Summary of main revisions to the manuscript

We would like to thank the referees and the editor again for their time in reviewing this work and providing valuable feedback that helped us improve the manuscript. In the revised manuscript, we have made numerous changes to address the referee comments and clarify our emulator methodology and its applications. Most notable are the changes to the Introduction and to Section 6.

We begin by providing a comprehensive list of all the changes we made in response to the reviewers' comments. We also include point by point answers to each question raised by the reviewers, because we are not sure what the expectations are for a review process structured as this one is. However, we want to stress that while our answers are deliberately very detailed and possibly long-winded, addressing them in the manuscript did not require much additional text because most of the context was already there.

In response to the comments and concerns of Referee #1, we have added language to the introduction to clarify that our emulator is not autoregressive, but rather an emulator of statistics directly; our emulator does not solve for the time evolution of climate variables, but rather predicts the regional statistics of climate variables as a function of global mean temperature **(lines 21-23, 47-51)** . It may be considered as an extension of pattern scaling that predicts the variance in addition to the mean of of regional climate variables as a function of global mean temperature. There are many applications for such an emulator, perhaps most notably in the realm of the rapid analysis of policy impacts. Following the reviewer's suggestion, we illustrate this in the now extended case study Section 6, which we have moved to the main text. In this case study, in addition to highlighting the use of the emulator to distinguish between regional effects of climate policies, we now also show how to use the emulator to estimate when the effects of a policy emerge from natural variability **(lines 462-507)**. This is very similar to measuring the time of emergence of a climate change signal, which our emulator can also be used to quantify.

We have also added language to clarify the innovation of our emulator methodology that can predict the evolution of both means and variances of climate variables over arbitrary regions. If one used linear regression to predict both means and variances, then the emulator could only predict means and variances at the grid resolution of the model it is trained on. We now highlight how the emulation of EOF covariance, one of the main innovations of our work, is necessary for the reconstruction of variances over arbitrary geographic areas **(lines 83-85, 516-519)**. Without this capability, the analysis shown in the case study (Section 6) would not be possible.

As Referee #1 notes, the utility of our emulator is extended even further when one thinks of it as being used within a larger ecosystem of modular emulators, which we now discuss in more detail **(lines 97-101, 112, 545-547)**. Indeed, some emulators of variables at daily resolution require monthly values as their input, as is the case with DiffESM (Bassetti et al., 2024). For such downstream emulators, our emulated variance provides the opportunity for probabilistic conditioning, allowing for a more robust exploration of possible outcomes. Finally, the emulation approach presented in this work can be easily applied to predict climate variables at temporal resolutions shorter than a month; it could be applied to predict the regional pattern of the mean and variance of climate variables at daily resolution as long as such data were available for the training (see **(lines 92-94)**).

In the interest of clarity, we have streamlined the introduction, moving the detailed discussions of the ESM data and the EOF methodology to the corresponding sections: Section 2 and 3, respectively, **(see lines 142-147, 193-196, 202-207)**. In Section 2, we have added a discussion of how many ensemble members might be considered sufficient for an emulator such as ours **(lines 142-147)**, drawing on the work in Lütjens et al. (2025). In Section 3, we have added a discussion of how this methodology can be applied to the joint emulation of two or more variables **see lines 123-124, 218-229**.

In response to the comments of Referee #2, we have added comments throughout the manuscript to clarify our considerations in designing the emulator. In particular, we have highlighted that (1) we use the EOF decomposition as a tool for dimensionality reduction to limit the computational cost of training and running the emulator **lines 81-83, 212-215, 537, 551-557**, as well as to minimize the assumptions required for emulation **lines 212-217, 285-286**, and (2) the choice of modeling the EOF covariance matrix to allow the prediction of means and variances on arbitrary regions to be selected by the emulator users **lines 84-86, 461-463, 516-519**. We believe that highlighting these design considerations, together with the extended case study, has helped to clarify some of the target applications of this emulator.

We would like to thank the referees and the editor for their time in reviewing this manuscript. We believe these edits have greatly improved the manuscript and clarified it for the audience of this journal. Below, we additionally include a point by point response to the original referee comments with descriptions of the associated edits, should this be useful in the review process.

Detailed Responses to Referee 1 Comments

This manuscript proposes a change of basis method to project spatially explicit means and covariances of monthly climate variables as a function of global average temperature change. There's nothing wrong I can see in the method and it is an approach to emulation I have not seen before. But it almost reads as not having been developed in close collaboration with a user group and I think suffers some serious shortcomings because of that. Ultimately, I am left asking 'who is this actually for, specifically? Who is going to pick this up, generate values, and use them? How specifically will they be used?' I'm confident the authors have something in mind for this, but I think the manuscript would benefit for a more explicit treatment of this question.

- The main contribution of our work has been to introduce a new approach to emulation that responds to two very practical needs: 1) a computationally cheap emulator, 2) an emulator that allows one to assess when the impact of climate intervention emerges from natural variability at a regional level. We now make this point more clearly in the introduction and by providing an explicit test case in the new Section 6.

Major issues

1. What is the use case for spatially resolved means and covariances of monthly variables? I'm not trying to be glib, I'm coming at this from the perspective of an impact modeler where I need time series of daily or monthly values of variables. I can't use the statistics you highlight reconstructing in section 4.2 and Section 5 + appendices.

- There is nothing about this emulator methodology that is specific to monthly values; indeed, it could be applied to any regional variable on any timescale. We choose monthly values as an illustrative example, but our colleagues have used a similar approach for daily maximum temperature (Wang et al., 2025). That being said, there are many cases when projections of monthly statistics are useful, particularly in agriculture and water management (Schlenker and Roberts, 2009; Mishra and Singh, 2010) or as priors for predicting daily values (Bassetti et al., 2024). Indeed, we are not the first to present an approach for the emulation of monthly statistics (e.g., Osborn et al., 2016; Nath et al., 2022). We have added language and citations to clarify these points in **lines 92-97**.
- However, we wish to reiterate that the key innovation of our work is the emulation of the covariances of climate variables (see new language in **lines 84-86, 461-463, 516-519**). As described in the manuscript in **lines 312-318**, it is these covariances that allow us to calculate variances over arbitrarily sized regions. Such spatially resolved statistics are essential for quantifying climate impacts on arbitrary regions of interest, as we illustrate in Section 6.

Am I missing something? Do you generate the time series and just demonstrate on statistics (Fig 6-8. E1) because those are more critical for validation?

- Our emulator is a direct statistical emulator that takes global mean temperature as input (either a single value or a sequence of values) and returns statistics corresponding to that mean climate. In other words, our emulator is not autoregressive, as is the case with some emulators in the literature (such as STITCHES), but rather purely statistical. There is significant precedent for statistical emulators, including statistical emulators of monthly values, in the literature from the 1990's onward. This genre of emulators has its roots with the widely used pattern scaling (Santer et al., 1990) and its various extensions. We have edited the manuscript to clarify these points (**lines 21-23, 47-51**).

If so, I think having at least an example plot in Appendix E showing an actual time series generated with this method is key, You may also want to consider extending Appendix E and moving it explicitly into the main body of the manuscript.

- We appreciate the referee's suggestion of extending Appendix E and moving it to the main section, which we have done. Showing individual timeseries for the purposes of validation would be misleading, because the emulator is not autoregressive and does not consider correlations in time. Inspired by that suggestion, though, we now show the evolution of January values of the regional variables with their uncertainty as a function of time in the case study scenario (see Figure 9 and **lines 464-471**). This better illustrates the kind of information that our emulator is designed to generate.
- We have also added to the case study an example of using the emulator to calculate the time of emergence of a climate signal, in this case, of the effects of a theoretical mitigation scenario (**lines 472-507**). We see questions such as these as being one of the main use cases of this emulator, which allows users (such as policymakers, planners, and educators) to quickly iterate over different scenarios and measure their impact on any second-order statistics of interest.

If I have to plug into another emulator like DiffESM to get daily values, why do I need this? A skim of the DiffESM paper shows they only need monthly averages and not covariances and other, simpler methods can give monthly averages. The STITCHES approach can generate a decent sized ensemble of time series of multiple variables jointly just from global temperature.

- It is true that our emulator can (and should!) be used in conjunction with other emulators both downstream (e.g., DiffESM) and upstream (e.g., simple climate models such as FairGP (Bouabid et al., 2024)). Such a modular approach to emulation is not uncommon in the literature; the MESMER family Beusch et al., 2020; Nath et al., 2022; Quilcaille et al., 2022, 2023 is a great example. As for the specific example of coupling to DiffESM, the covariances generated by our method allow one to draw samples from a distribution of monthly values, which can then be separately run through DiffESM, generating an ensemble of projections. Conditioning of DiffESM on only the average temperature would ignore the natural variability of the climate system at the monthly level. It is a key message of our manuscript that information about natural variability is essential for impact assessment studies and we illustrate that it can be provided at really modest computational costs. We discuss this modular/pipeline approach to emulation in the edits in **lines 97-101, 111, 545-547**.

2. The method seems to apply well to any individual gridded variables (demonstrated in the manuscript with surface temperature and surface relative humidity) but, unless I read it wrong, this doesn't extend to coherent joint emulation of multiple variables, right? There are certainly some impact models that only need temperature or relative humidity or precipitation, but many need all of those variables coherently together. I'm thinking of hydrology models especially. Can this method handle that? If not, what could be downsides of using independently generated time series of temperature, relative humidity, and precipitation together?

- There are indeed many applications where the joint behavior of climate variables is of great importance. Our emulator could be easily trained to handle joint variable emulation as well, though we do not demonstrate it in this work. We now explain in **lines 218-229** how such joint emulation would be implemented. The only difference between the single- and multiple-variable cases is the need for extra care in constructing the EOF basis. Since the EOF decomposition ranks modes in order of decreasing percent of variance explained, the leading modes will represent dominant patterns of joint-variable variability but not necessarily those of the individual variables. This may affect the number of modes that need to be retained, which should always be judged on a case by case basis.

3. The authors are clear an ESM needs to have provided a sufficiently large collection of runs to train from, but how large is large? You touch on the implications of this in your final paragraph but I think this needs to be expanded. Like most emulation techniques, this approach targets a single ESM. But many studies using outputs from emulation, say to study projections of a novel scenario, are also concerned with multi-model uncertainty, i.e. they would want multiple emulators each trained on a different ESM. Does the collection of ESMs that provided 'enough' training have uncertainty properties at all similar uncertainty characteristics as the full collection of models? I don't have any way of even roughly guessing because I don't know what the extent of ESMs providing enough data to be individually emulated is.

- The question of how many ensemble members is enough is a difficult, but important one, and we have endeavored to address it with an extended discussion in **lines 142-147**. We draw on work by Lütjens et al. (2025), where it was found that 10 ensemble members was generally sufficient to capture the natural variability of temperature and humidity. There are a number of ESMs out there with 10-member ensembles, spanning a fairly wide range of climate sensitivities and other metrics (Deser et al., 2020; Hausfather et al., 2022). More generally, the number of ensemble members must be sufficiently large to be representative of the internal variability of the climate variable of interest, and this number varies between models and variables. Domain knowledge will always be immensely helpful in this regard.

Missing relevant citations for exclusively emulators of the class trained to extend ESMs to arbitrary future scenarios:

Bassetti et al was actually published nearly a year ago https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/202

Tebaldi, C., Snyder, A., and Dorheim, K.: STITCHES: creating new scenarios of climate model output by stitching together pieces of existing simulations, Earth Syst. Dynam., 13, 1557–1609, https://doi.org/10.5194/esd-13-1557-2022, 2022.

Quilcaille et al https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2022GL099012

and Quilcaille et al https://esd.copernicus.org/articles/14/1333/2023/

- We thank the referee for the helpful references, which we have added in **lines 72, 98, 102**.

Referee 2 Comment

The authors use large-ensemble historical and scenario simulations of a single climate model to train a regression model that predicts means and covariances of 2-m temperature and relative humidity for each month as a quadratic function of global-mean (and ensemble-mean) temperature. The model trained on one scenario successfully approximates the means and covariances of the fields of interest over other two scenarios.

This is a reasonable strategy and implementation. I do have a relatively minor comment though that I would like to clarify. The authors apply the EOF decomposition to the full historical data, rather than anomalies, as is usually done. This is fine for the purposes of the data compression (the leading EOF would essentially give you mean field and the trailing EOFs would be close to the EOFs of anomalies) . Such choices are convenient sometimes (for example, when you use EOFs as a basis to project dynamical equations on (where you need mean state orthogonal to the basis of anomalies) but I am not sure this choice is super-convenient or most economical for the present purposes.

- Indeed, the primary purpose of the EOF projection in this work is for dimensionality reduction and data compression. We chose the EOF decomposition as an optimal basis for the projection of the data, which we clarify in **lines 76-81, 215-217**. In this context, using either the full variable or the anomaly is possible; we happened to choose the former.

I would compute instead the standard EOFs of the historical period (after subtracting the mean) [better yet - ensemble EOFs - why use the single realization?],

- The use of a single realization is a choice purely motivated by computational ease and expedience, which we clarify in **lines 212-215**. We designed this emulator to be lightweight so that it could be used cheaply and quickly in a variety of settings. A different choice of which realization to use to generate the EOFs does not make a difference for the performance of the emulator.

use pattern scaling for projecting the mean into the future (quadratic regression on global mean at each grid point), and then linear regression of eof amplitudes (or quadratic regression of EOF variances) on global mean temperature to project the covariance matrices. Since EOFs diagonalize covariance matrices, you would still get the positive-definite covariances as long as your projected EOF amplitudes remain positive. I think it is highly likely that this much simpler training procedure (which assumes a still diagonal future covariance matrix in the basis of historical EOFs) will give essentially the same results as a more complex method used by the authors - but, of course, I'd be willing and happy to hear the authors opinion and discuss further!

- If we understand the referee's suggestion correctly, the proposed approach would require additional assumptions of stationarity and intra-ensemble generalizability of the EOFs that we sought to avoid in our emulator design. We sought to design an emulator that is as generalizable as possible with minimal assumptions. We have extended our discussion of these motivations and choices in **lines 81-83, 212-217, 285-286**.

- We wish to take the reviewer's offer to discuss further the issue of generalizability of the emulator. The EOF amplitudes are only guaranteed to be bi-orthogonal over the dataset they are computed from, while the EOF basis functions are orthogonal in perpetuity since these only involve spatial integrals. The main issue for orthogonality to hold in general is the orthogonality of the EOF amplitudes, which involve temporal/ensemble averaging. For example, although the EOF amplitudes are orthogonal over the entire historical period for which they are computed, if we look at their monthly covariance as given by an ensemble average, this is no longer the case. This issue might be somewhat alleviated by, say, calculating the EOFs over the full range of the training data (historical and SSP5-8.5 scenario, full ensemble), but this would significantly increase the computational cost of training, which we sought to avoid, without guaranteeing full generalizability.

- In contrast, our approach maximizes the generalizability of the emulator by using the EOF basis solely as a projection basis and guaranteeing the positive-definiteness of the covariance matrix through its construction. Crucially, our regression method is agnostic to the choice of dimensionality reduction technique. We now clarify this in **lines 212-217**.

References

Bassetti, S., Hutchinson, B., Tebaldi, C., and Kravitz, B. (2024). Diffesm: Conditional emulation of temperature and precipitation in earth system models with 3d diffusion models. *Journal of Advances in Modeling Earth Systems*, 16(10):e2023MS004194.

Beusch, L., Gudmundsson, L., and Seneviratne, S. I. (2020). Emulating earth system model temperatures with mesmer: from global mean temperature trajectories to grid-point-level realizations on land. *Earth System Dynamics*, 11(1):139–159.

Bouabid, S., Sejdinovic, D., and Watson-Parris, D. (2024). Fairgp: A bayesian energy balance model for surface temperatures emulation. *Journal of Advances in Modeling Earth Systems*, 16(6):e2023MS003926.

Deser, C., Lehner, F., Rodgers, K. B., Ault, T., Delworth, T. L., DiNezio, P. N., Fiore, A., Frankignoul, C., Fyfe, J. C., Horton, D. E., et al. (2020). Insights from earth system model initial-condition large ensembles and future prospects. *Nature Climate Change*, 10(4):277–286.

Hausfather, Z., Marvel, K., Schmidt, G. A., Nielsen-Gammon, J. W., and Zelinka, M. (2022). Climate simulations: recognize the 'hot model'problem. *Nature*, 605(7908):26–29.

Lütjens, B., Ferrari, R., Watson-Parris, D., and Selin, N. E. (2025). The impact of internal variability on benchmarking deep learning climate emulators. *Journal of Advances in Modeling Earth Systems*, 17(8):e2024MS004619.

Mishra, A. K. and Singh, V. P. (2010). A review of drought concepts. *Journal of hydrology*, 391(1-2):202–216.

Nath, S., Lejeune, Q., Beusch, L., Seneviratne, S. I., and Schleussner, C.-F. (2022). Mesmer-m: an earth system model emulator for spatially resolved monthly temperature. *Earth System Dynamics*, 13(2):851–877.

Osborn, T. J., Wallace, C. J., Harris, I. C., and Melvin, T. M. (2016). Pattern scaling using climgen: monthly-resolution future climate scenarios including changes in the variability of precipitation. *Climatic Change*, 134:353–369.

Quilcaille, Y., Gudmundsson, L., Beusch, L., Hauser, M., and Seneviratne, S. I. (2022). Showcasing mesmer-x: Spatially resolved emulation of annual maximum temperatures of earth system models. *Geophysical Research Letters*, 49(17):e2022GL099012.

Quilcaille, Y., Gudmundsson, L., and Seneviratne, S. I. (2023). Extending mesmer-x: a spatially resolved earth system model emulator for fire weather and soil moisture. *Earth System Dynamics*, 14(6):1333–1362.

Santer, B. D., Wigley, T. M. L., Schlesinger, M. E., and Mitchell, J. F. B. (1990). Developing climate scenarios from equilibrium gcm results. Tech. rep., Max–Planck-Institut für Meteorologie, Hamburg, Germany.

Schlenker, W. and Roberts, M. J. (2009). Nonlinear temperature effects indicate severe damages to us crop yields under climate change. *Proceedings of the National Academy of sciences*, 106(37):15594–15598.

Wang, M., Souza, A. N., Ferrari, R., and Sapsis, T. (2025). Stochastic emulators of spatially resolved extreme temperatures of earth system models. *Authorea Preprints*.