

Mathison et al. describe a new climate emulator that is coupled with a land-based impact model, which together form PRIME. This paper describes a significant advance in the use of FalR, a commonly used 1D climate emulator because it allows users to simulate regional climate using patterns from many different climate models at relatively low computational cost. Furthermore, it allows for rapid impact assessment with its direct integration with an impact model.

I do not have major methodological concerns, but I have several questions and suggestions that I would like the authors to address before recommending publication. The more general ones are written here and several smaller line-by-line suggestions are listed below.

The underlying assumption of matching an ensemble of FalR output with an ensemble of climate model patterns is that there is no relationship between the two. In other words, the patterns in warm models do not look significantly different from the patterns in cold models. In some FalR papers, specific climate models are emulated by using tuned parameters (e.g. Leach et al. 2021, Figures 3 and 4), but here the full range of models of global mean T is then matched with a full range of patterns. To address this, I could imagine a supplemental figure that, for example, plots the pattern correlation between each model and the multi-model mean, against the global mean temperature in a future year (e.g. 2050 or 2100). There may be other ways to do this, but I think a small additional analysis to demonstrate that the pattern is not a strong function of sensitivity would be useful to demonstrate whether this is a limitation or a non-issue. I would be surprised if this was a major problem, but I think it would be worth characterizing for potential future users.

As for other users, I'm wondering whether the pattern scaling component is a stand-alone model. The schematic (Figure 1) shows patterns that include the ocean, whereas the rest of the figures are limited to the land. I could imagine many applications in which a user would want an ensemble of patterns-scaled climate (including the ocean) but is not interested in the land-based impacts. It would be helpful if you could address that kind of hypothetical use case in the text.

Lastly, I think the text about the land model needs to acknowledge that atmospheric carbon (and other variables) have no feedback with the changes in vegetation. Although I still see use in this application, I think this deserves more serious consideration in the text, especially as comparisons are made to CMIP6, where some models do simulate the terrestrial carbon cycle (line 398). This seems like a significant difference between PRIME and some GCMs.

My comment on line 164 about adding anomalies from a 1850-1889 baseline to 1901-1930 climatology may also require some changes if indeed I'm understanding the methodology correctly.

Minor:

Line 4: I think “global picture” is a confusing choice of words. Also there are other climate emulators that have a spatial element, so I’m not really sure what this sentence is trying to say.

Line 5: “general information” is vague

~60: the overview of ongoing efforts is great. I think the machine learning models like Climatebench probably deserve a mention for completeness, since they have similar objectives

66-68: I found this sentence confusing. “this type of input”? Not sure what input is being referenced

98-99: Choices affect the way patterns are selected? Do you just mean users choose the patterns? Also, please define Rose suite.

135: As stated repeatedly, part of the appeal of emulators is the low computational cost. However, the sampling done here is fairly sparse (Figure 3) and the reason given here is to “make the ensemble size manageable”. I think it would be worth dedicating a bit of text to explain this discrepancy.

136: “these are selected using one scenario so that scenarios can be compared against each other” is confusing

147-148: Do the linear regressions include an intercept? If so, how is it treated?
Please also specify one realization of each model (i.e. not an ensemble and not all available)

152: I think there needs to be a summary of this conversion from monthly to 3-hourly data. Not necessarily here, but somewhere.

159: “this is the method currently used in PRIME” - it’s not clear what is being referred to

161: I thought the methods were described quite clearly up until paragraph, where it starts to get confusing. It’s somewhat unclear what belongs to IMOGEN and what does not. I also don’t understand why the diurnal cycle calculation could cause numerical instabilities (166) when there are no feedbacks. And what part of this is 3 hourly and where is the monthly temperature from the pattern scaling coming into play? Is there only one diurnal cycle for each month?

164: Why are 1850-1889 anomalies added to 1901-1930 climatology? If it’s observationally limited (i.e. no data prior to 1900), then it seems like the anomalies should also be relative to 1901-1930? Maybe you could at least demonstrate that this is a reasonable approximation, but it seems inconsistent. Also why did you decide to add it to an observational climatology instead of a modeled one?

Equation 3: I think that 3 should be an exponent - please check the equation

194: typo - mode, not model

213: I don't think you mean "where ESMs have not been run" because the simulations exist

Figure 2 and S1. Caption should specify where the data are from. It says in Fig S1, but not in Fig 2. Could you also mention in the text again why you're subsampling? It hasn't been mentioned in several pages

Figure 4: I don't think the differences should be on the same scale as the anomalies. For example, 4c just shows that most differences are less than 0.6 degrees, but grouping 0.01 to 0.6 together does not make sense when almost all values are less than 0.6. Plus the colors are faint (for the same reason) so they are hard to see.

The figure captions (for 4 and accompanying SI figures) should also specify "anomalies" for the "predicted ensemble means"

Figure 5: I think adding a ratio column (MAE / IQR) would better make the point.

S15: All colors in all of the related figure are shown from light to dark, except pressure in S15. Although it's not wrong, I would recommend reversing the colors for clarity

307: This sentence and several sentences in this paragraph are a little wordy and could be edited for clarity. For example, the reference to Burton et al does not seem to fit well in the context, and "actual values" on line 315 is vague.

345: What does it mean for the pattern scaling technique to be well understood by the literature?

366: I'm a little confused here - you would not expect the values to be identical but "hope for a similar spread". I agree it seems to work impressively well, but there is some disconnect between saying there are several differences (e.g. only using one land model) and yet expecting similar spread. The rest of the paper is about emulating spread when using many models, so why would you expect the spread here to be similar when only using one model? A bit of clarifying text would be helpful.

Figure 9: The caption says "median and uncertainty ranges" but I don't see that in the figure. I am confused about what the lines are - can you clarify in the text and caption? Are they for different pattern scaling from each of the different climate models?

469: You mention that local land-atmosphere feedbacks are not included, but that seems secondary to me given the fact that the land model produces carbon-cycle relevant parameters like GPP that do not feed back into FaIR. As mentioned above, I think that qualification should be added.