Second-round review of “EGGS-WG: An open source global gridded stochastic weather generator derived from ERA5-Land”

I thank the authors for their response to my review. The revised manuscript is clearly improved. The main contribution is easier to identify, the choice of precipitation distribution is discussed more openly, the New Guinea subregions and additional regions are a useful diagnostic, the fitting and spatial-correlation procedures are better documented, and the practical guidance for users is a welcome addition.

Most of my first-round concerns have been dealt with adequately, and I would support publication after minor revisions. The points below do not require restructuring the paper, but several statements now need to be made more consistent with each other and with the revised figures.

Thank you for the time taken to review the paper and the supporting comments. We agree that several of the points below are significant and need to be addressed before publication (especially points 1, 3 and 4). We have done our best to address each of these comprehensively.

Remaining comments

1. Sect. 4.5 is a strong addition, but its conclusions currently point in two directions. Lines 563–564 state that the occurrence-correlation issue “must” be due to domain size or domain variation, while lines 581–582 conclude that “the particular climate of New Guinea is what causes the model issues”. The good North American results at 625 grid cells and the poorer Indian results suggest to me that the problem more likely comes from a large domain combined with the spatial heterogeneity of a monsoon-affected climate. Whatever interpretation the authors prefer, it should be stated once, consistently, and with hedged wording (“suggests”, “is consistent with”), since these runs are not a controlled attribution experiment. Related to this, lines 444–445 still describe the domain-size question as out of scope, and the abstract and conclusions still frame the validation around the original five regions. These sections should be updated to reflect the new analysis, or the new analysis should be described explicitly as an additional sensitivity test.

Thanks for this comment, in retrospect this section was somewhat incoherent. We have changed some of section 4.5 to make more direct arguments. We agree that the identified biases are clearly more driven by the monsoon climate than by the domain size and the text should now reflect this. Language has also been changed to add certainty. We have removed the out of scope comment and reviewed all of section 4 for similar comments to ensure consistency.

2. Lines 412–415 state that the spell results lack the long-dry-spell bias expected from a first-order Markov chain and that a deeper chain is unlikely to improve EGGS-WG. This no longer fits the revised Fig. 4, where the New Guinea dry-spell 99th percentile is 18 days in the simulation against 27 in ERA5-Land, or Sect. 4.5, which attributes the New Guinea spell problems to the simple occurrence model. I suggest limiting that statement to the New Zealand and Australian domains.

The language in this section has been updated to reflect the newer results related to the New Guinea region and clarify the limits of this result. Additionally, we have added a sentence describing some new results related to a greater depth implementation we have been exploring for future simulations. As these results are very preliminary, we have chosen not to include figures, but from what we have seen it does not appear to help our implementation very much to move to a depth 2 markov chain.

3. I would like the authors to confirm that restricting α to 0.5–0.9 does not constrain the fit. The response states that the lower bound has “no implications for the solution”. That would be true if it were only removing the label-switching symmetry, but Sect. 2.2 also imposes β1 ≥ β2. With both constraints, the larger-scale component must carry at least half the weight, which excludes mixtures in which a minority heavy-tail component sits on a majority of light-rain days. The unadjusted distribution shown in the response (too few days between roughly 1 and 3 mm, too many between 3 and 10 mm) looks consistent with this. If the ordering is not actually enforced in the code, the text should say so.

This is a good question and as far as we can tell your argument is correct. However, on reviewing our source code in preparation for our response, we realised we had made a major mistake in writing up this section. The approach described was one of the approaches we tested that performed consistently well in testing, however ultimately we decided to implement a slightly different approach then what is actually described here. This includes not fixing alpha to be greater than 0.5 but instead between 0.1 and 09. Because of this we have rewritten some of the key details of this section - and are very thankful for the reviewer for inadvertently identifying a major mistake. A review of other parts of the source code confirmed this was the only affected part of the manuscript.

4. The sentence at lines 423–425 describing the extreme-value discrepancy as relatively minor for general simulations should be qualified. It is fair for mean-state or broad weather-sequence uses, but not for flood-risk or high-return-period applications. The user guidance also needs checking against Fig. 5. It gives the bias at the 99.9th percentile as “around 10%”, but the figure suggests something closer to 30–40% for Canterbury and around 50% for Australia. It also frames the problem as starting above the 99.9th percentile, whereas the Results section, and the Conclusions at lines 599–600, place the visible deviations above the 99th percentile. Finally, the Introduction (lines 75–77) says that model and ERA5 biases in extremes “compound”, while the Conclusions correctly note that they go in opposite directions.

We agree that there has been some inconsistency in terms of our framing here, with some of the points you identify here also going from inconsistency through to fully erroneous. We thank you for identifying this issue - it is clearly the result of not appropriately correcting the entire document when we added content for revisions. Specifically, we have:

- The discussion around lines 423 has been updated to better reflect our results
- The 10% estimate has been removed, and a more regionally specific framing has been included in the user guidance
- The framing on 99 vs 99.9% comparisons has been updated to reflect the larger bias for lower rainfall regions
- The references to compounding has been updated in the introduction

5. Please state explicitly, in Sect. 3.7 or at the start of Sect. 4, whether the validation figures use adjusted or unadjusted output. The comparison added to the supplement is helpful, but the main manuscript should not require readers to infer which version of the output is being evaluated.

This has been clarified in the document.

6. The user guidance is now among the most useful parts of the paper. Two additions would help: a caution that monsoon-affected regions showed the weakest performance in this study, and a recommendation that users working in a new region first run a short validation against ERA5-Land or, where available, an independent precipitation product. The informal phrasing (“feel free to email the authors”, “we would be happy to do so”) would be better replaced by a pointer to the GitHub issue tracker or the correspondence address.

We agree that it is a very useful addition and thank the reviewer for suggesting it. We have added discussion of the two issues identified as an additional bullet point. We have also replaced the informal language with more appropriate wording.

Technical corrections and smaller comments

1. Line 347: the diurnal adjustment is said to make EGGS-WG “better match simulations”. Presumably this should be “observations”.

This has been corrected

2. Line 141: “modelling gamma functions” should be something like “using Gamma distributions”; the Gamma function is a different mathematical object.

Thank you for catching this - it is an embarrassing mistake to make

3. Lines 186–189: “again to chosen to match” and “To avoid confusion between each of these different kinds means we clarify” should be rephrased, as should “to ensure that the output the model produces spatially coherent output” (l. 160–161).

These have been reworded for clarity

4. The Zenodo DOI in the Introduction (10.5281/zenodo.15086533) differs from the one in the Code availability section (10.5281/zenodo.15307044). Please make these consistent and confirm that the version of record matches the revised code.

Thank you. All model numberings have been updated to the correct value and all links have been updated to the correct version.