipitation distribution. This holds true due to the estimate approaches used and range of precipitation distribution

shapes. The issues with data simulation of the New Guinea region are unrelated to the fitting of these precipitation parameters, so we hope the change in text makes this clear. There is a new section discussing the New Guinea Region that further explores this. We have also updated our late discussion of the fitting process as discussed in other comments. We have also added more discussion of limitations to the end of the introduction.

4. The Introduction presents the addition of dew point temperature as an improvement. However, it does not acknowledge that the validation results show a consistent underestimation of the cross-correlation between air temperature and dew point temperature. This limitation should be mentioned in the Introduction to provide a more balanced description of the model's performance.

Along with the above comment we have added some discussion of limitations to the introduction.

Model design section:

5. Line 105: "a single order chain has been show to fail to produce" - Please use "shown" instead of "show."

We have updated the line as suggested

Implementation section:

6. Line 205: "Expansion to other ERA5-Land variables... are planned"- use "is planned" instead of "are planned"

We have updated the line as suggested

7. Line 220: "This trend lines can then be inverted..." - Use "These trend lines" for multiple trend lines, or "This trend line" for a single trend line.

We have updated the line as suggested

Line 240: "Restricting α to a value between 0.5 and 0.9 has also shown to improve the results in our internal testing."- The authors state that restricting α to values between 0.5 and 0.9 improved the results in their internal testing. Was this testing conducted globally or only in selected regions? Please clarify the geographical scope of the testing and indicate whether this parameter range performs consistently across different climatic regions.

We agree that this was a currently under-described in the text. The constraint here is less complex than implied. Alpha must always be between 0 and 1 by definition. Because the precipitation function is a mixed exponential, constraining alpha to 0.5 to 1 will have no implications for the solution but does appear to improve the convergence of the solver. Restricting the value as less than 0.9 mostly ensure that the solver will not resolve to a

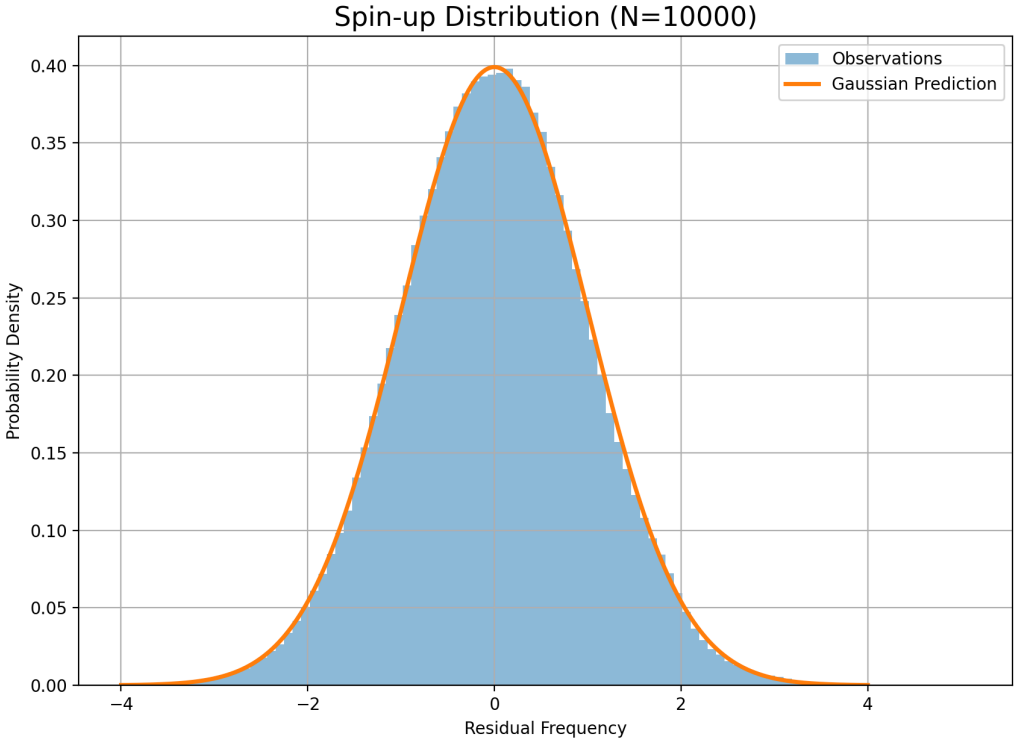
single exponential distribution. The corresponding text has been updated to clarify this as well as adding some details around the testing.

Line 295: “...adjustment is necessary for them to be adjusted”. - The phrase “for them to be adjusted” is redundant and should be removed or rephrased.

We have updated the line as suggested

Section 3.6: The model uses a fixed 1,000-day spin-up period to reach a random state, starting from completely dry conditions with temperature residuals set to zero. This simplified initialization may not represent the full range of realistic climatic starting conditions. The authors should justify this approach and discuss whether the results are sensitive to the chosen initial state and spin-up duration.

We agree that this is a sensible concern, so we have added the below figure to the supporting information and changed some corresponding text. This figure shows the distribution of the full set of temperature residuals from the end of the spin-up period from 10000 different spin-up simulations. What we see is that the starting residuals nearly perfectly follow a random Gaussian distribution. This means that from the perspective of the temperature fields, our starting data will be perfectly random. However, as the data has been run through the simulator, it should also have a sensible physical structure without unphysical spikes you might see with purely random data.



Results and discussion section:

Line 320: “...generally, the simulation appears to be able to capture to the mean regional properties...”. - Remove the extra “to” before “the mean regional properties.”

We have updated the line as suggested

Line 340: “The spatial dynamics are the very similar for the amount values...”. - Remove “the” before “very similar.”

We have updated the line as suggested

Line 355: “There appears to be some issues that with the New Guinea simulation for long wet streaks...”- Remove “that” and revise “appears” to “appear” to agree with the plural noun “issues.”

Text has been updated to fix the grammar

Line 460: “...maximum values t2m are well generally represented by the model...”. - Change “well generally represented” to “generally well represented.”

We have updated the line as suggested

Line 495: “...individual model runs successfully reproducing the variability within the observations...”. - Add “are” after “runs.”

We have updated the line as suggested

Line 500: “...individual ensemble runs are able to provide sufficiently variable simulations that are unrelated o each other” - Add the missing “t” in “to.”

Typo has been fixed

Line 345: The Canterbury results show that some grid-cell pairs have much lower simulated correlations than the observed values. The authors should clarify whether these outliers are geographically clustered, for example near mountainous areas, or whether they are likely due to random statistical variation.

This is a fantastic idea that the authors did not consider before receiving this comment. Some cursory exploration has not turned up anything obvious enough to be included in the paper at this stage. However, I am currently working with a student on this issue to see if anything extra can be determined using this approach. At this point, any useful result is likely to be such a significant deviation that it probably couldn't be included in this paper. However, if we find something useful, we will make the information publicly available (likely on the GitHub page for the project). This could also possible spur further coding improvements within the model.

Conclusions:

The authors describe the impact of the model's poor representation of extreme values as "minor" because such events are rare. This claim should be tempered. Stochastic weather generators are often used in hydrological risk assessment and flood modeling, where the accurate representation of extreme events is especially important. The systematic overestimation of extremes in most regions, and their underestimation in New Guinea, therefore represents a substantial limitation. The Conclusion should clearly advise users to exercise caution when applying the model to analyses involving extreme precipitation or related risks.

The conclusions have been reworked a bit in response to this comment and comments from the other reviewer. This includes tempering expectations where issues are identified and providing direct user guidance. This includes identifying guidance regarding extreme precipitation. Interestingly, simulations over some additional regions that have been included suggest that these biases are highly regional and not particularly consistent over different climates.