



EGUsphere, referee comment RC2  
<https://doi.org/10.5194/egusphere-2022-914-RC2>, 2023  
© Author(s) 2023. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## **Comment on egusphere-2022-914**

Raphael Hébert (Referee)

---

Referee comment on "Understanding pattern scaling errors across a range of emissions pathways" by Christopher D. Wells et al., EGU sphere,  
<https://doi.org/10.5194/egusphere-2022-914-RC2>, 2023

---

### General Comments:

The paper by Wells et al. presents an analysis of the errors arising when using the mean component of MESMER for emulating climate model simulations for different emission scenarios. The paper is well structured and the analysis appears sound, and therefore I find it eligible for publications after revision.

I think it could generally be rewritten more concisely as there are detailed descriptions of results that are sometimes trivial, and description of figures that would be better placed in figure captions. In addition, I think the motivation and usage of the model could be better explained. Particularly, I understood from Beusch et al. (2020) that MESMER served to emulate a single model-scenario (i.e. self-emulation) to generate a large ensemble, e.g. we have model X with scenario SSPyyy and we emulate this single run to obtain a large ensemble with random internal variability. What is thus the purpose of understanding the cross scenario errors presented here? Is the goal to use MESMER for making regional projections? In which case, what is the point of using the 2020-2070 period to set up the emulator and then project the 2070-2100 period? Wouldn't we want to rather use historical and/or idealized simulations or observations to set up our emulator and analyse the errors induced by the historical pattern on emission scenarios (e.g. Geoffroy & St-Martin, 2014; Hébert & Lovejoy, 2018)? I initially expected that the aerosol and GHG patterns would be used to calibrate a two-pattern emulator, but rather we only get insights on the difference between the extrapolation of these two patterns, results which I thought were a bit trivial since we already know that aerosols have a localized impact. Wouldn't it be possible to use those two patterns to emulate future scenarios if we have a decomposition of global mean temperature into aerosol and GHG driven components? I think this would be a more powerful framework since we could then use those patterns to emulate any scenarios given the global mean temperature along with the aerosol and GHG forcing timeseries. It is not necessary for the authors to do this in this paper, but I wanted to outline what I think would be useful to broaden the scope of the study.

### Specific Comments:

Line 47: "forcer pattern" --- I'm unsure about the use of 'forcer' here and elsewhere, shouldn't it be 'forced pattern'?

Line 129: "This study utilises the mean response component of the MESMER model (Beusch et al., 2020), implementing pattern scaling to emulate the spatial annual mean temperature response in a scenario." --- Is it still the MESMER model if we use only the mean component? Then isn't it just a regression of the local temperature with respect to the global one?

Line 137: "This is performed to ensure the global average parameter is very close to 1 K/K, as it should be by definition, when predictor the model on an individual low-emission scenario such as SSP119." --- What global average parameter are we talking about? The global average of the local sensitivities? Why does it matter to smooth or not the local temperature for the regression to obtain an average close to 1? Also, review the formulation of sentence, 'when predictor the model' doesn't sound right. I would also explain here or somewhere else the units of K/K since at first one thinks why don't they just cancel, and well they do, but I understand you wanted to make explicit that this was a local sensitivity of the local temperature to the global one, right?

Line 145: "A given emulation consists of the predictor set – comprising one or many scenarios – and a target scenario." --- I think this could be better explained. Are we talking about a set of model simulations following certain emission scenarios that are used to estimate the pattern, and then one separate scenario with its own set of model simulation is used as target?

Line 156: "temperatures relative to pre-industrial times rise from 1 K in 2015 to 2 K" --- Maybe give the approximate year when the temperature reaches the 2k 'from 1k in 2015 to 2k in ?????'

Line 208: "In hist-aer, the land-ocean distinction is still clear, but the northern hemisphere land is particularly sensitive, due to the historical concentration of aerosol emissions within this region." --- Sensitive isn't exactly the right word right? Are the land region really more sensitive, or is it purely because of the higher aerosol emissions there that the regression slopes are higher? I would consider rephrasing the paragraph to clarify this.

Line 217: "Parts of the NHMLs exhibit a significantly more sensitive response to hist-aer, including the USA, Europe, and east Asia, and the Southern Hemisphere oceans are significantly less sensitive." --- Again, this sounds like sensitivity to aerosols is a local property of the system, but really, the pattern of the response just corresponds to the sources of aerosol emissions. This would likely be outside the scope of this study as it might require more data about the spatial distribution and dispersion of aerosols, but it would be interesting to quantify the actual sensitivity to aerosols taking into account the pattern of emissions (and their dispersion).

Figure 2ab,3ab: I'm not sure the divergent colour palette is appropriate since there is no fundamental difference between values below 1 and above right?

Figure 2d,3d: Wouldn't it be more informative to look at the ratio of the absolute difference over the inter-model spread?

Line 228: "Figure 3 shows the same analysis for SSP119 and SSP585 in a similar way to Figure 2." --- I would complete this sentence with a restatement of what is calculated, something like, if I understand well: '...similar way to Figure 2, i.e. the local temperature series are regressed with the global mean temperature to extract a local sensitivity to global temperature changes.' (could be written more concisely).

Line 251: "Clear, significant differences are therefore found between the temperature response patterns attributable to different historical forcings, consistent with their different spatial patterns." --- I hate to be this guy, but if you say significant, the reader expects a p-value, and you should explain the statistical test used, the null hypothesis considered, etc.

Figure 3: What period is used to train the pattern? In the methods it is said that the first 50 years are used, but I don't think it is said which period the SSP simulations cover. In any case, I think it would be helpful to explicitly state the time period used for training the model.

Line 271: "since the aerosol pattern is more sensitive here than the GHG response, and the Southern Ocean is conversely under-sensitive" --- Again, I'm really not convinced by the usage of sensitivity when it comes to the aerosol pattern. It's not about the sensitivity

of aerosols, but rather the strength and spatial distribution of the aerosol forcing. The Southern Ocean is not less sensitive to aerosol forcing, there are just much less aerosol emissions reaching that region.

Line 305: "Errors are significant in the out-of-sample emulations" --- Again, if you say significant, we expect a p-value, if you don't want to give a p-value, use larger instead, otherwise we would also like to know if the smaller errors of the self-emulation are significant or not, just because they are smaller doesn't mean they might not be significant.

Line 358: "Note the smaller scale on the timeseries error plot." --- Might be more useful to have this statement in the caption.

Line 365: "note the slight variations in the SSP119 column compared to the SSP585 one." --- Do you mean 'smaller' variations rather than 'slight'?

Line 387: "The patterns are similar between SSP119 and SSP126, indicating some consistency between scenarios in this effect." --- Why are only those two scenarios considered for this comparison? Wouldn't it also be interesting to see the pattern for SSP245 with a later peak and drop?

Figure 5: Unclear on what period the patterns used for emulation were calculated.

Line 390: A lot of this paragraph could belong to the caption instead. There were several such instances where the figure was described in the text rather than in the captions, I would consider improving the captions and shortening the text to the results only and avoiding the description of the figures there.

Line 390: I would motivate why those specific region-model-scenario are used, I guess simply to explore problematic behaviours?

Figure 9: What period is used to train the predictor?