

Geosci. Model Dev. Discuss., referee comment RC2
<https://doi.org/10.5194/gmd-2022-144-RC2>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on gmd-2022-144

Anonymous Referee #2

Referee comment on "A new precipitation emulator (PREMU v1.0) for lower-complexity models" by Gang Liu et al., Geosci. Model Dev. Discuss.,
<https://doi.org/10.5194/gmd-2022-144-RC2>, 2022

General Comments:

The authors present a novel method for precipitation emulation (PREMU) that uses gridded temperature patterns over time to (re)construct global and local precipitation time series by means of principle component analysis (PCA). The method provides high accuracy against existing precipitation estimates, suggesting its potential power for situations where a temperature projection can be easily constructed but a precipitation projection cannot, such as with lower complexity models (LCMs) or novel projections of temperature outside of the most common future scenarios in more complex earth system models (ESMs). The organization of the text does not make this message as clearly as it needs to, though, and there are a few outstanding questions about the methodology as well that need elaboration.

The effort put into defending the ability of PREMU to accurately reconstruct ESM precipitation overwhelms any description of the applicability of the method to LCM temperature data, making that reproducibility appear to be the main message instead. While the accuracy of PREMU in this respect is impressive, it minimizes the importance of this novel method because it gives the impression that no new precipitation has been added. More emphasis on the potential application of PREMU is needed, and better framing of the ESM matching would greatly improve the article as well.

In particular, with PREMU being much less computationally intense than an ESM simulation, this paper should include at least one application of PREMU to an LCM projection of future temperature as an example. The example should also include a precipitation simulation based on that LCM using the traditional linear method so that the audience can see what differences (in theory, improvements) there are between the linear method and PREMU.

Some specific suggestions for providing better framework and organization for the core message are presented below, along with questions and comments about the methodology. Overall, I do believe that this is a valuable new tool for the community that should be shared; it just needs slightly more explanation and a better presentation.

Specific Comments:

Abstract: The mention of LCMs should be mentioned much earlier, along with the issue with linear scaling for precipitation. Then present PREMU as a solution to that issue, followed by the justification statistics from the historical and ESM comparisons. L22 onward can remain as the conclusion sentence. L12 "better estimate and represent precipitation simulated by Earth system models (ESMs)" wording should be changed as there is no justification in the paper suggesting the PREMU simulations are better than the ESM data they are recreating.

Introduction: Generally good, though precipitation is not mentioned until L47. This section should be rearranged to better emphasize the problems associated with precipitation estimation outside of the common ESM future scenarios; it does not have to lead the section but should be mentioned higher up and expounded upon. The final paragraph also needs to be fleshed out more, particularly in the sentence describing section 4 where you need to draw the connection between the presented ESM validation and the potential use with LCMs. The potential addition of a direct LCM-PREMU precipitation example would also be mentioned in these last two sentences.

Data & Methods: With the existing structure in this section, it appears that the Methods might benefit from being its own top-level section, as Methods is the only section of the paper with third-level subsections. There are also many choices (listed in detail below) made in the methodology that are ultimately justified in the Discussion section but present lingering questions when presented alone in this section. Full justification of these choices cannot be done before the Results section, but you should telegraph some of the comments that will be made in the Discussion section so that your audience recognizes you are already aware of some of the potential issues with your assumptions and choices.

L89-90: Calibrating to your endpoint (here, the hottest future temperature scenario) is good in the sense that your validation will only involve interpolation instead of extrapolation, but since we know that the atmosphere can respond nonlinearly with warming, this introduces the concern that this extreme warming scenario may not produce precipitation patterns that are representative of cooler scenarios. This is addressed later (e.g. L203-205, L271-292) but should be acknowledged here.

L91 "we constructed the emulator for each ESM respectively" On first reading this, it sounds like you performed the entire PCA individually for ESM. Upon reading further, it instead looks like only the coefficients differ for each ESM, as Fig. S2+S3 state that the same ten temperature modes are used for all ESMs. This should be stated clearer.

L95 Calibration: This is another section that should be split in half, as most of the first half of the section (L96-177?) focuses on the temperature PCA itself. This description should also use a little bit more elaboration for readers who may not be as familiar with the mathematics of the process.

L107 "used in climate research" requires at least two example references.

L111-112: You further investigate the effect of different lags later in the text, but this statement alone requires some elaboration and at least one reference to justify how earlier months must be considered.

L116-117: something like this should have been mentioned around L89-90, and similar statements (i.e. "discussed in Sect. 4") are needed for other issues.

L118: It's great to have the table for reference, but more of this information should also be in the text itself so readers don't have to go back and forth between the table and the text repeatedly in order to understand the equations.

L127-128: More justification is needed as to why both the historical and ESM versions of PREMU were constructed with 10 modes of temperature. From the four figures cited here, in addition to S13+S14, it's clear that the amount of warming in the chosen temperature data greatly affects how many modes are needed to hit any particular threshold for variance explained. This also goes back to the earlier concern of how the 8.5 scenario, which is so strongly dictated by worldwide warming, might not give reasonable precipitation results for scenarios with more varied temperature modes. This is later somewhat addressed in L275-278, but the stark differences between the first four figures mentioned in L128 will raise questions that should be partially addressed here.

L136,L141: The similarities between these two equations suggest that it should be possible to construct the global coefficients from the gridded ones, i.e. instead of doing the global analysis alone. I know some averages are computed and compared later, but unsure if that is the same as what I'm thinking here.

L159-164: if you say you found a slight difference, you should be more explicit as to what the difference, and potentially offer an explanation for why the difference is present. Without a physical reason given, this subsequent ratio correction feels like a poorly justified mathematical band-aid.

Results: This section seems thorough, but it often states a large number of different quantities in quick succession in many paragraphs, making it somewhat difficult to follow and occasionally feel like not every comparison is being directly stated. I would like to see a more organized pattern of describing results in each paragraph and/or another table stating all values of interest (average precip/yr, precip trend, year to year variance, error, correlation, proportion of grid cells with given error, etc.) for each set of simulations, both trainings and experiments, obs vs linear vs PREMU for historic and each ESM vs PREMU for future.

Fig. 3: correlation and error look like they might be strongly influenced by the extreme changes in the last year or two of the experiment time period. Out of curiosity, since PREMU is not too computationally intensive, could you do some sensitivity analysis where you vary the start and end years for both your trainings and experiments? I'm guessing it might not change much for your results and thus might not need to be put into a later version of this article, but it would be interesting to see.

L190-191: It feels very weird to end a section saying a particular part of the simulation is inaccurate - why not try to compare it to the linear scaling? Is it worse than the linear method, and that's why the linear scaling isn't mentioned here?

L194: This is the first time mentioning the possibility of PREMU working with novel trajectories for GHGs; something that should be mentioned earlier in the paper alongside the extra emphasis on using PREMU with LCM temperatures.

L201-202: While you later suggest that some of better variation simulation with ESMs might be due to topographic complication or aerosols, (L243-L252,) one thing that comes to mind with R values is the standard deviations in the data sets you're comparing. I would like to see some mention of the underlying statistics here.

L204-205: This speaks to the previously mentioned issue of training to the 8.5 scenario, and you do address it later in the Discussion section. Nothing necessarily needs to be added here if you properly acknowledge the issue earlier in the text.

L207-209: The issue of over half of your ESMs doing visibly worse than the others deserves more than the passing comment here, especially as you've devoted a figure (6) to show it. There is a much larger discussion in L254-279, but if you don't say more here, you should at least mention that the discussion is coming later. (Again, telegraph your discussion points so that if a reader starts to question a methodological point or an odd result, they know you're aware of it and are planning to address that concern.)

L212-223: Another paragraph that feels disorderly with the amount of widely varying results strung together.

Discussion: This is a great consideration of all potential issues with the methodology and interesting components of the results previously shown. As stated before, several of these need to be telegraphed earlier in the paper so your audience isn't reading through your results with too many questions. It has now been over six pages since "LCMs" was previously mentioned, which is why the article so far feels like the ESM comparison was the main goal, as opposed to the actual potential for use with temperatures from LCMs and other novel scenarios. Also, your audience should not get to this point without knowing exactly how one of these other experiments would be run, e.g. what would be the "training" precipitation data for an LCM experiment?

L254-279: These are all good suggestions for why some ESMs are better simulated with the PREMU method than others. However, it does raise the question: if we know different ESMs have different schemes relating atmospheric circulation to their precipitation, why was a set of 10 temperature modes chosen to use for all ESM experiments with the only differences being the subsequent coefficients, especially when we know the ESM temperature patterns themselves are not consistent between different ESMs to begin with? You could potentially better capture model-specific ENSOs and other such features this way.

L275-276: This paragraph is a great way to address some of the previously mentioned concerns about training to a scenario with such extreme warming. This particular sentence, though, still doesn't seem entirely reasonable considering how different the 8.5 modes and 2.6 modes are. You do later state that the mode order is slightly shuffled, implying that you have clearly identified similar pairs of modes between the two scenarios; it would be nice to show a side-by-side comparison justifying to your audience that these two sets of modes are indeed similar enough to explain the >95% similarity in coefficients.

L283-292: More discussion of the issue of how much warming in your training affects the fit during the experimental phase, which is good to see. You point out that using a historical/cool training for a warmer scenario would be unwise; it might be better to state the opposite in acknowledgement of how you already pointed out that the 8.5 training generally produces worse results as you go from the 7.0 experiment to the 2.6 experiment. The last sentence here also is potentially very instructive for if/when you include an LCM-PREMU example, i.e. you would train PREMU based on the SSP whose future temperature most closely resembles the future temperatures from the LCM simulation in question.

L304-305: While this is one possible conclusion to draw from the ESM results being more robust versus lag than the historical results were, you did previously posit "an alternative argument" suggesting that apparently better results in ESMs might be due to underrepresentation of topography and aerosols. As these are potential sources of variation, the "robustness" versus lag with ESMs could also be due to the ESMs not showing enough variation for the lag differences to matter.

L309-310: This could also be a potential concern since SSTs have a much larger influence on total atmospheric water vapor than land temperatures do, both by being much more surface area and by being the main source of evaporation. With that consideration, it should follow that PREMU should be at least slightly worse in its precipitation simulation – potentially still very good, but still not as good as when ocean area is also considered. I would like to see the differences between the normal emulator and emulator-land explicitly shown in some manner. You do state that including the ocean is still the preferred method for now; it would be good to expound on that.

Table 3: You have space at the upper left for a proper title and/or stating the units; the latter in particular should be easily read somewhere outside of just the caption.

Figure 4: I would love to see difference plots e-a, e-c, f-b, and f-d.

Figure 5 onward: From here on out, your maps are smaller than earlier; between the small size and the fact that the first red+blue from the middle are similar to the grey, all of the maps here and onward are hard to read. I would strive to rearrange all of these figures so that you are no more than two maps wide per page – and potentially darken those first lighter colors a bit as well.

Figure 7: In trying to make this figure no more than two maps wide, try visualizing it as a 2x2 grid where each quadrant has three maps stacked vertically.

Figure 8: I understand that 8.5 is the training scenario, then followed by 2.6 to 7.0 in that order as the experiments, but it still looks odd that the overall order isn't uniformly increasing or decreasing.

Technical Corrections:

There seems to be a minor inconsistency in the paper between whether 2015 or 2016 is the starting point for the ESM data, L125 vs L132. Please make sure these are better aligned

L117: remove “the” form “discussed in the Sect. 4”