Review for

METEORv1.0.1: A novel framework for emulating multi-timescale regional climate responses

This manuscript describes the new spatial climate emulator METEOR. The model represents the spatial responses to greenhouse gas forcings and aerosols forcings separately using impulse response functions & patterns. The model itself is an improvement of classic pattern scaling approaches, thus filling in an important research gap. The method and the validation are overall good, albeit some minor flaws/suggestions. The manuscript is written clearly, especially its storyline. The figures are quite complete, and easy enough to understand quickly. I think that this manuscript would completely fit the scope of GMD, and could be published with minor revisions. I will describe below the suggested revisions.

Minor flaws:

My principal point would be on the validation. The current framework is (1) deconvolving the global signals, (2) validation with the Pearson coefficients (Table 1), (3) deducing spatial patterns. The validation is indeed very good globally, but there is for now no validation of the spatial ouputs of the whole modelling chain. To evaluate the spatial decomposition, I would appreciate a map of the R2 for in-sample and out-of-sample scenarios, for comparison of the original data to the emulated data.

L85-90 & L256-257: training responses on single experiments is somewhat of a risk, reducing the domain of validity of the emulator. For instance, although IRFs are reasonable approximations[1], IRFs for the response atmospheric fraction of CO2 to a pulse of CO2 emissions is known to depend on its calibration under preindustrial and current conditions[2]. This is mostly because the preindustrial carbon cycle does not behave exactly like a perturbed carbon cycle. Here, some IRFs of METEOR are calibrated with *abrupt-4xCO2*, thus starting under a preindustrial climate, up to a disturbed climate.

Have you tried training on multiple experiments instead, using both experiments under past and current conditions? For instance with the variants of *ssp245* as well for GhG, and some variants of *historical* with only aerosols?

Related question L96-100 and L181-199, in particular equation 9: Why use the difference when there is *ssp245-aer*? Besides *ssp245-GHG, ssp245-CO2, ssp245-stratO3*, etc? Quick insight, there may some differences if using different combinations of experiments. For instance, using *hist-aer* would lead to different results than using *historical – hist-PiAer* or *historical – hist-piNTCF*. Because the temporal response of a forcing may depend on the other forcings. The response under *hist-aer* has much less warning, different atmospheric chemistry for aerosols than what we would see under *historical – hist-PiAer*. Though, I agree that it is a second order effect, tough to include in this framework. Thus I would simply suggest you to mention this limitation.

L101-110: I would be careful about summarizing aerosols with sulfates. Each aerosol would have its own specificities in terms of radiative effects, atmospheric chemistry, lifetimes and transport. Some experiments that would be useful here would be: *hist-aer, hist-piNTCF, hist-stratO3, hist-piAer*. I am not asking to recalibrate the model, that would represent a massive additional work to account for different aerosol species. But it could be noted as a potential limitation for future research.

L144-150 + 199-201: I appreciate the technique of separating the timescales into these bins, and great work to evaluate the adequate numbers of timescales in Appendix A. Though I have a question on the choice of bounds. For now, the minimum τ is 1 year. It assumes that there is a non-immediate stabilization of the response to the forcing, which is physically quite robust in this context. Though, it is not unlikely that there may be one mode for the response below 1 year, for a very rapid stabilization. Of course the model runs at an annual resolution, but it would give some flexibility to the response at t=1 and the asymptote (equation 2). Maybe even more relevant for aerosols? So my question would be: would there be a significant gain in performance by including another mode below 1 year?

By the way, looking at Table B1, there are lots of τ_1 at 1.0 year, which could be a sign that the minimization algorithm forced the τ_1 at the very limit of what it could do given the user-defined bounds, in other words that the error function could be minimized by relaxing these bounds. Overall, about half of the tau of the table seems to hit their bounds. I think that it should be investigated.

Suggestions:

L24: can add as well RCMIP phase 2, probably more relevant than RCMIP phase 1.

L26-27: the uses for fast spatial climate modelling frameworks is broader than that, see for instance[3-6]

L35-38: Important point on probabilistic spatial climate information & pattern scaling. Pattern scaling provides only a deterministic response, the mean of the climate field. Having a probabilistic information may come either from the uncertainty in modelling or through natural variability. Pure pattern scaling like PRIME does not include natural variability. MESMER does represent the natural variability obtained through temporal auto-regressions with spatially correlated innovations. Then, regarding precipitations, this is obtained through a more elaborated approach[7]. Finally, pattern scaling is generalized with non-stationary distributions with MESMER-X[8,9], while also removing the assumption of linearity, eg for soil moisture. For the sake of transparency, I'm the author of the two latter papers, and I'm not asking the authors to add them as references.

My point is that there is a need for clarification, that pure pattern scaling is limited to uncertainties in modelling for probabilistic assessment, and going further with natural variability requires additional statistical tools.

L38-41: In STITCHES, the portions of existing simulations are found not only through the median in global mean temperature, but also by its derivative.

L56-58: The feedbacks of JULES don't feedback in FaIR or PRIME (yet).

L17-137: At resolution of the ESM? Is there any rescaling?

L140-143: decomposing x_{global} with IRFs is a good idea, it simplifies the modelling[10]. Another approach is to decompose the local effects with IRFs, e.g. $\boldsymbol{T_{GHG}}$, as conducted in Womack et al 2025[11]. In my opinion, both approaches have pros and cons, I'm not sure which one performs best, but Womack et al, 2025 should be mentioned.

L211: the link between the Moore-Penrose pseudoinverse and Barata and Hussen, 2012[12] should be clearer. In the text, the paper was mentioned without the method, and here the method is mentioned without the paper.

Figures 4 & 5: typos in *historical*, and later the the.

Figures 3 & 4: would you have any insights why the signal for polar amplification is so clear for aerosols (fig 4) but not that much for GhG (fig 3)?

1 Myhre, G. *et al.* PDRMIP: A Precipitation Driver and Response Model Intercomparison Project—Protocol and Preliminary Results. *Bulletin of the American Meteorological Society* **98**, 1185-1198 (2017). https://doi.org/https://doi.org/10.1175/BAMS-D-16-0019.1

2 Joos, F. *et al.* Carbon dioxide and climate impulse response functions for the computation of greenhouse gas metrics: A multi-model analysis. *Atmospheric Chemistry and Physics* **13**, 2793-2825 (2013).

3 van Vuuren, D. P. *et al.* A comprehensive view on climate change: coupling of earth system and integrated assessment models. *Environmental Research Letters* **7**, 024012 (2012). https://doi.org/10.1088/1748-9326/7/2/024012

4 Kikstra, J. S. *et al.* The social cost of carbon dioxide under climate-economy feedbacks and temperature variability. *Environmental Research Letters* **16**, 094037 (2021). https://doi.org/10.1088/1748-9326/ac1d0b

5 Dietz, S., Rising, J., Stoerk, T. & Wagner, G. Economic impacts of tipping points in the climate system. *Proceedings of the National Academy of Sciences* **118**, e2103081118 (2021). https://doi.org/10.1073/pnas.2103081118

6 Rising, J., Tedesco, M., Piontek, F. & Stainforth, D. A. The missing risks of climate change. *Nature* **610**, 643-651 (2022). https://doi.org/10.1038/s41586-022-05243-6

7 Schöngart, S. *et al.* Introducing the MESMER-M-TPv0.1.0 module: Spatially Explicit Earth System Model Emulation for Monthly Precipitation and Temperature. *EGUsphere* **2024**, 1-51 (2024). https://doi.org/10.5194/egusphere-2024-278

8 Quilcaille, Y., Gudmundsson, L., Beusch, L., Hauser, M. & Seneviratne, S. I. Showcasing MESMER-X: Spatially Resolved Emulation of Annual Maximum Temperatures of Earth System Models. *Geophysical Research Letters* **49**, e2022GL099012 (2022). https://doi.org/10.1029/2022GL099012

9 Quilcaille, Y., Gudmundsson, L. & Seneviratne, S. I. Extending MESMER-X: a spatially resolved Earth system model emulator for fire weather and soil moisture. *Earth Syst. Dynam.* **14**, 1333-1362 (2023). https://doi.org/10.5194/esd-14-1333-2023

10 Proistosescu, C. & Huybers, P. J. Slow climate mode reconciles historical and model-based estimates of climate sensitivity. *Science Advances* **3**, e1602821 (2017). https://doi.org/10.1126/sciadv.1602821

11 Womack, C. B., Giani, P., Eastham, S. D. & Selin, N. E. Rapid Emulation of Spatially Resolved Temperature Response to Effective Radiative Forcing. *Journal of Advances in Modeling Earth Systems* **17**, e2024MS004523 (2025). https://doi.org/https://doi.org/10.1029/2024MS004523

12 Barata, J. C. A. & Hussein, M. S. The Moore–Penrose Pseudoinverse: A Tutorial Review of the Theory. *Brazilian Journal of Physics* **42**, 146-165 (2012). https://doi.org/10.1007/s13538-011-0052-z