An EOF-Based Emulator of Means and Covariances of Monthly Climate Fields

Gosha Geogdzhayev[1], Andre N. Souza[1], Glenn R. Flierl[1], and Raffaele Ferrari[1]

[1]Massachusetts Institute of Technology, Cambridge, MA, USA

Correspondence: Gosha Geogdzhayev (geogdzh@gmail.com)

Abstract. Fast emulators of comprehensive climate models are often used to explore the impact of anthropogenic emissions on future climate. A new approach to emulators is introduced that predicts means and covariances of monthly averaged climate variables. The emulator is trained with output from a state-of-the-art climate model and serves as a good first-order representation for the evolution of spatially resolved climate variables and their variability. For illustrative purposes, the emulator is applied to predict changes in the mean and variability of monthly values of both temperature and relative humidity as a function of global mean temperature changes. However, the approach can be applied to any other variable of interest.

1 Introduction

In the study of climate change, it is crucial to explore the response of the Earth system to a variety of possible future greenhouse gas emission scenarios and quantify the uncertainties associated with future projections. Despite their limitations, state-of-the-art Earth System Models (ESMs), such as those participating in the Climate Model Intercomparison Project (CMIP, Eyring et al., 2016), remain our most reliable tools for estimating the Earth system response to increased greenhouse gas concentrations. These large-scale models aim to represent as many aspects of the climate system as faithfully as possible. However, because of the high computational and material costs of running ESMs, these models can only simulate the Earth system response to a few potential future scenarios (Tebaldi et al., 2021a). On the other hand, studies of climate mitigation and adaptation strategies often seek to explore a wide range of possible solutions, creating a need for methods to compare localized impacts across a wide range of emissions scenarios (O'Neill et al., 2016b; Waidelich et al., 2024). Our work provides a framework for the direct emulation of ESMs that captures spatially coherent changes in both the mean and variance of climate variables in response to different emission scenarios.

Emulators of ESMs have been gaining popularity as a way to facilitate access to climate projections. Here, we consider the class of emulators trained on the output of ESMs to cheaply and quickly predict the regional climate response to arbitrary climate change scenarios. We will not consider other types of emulators that are being actively pursued for other applications, including climate model downscaling, parametrization of subgrid-scale processes, and model parameter calibration (e.g, Doury et al., 2023; Li et al., 2019; Lai et al., 2025; Peatier et al., 2022; Tebaldi et al., 2025). From here on, the term emulator will refer exclusively to the class of emulators trained to extend ESMs to arbitrary future scenarios. The importance of these emulators is likely to rise due to the increasing computational demands of ESMs: their ever refining spatial resolution, their

growing complexity (as embodied by the number of model components and their sophistication), the use of more accurate numerical methods (Griffies et al., 2020; Souza et al., 2023; Taylor et al., 2023; Silvestri et al., 2024a, b), and the push to run larger ensembles of simulations differing in initial conditions (Schneider et al., 2024, 2023). Emulators are necessary to both compress existing information into a more manageable form as well as to bridge the gap between the computational demand of running a full ESM and the computational hardware available to everyday consumers of climate information. This need is best illustrated by the evidence that projections of future emissions are updated every year, while a new deck of ESM simulations by CMIP takes close to a decade to generate and organize for community sharing.

The simplest and most common emulation technique is pattern scaling (Santer and Wigley, 1990; Huntingford and Cox, 2000; Mitchell, 2003). Pattern scaling estimates spatially resolved changes in climate variables by linearly regressing local variables on global mean temperature or cumulative emissions. While pattern scaling has been shown to have remarkable skill in emulating projected changes in local mean temperatures (Santer et al., 1990; Lütjens et al., 2024), it has no inherent probabilistic component and is less accurate for other climate variables (Tebaldi and Arblaster, 2014; Tebaldi and Knutti, 2018; Kravitz and Snyder, 2023). Recent work has been dedicated to constructing machine learning-based climate emulators (e.g. Watson-Parris et al., 2022; Kitsios et al., 2023; Yu et al., 2024; Christensen et al., 2024) to move beyond the linear assumption inherent in pattern scaling. However, questions remain about their overall skill compared to linear pattern scaling for certain variables (Lütjens et al., 2024). Additionally, their lack of interpretability and the computational cost of model training may limit their applications.

Significant work has also been dedicated to the extension of linear pattern scaling to account for internal variability in the climate system and structural uncertainty in ESMs. A shift in a climate variable is only meaningful insofar as it lies outside the bounds of internal variability. A common approach is to take a large ensemble of climate simulations, either from a single model or from a multi-model set, and apply pattern scaling separately to each simulation, deriving a measure of uncertainty from the resulting ensemble spread (Tebaldi and Knutti, 2018; Zelazowski et al., 2018; Gao et al., 2023; Kravitz and Snyder, 2023). Examining ensemble spread within a *single* model is among the most reliable ways of quantifying the internal variability of the climate system, while inter-model spread includes structural uncertainty (Collins and Allen, 2002; Tebaldi and Knutti, 2007; McKinnon and Deser, 2018; Tebaldi et al., 2021b; Lehner et al., 2020). Internal variability has been shown to dominate uncertainty in climate projections over a decade for regional temperature and much longer for other variables like regional precipitation (Hawkins and Sutton, 2009, 2011). The concept of *Time of Emergence* (ToE) can be used to quantify when a climate signal becomes statistically distinguishable from internal variability (Hawkins and Sutton, 2012; Tebaldi and Friedlingstein, 2013). In its simplest form, ToE is defined as the first time a climate variable (e.g., ensemble mean temperature) exceeds a threshold derived from internal variability (often the ensemble variance), which may itself evolve under climate change. The interest in ToE underpins the need for emulators to capture both mean trends and internal variability.

The importance of representing internal variability alongside the forced response has spurred innovation in climate model emulation techniques. Unlike traditional pattern-scaling that yields a single deterministic outcome for a given forcing scenario, new statistical emulators explicitly include stochastic components to mimic the climate system's inherent variability (Alexeeff et al., 2018; Beusch et al., 2020; Link et al., 2019; Nath et al., 2022). For example, researchers have augmented

pattern-scaling frameworks with residual resampling or autoregressive processes, ensuring that emulator outputs exhibit the same spatio-temporal covariance structure and ensemble spread as the original ESM runs (Link et al., 2019; Nath et al., 2022). Such probabilistic emulators can generate many realizations of future climate—at annual or even monthly resolution—each realization capturing different plausible sequences of internal fluctuations superimposed on the long-term trend (Link et al., 2019; Nath et al., 2022). This approach has been demonstrated in recent studies, from emulating seasonal temperature variability in single-model large ensembles (Alexeeff et al., 2018; Goodwin et al., 2020) to producing large multi-model "superensembles" of regional climate trajectories (Beusch et al., 2020; Alexeeff et al., 2018; Link et al., 2019). The MESMER emulators (Beusch et al., 2020; Nath et al., 2022) have applied approximate methods (Gaussian copulas) to compute the covariances among different climate variables and locations. Covariances between different variables are essential for quantifying the impact of internal variability on climate projections and to assess compound risks (e.g. high temperature and relative humidity is more problematic than high temperature and low relative humidity). Covariances between locations are required to estimate variability on a region beyond the size of an original model grid cell (e.g. the variability over Sub-Saharan Africa is much larger if all regions experience the same temperature fluctuations than if they are all independent).

Our emulator design for averaged climate fields distills the (often complex) approaches described above. We aim to marry a robust representation of forced changes with a statistical model of internal variability that accounts for spatial correlations. We introduce a generative data-driven emulation method for spatially resolved climate fields and apply it to monthly temperature and relative humidity. The main innovation of our method is the balance of simplicity alongside robustness in modeling the lower order statistics of the system. This includes modeling coherent changes in spatial correlations and variances of climate variables. Importantly, our method is fast, flexible, and interpretable.

A sufficiently large ensemble of initial conditions is necessary to infer non-stationary statistical information. For this reason, we choose to emulate the evolution of climate variables generated by a CMIP-class model, specifically MPI-ESM1.2 LR (v1.2.01p7) (Mauritsen et al., 2019; Olonscheck et al., 2023), that ran a large ensemble (50 members) of simulations for a number of emissions scenarios (see Section 2 for details). We emphasize that throughout this work, when referring to an ensemble or ensemble mean, we are considering only a single-model ensemble (e.g. an initial-condition ensemble). We do not intend for these methods to be applied to multi-model ensembles.

Our approach assumes that the internal variability of the climate system is well-approximated by a finite number of spatial modes. We define these modes using Empirical Orthogonal Function (EOF) decomposition (Lorenz, 1956; Kutzbach, 1967; Barnston and Livezey, 1987a); see Hannachi et al. (2007) for a comprehensive review of the technique's history. EOF modes have been shown to effectively capture the dominant patterns of variability of the Earth system (Barnston and Livezey, 1987b; Hannachi et al., 2007). The modes are ranked according to the fraction of overall variability they capture. The leading EOF modes represent patterns that span large geographical regions. While EOF modes can, with some limitations (Monahan et al., 2009), be interpreted physically, here they are only used as a data reduction technique. Using a subset of leading EOF modes (or basis functions) as a fixed-in-time orthogonal basis for the projection of ESM data, we model the statistics of EOF amplitude coefficients as a function of global mean temperature (similar to pattern scaling). We further model the covariance of all the

95   EOF coefficients as a function of global mean temperature, thus also addressing correlations among the spatial modes, and therefore modeling a coherent spatial structure of the variables of interest.

We illustrate our approach for two variables: surface temperature and surface relative humidity. Still, the approach is agnostic to the variables being emulated. It can easily be applied to any monthly variables from any single ESM ensemble, so long as their EOF amplitude statistics are mostly described by their means and covariances. Furthermore, we show how to convert

100   the means and covariances of EOF amplitudes to point-wise means and variances at a fixed location on Earth as well as those averaged over larger regions. Our probabilistic emulator is computationally efficient and, once trained, can be run many times at little additional cost on modest hardware such as single-core CPUs. This computational expedience allows us to generate a synthetic large ensemble for the exploration of internal variability of the climate system (as represented by the ESM), similar to Castruccio et al. (2019).

105   We apply our emulator to monthly data. Monthly climate projections are important for understanding detailed impacts of climate change, such as changes in the seasonal cycle and other phenomena of agricultural relevance (Guo et al., 2002; Odjugo, 2010; Osborn et al., 2016a; Kemsley et al., 2024). Most emulators to date have been developed to reproduce annual (e.g. Goodwin et al., 2020; Beusch et al., 2020) and seasonal (e.g. Holden et al., 2014; Alexeeff et al., 2018) averages of climate variables, while few have looked at reproducing monthly data (Osborn et al., 2016b; Castruccio et al., 2019; Nath et al.,

110   2022). Impact models sometimes require even higher temporal and spatial fidelity, in which case the emulator presented herein provides a framework that can be applied towards those cases as well.

We condition our emulator on the global, ensemble, and yearly mean temperature, which we refer to as "global mean temperature" for short. This procedure is in line with the precedent among other emulators of spatially-resolved climate variables, which are commonly conditioned on global mean temperature (e.g., Osborn et al., 2016b; Alexeeff et al., 2018; Goodwin et al.,

115   2020). Global mean temperature is generally understood to be approximately linear in cumulative emissions (Matthews et al., 2009; Masson-Delmotte et al., 2021), given a smoothly-changing system and ignoring, e.g., the time-lagged response to radiative forcing, or the impact of short-lived aerosols and nonlinear feedbacks like those from melting ice. This simple relationship underpins the utility of a global mean temperature-based emulator. However, one could easily go further by relying on Simple Climate Models (SCMs, e.g. Meinshausen et al. (2011); Lembo et al. (2020); Leach et al. (2021); Bouabid et al. (2024);

120   Dorheim et al. (2024)) to translate arbitrary emission pathways into novel trajectories of global mean temperature (beyond the ones represented by the ESM runs used for training). Our emulator would then use these trajectories to generate realizations of spatially resolved monthly temperatures and relative humidity under those emission scenarios. We comment that often the global mean temperature anomaly from a preindustrial period is used rather than the actual global mean temperature, but here we will use the absolute global mean temperature.

125   Our paper is organized as follows: in Section 2, we introduce the dataset used in this work. Section 3 discusses the EOF-based coarse-graining procedure applied to the data. In Section 4 we discuss the details of the emulator design and the regression problem. In Section 5, we show the emulator's ability to replicate the data's statistics under climate change. Finally, in Section 6, we discuss the broader implications of this work and propose future directions for constructing complementary emulator models.
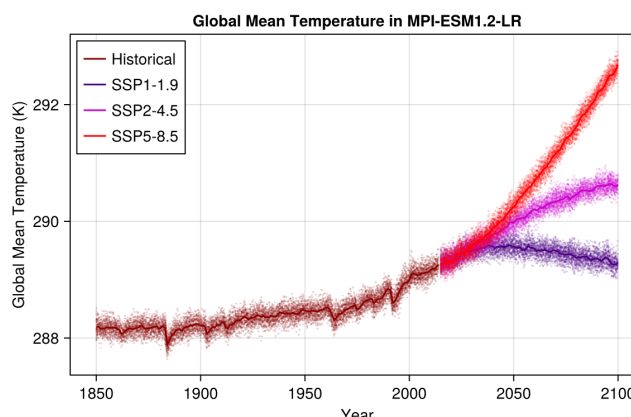
Figure 1. Yearly global mean temperature in the MPI-ESM1.2-LR model ensemble. Each dashed line represents one of 50 initial-condition ensemble members, and the solid line shows the ensemble mean. Different colors correspond to the historical 1850–2014 period and the three future scenarios considered in this study: SSP5-8.5, SSP2-4.5, and SSP1-1.9. The future period spans 2015–2100. The historical period lasts 165 years, and the future period—86 years.

2 Data

To train our emulator, we use the output from the MPI-ESM1.2-LR (v1.2.01p7) ESM model (Mauritsen et al., 2019; Olonscheck et al., 2023), which contributed to CMIP6. We chose this model because of the large number of simulations (ensemble members) it ran for each emission scenario: 50 simulations are run for each scenario, differing only in their initial conditions. A large ensemble is necessary to separate a model's internal variability from the anthropogenic signal (Collins and Allen, 2002; Tebaldi et al., 2021b; Lütjens et al., 2024). In the CMIP6 model set, only three ESMs submitted ensembles of 30 or more members: MPI-ESM1.2-LR, EC-Earth3, and CanESM5. Among these, the MPI model is the only one with an equilibrium climate sensitivity to greenhouse gas emissions within the "likely" range determined by multiple lines of evidence (Hausfather et al., 2022) and with the entire ensemble available for open download. Analysis performed in Lütjens et al. (2024) on the same MPI dataset suggests that an emulation method that uses polynomial fitting, such as ours, captures the evolution of the ensemble mean climate variables even if trained on few ensemble members (in terms of emulation error). However, an ESM with many ensemble members is necessary to train the emulation of internal variability (Tebaldi et al., 2021b).

Each MPI-ESM1.2-LR ensemble member is run for the historical period, spanning 165 years between 1850–2014, and for various future warming scenarios spanning the 86-year future period 2015-2100. We consider output from simulations of three future scenarios from the ScenarioMIP experiments: SSP5-8.5, SSP2-4.5, and SSP1-1.9. The ScenarioMIP experiments are plausible futures corresponding to different climate mitigation and cooperation narratives (O'Neill et al., 2016a; Tebaldi et al., 2021a). Figure 1 reports the global mean temperature profiles of the historical period and the three future scenarios considered in this work for each of the 50 MPI-ESM1.2-LR ensemble members.
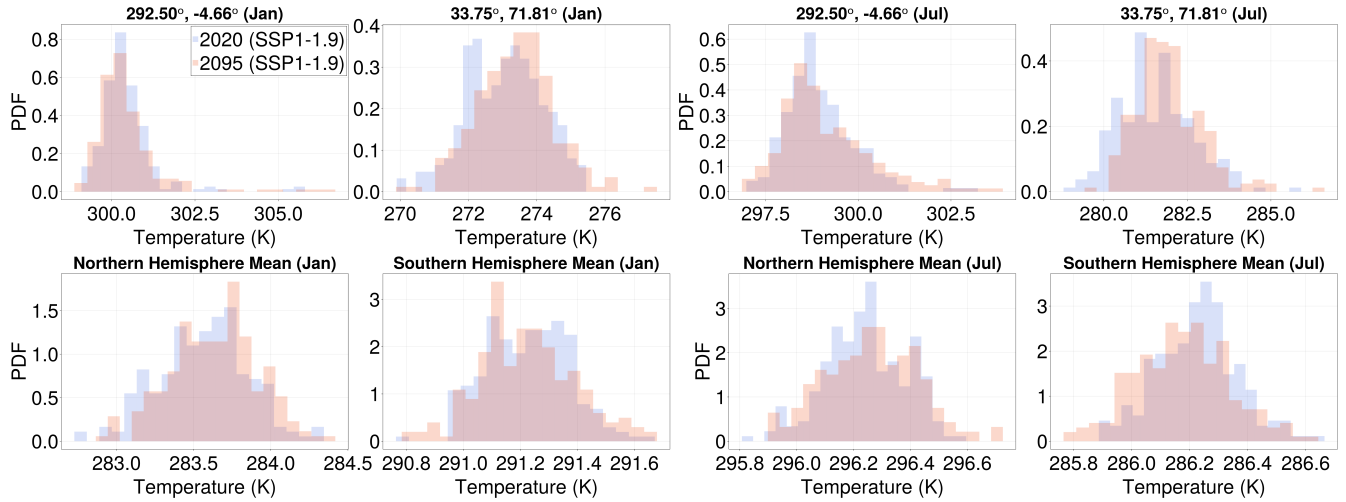
Figure 2. Surface temperature histograms of the non-monotonic SSP1-1.9 scenario corresponding to similar global mean temperatures ($\overline{T}_g = 289.3$ K) but for two different time intervals: $2018-2022$ (blue bars) and $2093-2097$ (red bars). Histograms are constructed for selected locations (top row) and hemisphere averages (bottom row) and for January (left) and July (right). The histograms mostly overlap, lending support to the choice of parameterizing the distributional change (and hence lower-order statistical changes) for a fixed month with only $\overline{T}_g$.

Our goal is to develop an emulator that predicts changes in the mean and covariance of climate fields as a function of emissions. The first step in the process is to represent the statistics as a function of (or, more appropriately, conditioned on)

150   global, ensemble, and yearly mean temperature and hence cumulative emissions (see Masson-Delmotte et al. (2021)). As mentioned above, the ensemble mean is calculated over a single model with multiple realizations. Following standard practice, we will refer to the global, ensemble, and yearly mean surface temperature as "global mean temperature" throughout the text, denoted $\bar{T}_g$.

We select the historical experiment and the SSP5-8.5 high-warming scenario for training the emulator because together they

155   span the widest range of global mean temperatures. This leaves SSP2-4.5 and SSP1-1.9 as the validation sets for regression, see Section 5. SSP2-4.5 features milder monotonic warming that levels off at the end of the century (through the elimination of emissions), while SSP1-1.9 features a peak in global mean temperature around mid-century followed by a decrease to end-of-historical-period temperatures by 2100 (as a consequence of negative emissions), see Figure 1. Because we are developing an emulator conditioned only on global mean temperature from a scenario with monotonically increasing emissions

160   (without accounting for emissions history or other memory effects), it is important to test its performance in scenarios with non-monotonic emissions. Such scenarios are of ever-increasing interest to mitigation and adaptation studies and policy. We emphasize that our approach is purely data-driven and should not be used to extrapolate statistics in climates outside of its global mean temperature training range.

To presume that time-dependent climate statistics for different emissions scenarios can be parameterized by a state-dependent
(i.e., time-independent or history-independent) scalar quantity is a strong assumption but one that is justified a posteriori for
the cases considered in this work, see Figure 2 and, later on, Figures 7 and 8. Figure 2 shows histograms of surface temperature
data in the MPI model for selected locations (top row) and hemisphere averages (bottom row) in January (left) and July (right).
Two 5-year windows in the SSP1-1.9 future scenario are compared: $2018 - 2022$ (in blue) and $2093 - 2097$ (in red). The SSP-
1.1.9 scenario is non-monotonic, and these two periods correspond to similar global mean temperatures of around $\overline{T}_g = 289.3$
K, separated by an intermediate 70-year period of higher temperatures. We observe that, for the most part, the histograms for
the two periods overlap. We thus conclude that a state-dependent parametrization of climate statistics is not unreasonable for
our purposes.

Before delving into the specific details of the emulator formulation, it is crucial to first understand the underlying rationale
guiding our approach. Our key assumption is that the statistics of climate system fields, in our case the EOF modal amplitudes
$\boldsymbol{a}$ (see Section 3), can be represented by a probability density for every possible state, with time and emissions history replaced
by the global mean temperature $\overline{T}_g$ and seasonal information such as the month m:

$$\rho(\boldsymbol{a}, t | \text{emissions}) \rightarrow \rho(\boldsymbol{a} | \overline{T}_g, m). \tag{1}$$

This assumes that conditioning on global mean temperature serves as an informative parametric form to characterize the
changing distribution of the climate-relevant quantities. In the application described in the following sections, we do not
consider the full probability distribution and instead assume that only the first two moments—means and covariances—can be
represented as functions of global mean temperature.

We illustrate our approach for monthly mean surface (2m) temperature and monthly mean surface (2m) relative humidity.
These are the 'tas' and 'hurs' variables, respectively, in the CMIP6 nomenclature. The MPI-ESM1.2-LR model has a horizontal
resolution in the atmosphere of $1.8°$. We use the model output on its original 192×96 lat-lon grid. We calculate the global
mean temperature (used to condition the emulator) by globally averaging the 2m temperature variable.

3 Coarse-Graining

We coarse-grain the representation of the temperature and relative humidity fields, using Empirical Orthogonal Functions
(EOFs) as a dimensionality reduction technique. EOF decomposition has been used in previous emulator work for both di-
mensionality reduction (Holden and Edwards, 2010; Holden et al., 2015; Yuan et al., 2021) and more generally as a method
of generating an uncorrelated projection basis (Link et al., 2019). More recently, Falasca et al. (2024a) and Falasca et al.
(2024b) have demonstrated how modal amplitudes of EOFs (under the additional assumption that they can be approximated
as multivariate Gaussian distributions) can be used to interpret patterns of variability and teleconnections through data-driven
approaches.

We compute the EOF basis ϕ through a singular value decomposition of our data in the historical period of one of the
ensemble members. The resulting basis constitutes $165 \times 12 = 1980$ EOFs ordered by the magnitude of the singular values. We

discard the last 980 basis functions, leading to a total of 1000 basis functions $\phi_i(\boldsymbol{x}), i \in [1, 1000]$. The notation $(\boldsymbol{x})$, which we use throughout this work, indicates that EOF basis functions are spatially resolved, i.e. functions of the location $\boldsymbol{x}$. We use the same basis set for all months and compute EOFs separately for each variable of interest. We project data from every scenario and every ensemble member onto our constant basis.

200   The choice to work with the leading EOFs is not only one of convenience to reduce the data size. It is also motivated by evidence that the projections of climate models are more reliable at larger scales (Gleckler et al., 2008). Finer-scale details captured by high order EOFs–such as full temperature distributions at a point–are better modeled using different approaches, such as downscaling from coarse-grained information. In probabilistic jargon, we divide the set of modal amplitudes $\boldsymbol{a}$ into modes corresponding to large scale coarse structures $\boldsymbol{a}_C$ and "fine scale modes" $\boldsymbol{a}_F$. The probability distribution for climate

205   variables (for a fixed global mean temperature $\overline{T}_g$ and month m) can then be decomposed as the product of two conditional distributions:

$$\rho(\boldsymbol{a}_C, \boldsymbol{a}_F | \overline{T}_g, m) = \rho(\boldsymbol{a}_F | \boldsymbol{a}_C, \overline{T}_g, m)\rho(\boldsymbol{a}_C | \overline{T}_g, m). \tag{2}$$

Our work focuses on approximating the means $\boldsymbol{\mu}$ and covariances $\boldsymbol{\mathcal{C}}$ of $\rho(\boldsymbol{a}_C | \overline{T}_g, m)$, the distribution of $\boldsymbol{a}_C$ as a function of the global mean temperature T_g and month m. Armed with the means and covariances of the modes $\boldsymbol{a}_C$, one can then reconstruct

210   means and variances of the original fields both locally or averaged over some area, as we will see in Section 4.2. In addition, if the statistics of the variables of interest are Gaussian, means and covariances are sufficient to reconstruct whole probability density functions (see Appendix A).

Approximating fine-scale structures conditioned on coarse-grained variables, i.e., approximating $\rho(\boldsymbol{a}_F | \boldsymbol{a}_C, \overline{T}_g, m)$, is left to future work. In downscaling excersises, it is typically assumed that

215   $$\rho(\boldsymbol{a}_F | \boldsymbol{a}_C, \overline{T}_g, m) \approx \rho(\boldsymbol{a}_F | \boldsymbol{a}_C), \tag{3}$$

i.e., information about the coarse scales may be sufficiently informative to parameterize the distribution of the fine scales.

4   Regression

After projecting the ESM data into the EOF space, we model the EOF coefficients as a function of global mean temperature (see also Holden and Edwards, 2010; Bruckner et al., 2003) and proceed to emulate their means and covariances. The method can

220   be applied to any variables regardless of their distributions insofar as one is only interested in the means and covariances of any linear functional of the variable field, see Section 4.2. Despite not requiring a Gaussian approximation, we will often overlay histograms of variables from the MPI model with Gaussian histograms with the means and covariances generated by our emulator. This is an arbitrary, yet somewhat constrained choice because a Gaussian distribution is among the most natural ways to visualize a distribution parameterized solely by its mean and variance. It so happens, however, that the EOF amplitudes

225   of monthly average temperature and relative humidity have approximately Gaussian distributions. This is to be expected from the multivariate version of the central limit theorem which states that any variable that is the average of many independent

realizations of the same process has a Gaussian distribution. In our case we average in time, over a month, and in space by the coarse-graining implied by the EOF decomposition.

We are now ready to describe our emulation approach in detail. Section 4.1 describes the procedure for fitting to data, and
230 Section 4.2 describes how to utilize the emulator and its relation to pattern scaling.

4.1 Emulator Details

Following the EOF-based dimensionality reduction of the ESM training data, we develop and train a stochastic emulator of spatially resolved monthly temperature and relative humidity fields. As stated previously, obtaining the set of estimated EOF coefficients $\hat{a}$ requires estimating a mean $\hat{\mu}$ and covariance $\hat{\mathcal{C}}$ as a function of global mean temperature $\bar{T}_g$ and month index
235 m, i.e.,

$$\hat{\mu}(\bar{T}_g, m) \quad \text{and} \quad \hat{\mathcal{C}}(\bar{T}_g, m). \tag{4}$$

Since each month is modeled separately, we will drop the notation m with the implicit understanding that any regression is for a fixed month. Throughout this work, we use the notation $\hat{\cdot}$ to denote an emulator-derived estimate of a quantity, in contrast to the ESM-derived "ground truth". The large ensemble of MPI model simulations offers a robust way to estimate the means and
240 covariances since we view, for a fixed month and year, every individual ensemble member as a realization of a high-dimensional distribution parameterized solely by the month and global mean temperature.

The dependence of the means $\hat{\mu}$ on $\bar{T}_g$ is modeled as a quadratic function

$$\hat{\mu} = \hat{\mu}_0 + \hat{\mu}_1 \bar{T}_g + \hat{\mu}_2 \bar{T}_g^2. \tag{5}$$

This approximation yields a better overall fit than a linear function, see Figure 3 and Appendix B. Higher-order polynomial fits
245 or neural networks could be used to improve on the results presented here but may also overfit the data.

Modeling the covariance of the EOF coefficients as a function of global mean temperature requires more care. When fitting the mean of each EOF mode, one can use standard methods for curve fitting, such as least squares. Parameterizing a covariance matrix as a function of global mean temperature is more subtle since all the matrix entries must conspire together to yield a symmetric positive definite matrix. Thus we choose to represent the dependence of the covariance matrices $\hat{\mathcal{C}}$ on $\bar{T}_g$ as

250 $$\hat{\mathcal{C}} = \hat{L}\hat{L}^T \text{ and } \hat{L} = \hat{L}_0 + \hat{L}_1 \bar{T}_g. \tag{6}$$

This functional form guarantees that $\hat{\mathcal{C}}$ is symmetric positive definite because it is the product of a matrix $\hat{L}$ and its transpose. In Equation 6 each entry of $\hat{L}$ is modeled as a linear function of $\bar{T}_g$, leading to a quadratic model for the covariance:

$$\hat{\mathcal{C}}(\overline{T}_g) = \hat{L}_0\hat{L}_0^T + \left(\hat{L}_0\hat{L}_1^T + \hat{L}_1\hat{L}_0^T\right)\bar{T}_g + \hat{L}_1\hat{L}_1^T \left(\bar{T}_g\right)^2. \tag{7}$$

As with the means, it is possible to go beyond a linear representation for $\hat{L}$, though one would run the risk of overfitting. This
255 parametric form yields a representation of how the internal variability changes as a function of global mean temperature, and hence time and emissions scenario (under appropriate assumptions).
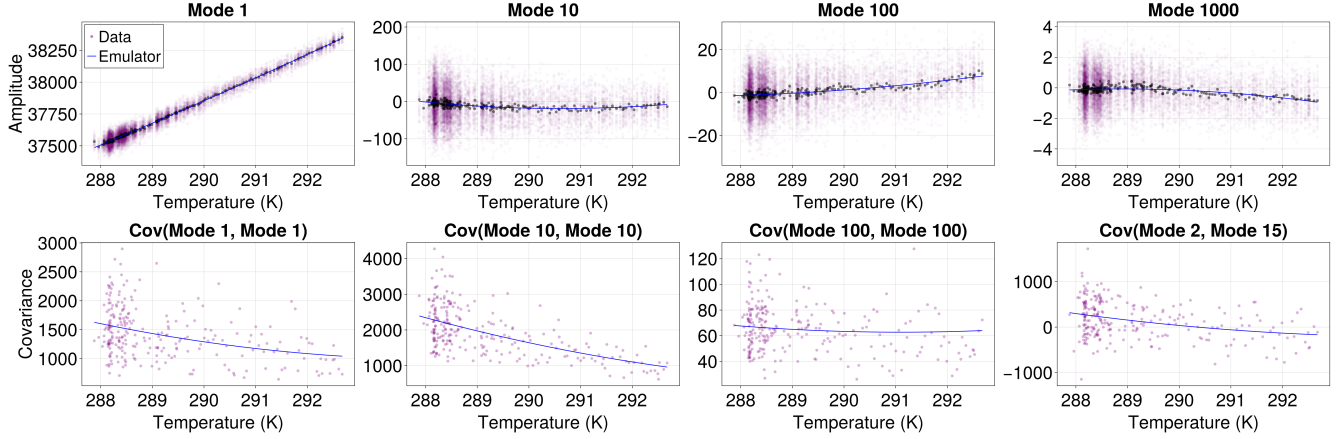
Figure 3. Regression of mean (top row) and covariance (bottom row) of the surface temperature EOF amplitudes as a function of global, ensemble, and yearly mean temperature in January. Estimates from the MPI model ensemble simulations are shown as purple points, while the blue lines are the emulator fits. The quadratic fits capture both the linear trends and the slight curvature.

To find $\hat{\boldsymbol{L}}_0$ and $\hat{\boldsymbol{L}}_1$ we minimize the loss function

$$\text{Loss}(\hat{\boldsymbol{L}}_0, \hat{\boldsymbol{L}}_1) = \sum_{\bar{T}_g} \|\hat{\boldsymbol{\mathcal{C}}}(\bar{T}_g) - \boldsymbol{\mathcal{C}}(\bar{T}_g)\|^2, \tag{8}$$

where $\hat{\boldsymbol{\mathcal{C}}}$ is given by Eq.7 and $\|\cdot\|$ is an appropriately chosen norm. In our case, we used a Frobenius norm (minimizing

260   the square distance between each matrix entry), but other choices would likely yield good answers as well. This minimization was performed in JAX (Bradbury et al., 2018), on an H100 Nvidia GPU using automatic differentiation and Kingma and Ba (2014)'s "ADAM" for optimization. The minimization was initialized with a constant covariance matrix, i.e., $\overline{\boldsymbol{\mathcal{C}}} = \frac{1}{251} \sum_{year=1850}^{2100} \boldsymbol{\mathcal{C}}(year)$. This choice was implemented by taking $\hat{\boldsymbol{L}}_1 = 0$ and obtaining $\hat{\boldsymbol{L}}_0$ from the Cholesky factorization of $\overline{\boldsymbol{\mathcal{C}}}$. To perform the regression for the covariance matrix, we used the fact that the covariance can be computed separately for

265   each year (over the set of all ensemble realizations) and each year has an associated global mean temperature.

We illustrate the result of the regression procedure for surface temperature in Figure 3. The top row represents the regression for the ensemble mean EOF coefficients, and the bottom row shows the regression problem for the covariance matrix between EOF modes, both for January. We first describe the top row. The purple dots are the projected modal amplitudes for a few sample EOFs plotted at each year for all ensemble members of the MPI data for the historical and SSP5-8.5 scenario. The bottom

270   row is obtained by calculating the covariance between sample modal amplitudes for each year separately using all ensemble members. These data are then regressed against each year's global, ensemble, and temporal mean of surface temperature ($\bar{T}_g$). From Figure 3, we see that the trends are well captured by performing the regression (in blue). As mentioned before, we use a quadratic model for the mean of the EOF coefficients and a quadratic fit for the entries of the covariance matrix. The covariance data are much noisier but still display overall trends captured through the regression.

4.2 Reconstructing Statistics of Interest from EOF Amplitudes

We can reconstruct spatially resolved statistics of the variables of interest for a fixed month m and global mean temperature $\overline{T}_g$ by making use of the EOF basis $\phi(\boldsymbol{x})$. In this basis, any variable of interest, such as surface temperature T, is given by

$$\hat{T}(\boldsymbol{x}) = \sum_{i=1}^{N} \hat{a}_i \phi_i(\boldsymbol{x}) \tag{9}$$

where N is the total number of EOF modes used (in our case $N = 1000$). Our emulator generates the means and covariances of the EOF amplitudes $\hat{a}_i$. The ensemble mean of $\hat{T}$ at a fixed location $\boldsymbol{x}$ is therefore given by

$$\langle \hat{T}(\boldsymbol{x}) \rangle = \sum_{i=1}^{N} \hat{\mu}_i \phi_i(\boldsymbol{x}) \tag{10}$$

and its variance at the same point is given by

$$\langle \hat{T}(\boldsymbol{x})^2 \rangle - \langle \hat{T}(\boldsymbol{x}) \rangle^2 = \sum_{ij} \hat{\mathcal{C}}_{ij} \phi_i(\boldsymbol{x}) \phi_j(\boldsymbol{x}). \tag{11}$$

In fact, for any linear functional $\mathcal{L}$ acting on the temperature field $\hat{T}$, e.g., a regional average / zonal average / average of a patch of land such as North America or Africa, the ensemble mean and variance is given by

$$\langle \mathcal{L}[\hat{T}] \rangle = \sum_{i=1}^{N} \hat{\mu}_i \mathcal{L}[\phi_i] \quad \text{and} \quad \langle \mathcal{L}[\hat{T}]^2 \rangle - \langle \mathcal{L}[\hat{T}] \rangle^2 = \sum_{ij} \hat{\mathcal{C}}_{ij} \mathcal{L}[\phi_i] \mathcal{L}[\phi_j]. \tag{12}$$

Thus, the mean and variance of any linear functional of temperature (or whatever field is being emulated) can be computed from the mean and covariance of all the EOF coefficients and the action of the linear functional on the basis functions. It is worth repeating that the EOF covariances are essential to estimate the variance of arbitrary functionals. Fitting the variance of each EOF as a function of global mean temperature is not sufficient. Equation 12 illustrates that the entire covariance structure of the EOF amplitudes is key for computing regional statistics beyond the mean.

Extending the calculation to higher-order statistics or nonlinear functionals requires additional assumptions. For example, assuming that the EOF coefficients can be modeled as a multivariate Gaussian would supply the necessary information to extend means and covariances to higher-order statistics; see discussion in Appendix 4.1. We leave such an extension for future work.

4.3 Relationship to Pattern Scaling

It is instructive to compare our approach to linear pattern scaling. Linear pattern scaling predicts the temperature at every location as a linear function of the global, yearly, and ensemble mean temperature (Santer and Wigley, 1990). In equations, this means the spatial temperature field is a linear function

$$\langle \hat{T}(\boldsymbol{x}) \rangle = T_0(\boldsymbol{x}) + \overline{T}_g T_1(\boldsymbol{x}) \tag{13}$$

for appropriately chosen T_0 and T_1. These spatial patterns are often obtained explicitly by performing pointwise linear regression; $T_1(\boldsymbol{x})$ is the fixed spatial anomaly "pattern" of the namesake.

We now show how to connect the ensemble mean of our emulator to pattern scaling. Summing over all EOF modes ϕ_i, our emulator for fixed temperature at a location $\boldsymbol{x}$ yields,

$$\langle \hat{T}(\boldsymbol{x}) \rangle = \sum_{i=1}^{N} \hat{\mu}_i \phi_i(\boldsymbol{x}) = \sum_{i=1}^{N} [\mu_{0,i} + \overline{T}_g \mu_{1,i} + \left(\overline{T}_g\right)^2 \mu_{2,i}] \phi_i(\boldsymbol{x}) \tag{14}$$

$$= \sum_{i=1}^{N} \mu_{0,i} \phi_i(\boldsymbol{x}) + \overline{T}_g \sum_{i=1}^{N} \mu_{1,i} \phi_i(\boldsymbol{x}) + \left(\overline{T}_g\right)^2 \mu_{2,i} \sum_{i=1}^{N} \mu_{2,i} \phi_i(\boldsymbol{x}) \tag{15}$$

$$\equiv T_0(\boldsymbol{x}) + \overline{T}_g T_1(\boldsymbol{x}) + \left(\overline{T}_g\right)^2 T_2(\boldsymbol{x}), \tag{16}$$

where

$$T_0(\boldsymbol{x}) = \sum_{i=1}^{N} \mu_{0,i} \phi_i(\boldsymbol{x}), \, T_1(\boldsymbol{x}) = \sum_{i=1}^{N} \mu_{1,i} \phi_i(\boldsymbol{x}) \text{ and } T_2(\boldsymbol{x}) = \sum_{i=1}^{N} \mu_{2,i} \phi_i(\boldsymbol{x}). \tag{17}$$

Thus our emulator would reduce to linear pattern scaling for the ensemble mean temperature at a point if we were to sum over all EOF modes and choose a linear fit, rather than a quadratic one, for the mean of their amplitudes (i.e. force T_2 to be zero); we illustrate this in Appendix C.

Our approach, however, expands over pattern scaling in three important ways. First, our focus is on performing regression on modal amplitudes of the largest and most robust spatial patterns of the climate model output, which are given by suitably coarse-grained basis functions rather than pointwise measurements. We use a sufficiently large basis such that we can still represent pointwise values but ignore small-scale spatial patterns, whose representations in ESMs are typically quite uncertain.

Second, our emulator produces more than the ensemble mean since it also accounts for covariances across EOF modes. This is no trivial point since the value of a physical variable is meaningless without an estimate of its uncertainty, or internal variability. For example, as mentioned in the Introduction, estimates of the time when the forced climate response of a climate variable emerges from internal variability (the ToE) require emulators of both means and variances. It may be argued that the linear pattern scaling approach can also be used to predict the temperature variance at each location as a linear function of the global, yearly, and ensemble mean temperature. However, a new linear fit must be computed if one wishes to estimate the variance over a wider region (for example) because it requires knowing correlations between different points. With our emulator, we can reconstruct a much larger range of statistics of the field in question from the mean and covariance estimates of the modal amplitudes without any additional fitting.

Third, our emulator can "generate" higher order statistics under the assumption that they are functions of lower-order statistics such as means and covariances. For example, all statistics of multivariate Gaussian probability distributions are described by knowing its mean and covariance.

5 Validation

330 Now that we have described the design of our emulator, it is time to validate its performance. The emulator is trained on the historical period and the SSP5-8.5 future scenario of the MPI ESM-1.2-LR model. The testing is performed by (a) applying it to other scenarios run by the MPI model that were not used in training and (b) using it to compute various statistics of interest. We use the SSP2-4.5 and SSP1-1.9 as our "test" scenarios.

5.1 Spatial Error

335 We begin by exploring the patterns of error in the emulator and comparing them across training and test scenarios. To understand the spatial distribution of error, we show in Figure 4 the averaged absolute difference between our emulator-predicted mean temperature and the ensemble mean of the data over the historical period, SSP5-8.5, SSP2-4.5, and SSP1-1.9. In formulas, this is

$$\text{Error}(\boldsymbol{x}) = \frac{1}{\text{duration}} \int\limits_{\text{scenario start}}^{\text{scenario end}} \left| \frac{1}{12} \sum_m \left(\langle T(\boldsymbol{x}, m, t, \omega) \rangle - \hat{\mu}(\boldsymbol{x}, m, \overline{T}_g(t)) \right) \right| dt \qquad (18)$$

340 where $\langle \cdot \rangle$ denotes an average over ensemble members ω. We have explicitly written down the dependence of the ESM temperature T on the month m, spatial location $\boldsymbol{x}$, and year t and the emulator estimate $\hat{T}$ on the same variables, but with time t replaced with the appropriate global mean temperature $\overline{T}_g$ evaluated at the specified year. The integral is taken over a specified future scenario across all the years encompassed by that scenario.

The emulator is quite accurate across the majority of the globe, especially when assessed relative to the internal variability of the system. For all scenarios, the largest errors are found in the high latitudes. There are also additional errors over Africa, India, and the southeast tip of Australia. Relative to the spatially-varying temperature variability, however, only some of these errors are significant. Figure 5 illustrates the averaged pointwise standard deviation of surface temperature for each scenario shown in Figure 4. The standard deviation is calculated across ensemble members, then averaged over time for each scenario. The variability of temperature is by far the largest in the polar regions; in these regions it is much larger than the errors of the emulator. Likewise, the errors over India and particularly the Himalayas coincide with regions of high temperature variance. Thus, when assessed relative to the internal variability of the ESM, the only significant emulation errors are those in equatorial Africa, in south-eastern Australia, and in the North Atlantic (for low-emissions scenarios).

We expect the errors in spatial patterns to change upon using non-polynomial regression for the mean or a different set of basis functions. However, we argue against such an approach because we suspect that these errors reflect regional changes that depend on variables beyond global mean temperature such as changes in dynamics due to the disappearance of sea ice in the northern hemisphere and changes in the hydrological cycle in the low latitudes. A more promising strategy would be to formulate an emulator that takes large-scale variables beyond the global mean temperature as input, but this is left for future work. For present purposes, we are encouraged by the fact that the spatial errors are largely consistent across all future scenario cases. This indicates that our emulator is effectively generalizing to scenarios for which it wasn't specifically trained.
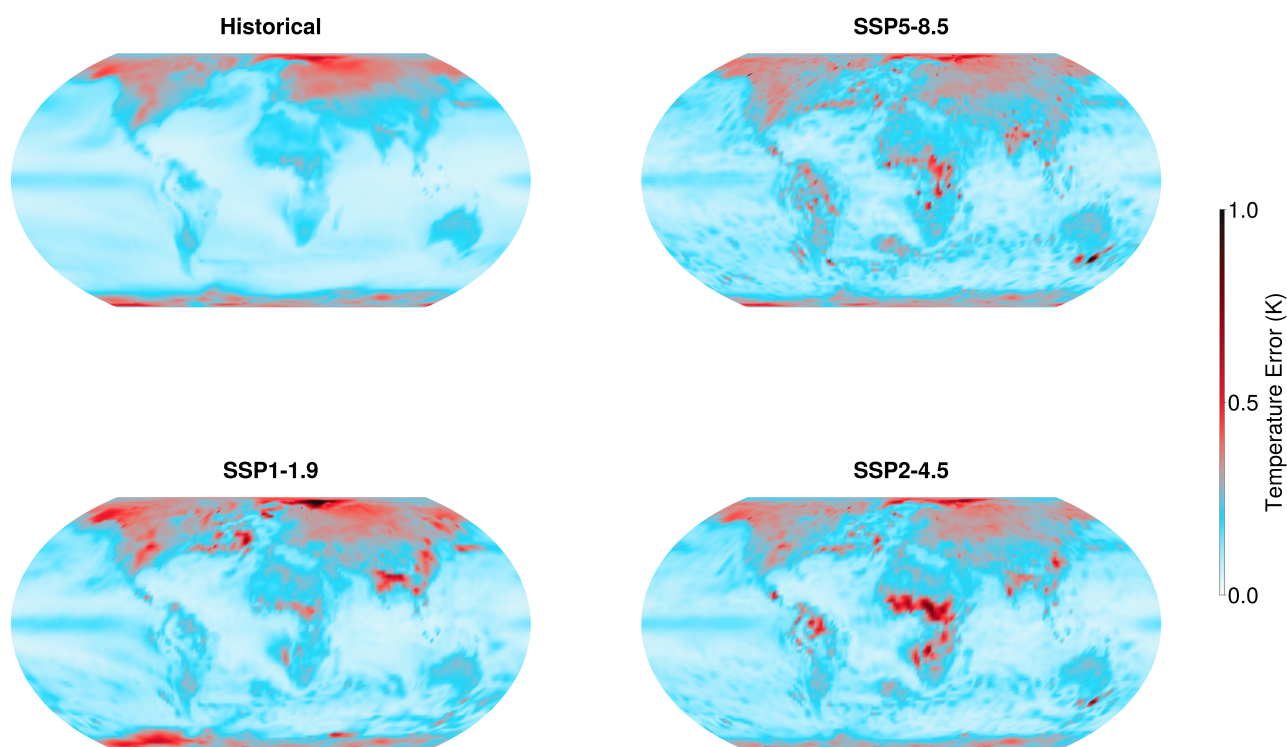
Figure 4. Absolute difference between pointwise emulated temperature and the ensemble mean from the MPI model averaged over all time for four different scenarios: the historical period (top left), SSP5-8.5 (top right), SSP1-1.9 (bottom left), and SSP2-4.5 (bottom right). The maximum temperature difference in the period 2015 to 2100 is 1.4 K for SSP5-8.5, 0.3 K for SSP1-1.9 and 1.4 K for SSP2-4.5. The emulator was trained on the historical period and SSP5-8.5; SSP2-4.5 and SSP1-1.9 are "test" scenarios.

360   In Appendix B we additionally explore the temporal error of the emulator and reconfirm that the emulator skill is improved significantly by our choice of using a quadratic, rather than linear, regression of EOF mean amplitudes on global mean temperature. In Appendix C we report on the temporal error of the emulator's representation of temperature standard deviation. While these tests are commonly used in the assessment of emulators, they are limited in scope. In the next section we illustrate a major advantage of our emulator: its ability to quantify the significance of shifts in the mean and variability of climate variables

365   averaged over different spatial regions.
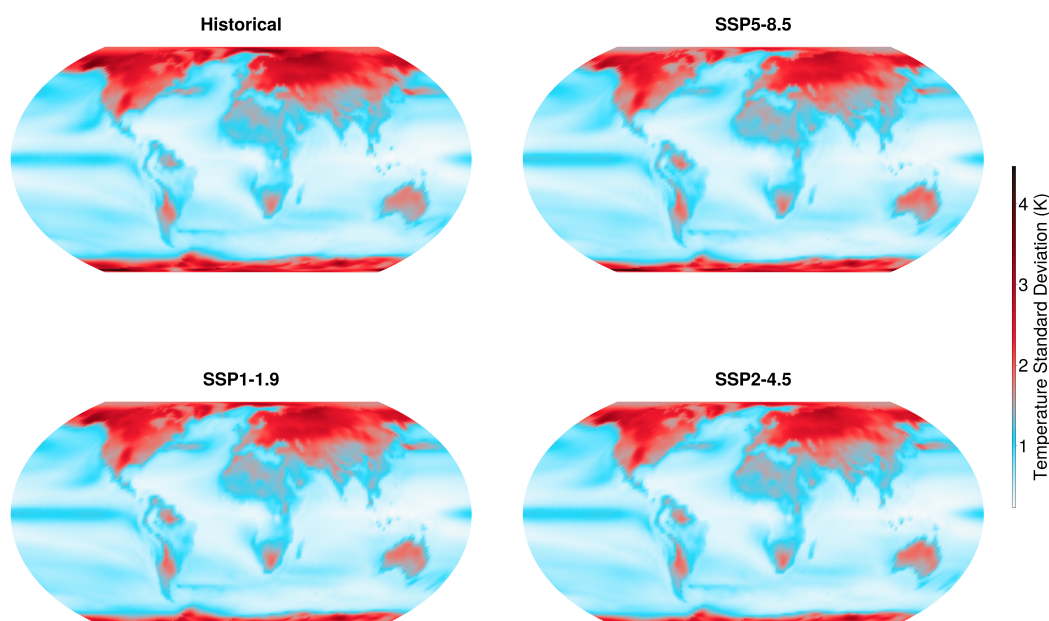
Figure 5. Pointwise standard deviation of surface temperature in the MPI model averaged over all months and over each scenario, laid out as in Figure 4. The standard deviation is calculated across ensemble members for each time step. Note the high variability in the polar regions.

5.2 Spatial Statistics

As stated in Section 4.2, it is possible to reconstruct spatial statistics of any observable of our system as long as it is a linear function of the variables we emulate. The statistics of the zonal average at a fixed latitude for temperature in January from the MPI model (purple) are reconstructed in Figure 6 for a range of similar global mean temperatures taken over the historical period to have higher fidelity statistics. The corresponding predictions from the emulator are in blue. The emulator reconstructs very accurately the mean and variance for all latitudes (top row). In the bottom row, we compare the whole distributions of zonally averaged temperature at a few sample latitudes. Here, the emulated distributions are derived using the mean and covariances by assuming a multivariate Gaussian distribution for the EOF coefficients (see Appendix A for an extended discussion on Gaussianity). It is remarkable that such close agreement is achieved, given that the regression focuses on the mean and covariance of the EOF amplitudes rather than directly on the zonal averages. We emphasize, once again, that the emulator must predict the full covariance matrix for all EOF amplitudes–the main innovation of our work–to accurately capture the zonally averaged statistics of surface temperature at each latitude, as demonstrated in Equation 17.

Until now, we have focused on surface temperature statistics, but applying the methodology to other variables is straightforward. As an example, we apply the method to surface relative humidity. We show the emulator prediction and the MPI data
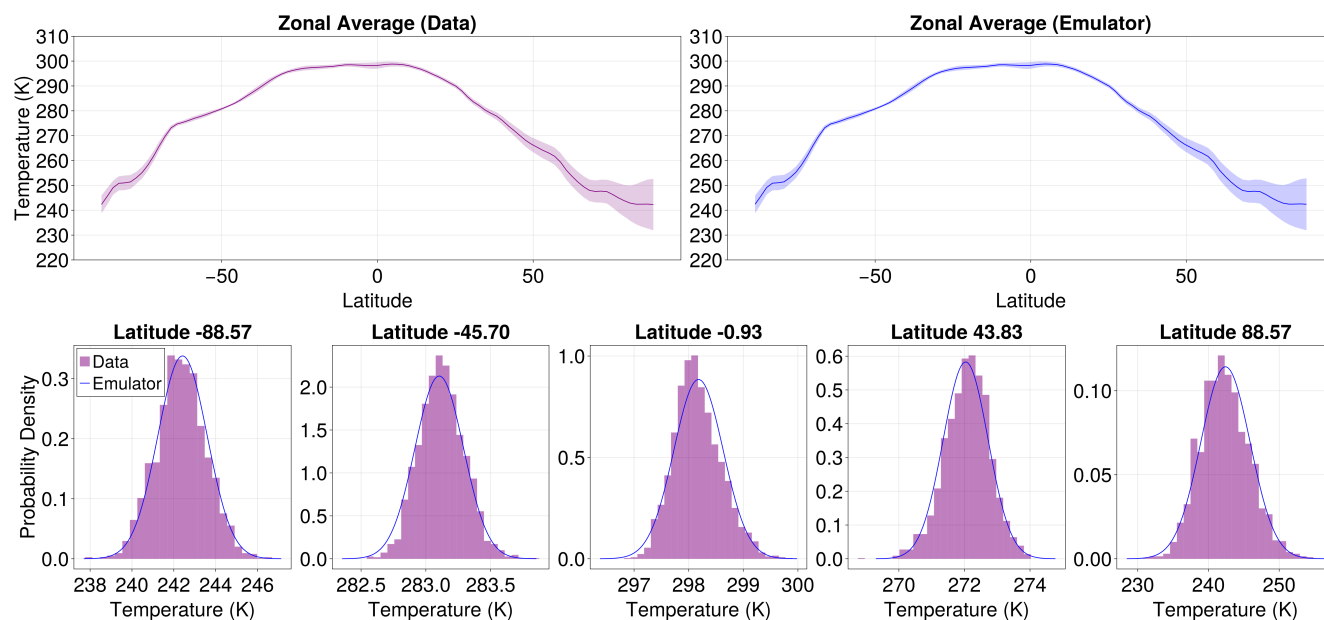
Figure 6. Spatially coarse emulator statistics. The purple color indicates data coming from the MPI model over the historical training period with similar global mean temperatures, and in blue, the emulator prediction for the same period. The top panel shows the ensemble mean and variance of the zonally averaged surface temperature field at each latitude, where the shading corresponds to three standard deviations. In the bottom panel, we show histograms at several fixed latitudes and compare the empirical distribution of the MPI data to the emulator prediction with an additional Gaussian assumption (for plotting purposes).

380 in Figure 7 for two of the twelve months. We show the probability distributions of surface temperature (top row) and relative humidity (bottom row) averaged over land and ocean separately, where, to visualize the emulator estimates, we have assumed a multivariate Gaussian form for the distributions. We re-emphasize that the assumption of Gaussianity is only necessary for generating full distributions, but the emulation of mean and covariances remains valid even for variables with non-Gaussian statistics. For example, our emulator can be used to emulate the mean and covariance of pointwise precipitation, but a Gaussian

385 approximation would fail to reproduce its full distribution, which is known to depart strongly from Gaussian. We compare the temperature and relative humidity distributions for the years 1850 and 2050, where the latter is taken from the SSP2-4.5 scenario. While this scenario was not used in training, the emulator prediction for 2050 closely matches the MPI data. This further supports the generalizability of the emulator to arbitrary future scenarios. See Figure D2 in Appendix D for the same statistics as Figure 7 but for the SSP5-8.5 training scenario. Inspection of probability distributions can be used to determine whether

390 the shifts in the mean values of temperature and relative humidity over this century are significant or not. For temperature, we see that the distribution shifts towards a warmer climate are outside the internal variability. For relative humidity, the shifts in its mean value are in accordance with the expectation that relative humidity will decrease over land in a warmer climate
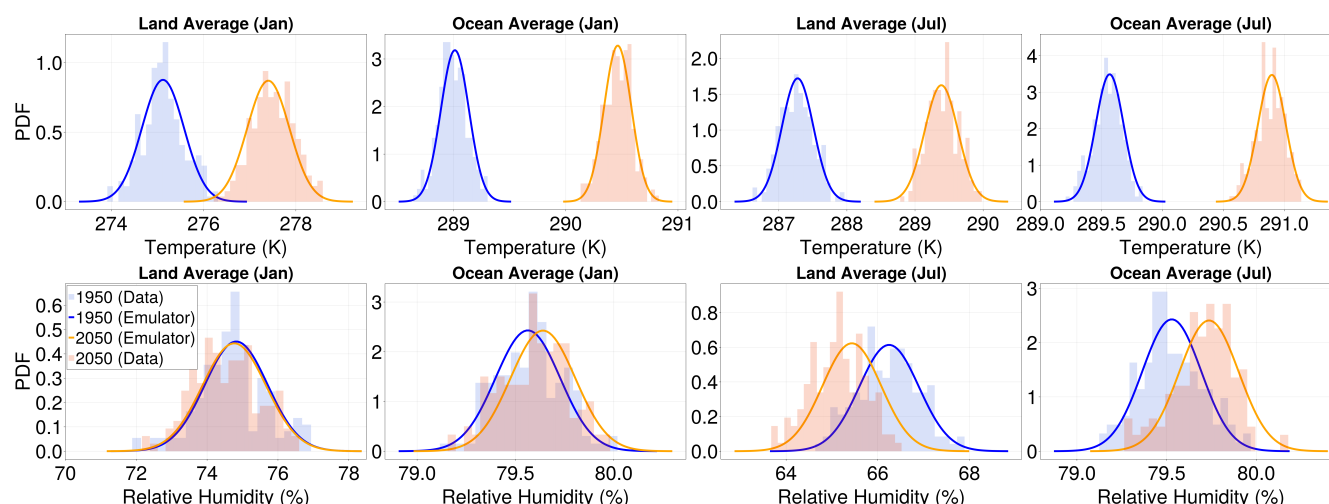
Figure 7. Distribution of temperature (top row) and relative humidity (bottom row) for the months of January (left) and July (right) under historical and SSP2-4.5 scenarios averaged over land and ocean. The emulator (solid line) captures the shift under climate change in mean and variance of the data distributions (histograms). For plotting purposes, we use our emulator's mean and variance prediction as input to a Gaussian distribution.

and increase over the ocean (Byrne and O'Gorman, 2016). However, the distributions shows that these shifts in mean relative humidity are outside the internal variability in July, but not in January.

395   Having shown that our emulator both captures spatial statistics and generalizes to previously unseen future scenarios, we proceed to illustrate how it might be used in practice. Figure 8 repeats the analysis of Figure 7 but for pointwise statistics at two locations. We can use the emulator to observe the localized effects of climate change under SSP2-4.5. We see apparent shifts in the pointwise temperature distributions (top row), but less so for relative humidity (bottom row), where in all cases, the shifts in mean are well within the variance of internal variability. The relative heights of the distributions within a given

400   panel offer a quick way to assess whether the variance has shifted. For example, there seems to be an increase in variance in the top left panel, and a decrease in variance in the top right panel. For both temperature and relative humidity, we see that shifts of means and variances are well captured. We re-emphasize that our emulator produces means and covariances which, for convenience, we plot as Gaussian distributions. Even in cases where the underlying distribution is non-Gaussian, the emulator captures the evident trends in mean and variance. Overall, the pointwise variance is well-represented by our model; see Figure

405   C2 in Appendix C for a quantitative metric. Appendix E further illustrates how the emulator can be used to study an arbitrary warming scenario outside of the SSP framework.
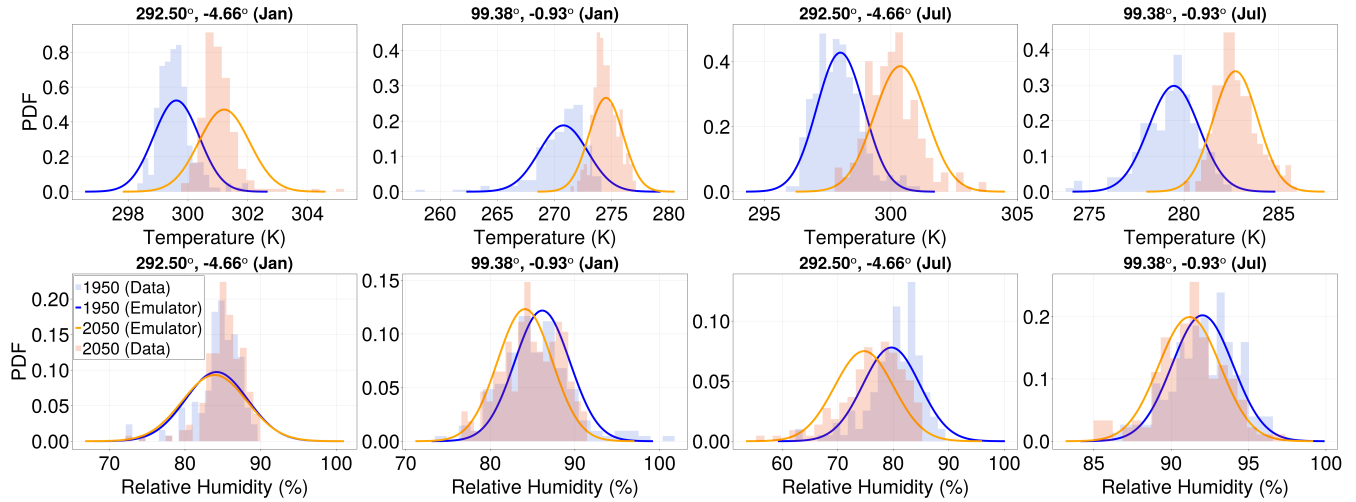
Figure 8. Similar to Figure 7, but for pointwise statistics at different locations on Earth rather than averaged over ocean and land. The emulator represents the overall trends in mean and variances even when the distributions are non-Gaussian.

6 Conclusions

We introduced an emulator for the evolution of the means and covariances of climate variables and applied it to spatially resolved monthly averaged temperature and relative humidity. This emulator provides a simple and computationally efficient method for extending the MPI-ESM1.2-LR global climate model to arbitrary warming scenarios while retaining the ability to quantify the significance of trends by modeling covariances. Computational efficiency is achieved by compressing the data onto the leading 1000 EOF modes of variability, reducing its memory footprint by two orders of magnitude compared to the training dataset. This dimensionality reduction is not a significant limitation, as Earth System Models have substantially greater fidelity at large scales than at fine spatial resolution.

The emulator is designed to capture changes in the mean and covariance structure of climate fields as a function of global mean temperature. We assume that EOF covariances vary smoothly with temperature. However, this assumption may break down in regions or timescales where localized processes—such as abrupt sea ice loss or monsoon regime shifts—introduce discontinuities in the covariance structure. Capturing such transitions may require future extensions that incorporate non-smooth, regime-dependent components.

Several straightforward extensions of the proposed emulator are possible. These include applying the method to additional climate fields, adopting alternative basis functions, using higher-order regression schemes, or incorporating additional predictors beyond global mean temperature. The framework can also be generalized to capture correlations between different variables (e.g., temperature and relative humidity) by modeling joint EOF amplitudes or post hoc correlations. Temporal coherence could similarly be introduced, for example, by estimating autocovariances of EOF amplitudes across months.
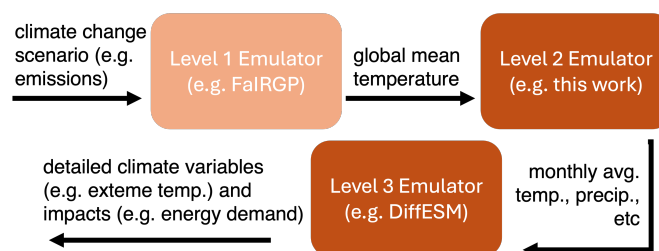
Figure 9. Potential ecosystem for coupled emulator models. FairGP (Bouabid et al., 2024) is an example of a simple climate model (SCM) that translates emissions scenarios to global mean temperature. DiffESM (Bassetti et al., 2024) is an example of a model that emulates daily values of climate fields conditioned on monthly mean temperature.

425   The emulator framework we introduced cannot be easily extended to statistics beyond mean and covariances. Estimating even the three-point correlation of a high-dimensional distribution becomes cumbersome, requiring the computation and storage of a tensor with N^3 points if one uses a basis of N EOF amplitudes. Generally, a multivariate distribution of size N necessitates the storage of N^n points for an n-point correlation, rapidly becoming intractable for large N or n. To emulate higher order statistics, like weather extremes, generative models like diffusion-based emulators (Song et al., 2020) may be better suited.

430 These approaches are well equipped to capture higher-order correlations. However, this added flexibility comes with increased cost and training requirements which may limit their use in applications that demand fast inference and interpretability on limited hardware.

     More generally, the present work lays the foundation for specialized downstream emulators. For example, one could condition a high-resolution or extreme-event emulator on monthly temperature fields produced by our framework. This would

435 integrate well with existing techniques, including Generalized Extreme Value theory or machine-learning-based generative models. Recent work such as Bassetti et al. (2024) and Christensen et al. (2024) illustrates how such hybrid frameworks might be developed. Figure 9 sketches a potential ecosystem of interconnected emulators. A simple climate model (e.g., FaIR-GP, Bouabid et al. 2024) translates emissions scenarios into global mean temperature trajectories. These are passed to an emulator of monthly-averaged spatial fields such as the one developed here. The final stage would use a downscaling model to recon-

440 struct finer-grained or daily fields. Limitations of this pipeline framework include the assumption that all regional statistics are determined by the global mean temperature and the challenge of capturing spatially heterogeneous forcing agents such as aerosols.

     Projections accounting for model spread, such as those emulated in this work, are vital to designing effective solutions to climate adaptation (Hansen et al., 2012; Deser et al., 2012; Woodruff, 2016). While the present emulator is trained on a par-

445 ticular Earth System Model (MPI-ESM1.2-LR), the methodology is broadly applicable to any model ensemble with sufficient realizations. Moreover, such model-trained emulators can serve as physically consistent priors that are further constrained with observational data to reduce structural model bias, as demonstrated in Immorlano et al. (2025).

Code availability. Analysis, plotting, and processing scripts may be found in https://github.com/sandreza/GaussianEarth.

Data availability. The MPI-ESM1.2-LR data may be downloaded from any of the public CMIP6 nodes listed here: https://pcmdi.llnl.gov/CMIP6/.

450 *Author contributions.* GG: data curation, investigation, software, writing (original draft). ANS: conceptualization, investigation, software, writing (original draft). GRF: conceptualization, writing (review and editing). RF: conceptualization, funding acquisition, writing (review and editing).

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This work acknowledges support by Schmidt Sciences, LLC, through the Bringing Computation to the Climate Chal-
455 lenge, an MIT Climate Grand Challenge Project. The authors would also like to thank Shahine Bouabid, Björn Lütjens, Iris de Vries, Mengze Wang, Paul O'Gorman, Themis Sapsis, and Claudia Tebaldi for their help in improving a preliminary version of this manuscript.

References

Alexeeff, S. E., Nychka, D., Sain, S. R., and Tebaldi, C.: Emulating mean patterns and variability of temperature across and within scenarios in anthropogenic climate change experiments, Climatic Change, 146, 319–333, 2018.

460 Barnston, A. G. and Livezey, R. E.: Classification, Seasonality and Persistence of Low-Frequency Atmospheric Circulation Patterns, Monthly Weather Review, 115, 1083 – 1126, https://doi.org/10.1175/1520-0493(1987)115<1083:CSAPOL>2.0.CO;2, 1987a.

Barnston, A. G. and Livezey, R. E.: Classification, seasonality and persistence of low-frequency atmospheric circulation patterns, Monthly weather review, 115, 1083–1126, 1987b.

Bassetti, S., Hutchinson, B., Tebaldi, C., and Kravitz, B.: DiffESM: Conditional Emulation of Temperature and Precipitation in Earth System
465 Models with 3D Diffusion Models, Authorea Preprints, 2024.

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

Bouabid, S., Sejdinovic, D., and Watson-Parris, D.: FaIRGP: A Bayesian energy balance model for surface temperatures emulation, Journal of Advances in Modeling Earth Systems, 16, e2023MS003 926, 2024.

470 Bradbury, J., Frostig, R., Hawkins, P., Johnson, M. J., Leary, C., Maclaurin, D., and Wanderman-Milne, S.: JAX: Composable transformations of Python+NumPy programs, http://github.com/google/jax, 2018.

Bruckner, T., Hooss, G., Füssel, H.-M., and Hasselmann, K.: Climate system modeling in the framework of the tolerable windows approach: the ICLIPS climate model, Climatic Change, 56, 119–137, 2003.

Byrne, M. P. and O'Gorman, P. A.: Understanding Decreases in Land Relative Humidity with Global Warming: Conceptual Model and GCM
475 Simulations, Journal of Climate, 29, 9045 – 9061, https://doi.org/10.1175/JCLI-D-16-0351.1, 2016.

Castruccio, S., Hu, Z., Sanderson, B., Karspeck, A., and Hammerling, D.: Reproducing internal variability with few ensemble runs, Journal of Climate, 32, 8511–8522, 2019.

Christensen, K., Otto, L., Bassetti, S., Tebaldi, C., and Hutchinson, B.: Diffusion-Based Joint Temperature and Precipitation Emulation of Earth System Models, arXiv preprint arXiv:2404.08797, 2024.

480 Collins, M. and Allen, M. R.: Assessing the relative roles of initial and boundary conditions in interannual to decadal climate predictability, Journal of Climate, 15, 3104–3109, 2002.

Deser, C., Phillips, A., Bourdette, V., and Teng, H.: Uncertainty in climate change projections: the role of internal variability, Climate dynamics, 38, 527–546, 2012.

Dorheim, K., Gering, S., Gieseke, R., Hartin, C., Pressburger, L., Shiklomanov, A. N., Smith, S. J., Tebaldi, C., Woodard, D. L., and Bond-
485 Lamberty, B.: Hector V3.2.0: functionality and performance of a reduced-complexity climate model, Geoscientific Model Development, 17, 4855–4869, 2024.

Doury, A., Somot, S., Gadat, S., Ribes, A., and Corre, L.: Regional climate model emulator based on deep learning: Concept and first evaluation of a novel hybrid downscaling approach, Climate Dynamics, 60, 1751–1779, 2023.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercom-
490 parison Project Phase 6 (CMIP6) experimental design and organization, Geoscientific Model Development, 9, 1937–1958, 2016.

Falasca, F., Basinski-Ferris, A., Zanna, L., and Zhao, M.: Diagnosing the pattern effect in the atmosphere-ocean coupled system through linear response theory, arXiv preprint arXiv:2408.12585, https://doi.org/10.48550/arXiv.2408.12585, 2024a.

Falasca, F., Perezhogin, P., and Zanna, L.: Data-driven dimensionality reduction and causal inference for spatiotemporal climate fields, Physical Review E, 109, 044 202, 2024b.

495   Gao, X., Sokolov, A., and Schlosser, C. A.: A Large Ensemble Global Dataset for Climate Impact Assessments, Scientific Data, 10, 801, https://doi.org/10.1038/s41597-023-02708-9, 2023.

Giorgini, L. T., Deck, K., Bischoff, T., and Souza, A.: Response Theory via Generative Score Modeling, Phys. Rev. Lett., 133, 267 302, https://doi.org/10.1103/PhysRevLett.133.267302, 2024.

Gleckler, P. J., Taylor, K. E., and Doutriaux, C.: Performance metrics for climate models, Journal of Geophysical Research: Atmospheres,
500   113, https://doi.org/10.1029/2007JD008972, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2007JD008972, 2008.

Goodwin, P., Leduc, M., Partanen, A.-I., Matthews, H. D., and Rogers, A.: A computationally efficient method for probabilistic local warming projections constrained by history matching and pattern scaling, demonstrated by WASP–LGRTC-1.0, Geoscientific Model Development, 13, 5389–5399, 2020.

Griffies, S. M., Adcroft, A., and Hallberg, R. W.: A Primer on the Vertical Lagrangian-Remap Method in Ocean Models Based
505   on Finite Volume Generalized Vertical Coordinates, Journal of Advances in Modeling Earth Systems, 12, e2019MS001 954, https://doi.org/https://doi.org/10.1029/2019MS001954, e2019MS001954 10.1029/2019MS001954, 2020.

Guo, S., Wang, J., Xiong, L., Ying, A., and Li, D.: A macro-scale and semi-distributed monthly water balance model to predict climate change impacts in China, Journal of hydrology, 268, 1–15, 2002.

Hannachi, A., Jolliffe, I. T., Stephenson, D. B., et al.: Empirical orthogonal functions and related techniques in atmospheric science: A review,
510   International journal of climatology, 27, 1119–1152, 2007.

Hansen, J., Ruedy, R., Sato, M., and Lo, K.: Global surface temperature change, Reviews of geophysics, 48, 2010.

Hansen, J., Sato, M., and Ruedy, R.: Perception of climate change, Proceedings of the National Academy of Sciences, 109, E2415–E2423, 2012.

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

Hawkins, E. and Sutton, R.: The potential to narrow uncertainty in regional climate predictions, Bulletin of the American Meteorological Society, 90, 1095–1108, 2009.

Hawkins, E. and Sutton, R.: The potential to narrow uncertainty in projections of regional precipitation change, Climate Dynamics, 37, 407–418, https://doi.org/10.1007/s00382-010-0810-6, 2011.

520   Hawkins, E. and Sutton, R.: Time of emergence of climate signals, Nature Climate Change, 2, 646–649, https://doi.org/10.1038/nclimate1585, 2012.

Holden, P., Edwards, N., Garthwaite, P., Fraedrich, K., Lunkeit, F., Kirk, E., Labriet, M., Kanudia, A., and Babonneau, F.: PLASIM-ENTSem v1. 0: a spatio-temporal emulator of future climate change for impacts assessment, Geoscientific Model Development, 7, 433–451, 2014.

Holden, P. B. and Edwards, N. R.: Dimensionally reduced emulation of an AOGCM for application to integrated assessment modelling,
525   Geophysical Research Letters, 37, 2010.

Holden, P. B., Edwards, N. R., Garthwaite, P. H., and Wilkinson, R. D.: Emulation and interpretation of high-dimensional climate model outputs, Journal of Applied Statistics, 42, 2038–2055, 2015.

Huntingford, C. and Cox, P.: An analogue model to derive additional climate change scenarios from existing GCM simulations, Climate dynamics, 16, 575–586, 2000.

530   Immorlano, F., Eyring, V., le Monnier de Gouville, T., Accarino, G., Elia, D., Mandt, S., Aloisio, G., and Gentine, P.: Transferring climate change physical knowledge, Proceedings of the National Academy of Sciences, 122, e2413503 122, https://doi.org/10.1073/pnas.2413503122, 2025.

Kemsley, S. W., Osborn, T. J., Dorling, S. R., and Wallace, C.: Pattern scaling the parameters of a Markov-chain gamma-distribution daily precipitation generator, International Journal of Climatology, 44, 144–159, https://doi.org/https://doi.org/10.1002/joc.8320, 2024.

535   Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, arXiv preprint arXiv:1412.6980, https://arxiv.org/abs/1412.6980, 2014.

Kitsios, V., O'Kane, T. J., and Newth, D.: A machine learning approach to rapidly project climate responses under a multitude of net-zero emission pathways, Communications Earth & Environment, 4, 355, 2023.

Kravitz, B. and Snyder, A.: Pangeo-Enabled ESM Pattern Scaling (PEEPS): A customizable dataset of emulated Earth System Model output,

540   PLOS Climate, 2, e0000 159, 2023.

Kutzbach, J. E.: Empirical eigenvectors of sea-level pressure, surface temperature and precipitation complexes over North America, Journal of Applied Meteorology and Climatology, 6, 791–802, 1967.

Lai, C.-Y., Hassanzadeh, P., Sheshadri, A., Sonnewald, M., Ferrari, R., and Balaji, V.: Machine Learning for Climate Physics and Simulations, Annual Review of Condensed Matter Physics, 16, 343–365, https://doi.org/https://doi.org/10.1146/annurev-conmatphys-043024-114758,

545   2025.

Leach, N. J., Jenkins, S., Nicholls, Z., Smith, C. J., Lynch, J., Cain, M., Walsh, T., Wu, B., Tsutsui, J., and Allen, M. R.: FaIRv2.0.0: a generalized impulse response model for climate uncertainty and future scenario exploration, Geoscientific Model Development, 14, 3007–3036, https://doi.org/10.5194/gmd-14-3007-2021, 2021.

Legates, D. R.: An evaluation of procedures to estimate monthly precipitation probabilities, Journal of Hydrology, 122, 129–140, 1991.

550   Lehner, F., Deser, C., Maher, N., Marotzke, J., Fischer, E. M., Brunner, L., Knutti, R., and Hawkins, E.: Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6, Earth System Dynamics, 11, 491–508, https://doi.org/10.5194/esd-11-491-2020, 2020.

Lembo, V., Lucarini, V., and Ragone, F.: Beyond Forcing Scenarios: Predicting Climate Change through Response Operators in a Coupled General Circulation Model, Scientific Reports, 10, 8668, https://doi.org/10.1038/s41598-020-65297-2, 2020.

555   Li, S., Rupp, D. E., Hawkins, L., Mote, P. W., McNeall, D., Sparrow, S. N., Wallom, D. C., Betts, R. A., and Wettstein, J. J.: Reducing climate model biases by exploring parameter space with large ensembles of climate model simulations and statistical emulation, Geoscientific Model Development, 12, 3017–3043, 2019.

Link, R., Snyder, A., Lynch, C., Hartin, C., Kravitz, B., and Bond-Lamberty, B.: Fldgen v1.0: an emulator with internal variability and space–time correlation for Earth system models, Geoscientific Model Development, 12, 1477–1489, https://doi.org/10.5194/gmd-12-

560   1477-2019, 2019.

Lorenz, E. N.: Empirical orthogonal functions and statistical weather prediction, vol. 1, Massachusetts Institute of Technology, Department of Meteorology Cambridge, 1956.

Lütjens, B., Ferrari, R., Watson-Parris, D., and Selin, N.: The impact of internal variability on benchmarking deep learning climate emulators, arXiv preprint arXiv:2408.05288, 2024.

565   Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B.: Summary for Policymak-

ers, Cambridge University Press, Cambridge, United Kingdom, https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_ SPM.pdf, contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, 2021.

Matthews, H. D., Gillett, N. P., Stott, P. A., and Zickfeld, K.: The proportionality of global warming to cumulative carbon emissions, Nature, 459, 829–832, 2009.

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., Brovkin, V., Claussen, M., Crueger, T., Esch, M., et al.: Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1. 2) and its response to increasing CO2, Journal of Advances in Modeling Earth Systems, 11, 998–1038, 2019.

McKinnon, K. A. and Deser, C.: Internal variability and regional climate trends in an observational large ensemble, Journal of Climate, 31, 6783–6802, 2018.

Meinshausen, M., Raper, S. C. B., and Wigley, T. M. L.: Emulating coupled atmosphere-ocean and carbon cycle models with a simpler model, MAGICC6 – Part 1: Model description and calibration, Atmospheric Chemistry and Physics, 11, 1417–1456, 2011.

Mitchell, T. D.: Pattern scaling: an examination of the accuracy of the technique for describing future climates, Climatic change, 60, 217–242, 2003.

Monahan, A. H., Fyfe, J. C., Ambaum, M. H., Stephenson, D. B., and North, G. R.: Empirical orthogonal functions: The medium is the message, Journal of Climate, 22, 6501–6514, 2009.

Nath, S., Lejeune, Q., Beusch, L., Seneviratne, S. I., and Schleussner, C.-F.: MESMER-M: an Earth system model emulator for spatially resolved monthly temperature, Earth System Dynamics, 13, 851–877, https://doi.org/10.5194/esd-13-851-2022, 2022.

Odjugo, P. A.: General overview of climate change impacts in Nigeria, Journal of human ecology, 29, 47–55, 2010.

Olonscheck, D., Suarez-Gutierrez, L., Milinski, S., Beobide-Arsuaga, G., Baehr, J., Fröb, F., Ilyina, T., Kadow, C., Krieger, D., Li, H., et al.: The new Max Planck Institute grand ensemble with CMIP6 forcing and high-frequency model output, Journal of Advances in Modeling Earth Systems, 15, e2023MS003 790, 2023.

O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J.-F., Lowe, J., Meehl, G. A., Moss, R., Riahi, K., and Sanderson, B. M.: The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6, Geoscientific Model Development, 9, 3461–3482, https://doi.org/10.5194/gmd-9-3461-2016, 2016a.

O'Neill, B. C., Tebaldi, C., Van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J.-F., Lowe, J., et al.: The scenario model intercomparison project (ScenarioMIP) for CMIP6, Geoscientific Model Development, 9, 3461–3482, 2016b.

Osborn, T. J., Wallace, C. J., Harris, I. C., and Melvin, T. M.: Pattern scaling using ClimGen: monthly-resolution future climate scenarios including changes in the variability of precipitation, Climatic Change, 134, 353–369, https://doi.org/10.1007/s10584-015-1509-9, 2016a.

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

O'Neill, B. C., M Done, J., Gettelman, A., Lawrence, P., Lehner, F., Lamarque, J.-F., Lin, L., J Monaghan, A., Oleson, K., Ren, X., et al.: The benefits of reduced anthropogenic climate change (BRACE): a synthesis, Climatic Change, 146, 287–301, 2018.

Peatier, S., Sanderson, B. M., Terray, L., and Roehrig, R.: Investigating parametric dependence of climate feedbacks in the atmospheric component of CNRM-CM6-1, Geophysical Research Letters, 49, e2021GL095 084, 2022.

Santer, B. D. and Wigley, T.: Regional validation of means, variances, and spatial patterns in general circulation model control runs, Journal of Geophysical Research: Atmospheres, 95, 829–850, 1990.

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

605 Schneider, T., Bischoff, T., and Płotka, H.: Physics of Changes in Synoptic Midlatitude Temperature Variability, Journal of Climate, 28, 2312–2331, https://doi.org/10.1175/JCLI-D-14-00632.1, 2015.

Schneider, T., Behera, S., Boccaletti, G., Deser, C., Emanuel, K., Ferrari, R., Leung, L. R., Lin, N., Müller, T., Navarra, A., Ndiaye, O., Stuart, A., Tribbia, J., and Yamagata, T.: Harnessing AI and computing to advance climate modelling and prediction, Nature Climate Change, 13, 887–889, https://doi.org/10.1038/s41558-023-01769-3, 2023.

610 Schneider, T., Leung, L. R., and Wills, R. C. J.: Opinion: Optimizing climate models with process-knowledge, resolution, and AI, EGUsphere, 2024, 1–26, https://doi.org/10.5194/egusphere-2024-20, 2024.

Schär, C., Vidale, P., Lüthi, D., Frei, C., Häberli, C., Liniger, M. A., and Appenzeller, C.: The role of increasing temperature variability in European summer heatwaves, Nature, 427, 332–336, https://doi.org/10.1038/nature02300, 2004.

Silvestri, S., Wagner, G. L., Campin, J.-M., Constantinou, N. C., Hill, C. N., Souza, A., and Ferrari, R.: A New WENO-Based Momentum Ad-
615 vection Scheme for Simulations of Ocean Mesoscale Turbulence, Journal of Advances in Modeling Earth Systems, 16, e2023MS004 130, https://doi.org/https://doi.org/10.1029/2023MS004130, e2023MS004130 2023MS004130, 2024a.

Silvestri, S., Wagner, G. L., Constantinou, N. C., Hill, C. N., Campin, J.-M., Souza, A. N., Bishnu, S., Churavy, V., Marshall, J. C., and Ferrari, R.: A GPU-based ocean dynamical core for routine mesoscale-resolving climate simulations, Tech. rep., https://doi.org/10.22541/essoar.171708158.82342448/v1, 2024b.

620 Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B.: Score-based generative modeling through stochastic differential equations, arXiv preprint arXiv:2011.13456, 2020.

Souza, A. N.: Representing turbulent statistics with partitions of state space. Part 1. Theory and methodology, Journal of Fluid Mechanics, 997, A1, https://doi.org/10.1017/jfm.2024.658, 2024a.

Souza, A. N.: Representing turbulent statistics with partitions of state space. Part 2. The compressible Euler equations, Journal of Fluid
625 Mechanics, 997, A2, https://doi.org/10.1017/jfm.2024.657, 2024b.

Souza, A. N., He, J., Bischoff, T., Waruszewski, M., Novak, L., Barra, V., Gibson, T., Sridhar, A., Kandala, S., Byrne, S., Wilcox, L. C., Koz-don, J., Giraldo, F. X., Knoth, O., Marshall, J., Ferrari, R., and Schneider, T.: The Flux-Differencing Discontinuous Galerkin Method Applied to an Idealized Fully Compressible Nonhydrostatic Dry Atmosphere, Journal of Advances in Modeling Earth Systems, 15, e2022MS003 527, https://doi.org/https://doi.org/10.1029/2022MS003527, e2022MS003527 2022MS003527, 2023.

630 Taylor, M., Caldwell, P., Bertagna, L., Clevenger, C., Donahue, A., Foucar, J., Guba, O., Hillman, B., Keen, N., Krishna, J., Norman, M., Sreepathi, S., Terai, C., White, J. B., Salinger, A. G., McCoy, R. B., y. R. Leung, L., Bader, D. C., and Wu, D.: The Simple Cloud-Resolving E3SM Atmosphere Model Running on the Frontier Exascale System, in: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, SC '23, Association for Computing Machinery, New York, NY, USA, ISBN 9798400701092, https://doi.org/10.1145/3581784.3627044, 2023.

635 Tebaldi, C. and Arblaster, J. M.: Pattern scaling: Its strengths and limitations, and an update on the latest model simulations, Climatic Change, 122, 459–471, 2014.

Tebaldi, C. and Friedlingstein, P.: Delayed detection of climate mitigation benefits due to climate inertia and variability, Proceedings of the National Academy of Sciences, 110, 17 229–17 234, https://doi.org/10.1073/pnas.1300005110, 2013.

Tebaldi, C. and Knutti, R.: The use of the multi-model ensemble in probabilistic climate projections, Philosophical transactions of the royal
640 society A: mathematical, physical and engineering sciences, 365, 2053–2075, 2007.

Tebaldi, C. and Knutti, R.: Evaluating the accuracy of climate change pattern emulation for low warming targets, Environmental Research Letters, 13, 055 006, 2018.

Tebaldi, C., Debeire, K., Eyring, V., Fischer, E., Fyfe, J., Friedlingstein, P., Knutti, R., Lowe, J., O'Neill, B., Sanderson, B., et al.: Climate model projections from the scenario model intercomparison project (ScenarioMIP) of CMIP6, Earth System Dynamics, 12, 253–293, 2021a.

Tebaldi, C., Dorheim, K., Wehner, M., and Leung, R.: Extreme metrics from large ensembles: investigating the effects of ensemble size on their estimates, Earth System Dynamics, 12, 1427–1501, 2021b.

Tebaldi, C., Selin, N., Ferrari, R., and Flierl, G.: Emulators of Climate Model Output, Annual Review of Environment and Resources, https://doi.org/https://doi.org/10.1146/annurev-environ-012125-085838, 2025.

Waidelich, P., Batibeniz, F., Rising, J., Kikstra, J. S., and Seneviratne, S. I.: Climate damage projections beyond annual temperature, Nature Climate Change, pp. 1–8, 2024.

Watson-Parris, D., Rao, Y., Olivié, D., Seland, Ø., Nowack, P., Camps-Valls, G., Stier, P., Bouabid, S., Dewey, M., Fons, E., et al.: ClimateBench v1. 0: A benchmark for data-driven climate projections, Journal of Advances in Modeling Earth Systems, 14, e2021MS002 954, 2022.

Woodruff, S. C.: Planning for an unknowable future: Uncertainty in climate change adaptation planning, Climatic Change, 139, 445–459, 2016.

Yu, S., Hannah, W., Peng, L., Lin, J., Bhouri, M. A., Gupta, R., Lütjens, B., Will, J. C., Behrens, G., Busecke, J., et al.: ClimSim: A large multi-scale dataset for hybrid physics-ML climate emulation, Advances in Neural Information Processing Systems, 36, 2024.

Yuan, X.-C., Zhang, N., Wang, W.-Z., and Wei, Y.-M.: Large-scale emulation of spatio-temporal variation in temperature under climate change, Environmental Research Letters, 16, 014 041, 2021.

Zelazowski, P., Huntingford, C., Mercado, L. M., and Schaller, N.: Climate pattern-scaling set for an ensemble of 22 GCMs–adding uncertainty to the IMOGEN version 2.0 impact system, Geoscientific Model Development, 11, 541–560, 2018.

Appendix A: Statistics of EOF Amplitudes and Pointwise Quantities

We mention in Section 3 that under certain assumptions, it is possible to translate the means and covariances produced by our emulator into higher-order statistics. For example, leading EOFs represent averages of the original variables over large swaths of Earth and, as such, are generally well-represented as multivariate Gaussians from central limit theorem arguments. Both monthly and spatial averaging make the multivariate statistics of the EOF amplitudes more Gaussian than the original variables. To substantiate this ansatz, we leverage evidence from the literature that spatially coarse-grained and monthly mean temperatures follow a Gaussian distribution (e.g., Schär et al. (2004); Hansen et al. (2010); Schneider et al. (2015); Falasca et al. (2024a, b)). The multivariate Gaussian assumption is more appropriate for smoothly varying variables like temperature and relative humidity than for variables like precipitation, which have a much more nonlinear response to temperature fluctuations and non-Gaussian statistics (Legates, 1991).

Although not necessary for the work in the main text, we will show that we can represent many statistics of interest by assuming that $\boldsymbol{a}_C$ is well represented by a multivariate Gaussian, i.e.

$$\rho(\boldsymbol{a}_C | \overline{T}_g, m) \approx \mathcal{N}(\boldsymbol{\mu}(\overline{T}_g, m), \boldsymbol{\mathcal{C}}(\overline{T}_g, m)), \tag{A1}$$

where the coarse statistical variables $\boldsymbol{a}_C$ are approximated as a Gaussian distribution, denoted $\mathcal{N}$, with means $\boldsymbol{\mu}$ and covariances $\mathcal{C}$ as a function of the global mean temperature $\overline{T}_g$ and month m.

For example, we show in Figure A1 the distributions of EOF modes at selected locations of surface temperature in purple, chosen from a subset of the historical period of the MPI ensemble with similar global mean temperatures. The figure illustrates
680   the four most "non-Gaussian" modes/locations and one "most Gaussian" mode/location. Specifically, the modes and locations were selected by constructing histograms for every location and mode, finding the locations/modes with the most positive/negative kurtosis and skewness (four total) and one location with skewness and kurtosis closest to zero. We show in blue the result of the emulator mean and variance prediction for the statistics under a Gaussian assumption.

We see from Figure A1 that even the most "non-Gaussian" EOF coefficients (top row) display a familiar bell-shaped curve,
685   whereas the different locations for pointwise statistics display non-Gaussian features (bottom row). *A subtle point now arises.* All distributions of the EOF coefficients appear to be quasi-Gaussian. Furthermore, point statistics can be reconstructed from the EOF mode statistics and the EOF basis through a linear sum. Lastly, sums of Gaussian random variables are Gaussian. Reconciling these three facts with the non-Gaussian point statistics of the bottom row in Figure A1 requires non-Gaussian higher-order correlations between the different EOF modes. These non-Gaussian correlations ought to be captured to emulate
690   the tails of the distributions at a location; this could be achieved with other data-driven methods such as "score-matching" or Markov models (see Souza, 2024a, b; Giorgini et al., 2024; Bassetti et al., 2024; Christensen et al., 2024). In this work, we focus on robust spatially coarse-grained and lower-order statistics. This focus allows us to ignore these higher-order moment effects.

In summary, while a multivariate Gaussian assumption is not necessary for our emulator, it can be justified for the variables
695   considered. The addition of this assumption would allow for the inference of higher-order statistics beyond the variance.

Appendix B: Linear emulator

In constructing our emulator we used a quadratic function to regress the mean EOF coefficients onto $\bar{T}_g$, see Equation 5. However, we also experimented with using a linear function to perform the same regression:

$$\hat{\boldsymbol{\mu}} = \hat{\boldsymbol{\mu}}_0 + \hat{\boldsymbol{\mu}}_1 \bar{T}_g. \tag{B1}$$

700   We found the linear emulator to result in a slightly worse fit, which we illustrate in Figure B1. In the top row of Figure B1, we plot the difference in the yearly and ensemble-averaged global temperature emulation error between the linear and quadratic emulators as a function of time. Emulators using an increasing number of EOF modes (10, 100, 1000) are plotted in different colors. At each time along the horizontal axis, the errors for both emulators are computed with respect to $\overline{T}_g$ from the MPI ensemble for the year in question; the error of the quadratic emulator is then subtracted from that of the linear emulator. In
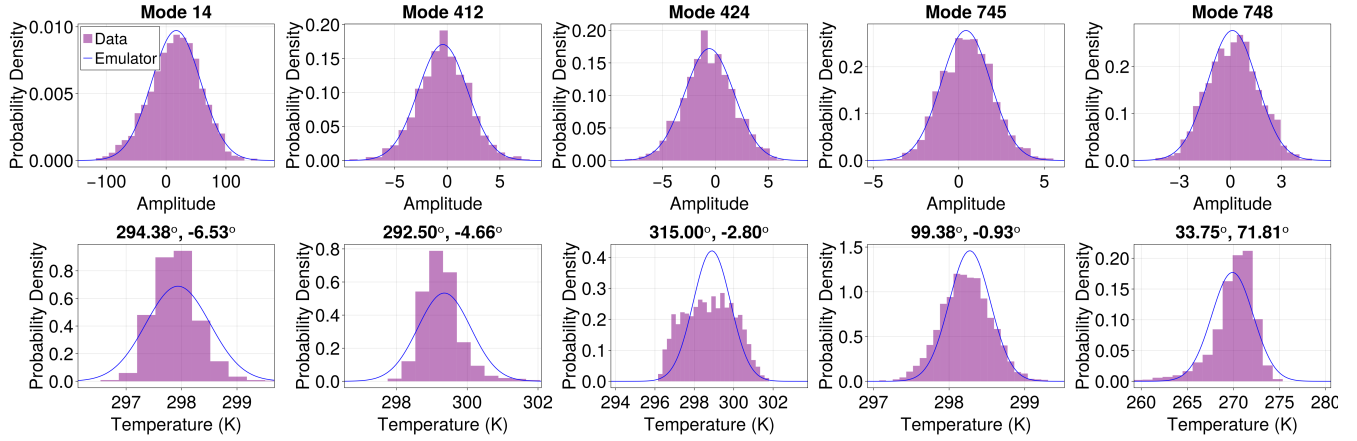
Figure A1. Statistics of surface temperature EOF modes and selected locations. In purple, we show the histograms of data collected over the historical period of the MPI ensemble with similar global mean temperatures. In blue, we show the fit given by the emulator described in this work. The EOF amplitudes populate the top row, and the bottom row constitutes point locations. The locations and modes were selected for being the least Gaussian, except for Mode 412 and location $99.39°, -0.93°$ which are the most Gaussian.

formulas, we are comparing (for each year) between the two emulators

$$\text{temporal error}(T, \hat{T}) =$$

$$\sqrt{\frac{1}{4\pi} \int\limits_{\theta=0}^{2\pi} \int\limits_{\phi=0}^{\pi} \left| \frac{1}{12} \sum_m \left(\langle T(\boldsymbol{x}, m, t, \omega) \rangle - \hat{\mu}(\boldsymbol{x}, m, \overline{T}_g(t)) \right) \right|^2 \sin(\phi) d\theta d\phi,} \tag{B2}$$

where $\langle \cdot \rangle$ to denotes an ensemble average. The spatial average is taken over the surface of Earth, with $x = (\theta, \phi)$, and t here denotes a fixed year. We emphasize that the training for both emulators is performed on the historical period and SSP5-8.5, whereas our "test" is with respect to SSP1-1.9 and SSP2-4.5.

Note that for a small number of modes, the linear and quadratic emulators perform about the same. As we increase the number of modes, however, the quadratic emulator has significantly lower errors in future scenarios where nonlinear effects become more prominent. The only significant exception is the historical period, in which the quadratic emulator appears to overfit to large volcanic events. The bottom row of Figure B1 illustrates the origin of the improvement in fit yielded by the quadratic emulator. Here we repeat the plots in Figure 3 for the three "most quadratic" modes (where the regression error difference between the linear and quadratic fit was maximal). The EOF coefficients of these modes—and many others—have a distinct quadratic dependence on $\overline{T}_g$, which is not captured by a linear emulator.

We additionally compare the spatial error, as defined in Equation 18, of the linear and quadratic emulators. Figure B2 shows the same spatial errors over the different scenarios as Figure 4, but for the linear emulator. Comparing the two figures, one observes a significant reduction in errors for the quadratic emulator, most notably in the Arctic, the North Atlantic, South Asia, and Sub-Saharan Africa.
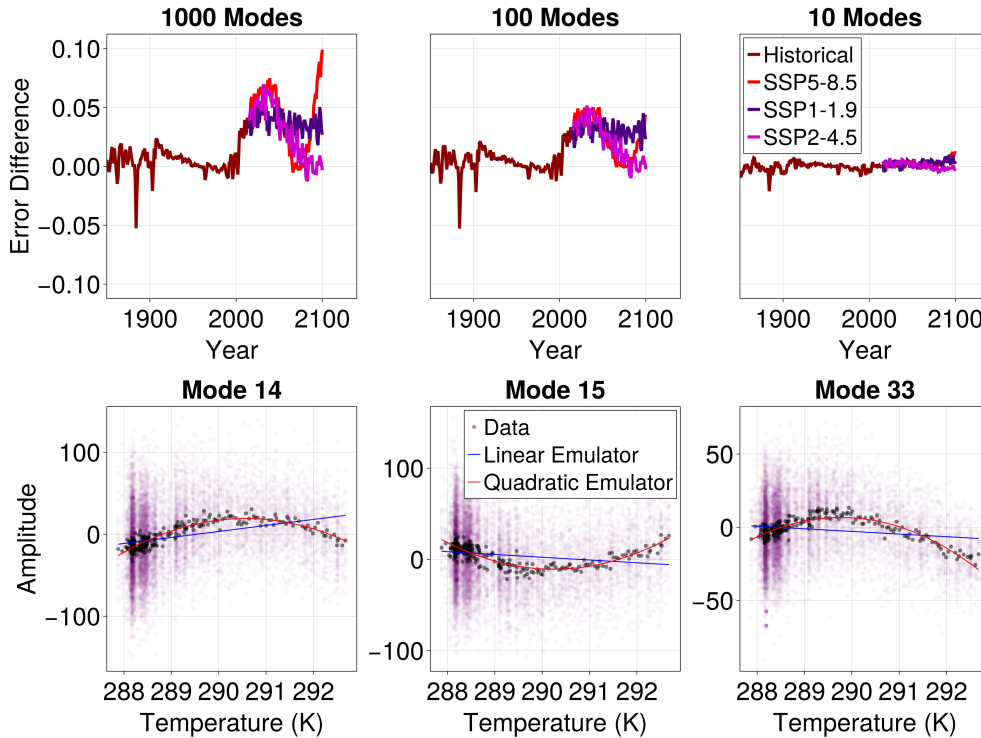
Figure B1. Difference in regression error between linear and quadratic emulators. Top row: difference in regression (emulation) error for ensemble and yearly averaged surface temperature as a function of time for different scenarios. Error of the quadratic emulator is subtracted from that of the linear emulator, so positive values indicate higher error in the linear emulator. Bottom row: fit comparison between quadratic and linear emulators for the three "most quadratic" EOF modes, analogous to Figure 3.

Appendix C: Comparison to linear pattern scaling

To demonstrate that the linear version of our emulator, as described above in Appendix B, reduces to a form of a linear pattern scaling emulator for mean temperature at a single location, we plot in Figure C1 the temporal error defined in Equation B2
725 as a function of time for an increasing number of EOF modes used in the emulator (10, 100, 1000). We compare the error of performing linear regression pointwise (i.e. applying pattern scaling) on the ensemble mean in the historical period and SSP5-8.5 to the error of the mean generated by our linear emulator. We re-emphasize that the training for both emulators is performed on the historical period and SSP5-8.5, whereas our "test" is with respect to SSP1-1.9 and SSP2-4.5.

As we increase the number of modes, the error in the approximation becomes similar to the pointwise error when utilizing
730 pattern scaling. In the limit of using all EOF modes, we expect this error to match that of pattern scaling. This is to be expected: a few modes corresponding to large-scale patterns cannot represent the ensemble mean's spatial structure in scenarios outside the historical period. This error is due to a combination of two factors. The basis functions are constructed over the historical period, and secondly, even though we fit to SSP5-8.5 data for the EOF amplitudes, there is less data corresponding to warmer
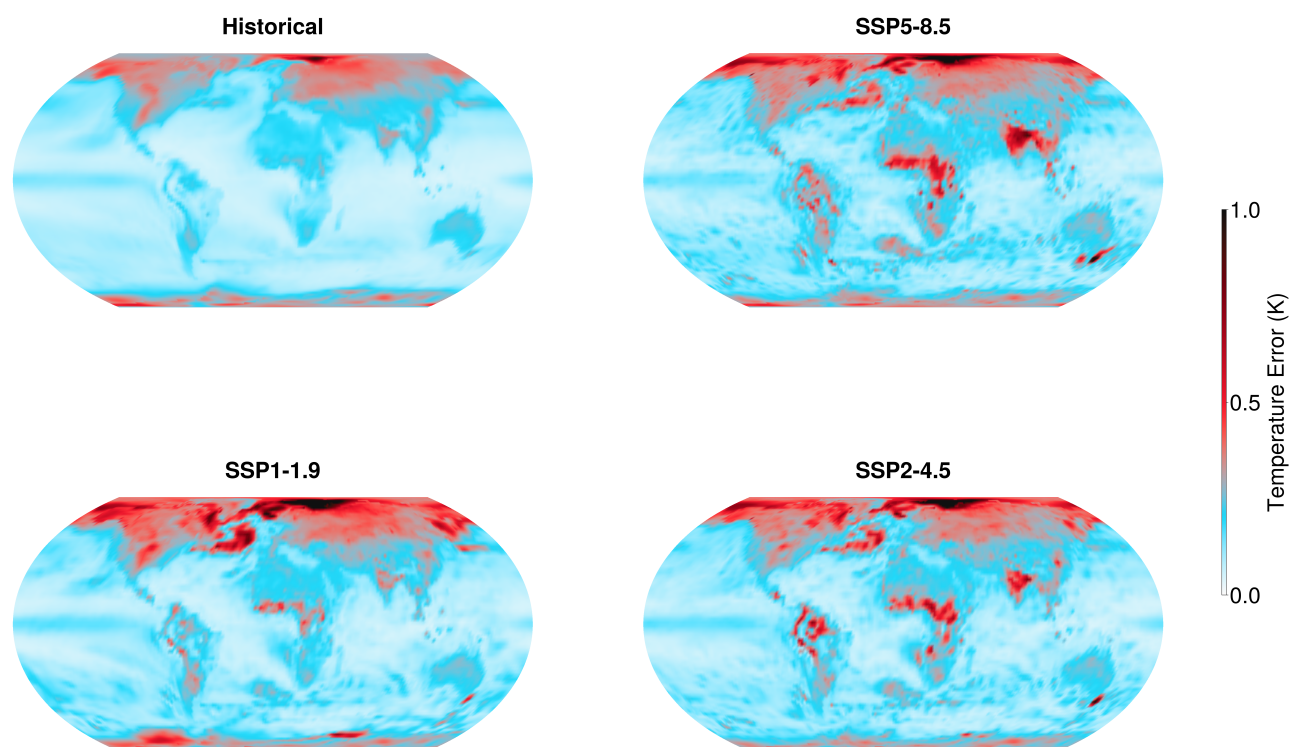
Figure B2. Average regression error of a linear emulator for ensemble and yearly averaged surface temperature as a function of space for different scenarios. Same as Figure 4, but for a linear emulator instead of a quadratic one.

temperatures. Thus, the emulator underperforms when it sees less data. With more modes (and hence a more complete basis
735   for representing functions), we see that the generalization error of going to different SSP scenarios matches the error of the historical period. Additionally, the quadratic version of the emulator described in the main text improves upon the errors shown in Figure C1, see Figure B1. We illustrate the error of the linear emulator here for the sake of an accurate comparison with pattern scaling.

As discussed in the main text, a major advantage of our emulator over a method like pattern scaling is the direct emulation
740   of covariances. To help illustrate this, we additionally plot in Figure C2 the temporal error of the ensemble-averaged standard deviation of January surface temperature as a function of time and for an increasing number of modes. In formulas, this is
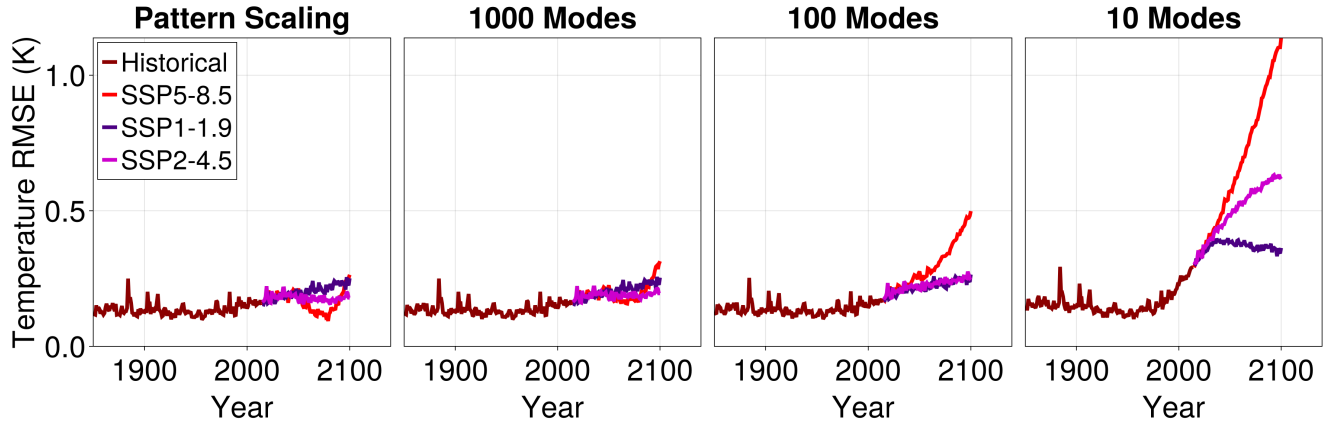
Figure C1. Regression error for ensemble and yearly averaged surface temperature as a function of time for different scenarios for a linear emulator. Different colors correspond to the historical 1850–2014 period (maroon) and the three future scenarios considered in this study: SSP5-8.5 (red), SSP2-4.5 (pink), and SSP1-1.9 (purple). The future period spans 2015–2100. The historical period lasts 165 years, and the projected period—86 years. We show the root mean square error of pattern-scaling on the left and increasing the number of modes used in the linear emulator in the subsequent rightward panels. As we increase the number of modes, the error in capturing pointwise statistics decreases.

analogous to Equation B2:

$$\text{Temporal Error}(\sigma_{T,m}, \hat{\sigma}_{T,m}) = \sqrt{\frac{1}{4\pi} \int\limits_{\theta=0}^{2\pi} \int\limits_{\phi=0}^{\pi} |\sigma_{T,m} - \hat{\sigma}_{T,m}|^2 \sin(\phi) d\theta d\phi}. \tag{C1}$$

For a given number of modes, the standard deviation errors are approximately constant as a function of time. This illustrates
745 that almost the entirety of the emulator's error in representing variance comes from the EOF approximation. Accordingly, as the number of EOF modes included increases, the error decreases.

Appendix D: Extended validation

In this appendix we provide additional figures in support of the validation of our emulator design, see Section 5 in the main text.
750 Since our emulator captures each month separately, we can investigate a-posteriori shifts in the ensemble average seasonal temperature cycle. In Figure D1, we show the seasonal emulator prediction for the upper and southern hemisphere averages separately, where the blue corresponds to a global mean temperature in the historical period and the orange corresponds to the end of the SSP5-8.5 scenario. The amplitude of seasonal variation changes by approximately one Kelvin in the northern hemisphere and is smaller in the southern hemisphere. This asymmetry reflects the larger fraction of land in the northern
755 hemisphere (land warms more than the ocean because it is drier and less efficient at cooling through latent heat release).
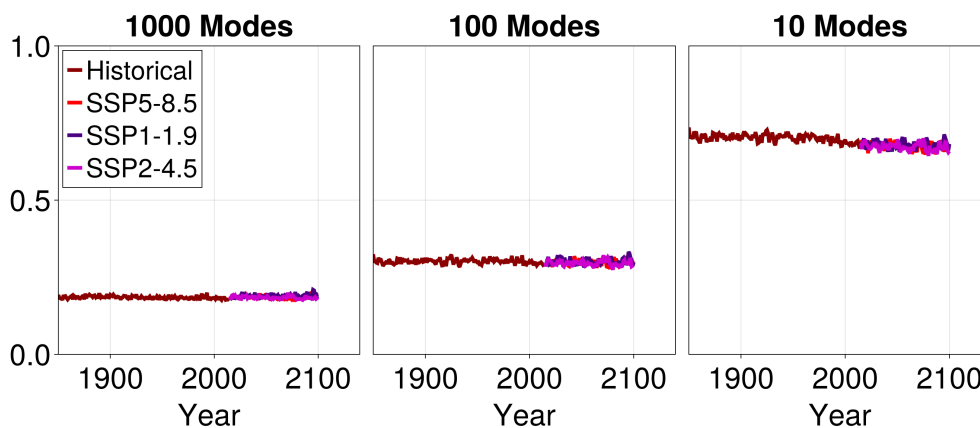
Figure C2. Regression error for ensemble averaged January surface temperature standard deviation as a function of time for different scenarios. Same as last three panels of Figure C1, but for the standard deviation of temperature in January.
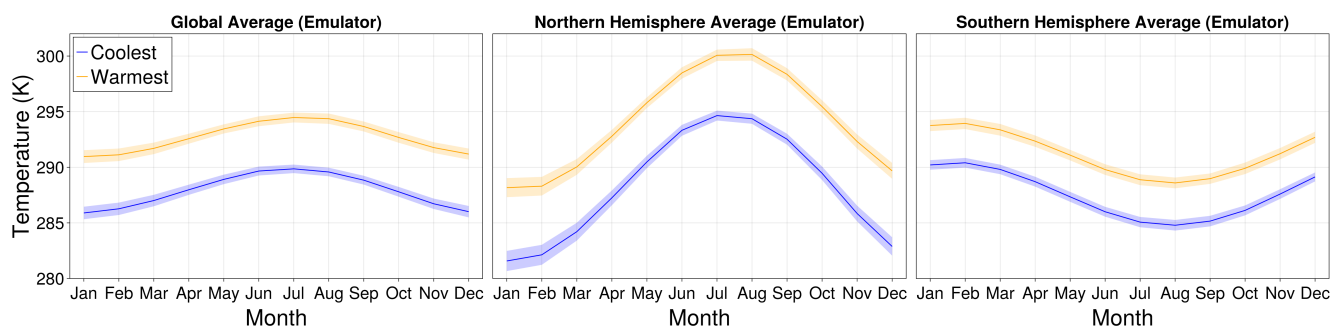
Figure D1. Monthly emulator output for global quantities. We show the emulator prediction for the global average (left), northern hemisphere average (middle), and southern hemisphere average (right), as well as two global mean temperatures, $\overline{T}_g = 288(K)$ (blue) and $\overline{T}_g = 293(K)$ (orange). The solid line indicates the ensemble average, and the shaded region indicates three standard deviations. We see a shift in the seasonal cycle for a warmer climate. In particular, in the past northern hemisphere, January has become similar to April.
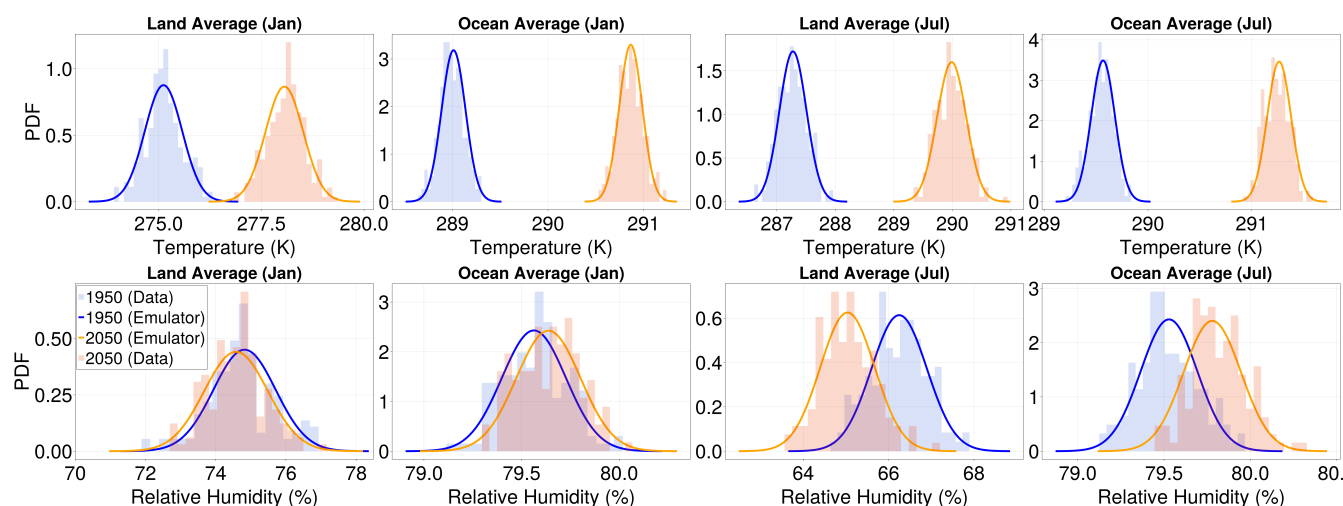
Figure D2. Distributional shifts under climate change for land and ocean spatial averages in SSP5-8.5. Same as Fig. 7, but here we compare the historical and SSP5-8.5 scenarios.

Additionally, Figures D2 and D3 show the same statistics as Figures 7 and 8, respectively, but for the training scenario SSP5-8.5.

Finally, Figures D4 and D5 illustrate, for surface temperature, the differences between the emulator and the original MPI data in 2100 across the three future scenarios. Figure D4 shows the year-average mean surface temperature, while Figure D5 shows the standard deviation of surface temperature in January. These spatial patterns are similar between the emulator and the data.

Appendix E: Emulator Usage and Case Study

Thus far we have only used pre-defined SSP scenarios to illustrate the generalizability of the emulator. In practice, however, the main application of an emulator is in quickly and cheaply inferring the behavior of an ESM in arbitrary scenarios. Figure E1 illustrates a case study for such an application. We define an arbitrary new climate change scenario in terms of the global mean temperature. This scenario, which is shown in the first column, grows exponentially at a rate significantly slower than that of SSP5-8.5, with an end of century global mean temperature 3 degrees Celsius above the 1850-1900 average. One could imagine that this scenario is the result of a slight reduction of emissions, driven by changing economic incentives but without strong international cooperation on climate goals. If we follow the course set out by this scenario, how soon would we be able to see the effects of emission reduction relative to the worst-case path of SSP5-8.5? In other words, when will the diverging scenarios yield significantly different climate statistics, and how will this vary across different locations on Earth? Questions such as these are relevant as they can help quantify the timescale of local changes from global climate policy. Similar questions have been explored in the BRACE project (O'Neill et al., 2018), but focusing only on two pre-defined scenarios.
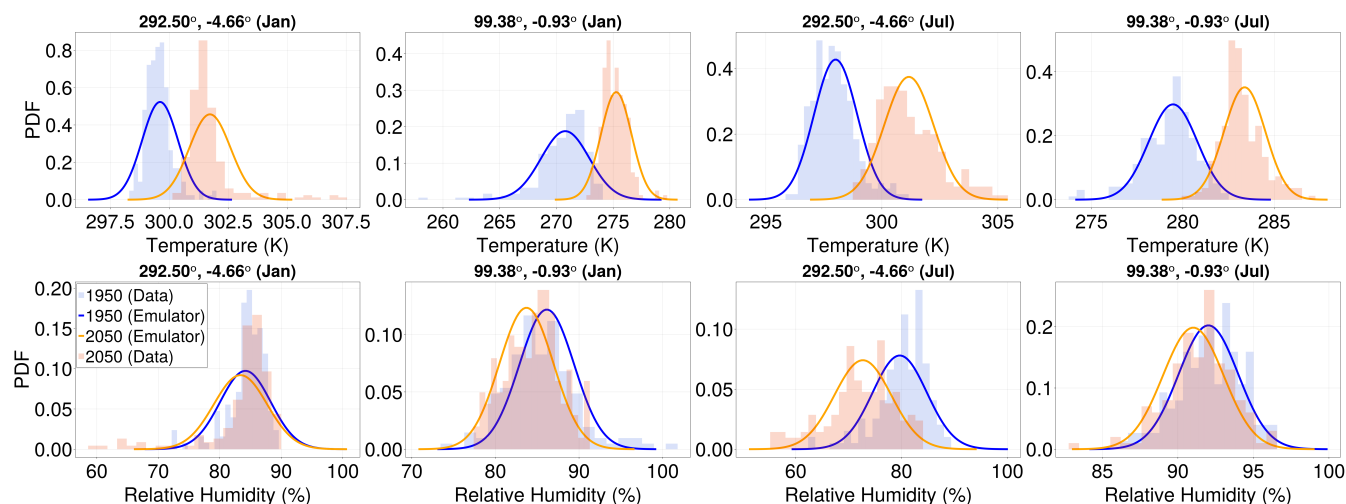
Figure D3. Distributional shifts under climate change for different locations in SSP5-8.5. Same as Fig. 8, but here we compare the historical and SSP5-8.5 scenarios.

To answer these questions, we examine the statistics for temperature and relative humidity at two geographic locations:
775 (88.1W, 40.1N) in the Midwestern United States, near Chicago, and (84.4E, 23.3N) in Jharkand in northwestern India. In the second and third columns of Figure E1 we show, for the Midwestern US and northwestern India, respectively, the distributions of temperature (top) and relative humidity (bottom) at the end of the century (in 2100) for all months. We plot, for both SSP5-8.5 and the new scenario, the mean monthly values $\pm$ 2 standard deviations as estimated by the emulator. For both locations, the variance provides critical context for assessing whether a change in mean is significant. Note that, for both variables, the
780 difference between the two scenarios is much more prominent in NW India. In contrast, in the Midwest US the differences between the scenarios are largely within the range of internal variability of the model; e.g., although there is a change in mean in relative humidity in August, accounting for variance suggests that that change is not statistically significant. The emulator additionally resolves seasonal differences in the influence of the mitigated scenario; e.g., for relative humidity in NW India, there is a significant difference between the scenarios only in the winter months November-April.
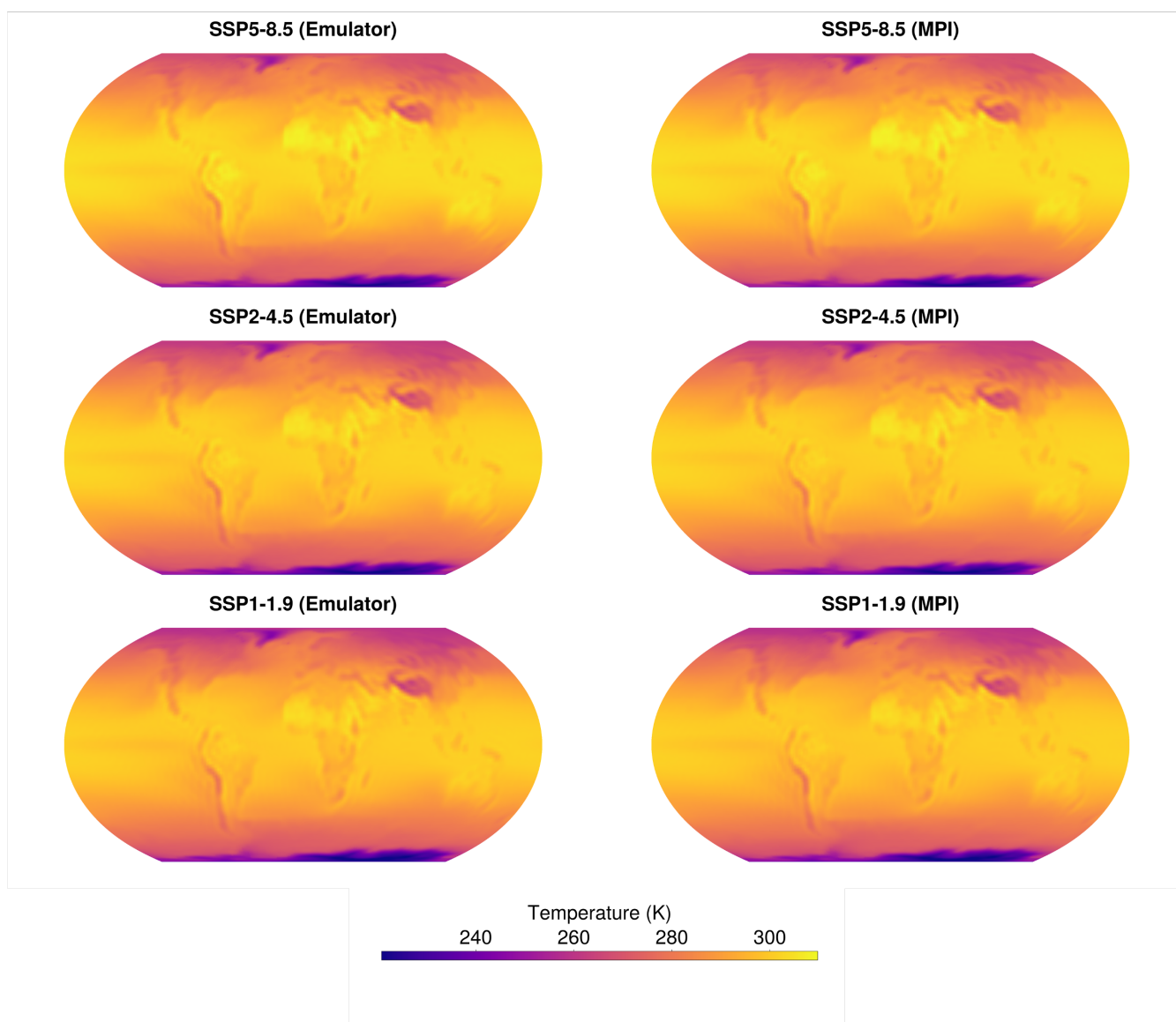
Figure D4. Emulator representation of yearly-average mean surface temperature in 2100. Emulator representation on the left, MPI model data on the right for comparison.

SSP5-8.5 (Emulator)   **SSP5-8.5 (MPI)**

SSP2-4.5 (Emulator)   **SSP2-4.5 (MPI)**

SSP1-1.9 (Emulator)   **SSP1-1.9 (MPI)**

Standard Deviation (K)
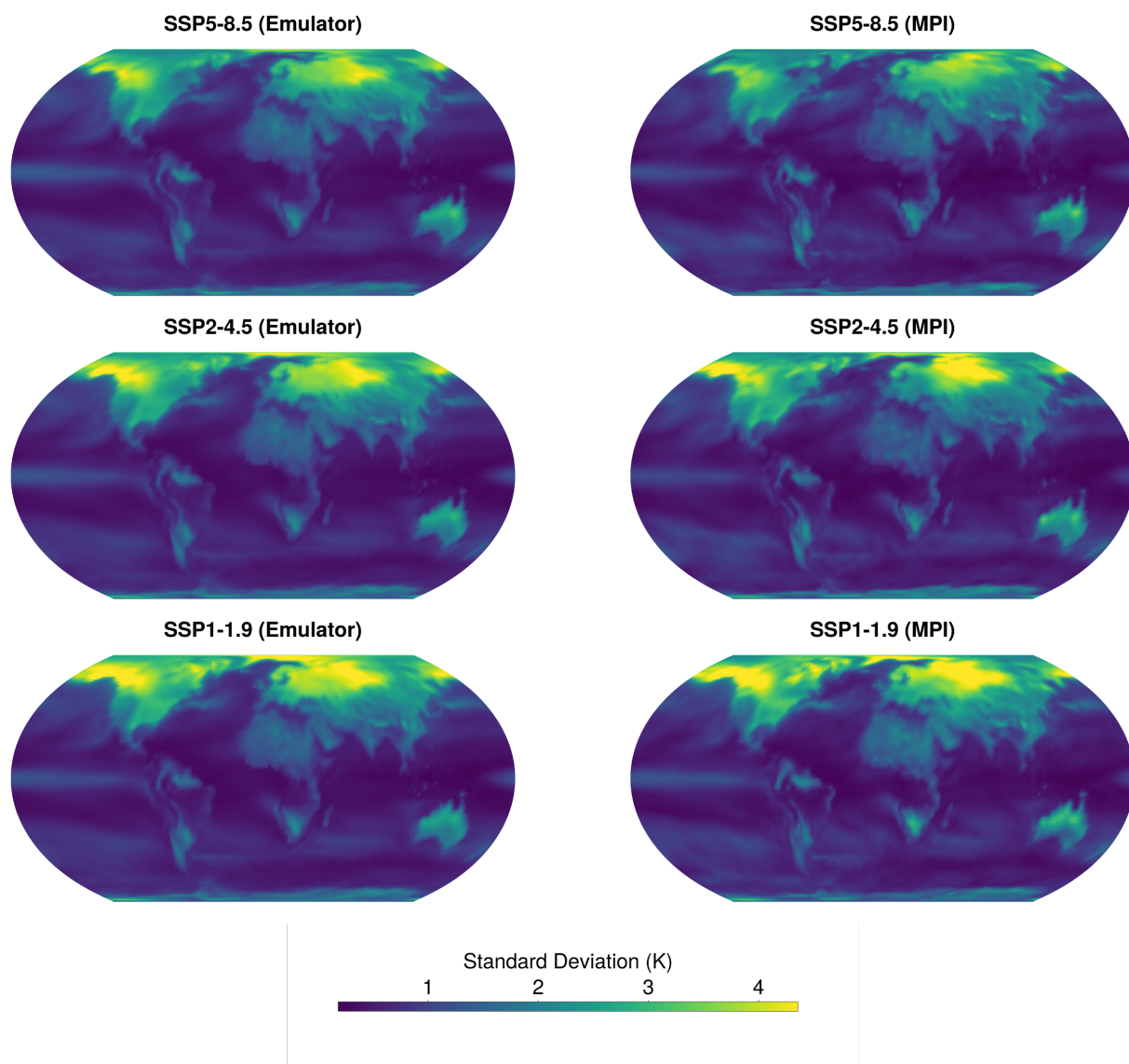
Figure D5. Emulator representation of January standard deviation of surface temperature in 2100. Emulator representation on the left, MPI model data on the right for comparison.
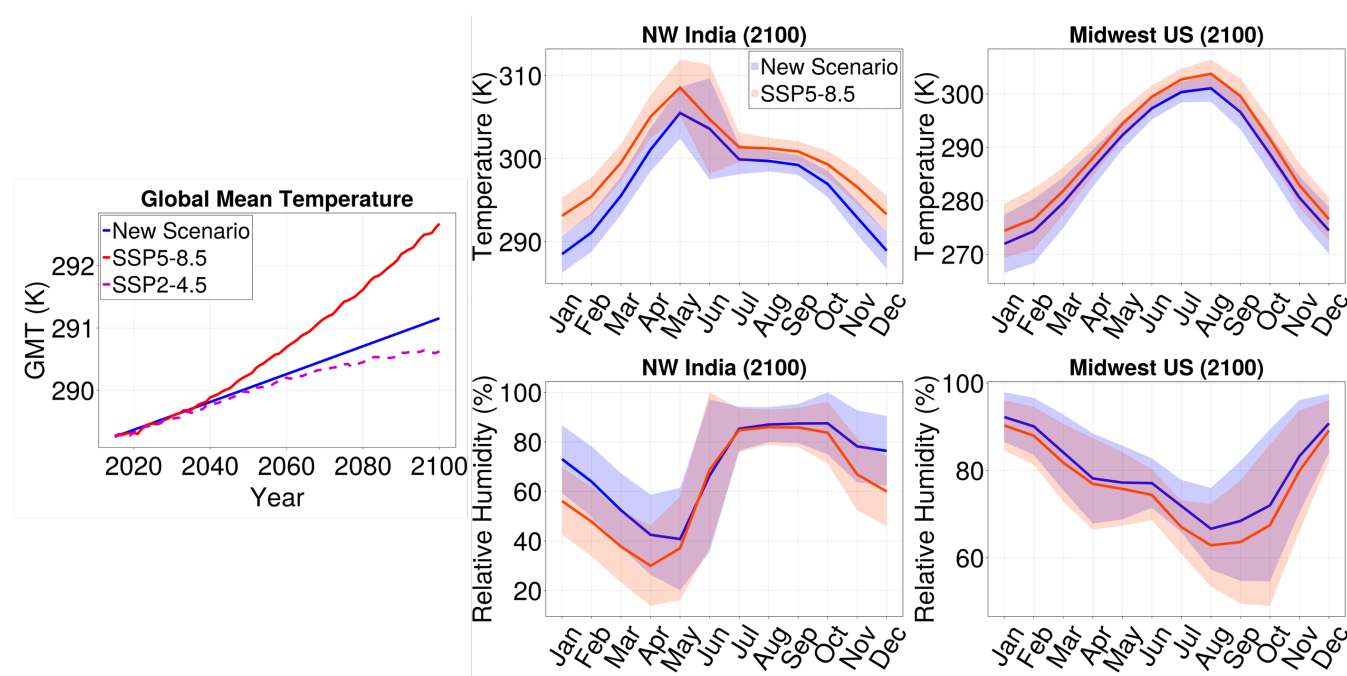
Figure E1. Case study: Using the emulator to assess impacts of emissions reduction. On the left we define a new scenario (blue) in global mean temperature. In the second and third columns ribbon plots show monthly mean temperature (top) and relative humidity (bottom) $\pm2\sigma$ in 2100 in SSP5-8.5 and the new scenario for two locations: NW India and Midwest US. This illustrates the importance of internal variability in climate assessment.