A Statistical Emulator Design for Averaged Climate Fields

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

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

Key Points:

- Assessing climate change requires predictions of how distributions of climate variables shift in time.
- Gaussian assumptions on changing statistics yield simple algorithms that capture global warming trends.
- Modeling variables as Gaussian distributions requires care and validation.

Corresponding author: Gosha Geogdzhayev, `geogdzh@gmail.com`

Abstract

Fast emulators of comprehensive climate models are used to explore the impact of anthropogenic emissions in future climate. A new approach to emulators is introduced that predicts distributions of coarse-grained monthly averaged variables as a multivariate Gaussian distribution. The emulator is trained with a state-of-the-art climate model and serves as a good first-order representation for many statistics of future climates. The emulator is applied to statistics of surface temperature and relative humidity for illustrative purposes, but the approach can be applied to any other variable of interest as long the multivariate Gaussian approximation captures the bulk of the distribution. Importantly the emulator accounts for the internal variability of the system, allowing one to examine shifts in distributions of climate variables. In this sense the work can be considered as an extension of pattern scaling emulators that focus on the evolution of the mean rather than the distribution of climate variables.

Plain Language Summary

Assessing how the climate changes as a consequence of human emissions of greenhouse gases requires modeling how ranges of temperatures and their likelihoods can change over time. Climate models serve as the best guess on how humans affect the climate, but they do not explore every possible future scenario that could be of interest. To this end, we develop a data-driven method that can serve as a fast and cheap surrogate to evaluate likely changes in variables like surface temperature and relative humidity at a regional scale in future climates. This work extends previous approaches in that it predicts not only the evolution of the mean of those variables, but also of their fluctuations due to internal variability in the climate system.

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. State-of-the-art Earth System Models (ESMs), such as those participating in the Climate Model Intercomparison Project (CMIP, Eyring et al. (2016)), are arguably our best approach for quantifying 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 cost of running ESMs, these models can only simulate the Earth system response to a few potential future scenarios (Tebaldi, Debeire, et al., 2021). 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, Tebaldi, Van Vuuren, et al., 2016; Waidelich et al., 2024).

In recent years, emulators of climate models have been gaining popularity as a way to extend the utility of ESMs. Climate emulators are simplified models trained to cheaply and quickly recreate the behavior of ESMs. The importance of emulators is likely to rise due to increasing and competing computational demands from the ever refining spatial resolution, complexity as embodied by the number of model components and their sophistication, the interest in using more accurate numerical methods (and hence computational grids), and the need to run initial condition ensembles, besides alternative scenarios (Nair & Toy, 2016; Griffies et al., 2020; Souza et al., 2023; Taylor et al., 2023; Silvestri, Wagner, Campin, et al., 2024; Silvestri, Wagner, Constantinou, et al., 2024; Schneider et al., 2024, 2023). The necessity of emulators is 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 with computational hardware available to everyday consumers. While emulators are most commonly used to extend ESMs to arbitrary climate change scenarios, emulators have also been developed for other applications, including climate model downscaling (Doury et al., 2023), parametrization of subgrid-scale processes (Li et al., 2019), and model parameter calibration (Peatier et al., 2022). This work focuses on the class of emulators trained to extend ESMs to arbitrary future scenarios. The simplest and most common emulation technique in this area is pattern scaling (Santer & Wigley, 1990; Huntingford & Cox, 2000; Mitchell, 2003). Pattern scaling estimates spatially resolved changes in climate variables by regressing local variables on global mean temperature. While pattern scaling performs well for projecting local mean temperatures (Santer et al., 1990; Lütjens et al., 2024), it has no inherent probabilistic component and is significantly less accurate for other climate variables (Tebaldi & Arblaster, 2014; Tebaldi & Knutti, 2018). This work focuses on the probabilistic component. After all, a shift in a climate variable is only significant if it is outside the realm of natural variability of the system.

Work over the past two decades has augmented pattern scaling with various representations of uncertainty (Zelazowski et al., 2018; Alexeeff et al., 2018; Goodwin et al., 2020; Gao et al., 2023) and introduced more complex statistical emulators (Castruccio et al., 2014; Beusch et al., 2020). Much recent work has also been dedicated to constructing machine learning-based climate emulators (e.g. Watson-Parris et al. (2022); Yu et al. (2024); Christensen et al. (2024)). While these varied approaches have improved upon the pattern scaling baseline by adding uncertainty quantification and better representation of nonlinear relationships, the need remains for the development of robust emulators addressing multiple variables (individually or jointly), at scales relevant to impacts, and able to represent effectively the internal variability of the model emulated.

In addition, for the case of deep-learning methods, questions remain about their overall skill compared to pattern scaling (Lütjens et al., 2024), the lack of emulator interpretability, and the computational cost of training. Furthermore, while many emulators have been developed to reproduce annual (Goodwin et al., 2020; Beusch et al., 2020) and seasonal (P. Holden et al., 2014; Alexeeff et al., 2018) averages of climate variables, few have looked at reproducing monthly data (Osborn et al., 2016a; Castruccio et al., 2019; Nath et al., 2022). 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; Kemsley et al., 2024; Osborn et al., 2016b). Impact models sometimes require even higher temporal and spatial fidelity, in which case the model presented herein is viewed as a first step towards those cases.

In this work, we develop a data-driven emulation method for spatially resolved monthly temperature and relative humidity. Our method is fast, flexible, interpretable, and probabilistic. In designing this methodology, we sought to represent not just the ensemble mean of the ESM but the entire ensemble distribution. Assessing ensemble spread is among the most reliable ways of quantifying the internal variability of the climate system as represented by ESMs (Collins & Allen, 2002; Tebaldi & Knutti, 2007; McKinnon & Deser, 2018; Tebaldi, Dorheim, et al., 2021; Lehner et al., 2020). It has also been noted that projections accounting for model spread are vital to improving climate adaptation (Hansen et al., 2012; Deser et al., 2012; Woodruff, 2016). A sufficiently large ensemble is necessary to infer distributions of internal variability from a set of individual realizations. For this reason, we choose to emulate the evolution of climate variables generated with a CMIP-class model, specifically MPI-ESM1.2 LR (v1.2.01p7) (Mauritsen et al., 2019), that ran a large ensemble (50 members) of simulations for a number of emissions scenarios (see Section 2 for details).

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 & Livezey, 1987a), and see Hannachi et al. (2007) for a comprehensive review of the technique's history. EOF modes have been shown to effectively capture the patterns of variability of the Earth system (Barnston & 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 and can, with some limitations (Monahan et al., 2009), be interpreted physically. Using a subset of leading EOF 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 coefficients as a multivariate Gaussian distribution, thus also addressing correlations among the spatial modes, and therefore modeling a coherent spatial structure of the variables of interest.

The Gaussian assumption for the EOF amplitudes may seem overly restrictive for many climate variables. However, the leading EOFs represent averages of the original variables over large swaths of the Earth. The monthly and spatial averaging makes the multivariate statistics of the EOF amplitudes more Gaussian than the original variables, but other coarse-graining techniques could be used to improve further the skill of the Gaussian approach described here, see Falasca, Basinski-Ferris, et al. (2024). 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 ESM ensemble, so long as their EOF amplitudes have approximately multivariate Gaussian statistics. 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, similar to Castruccio et al. (2019). Furthermore, the Gaussian assumption allows us to calculate the distributions for observables of interest in closed form.

We condition our emulator on the ensemble mean global mean temperature. Global mean temperature is generally understood to be approximately linear in cumulative emissions (H. D. Matthews et al., 2009; Masson-Delmotte et al., 2021), given a smoothly-changing system and ignoring, e.g., time-lagged response to radiative forcing, or the impact of short-lived aerosols and nonlinear feedbacks like those from melting ice. However, there are also more sophisticated models that can be used. Thus at a later time we can rely on Simple Climate Models (SCMs, e.g. Meinshausen et al. (2011); Lembo et al. (2020); Leach et al. (2021); Bouabid et al. (2024); Dorheim et al. (2024)) to translate arbitrary emission pathways into novel trajectories of global mean temperature (other than the one represented by the ESM runs we used for training) which can drive realizations of spatially resolved monthly temperatures and humidity under new scenarios of future emissions. 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. (2016a); Alexeeff et al. (2018); Goodwin et al. (2020)). For example, pattern scaling conditioned on global mean temperature has been shown to predict regionally resolved ensemble mean temperature (Lütjens et al., 2024). We comment that often the global mean temperature anomaly is used rather than the actual global mean temperature, but here we will use the global mean temperature.

Our paper is organized as follows: In Section 2, we introduce the dataset used in this work. Section 3 discusses the Gaussian assumption and coarse-graining procedure. In Section 4 we discuss the details of the emulator 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.

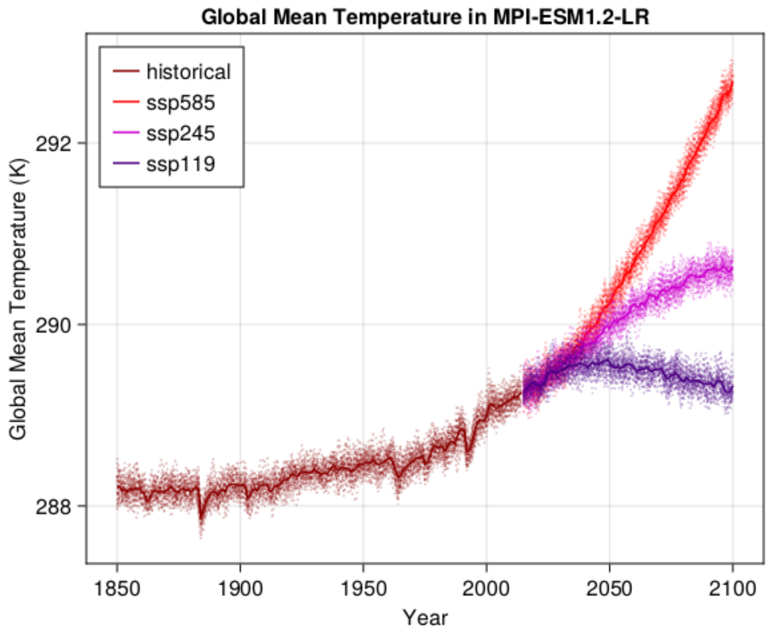

Figure 1. Global mean temperature in the MPI-ESM1.2-LR ensemble. Each dashed line represents one of 50 ensemble members, and the solid line shows the ensemble mean. Different colors correspond to the historical 1850–2014 period (maroon) and the three future scenarios considered in this study: SSP5-8.5 (red), SSP2-4.5 (pink), and SSP1-1.9 (purple). The future period spans 2015–2100. The historical period lasts 165 years, and the future period—86 years.

2 Data

We use the output from the MPI-ESM1.2 LR (v1.2.01p7) ESM model (Mauritsen et al., 2019), which contributed to CMIP6. We chose this model because of its large number of simulations (ensemble members) run for each emission scenario: 50 simulations are run for each scenario, differing only in their initial conditions. A large ensemble is necessary to separate the model's internal variability from the anthropogenic signal (Collins & Allen, 2002; Tebaldi, Dorheim, et al., 2021). 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. This is the same dataset used in Lütjens et al. (2024).

Each MPI-ESM1.2 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 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, Tebaldi, van Vuuren, et al., 2016; Tebaldi, Debeire, et al., 2021). 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. We select the historical experiment and the SSP5-8.5 high-warming scenario for training the emulator because together they span the widest range of global mean tem-

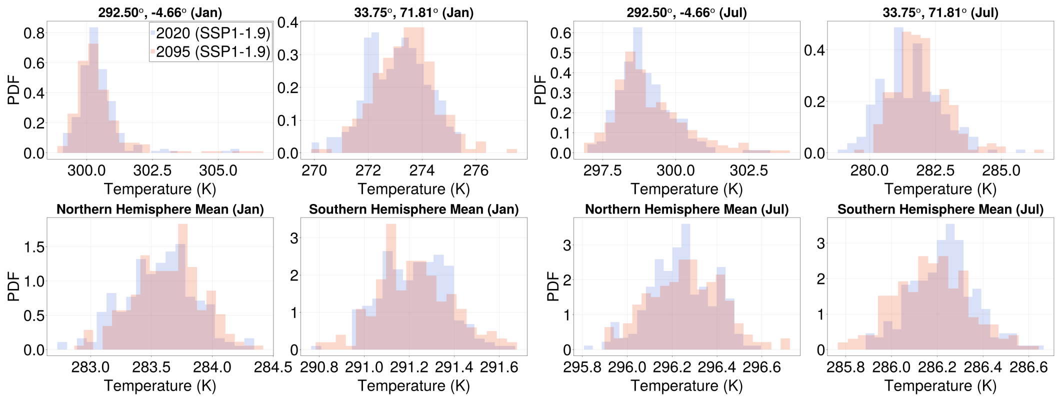

Figure 2. Statistics of Surface Temperature Global Averages and Selected Locations for SSP1-1.9 at years 2020 ± 2 **and** 2095 ± 2. Here, we show the surface temperature histograms of the SSP1-1.9 scenario corresponding to similar global mean temperatures ($\overline{T}_g = 289.3$ K) but different points in time. The histograms overlap, lending credence to parameterizing the distributional change for a fixed month with $\overline{T}_g$.

peratures. This leaves SSP2-4.4 and SSP1-1.9 as the validation sets for regression, see Section 4.2.

Our goal is to develop an emulator that predicts the changes in the multivariate probability density function of climate fields as a function of emissions. The first step in the process is to instead represent the distributions as a function of (or, more appropriately, conditioned on) global, ensemble, and yearly mean temperature and hence cumulative emissions (see Masson-Delmotte et al. (2021). Following standard practice, we will refer to global, ensemble, and yearly mean surface temperature as "global mean temperature" throughout the text.

To presume that time-dependent climate statistics for different emissions scenarios can be parameterized by a state-dependent (time-independent / 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 9 and 10. In symbols, we assume that the statistics of climate system fields, in our case the EOF model amplitudes $\boldsymbol{a} \in \ell_2$ (countably infinite), 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:

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

The hope is that conditioning on global mean temperature serves as an informative parametric form to characterize the changing distribution of the climate relevant quantities. Our formulation is well-posed. If no functional form relates $\overline{T}_g$ to a particular observable of the climate system, the probabilistic description implies that the conditioning information is uninformative. Thus, in the worst-case scenario, the conditional information reduces to the distribution, e.g., $\rho(\boldsymbol{a}|\overline{T}_g, m) \to \rho(\boldsymbol{a}|m)$. In such cases, our task is to find additional quantities that yield informative distributions.

We illustrate our approach for monthly mean surface (2m) temperature and monthly mean surface (2m) relative humidity. These variables are the 'tas' and 'hurs' variables 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 192 × 96 lat-lon grid. The emulator is conditioned on globally averaged ensemble mean surface temperature, which we calculate from the 2m temperature variable.

Our approach is purely data-driven and should not be used to extrapolate statistics outside its global mean temperatures training range. We use the two additional future scenarios to test the emulator performance: SSP2-4.5, which features milder monotonic warming that levels off at the end of the century (elimination of emissions), and SSP1-1.9, which 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 Fig. 1). Because we are developing an emulator conditioned only on global mean temperature from a scenario with exponentially increasing emissions (without accounting for emissions history or other memory effects), it is important to test its performance in scenarios with non-monotonic emissions, which are also of ever-increasing interest to mitigation and adaptation studies and policy.

3 Multivariate Gaussian Assumption and Coarse-Graining

Assuming that our data can be approximated as multivariate Gaussian random variables for every grid point for any given global mean temperature is an unrealistic assumption. Still, it is ameliorated by working with monthly and spatially averaged variables from whence a Gaussian distribution would follow from sufficient averaging and the multivariate central limit theorem (Hasselmann, 1976). 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, Basinski-Ferris, et al. (2024); Falasca, Perezhogin, & Zanna (2024)). We also emulate surface relative humidity (RH) statistics to future climates, because of its relevance for climate adaptation and impact studies (T. Matthews, 2018). Our multivariate Gaussian assumption applies better to smoothly varying variables like temperature and relative humidity but less so for variables like precipitation, which have a much more nonlinear response to temperature fluctuations and non-Gaussian statistics (Legates, 1991).

In this work, we choose to coarse-grain the representation of our fields with Empirical Orthogonal Functions (EOFs). The EOF decomposition has been used in previous emulator work for both dimensionality reduction (P. B. Holden & Edwards, 2010; P. B. 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, Perezhogin, & Zanna (2024) has demonstrated how modal amplitudes of EOFs (under the assumption that they can be approximated as multivariate Gaussian distributions) can be used to interpret patterns of variability and teleconnections recovered by data-driven approaches.

We compute the EOF basis through a singular value decomposition of our data in the historical period of one of the ensemble members. The resulting basis constitutes 165×12 EOFs ordered by the magnitude of the singular values. We discard the latter 980 basis functions, leading to a total of 1000 basis functions. We use the same basis set every month and compute EOFs separately for each variable of interest. We project data from every scenario and every ensemble member onto our original basis.

At this point, we return to our assumptions about the multivariate Gaussian nature of coarse-grained representations of our system. We show in Figure 3 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 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. In addition, we anticipate the result section and show the result of the emulator prediction for the statistics in blue.

We see from Figure 3 that even the most "non-Gaussian" EOF coefficients (top row) display a familiar bell-shaped curve, 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 3 requires non-Gaussian higher-order correlations between the different EOF modes. These non-Gaussian correlations ought to be captured to emulate the tails of the distributions at a location and this could be achieved with other data-driven methods such as "score-matching" or Markov models, see Souza (2023); Giorgini et al. (2024); Bassetti et al. (2023); Christensen et al. (2024). Here we focus on robust spatially coarse-grained statistics. As we will show, this focus allows us to ignore these non-Gaussian correlations. We return to this point later in the manuscript in Section 5, where we show that, despite the existence of non-Gaussian correlations, the bulk of the pointwise statistics are captured by the emulator.

Our thought process is as follows: Coarse-grained features constitute the most predictable aspects of the climate signal. As such, finer-scale details, such as temperature distributions at a point, are better modeled using different approaches, such as downscaling from coarse-grained information. It is therefore useful to express the climate state as a set of model amplitudes $\boldsymbol{a}$ where the vector itself can be decomposed into modes corresponding to large scale coarse structures $\boldsymbol{a}_C$ and "fine scale modes" $\boldsymbol{a}_F$. We then decompose the probability distribution for climate variables (for a fixed global mean temperature $\overline{T}_g$ and month m) as

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

Our work focuses on the approximation $\rho(\boldsymbol{a}_C | \overline{T}_g, m) \approx \mathcal{N}(\boldsymbol{\mu}(\overline{T}_g, m), \boldsymbol{\mathcal{C}}(\overline{T}_g, m))$, where the coarse statistical variables $\boldsymbol{a}_C$ are approximated as a Gaussian distribution, given by the symbol $\mathcal{N}$, with means $\boldsymbol{\mu}$ and covariances $\boldsymbol{\mathcal{C}}$ conditioned on the global mean temperature $\overline{T}_g$ and month m. Approximating fine-scale structures conditioned on larger coarse-grained variables, i.e., approximating $\rho(\boldsymbol{a}_F | \boldsymbol{a}_C, \overline{T}_g, m)$, is left to future work. In particular, we surmise that

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

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. Our approach is similar to that of P. B. Holden & Edwards (2010), which builds upon Bruckner et al. (2003). In P. B. Holden & Edwards (2010), the authors fit polynomial functions to EOF coefficients to emulate the annual temperature response to radiative forcing. They also assume a prior form for the shape of the ensemble distribution of yearly temperatures and use Bayesian estimation to emulate the ensemble variability. Instead, we model the EOF coefficients as multivariate Gaussians, which allows us to emulate both the mean and variability of the model directly. In other words, we model the system's statistics as a Gaussian process. We also model the EOF coefficients for each month separately, allowing for the emulation of monthly-resolution data.

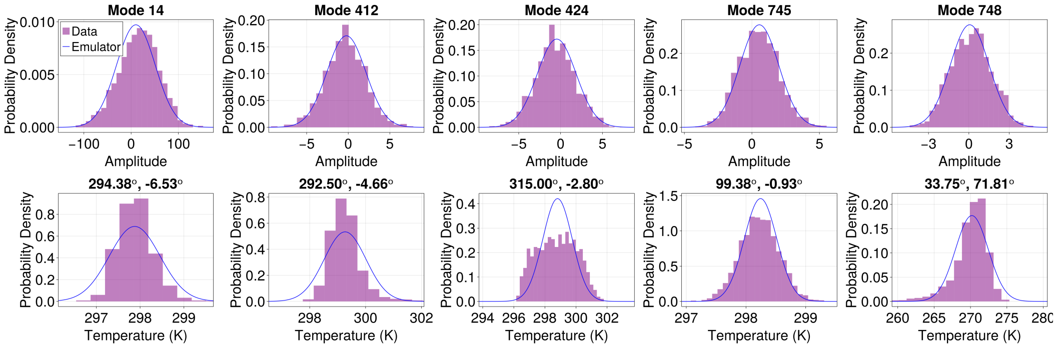

Figure 3. Statistics of Surface Temperature EOF Modes and Selected Locations. In purple, we show the histograms of data collected over the historical period of the MPI ensemble with similar global mean temperatures, and in blue, we show the fit as given by the emulator described in this work. The EOF amplitudes populate the top row, and the bottom row constitutes point locations. The locations and modes were selected according to their Gaussian/non-Gaussian behaviors. The most Gaussian cases are Mode 412 and location 99.39°, −0.94°.

Our work also shares some similarities with Nath et al. (2022), in which the authors augment an existing annual-average emulator with Gaussian processes to model monthly variability. In contrast, we make use of a Karhunen-Loève expansion to model the EOF coefficients as described in Fontanella & Ippoliti (2012) and assume Gaussianity in consideration of temporal (monthly) and spatial (EOFs) averaging. This approximate Gaussianity is motivated by the multivariate version of the central limit theorem, per Hasselmann (1976). Explicitly, we *are not doing Gaussian Process Regression* (Williams & Rasmussen, 1995), which requires making assumptions on the covariance structure of a kernel. Our assumption is that a finite rank approximation suffices to describe the covariance kernel and that the EOF basis functions serve as the eigenvectors of the covariance kernel. The method described herein can also be applied to variables that do not satisfy the Gaussian assumption if one is instead concerned with data over a larger period of time (M. Wang and T. Sapsis, personal communication, September 19, 2024). Finally, we emphasize that our method is data-driven and applies to any variables that meet the above mentioned criteria in Section 2.

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

4.1 Gaussian process emulator

Following the training data's EOF-based dimensionality reduction, we develop and train a Gaussian process-based stochastic emulator of regional monthly temperature and relative humidity fields. As stated previously, the set of EOF coefficients $\hat{\boldsymbol{a}}$ is modeled as multivariate Gaussian

$$\hat{\boldsymbol{a}} \sim \mathcal{N}(\hat{\boldsymbol{\mu}}(\bar{T}_g, m), \hat{\boldsymbol{\mathcal{C}}}(\bar{T}_g, m)). \tag{4}$$

as a function of global mean temperature $\overline{T}_g$ and the month index m. Since each month is modeled separately, we will drop the subscript m with the implicit understanding that any regression is for a fixed month. The large ensemble MPI model offers a robust way to estimate the means and covariances since we view, for a fixed month and year, an in-

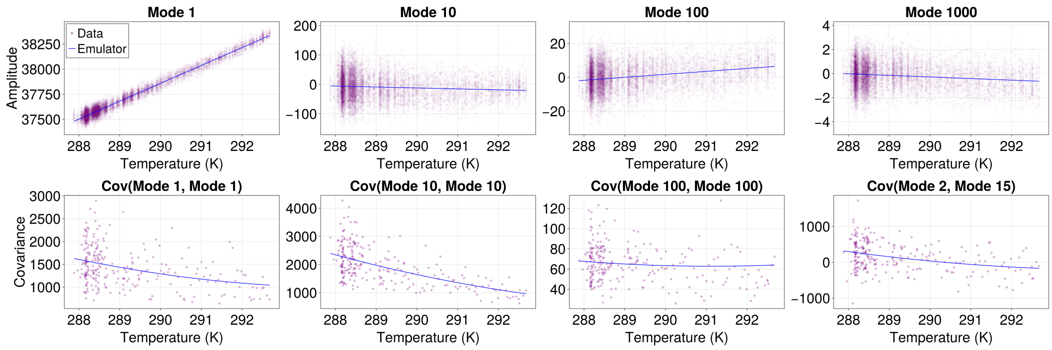

Figure 4. Surface Temperature EOF Amplitude Mean and Covariance Regression as a Function of Global, Ensemble, Yearly Mean Temperature in January. We show the projected and computed data of the MPI ensemble (purple) and the emulator fit (blue). We see that the fit to data captures overall trends.

dividual ensemble member of the MPI model as a realization of a multi-variate Gaussian distribution parameterized solely by the month and global mean temperature.

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

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

Higher-order polynomial fits 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 could 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. Our first attempt at solving this problem failed: fitting a linear function for each entry of the covariance matrix as a function of global mean temperature does not produce a positive definite matrix.

Our second attempt was to represent the dependence of the covariance matrices $\hat{\boldsymbol{\mathcal{C}}}$ on $\bar{T}_g$ as

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

This functional form guarantees that $\hat{\boldsymbol{\mathcal{C}}}$ is symmetric positive definite because it is represented indirectly via $\hat{\boldsymbol{L}}$, the product of a matrix and its transpose. In one dimension, this functional form represents the standard deviation as a linear function in global mean temperature $\overline{T}_g$. In Equation 6 each entry of $\hat{\boldsymbol{L}}$ is modeled as a linear function of $\bar{T}_g$, leading to a quadratic model for the covariance:

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

As with the means, it is possible to go beyond a linear representation for $\hat{\boldsymbol{L}}$.

However, it proved challenging to properly represent $\hat{\boldsymbol{L}}_0$ and $\hat{\boldsymbol{L}}_1$. We first performed linear regression on each entry of $\hat{\boldsymbol{L}}$ by computing a Cholesky factorization of $\hat{\boldsymbol{\mathcal{C}}}$, e.g. representing $\hat{\boldsymbol{\mathcal{C}}} = \hat{\boldsymbol{L}}\hat{\boldsymbol{L}}^T$, see Trefethen & Bau III (1997). This procedure led to inaccurate estimates of $\mathcal{C}$; in particular, the method underestimated the variance of the higher EOF modes.

The methodology that gave the highest fidelity results was formulating (and solving) an optimization problem. Thus, to find $\hat{\boldsymbol{L}}_0$ and $\hat{\boldsymbol{L}}_1$ we minimized 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 Equation 7 and $\|\cdot\|$ is an appropriately chosen norm. In our case, we used a Frobenius norm (minimizing 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 & Ba (2014)'s "ADAM" for optimization. The initial guess for iteration was 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 each year, and each year has an associated global mean temperature.

We illustrate the result of the regression procedure for surface temperature in Figure 4. 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 at each year for all ensemble members of the MPI data for the historical and SSP5-8.5 scenario. The bottom row is obtained by calculating the covariance between sample modal amplitudes each year separately using all ensemble members. These data are then regressed against each year's global, ensemble, and temporal mean of surface temperature. From Figure 4, we see that the trends are well captured by performing the regression (blue). As mentioned before, we use a linear model for the mean of the EOF coefficients and consistently 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 process.

4.2 Using the Emulator and Relation to Pattern Scaling

Upon performing the dimensionality reduction and the regression problem, we can reconstruct spatial fields for a fixed month m and global mean temperature $\overline{T}_g$ by making use of the basis functions and representing a field such at surface temperature T as

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

where $\hat{\boldsymbol{a}} \sim \mathcal{N}(\hat{\boldsymbol{\mu}}(\overline{T}_g, m), \hat{\boldsymbol{\mathcal{C}}}(\overline{T}_g, m))$ are the EOF coefficients sampled from a multivariate Gaussian distribution and $\phi_i(\boldsymbol{x})$ are our EOF basis functions. The ensemble average of $\hat{T}$ for a fixed location $\boldsymbol{x}$ is given by

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

and the variance at a point $\boldsymbol{x}$ 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 spatial average / zonal average for a fixed latitude / fixed location / average of a patch of land such as North America or Africa, we have that the ensemble average 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 function of temperature can be computed from the mean and covariance of all the EOF coefficients and the action of the linear functional on the basis functions. Similarly, any higher-order statistics can be computed by using the Gaussian assumption for the EOF amplitudes. Equation 12 illustrates that the entire covariance structure of the EOF amplitudes is key to compute temperature statistics beyond the mean.

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-averaged temperature Santer & Wigley (1990). If we sum over all EOF modes our emulator for temperature at a location is given by,

$$\langle \hat{T}(\boldsymbol{x}) \rangle = \sum_{i=1}^{N} \hat{\mu}_i \phi_i(\boldsymbol{x}) = \sum_{i=1}^{N} (\mu_{0,i} + \overline{T}_g \mu_{1,i}) \phi_i(\boldsymbol{x}) \tag{13}$$

$$= \sum_{i=1}^{N} \mu_{0,i} \phi_i(\boldsymbol{x}) + \overline{T}_g \sum_{i=1}^{N} \mu_{0,i} \phi_i(\boldsymbol{x}) \equiv T_0(\boldsymbol{x}) + \overline{T}_g T_1(\boldsymbol{x}) \tag{14}$$

where

$$T_0(\boldsymbol{x}) = \sum_{i=1}^{N} \mu_{0,i} \phi_i(\boldsymbol{x}) \text{ and } T_1(\boldsymbol{x}) = \sum_{i=1}^{N} \mu_{1,i} \phi_i(\boldsymbol{x}). \tag{15}$$

This confirms that our emulator does indeed reduce to linear pattern scaling for the surface temperature at a location.

It may be argued that the linear pattern scaling approach can also be used to predict any functional of temperature at each location as a linear function of the global, yearly, and ensemble-averaged temperature. However a new linear fit must be computed for any statistics of interest. The advantage of our emulator is that we can reconstruct any statistic of the field in question from the mean and covariance estimates in so far as the Gaussian assumption is satisfied. Pattern scaling would fail to do so since calculating the variance of a spatial average (for example) requires knowing correlations between different points.

To demonstrate that our Gaussian process emulator reduces to a form of a linear pattern scaling emulator for temperature at a location, we plot the yearly and ensemble-averaged global temperature emulation error as a function of time for an increasing number of modes (10, 100, 1000). At each time along the horizontal axis the errors are computed with respect to $\overline{T}_g$ computed from the MPI ensemble for the year in question. We compare the error of performing linear regression pointwise on the ensemble mean in the historical periodic and SSP5-8.5 to the ensemble mean of our Gaussian process emulator in Figure 5. 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. We do not use a "validation" dataset in the present case since our regression does not have any hyperparameters to tune. In formulas, we are comparing (for each year)

$$\text{temporal error}(\langle T \rangle_{t,\omega}, \langle \hat{T} \rangle_{t,\omega}) = \sqrt{\frac{1}{4\pi} \int_{\theta=0}^{2\pi} \int_{\phi=0}^{\pi} |\langle T \rangle_{t,\omega} - \langle \hat{T} \rangle_{t,\omega}|^2 \sin(\phi) d\theta d\phi}, \tag{16}$$

where $\langle \cdot \rangle_{t,\omega}$ to denotes an ensemble and yearly average and the spatial average is taken over the Earth's sphere. As we increase the number of modes, the error in the approximation becomes similar to the pointwise error when utilizing pattern scaling. A few modes

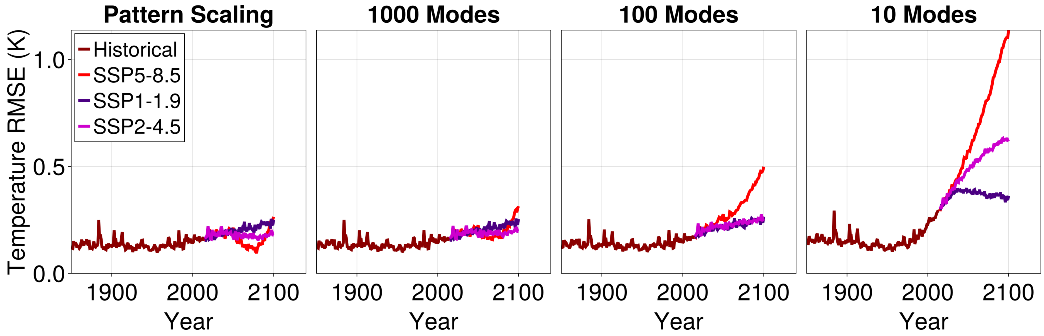

Figure 5. Regression Error for Ensemble and Yearly Averaged Surface Temperature as a Function of Time for Different Scenarios. Different colors correspond to the historical 1850–2014 period (maroon) and the three future scenarios considered in this study: SSP5-8.5 (red), SSP2-4.5 (pink), and SSP1-1.9 (purple). The future period spans 2015–2100. The historical period lasts 165 years, and the projected period—86 years. We show the RMS error of pattern-scaling on the left and increasing the number of modes used in the emulator in the subsequent rightward panels. As we increase the number of modes, the error in capturing pointwise statistics decreases.

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 SSP5-8.5 data for the EOF amplitudes, there is less data corresponding to warmer temperatures. Thus, the emulator underperforms where it has seen fewer data. With more modes (and hence a more complete basis for representing functions), we see that the generalization error of going to different SSP scenarios matches the error of the historical period.

To understand the spatial distribution of error, we average the absolute difference between our emulator predicted mean and the ensemble average of the data over the historical period, SSP5-8.5, and SSP1-1.9 in Figure 6. In formulas, this is

$$\text{spatial error}(\langle T\rangle_{t,\omega}, \langle \hat{T}\rangle_{t,\omega}) = \frac{1}{\text{scenario duration}} \int_{\text{scenario start}}^{\text{scenario end}} |\langle T\rangle_{t,\omega} - \langle \hat{T}\rangle_{t,\omega}| dt. \tag{17}$$

In all cases, most of the average error comes from the high latitudes. There are also additional significant errors over Africa, India, and the southeast tip of Australia. Overall the spatial errors look similar in the future scenario cases. We expect the errors in spatial patterns to change upon using nonlinear regression for the mean or a different set of basis functions; however, the error can perhaps be traced to a physical origin as the disappearance of sea ice in the northern hemisphere and desertification.

While these error estimates are commonly used in the assessment of emulators, they are quite limited. In the next section we illustrate that a major advantage of our emulator is its ability to quantify the significance in shifts in the distributions of climate variables as a function of global mean temperature.

5 Results

As stated in the previous section, it is possible to reconstruct spatial statistics of any observable of our system with simplified formulas for linear functionals of our state.

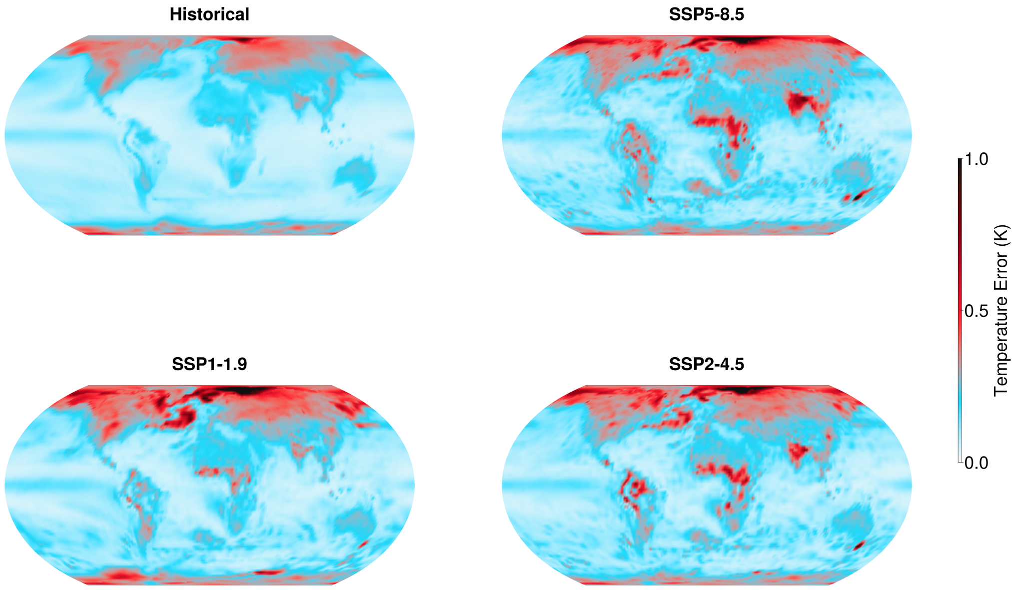

Figure 6. Average Regression Error for Ensemble and Yearly Averaged Surface Temperature as a Function of Space for Different Scenarios. We show the temporal average error for each point in four cases: the historical period (top left), SSP5-8.5 (top right), SSP1-1.9 (bottom left), and SSP2-4.5 (bottom right). The maximum temperature difference in the time period 2015 to 2100 in SSP5-8.5 and SSP1-1.9 is 3.4 K and 0.3 K, respectively. For SSP2-4.5, the maximum temperature difference in the time period 2015 to 2100 is 1.4 K

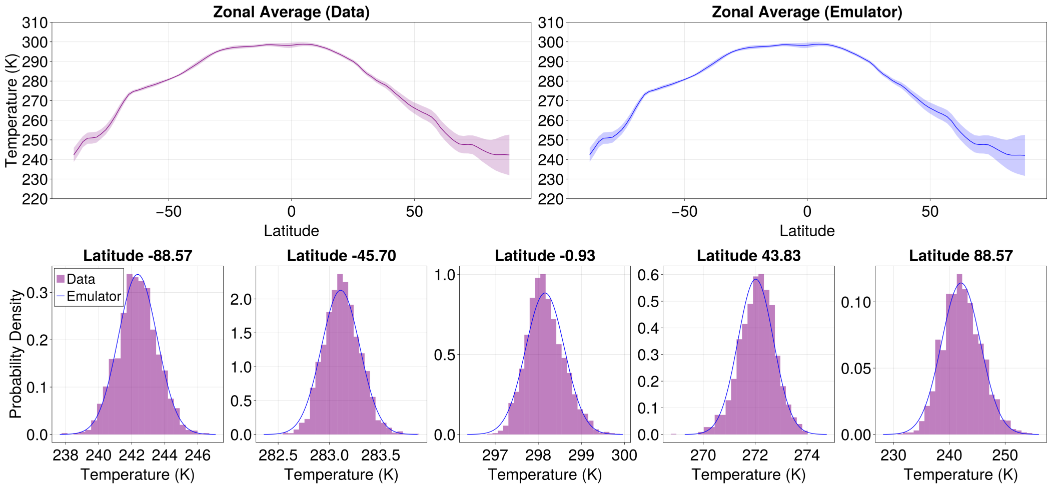

Figure 7. Spatially Coarse Emulator Statistics. The purple color indicates data coming from the MPI model over the historical period with similar global mean temperatures, and in blue, the emulator prediction. The top panel shows the ensemble mean and variance of the zonally averaged surface temperature 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.

In particular, the statistics of the zonal average at a fixed latitude for temperature in January are reconstructed in Figure 7 for a range of similar global mean temperatures taken over the historical period (to have higher fidelity statistics). The purple colors indicate data from the MPI model, and the blue represents the Gaussian emulator. In the top row, the data's zonal average mean and variance (left) are reconstructed well using the model (right). The two distributions look nearly identical in mean and variance. This similarity should not be taken for granted, since regression is performed on the mean and covariance of the EOF amplitudes rather than the averages directly. Furthermore, when we check the histograms for the zonal average at different latitudes (bottom row), we see that the distributions are well-represented by Gaussian distributions. The emulator's ability to capture the zonally averaged statistics surface temperature at each latitude comes directly from the representation of the covariance structure between EOF amplitudes, as necessitated by Equation 15. This test serves as an indirect validation of using multivariate Gaussian statistics for the EOF coefficients.

Since our emulator captures each month separately, we can investigate a-posteriori shifts in the ensemble average seasonal temperature cycle. In Figure 8, we show the emulator prediction for the the seasons for the upper and southern hemisphere averages separately, where the blue corresponds 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 hemisphere (land warms more than the ocean because it is drier and less efficient at cooling through latent heat release.)

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 in the top row of Figure 9 for two of the twelve months. In the top row, we show spatial averages

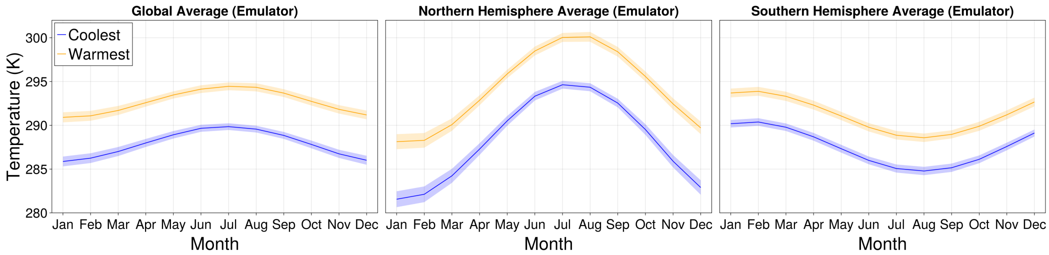

Figure 8. Monthly Emulator Output for Global Quantities. We show the emulator prediction for the global average (left), upper-hemisphere average (middle), and lower-hemisphere average (right), as well as two global mean temperatures, $\overline{T}_g = 288(K)$ (blue) and $\overline{T}_g = 293(K)$ (orange). 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.

of surface temperature and, in the bottom row, spatial averages of relative humidity. Accounting for the internal variability of the system helps us distinguish whether or not there are significant shifts due to climate change. For temperature, we see that the distribution shifts are outside the climate system's natural variability. In contrast, despite minor changes in the mean value, relative humidity is relatively unchanged when accounting for internal variability during January, while in July there is a more significant shift. The shifts are in accordance with the expectation that relative humidity will decrease over land in a warmer climate and increase over the ocean (Byrne & O'Gorman, 2016).

In addition, we can reconsider the assumptions of pointwise Gaussian statistics and see if global warming trends are captured for the pointwise statistics in Figure 10. For both temperature (top row) and relative humidity (bottom row), we see that, although the distribution shape is not well-approximated as Gaussian for some of the selected points, the trends in shifts of means and variances are well captured. Furthermore, we see an apparent change in the shifts in pointwise temperature distributions, but less so for relative humidity, where in all cases, the shifts in mean are well within the variance of internal variability. The relative heights of the Gaussian distributions within a given 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.

6 Conclusion

We have demonstrated a novel probabilistic emulator and applied it to spatially resolved monthly averaged temperature and relative humidity. This emulator provides a computationally efficient method for extending the MPI-ESM-1.2-LR global climate model to arbitrary warming scenarios while retaining the ability to separate trends from internal variability.

The Gaussian approximation serves as a foundational step, enabling us to represent changes in distributions by describing changes in means and covariances. While this simplified parametric family is effective for coarse-grained variables, it can be extended to a more expressive form, such as through diffusion models, to capture more complex distributions, (Song et al., 2020). Indeed, as we consider higher-order correlations, the appeal of neural networks becomes evident. Estimating even the three-point correlation of a high-dimensional distribution becomes cumbersome, requiring the computation and storage of a tensor with 1000^3 points if one uses a basis of 1000 EOF amplitudes. Gen-

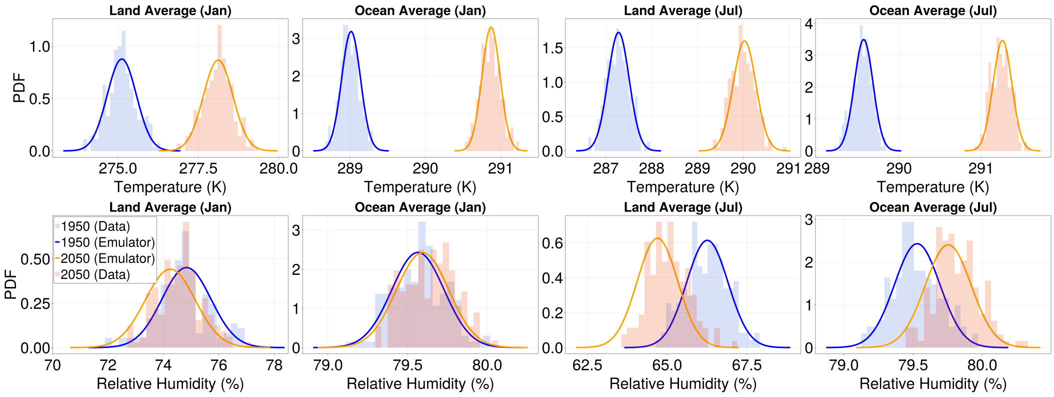

Figure 9. Distributional Shifts Under Climate Change for Land and Ocean Spatial Averages. Here, we show the shift in distribution for the temperature field (top row) and relative humidity (bottom row) for the months of January and July, and a land and ocean spatial average. We see that the emulator (solid line) captures the shift in mean and variance of the data distributions (histograms).

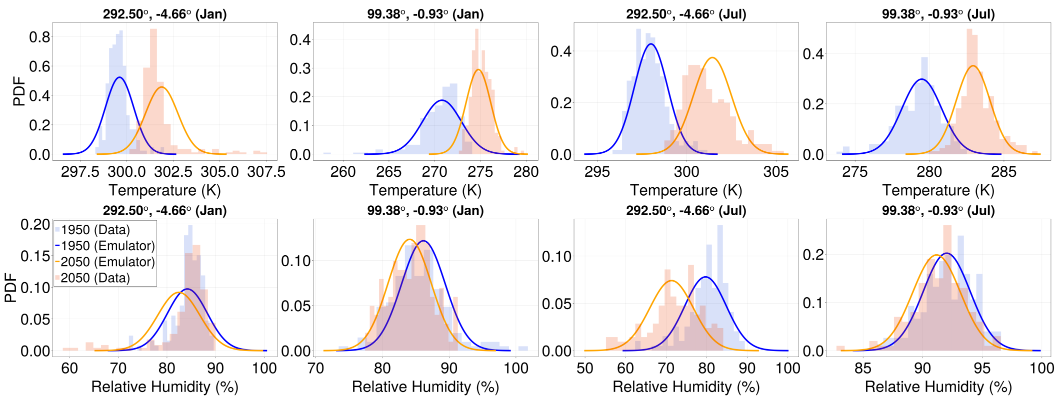

Figure 10. Distributional Shifts Under Climate Change for Different Locations. Similar to Figure 9, but for pointwise statistics at different locations on Earth. Even when the distributions are non-Gaussian, the model represents the overall trend in mean and variances.

erally, a multivariate distribution of size N_d necessitates the storage of $(N_d)^n$ points for an n-point correlation, which rapidly becomes intractable for large N_d or n; however, the added flexibility of using neural networks comes with a steep cost, necessitating larger training datasets, time, expertise, and computational resources. Furthermore, even a trained neural network can be slow for inference and our goal here was to create a computationally expedient emulator that works on today's hardware.

We found that coarse-grained statistics are more amenable to Gaussian representation than point-wise statistics, making them a useful starting point for conditional information. Earth System Models are expected to have significantly higher skill in representing coarse features than in capturing fine-scale details, reinforcing the utility of our approach. The emulator also benefits from a smaller memory footprint, whose dominant cost is storing EOF basis functions. In our work, the data reduction over the training dataset was over a factor of 100.

Simple extensions of the emulator include representing different fields, using different basis functions, using higher-order regression for EOF statistics, using more regression variables other than global mean temperature, capturing correlations between different fields, or capturing temporal correlations. The correlations between different fields can be represented by computing a joint EOF amplitude for quantities such as temperature and relative humidity or computing correlations between EOF coefficients after the fact. Lastly, temporal correlations between months can be calculated to emulate potential trajectories under a Gaussian assumption. This latter avenue allows one to have a predictive model for monthly temperature transition probabilities using a conditional Gaussian distribution.

While our emulator captures distributions only up to the second moment and thus is not suited for extreme events, it lays the groundwork for more specialized emulators. For example, one could condition a separate emulator on our monthly temperature outputs to study extremes or non-Gaussian variables like precipitation. This approach would couple well with existing methods, such as Generalized Extreme Value distribution modeling or generative AI, allowing for rapid emulation of climate extremes in future scenarios. Similar work has been done in Bassetti et al. (2024). A potential ecosystem of emulators is illustrated in Fig. 11. The hierarchy is to first develop a model for a predictive variable for characterizing climate change, such as global mean temperature, using cumulative emissions. The second step is to use an emulator for coarse-grained variables such as the work described here. The last step would be using a downscaling approach for finer-grained statistics. However some limitations to this pipeline should be acknowledged. First, it makes it difficult to capture the impact on global mean temperature of emissions with local rather than global impacts, like aerosols. Second, it assumes that all regional variables can be inferred from global mean temperature which is clearly an oversimplification.

The emulator described in the manuscript aims to learn the trends and internal variability of the climate system as represented by a particular ensemble of global climate model simulations. We do not delve into the accuracy of this ensemble compared to observations. However, note that our model-trained emulators can be used as priors to be further trained with available observations to remove model bias. While we chose the MPI model due to its large ensemble size, the methodology applies to any model with a sufficiently large ensemble.

References

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

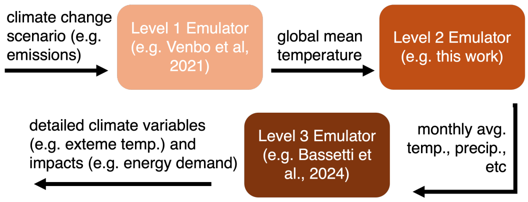

Figure 11. Potential Ecosystem for Coupled Emulator Models.

Barnston, A. G., & Livezey, R. E. (1987a). Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Monthly Weather Review*, *115*(6), 1083 - 1126. Retrieved from `https://journals.ametsoc.org/view/journals/mwre/115/6/1520-0493_1987_115_1083_csapol_2_0_co_2.xml` doi: 10.1175/1520-0493(1987)115⟨1083:CSAPOL⟩2.0.CO;2

Barnston, A. G., & Livezey, R. E. (1987b). Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Monthly weather review*, *115*(6), 1083–1126.

Bassetti, S., Hutchinson, B., Tebaldi, C., & Kravitz, B. (2023). Diffesm: Conditional emulation of earth system models with diffusion models. In *Iclr 2023 workshop on tackling climate change with machine learning.* Retrieved from `https://www.climatechange.ai/papers/iclr2023/42`

Bassetti, S., Hutchinson, B., Tebaldi, C., & Kravitz, B. (2024). Diffesm: Conditional emulation of temperature and precipitation in earth system models with 3d diffusion models. *Authorea Preprints*.

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

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

Bradbury, J., Frostig, R., Hawkins, P., Johnson, M. J., Leary, C., Maclaurin, D., & Wanderman-Milne, S. (2018). *Jax: Composable transformations of python+numpy programs.* Retrieved from `http://github.com/google/jax`

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

Byrne, M. P., & O'Gorman, P. A. (2016). Understanding decreases in land relative humidity with global warming: Conceptual model and gcm simulations. *Journal of Climate*, *29*(24), 9045 - 9061. Retrieved from `https://journals.ametsoc.org/view/journals/clim/29/24/jcli-d-16-0351.1.xml` doi: 10.1175/JCLI-D-16-0351.1

Castruccio, S., Hu, Z., Sanderson, B., Karspeck, A., & Hammerling, D. (2019). Reproducing internal variability with few ensemble runs. *Journal of Climate*, *32*(24), 8511–8522.

Castruccio, S., McInerney, D. J., Stein, M. L., Crouch, F. L., Jacob, R. L., & Moyer, E. J. (2014). Statistical emulation of climate model projections based on precomputed gcm runs. *Journal of Climate*, *27*(5), 1829–1844.

Christensen, K., Otto, L., Bassetti, S., Tebaldi, C., & Hutchinson, B. (2024). Diffusion-based joint temperature and precipitation emulation of earth system models. *arXiv preprint arXiv:2404.08797*.

Collins, M., & Allen, M. R. (2002). Assessing the relative roles of initial and boundary conditions in interannual to decadal climate predictability. *Journal of Climate*,

15(21), 3104–3109.
Deser, C., Phillips, A., Bourdette, V., & Teng, H. (2012). Uncertainty in climate change projections: the role of internal variability. *Climate dynamics*, *38*, 527–546.
Dorheim, K., Gering, S., Gieseke, R., Hartin, C., Pressburger, L., Shiklomanov, A. N., ... Bond-Lamberty, B. (2024). Hector V3.2.0: functionality and performance of a reduced-complexity climate model. *Geoscientific Model Development*, *17*(12), 4855–4869.
Doury, A., Somot, S., Gadat, S., Ribes, A., & Corre, L. (2023). Regional climate model emulator based on deep learning: Concept and first evaluation of a novel hybrid downscaling approach. *Climate Dynamics*, *60*(5), 1751–1779.
Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958.
Falasca, F., Basinski-Ferris, A., Zanna, L., & Zhao, M. (2024). Diagnosing the pattern effect in the atmosphere-ocean coupled system through linear response theory. *arXiv preprint arXiv:2408.12585*. Retrieved from `https://doi.org/10.48550/arXiv.2408.12585`
Falasca, F., Perezhogin, P., & Zanna, L. (2024). Data-driven dimensionality reduction and causal inference for spatiotemporal climate fields. *Physical Review E*, *109*(4), 044202.
Fontanella, L., & Ippoliti, L. (2012). Karhunen–loève expansion of temporal and spatio-temporal processes. In *Time series analysis: Methods and applications* (Vol. 30, pp. 379–398). Amsterdam, The Netherlands: Elsevier B.V. doi: 10.1016/B978-0-444-53858-1.00017-X
Gao, X., Sokolov, A., & Schlosser, C. A. (2023). A large ensemble global dataset for climate impact assessments. *Scientific Data*, *10*(1), 801. Retrieved from `https://doi.org/10.1038/s41597-023-02708-9` doi: 10.1038/s41597-023-02708-9
Giorgini, L. T., Deck, K., Bischoff, T., & Souza, A. N. (2024). Response theory via generative score modeling. *arXiv preprint arXiv:2402.01029*. Retrieved from `https://doi.org/10.48550/arXiv.2402.01029`
Goodwin, P., Leduc, M., Partanen, A.-I., Matthews, H. D., & Rogers, A. (2020). 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*(11), 5389–5399.
Griffies, S. M., Adcroft, A., & Hallberg, R. W. (2020). A primer on the vertical lagrangian-remap method in ocean models based on finite volume generalized vertical coordinates. *Journal of Advances in Modeling Earth Systems*, *12*(10), e2019MS001954. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001954` (e2019MS001954 10.1029/2019MS001954) doi: https://doi.org/10.1029/2019MS001954
Guo, S., Wang, J., Xiong, L., Ying, A., & Li, D. (2002). A macro-scale and semi-distributed monthly water balance model to predict climate change impacts in china. *Journal of hydrology*, *268*(1-4), 1–15.
Hannachi, A., Jolliffe, I. T., Stephenson, D. B., et al. (2007). Empirical orthogonal functions and related techniques in atmospheric science: A review. *International journal of climatology*, *27*(9), 1119–1152.
Hansen, J., Ruedy, R., Sato, M., & Lo, K. (2010). Global surface temperature change. *Reviews of geophysics*, *48*(4).
Hansen, J., Sato, M., & Ruedy, R. (2012). Perception of climate change. *Proceedings of the National Academy of Sciences*, *109*(37), E2415–E2423.
Hasselmann, K. (1976). Stochastic climate models part i. theory. *tellus*, *28*(6), 473–485.

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

Holden, P., Edwards, N., Garthwaite, P., Fraedrich, K., Lunkeit, F., Kirk, E., ... Babonneau, F. (2014). Plasim-entsem v1. 0: a spatio-temporal emulator of future climate change for impacts assessment. *Geoscientific Model Development*, *7*(1), 433–451.

Holden, P. B., & Edwards, N. R. (2010). Dimensionally reduced emulation of an aogcm for application to integrated assessment modelling. *Geophysical Research Letters*, *37*(21).

Holden, P. B., Edwards, N. R., Garthwaite, P. H., & Wilkinson, R. D. (2015). Emulation and interpretation of high-dimensional climate model outputs. *Journal of Applied Statistics*, *42*(9), 2038–2055.

Huntingford, C., & Cox, P. (2000). An analogue model to derive additional climate change scenarios from existing gcm simulations. *Climate dynamics*, *16*, 575–586.

Kemsley, S. W., Osborn, T. J., Dorling, S. R., & Wallace, C. (2024). Pattern scaling the parameters of a markov-chain gamma-distribution daily precipitation generator. *International Journal of Climatology*, *44*(1), 144-159. Retrieved from `https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.8320` doi: https://doi.org/10.1002/joc.8320

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. Retrieved from `https://arxiv.org/abs/1412.6980`

Kutzbach, J. E. (1967). Empirical eigenvectors of sea-level pressure, surface temperature and precipitation complexes over north america. *Journal of Applied Meteorology and Climatology*, *6*(5), 791–802.

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

Legates, D. R. (1991). An evaluation of procedures to estimate monthly precipitation probabilities. *Journal of Hydrology*, *122*(1-4), 129–140.

Lehner, F., Deser, C., Maher, N., Marotzke, J., Fischer, E. M., Brunner, L., ... Hawkins, E. (2020). Partitioning climate projection uncertainty with multiple large ensembles and cmip5/6. *Earth System Dynamics*, *11*(2), 491–508. Retrieved from `https://esd.copernicus.org/articles/11/491/2020/` doi: 10.5194/esd-11-491-2020

Lembo, V., Lucarini, V., & Ragone, F. (2020). Beyond forcing scenarios: Predicting climate change through response operators in a coupled general circulation model. *Scientific Reports*, *10*(1), 8668. Retrieved from `https://doi.org/10.1038/s41598-020-65297-2` doi: 10.1038/s41598-020-65297-2

Li, S., Rupp, D. E., Hawkins, L., Mote, P. W., McNeall, D., Sparrow, S. N., ... Wettstein, J. J. (2019). Reducing climate model biases by exploring parameter space with large ensembles of climate model simulations and statistical emulation. *Geoscientific Model Development*, *12*(7), 3017–3043.

Link, R., Snyder, A., Lynch, C., Hartin, C., Kravitz, B., & Bond-Lamberty, B. (2019). Fldgen v1. 0: an emulator with internal variability and space–time correlation for earth system models. *Geoscientific Model Development*, *12*(4), 1477–1489.

Lorenz, E. N. (1956). *Empirical orthogonal functions and statistical weather prediction* (Vol. 1). Massachusetts Institute of Technology, Department of Meteorology Cambridge.

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

Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., … Zhou, B. (2021). *Summary for policymakers* (V. Masson-Delmotte et al., Eds.). Cambridge, United Kingdom: Cambridge University Press. Retrieved from `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)

Matthews, H. D., Gillett, N. P., Stott, P. A., & Zickfeld, K. (2009). The proportionality of global warming to cumulative carbon emissions. *Nature*, *459*(7248), 829–832.

Matthews, T. (2018). Humid heat and climate change. *Progress in Physical Geography: Earth and Environment*, *42*(3), 391–405.

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., … others (2019). 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*(4), 998–1038.

McKinnon, K. A., & Deser, C. (2018). Internal variability and regional climate trends in an observational large ensemble. *Journal of Climate*, *31*(17), 6783–6802.

Meinshausen, M., Raper, S. C. B., & Wigley, T. M. L. (2011). Emulating coupled atmosphere-ocean and carbon cycle models with a simpler model, MAGICC6 – Part 1: Model description and calibration. *Atmospheric Chemistry and Physics*, *11*(4), 1417–1456.

Mitchell, T. D. (2003). Pattern scaling: an examination of the accuracy of the technique for describing future climates. *Climatic change*, *60*(3), 217–242.

Monahan, A. H., Fyfe, J. C., Ambaum, M. H., Stephenson, D. B., & North, G. R. (2009). Empirical orthogonal functions: The medium is the message. *Journal of Climate*, *22*(24), 6501–6514.

Nair, R., & Toy, M. (2016, 06). A high-order multiscale global atmospheric model (invited).. doi: 10.2514/6.2016-3888

Nath, S., Lejeune, Q., Beusch, L., Seneviratne, S. I., & Schleussner, C.-F. (2022). Mesmer-m: an earth system model emulator for spatially resolved monthly temperature. *Earth System Dynamics*, *13*(2), 851–877. Retrieved from `https://esd.copernicus.org/articles/13/851/2022/` doi: 10.5194/esd-13-851-2022

Odjugo, P. A. (2010). General overview of climate change impacts in nigeria. *Journal of human ecology*, *29*(1), 47–55.

O'Neill, B. C., Tebaldi, C., Van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., … others (2016). The scenario model intercomparison project (scenariomip) for cmip6. *Geoscientific Model Development*, *9*(9), 3461–3482.

O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., … Sanderson, B. M. (2016). The scenario model intercomparison project (scenariomip) for cmip6. *Geoscientific Model Development*, *9*(9), 3461–3482. Retrieved from `https://gmd.copernicus.org/articles/9/3461/2016/` doi: 10.5194/gmd-9-3461-2016

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

Osborn, T. J., Wallace, C. J., Harris, I. C., & Melvin, T. M. (2016b). Pattern scaling using climgen: monthly-resolution future climate scenarios including changes in the variability of precipitation. *Climatic Change*, *134*(3), 353–369. Retrieved from `https://doi.org/10.1007/s10584-015-1509-9` doi: 10.1007/s10584-015-1509-9

Peatier, S., Sanderson, B. M., Terray, L., & Roehrig, R. (2022). Investigating parametric dependence of climate feedbacks in the atmospheric component of cnrm-cm6-1. *Geophysical Research Letters*, *49*(9), e2021GL095084.

Santer, B. D., & Wigley, T. (1990). Regional validation of means, variances, and spatial patterns in general circulation model control runs. *Journal of Geophysical Research: Atmospheres*, *95*(D1), 829–850.

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

Schneider, T., Behera, S., Boccaletti, G., Deser, C., Emanuel, K., Ferrari, R., … Yamagata, T. (2023). Harnessing ai and computing to advance climate modelling and prediction. *Nature Climate Change*, *13*(9), 887–889. Retrieved from `https://doi.org/10.1038/s41558-023-01769-3` doi: 10.1038/s41558-023-01769-3

Schneider, T., Bischoff, T., & Płotka, H. (2015). Physics of changes in synoptic midlatitude temperature variability. *Journal of Climate*, *28*(6), 2312 - 2331. Retrieved from `https://journals.ametsoc.org/view/journals/clim/28/6/jcli-d-14-00632.1.xml` doi: 10.1175/JCLI-D-14-00632.1

Schneider, T., Leung, L. R., & Wills, R. C. J. (2024). Opinion: Optimizing climate models with process-knowledge, resolution, and ai. *EGUsphere*, *2024*, 1–26. Retrieved from `https://egusphere.copernicus.org/preprints/2024/egusphere-2024-20/` doi: 10.5194/egusphere-2024-20

Schär, C., Vidale, P., Lüthi, D., Frei, C., Häberli, C., Liniger, M. A., & Appenzeller, C. (2004). The role of increasing temperature variability in european summer heatwaves. *Nature*, *427*, 332–336. Retrieved from `https://doi.org/10.1038/nature02300` doi: 10.1038/nature02300

Silvestri, S., Wagner, G. L., Campin, J.-M., Constantinou, N. C., Hill, C. N., Souza, A., & Ferrari, R. (2024). A new weno-based momentum advection scheme for simulations of ocean mesoscale turbulence. *Journal of Advances in Modeling Earth Systems*, *16*(7), e2023MS004130. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2023MS004130` (e2023MS004130 2023MS004130) doi: https://doi.org/10.1029/2023MS004130

Silvestri, S., Wagner, G. L., Constantinou, N. C., Hill, C. N., Campin, J.-M., Souza, A. N., … Ferrari, R. (2024). *A gpu-based ocean dynamical core for routine mesoscale-resolving climate simulations* (Tech. Rep.). Retrieved from `http://dx.doi.org/10.22541/essoar.171708158.82342448/v1` doi: 10.22541/essoar.171708158.82342448/v1

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

Souza, A. N. (2023). Transforming butterflies into graphs: Statistics of chaotic and turbulent systems. *arXiv preprint arXiv:2304.03362*. (Subjects: Fluid Dynamics (physics.flu-dyn); Dynamical Systems (math.DS); Numerical Analysis (math.NA); Computational Physics (physics.comp-ph); Geophysics (physics.geo-ph))

Souza, A. N., He, J., Bischoff, T., Waruszewski, M., Novak, L., Barra, V., … Schneider, T. (2023). The flux-differencing discontinuous galerkin method applied to an idealized fully compressible nonhydrostatic dry atmosphere. *Journal of Advances in Modeling Earth Systems*, *15*(4), e2022MS003527. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2022MS003527` (e2022MS003527 2022MS003527) doi: https://doi.org/10.1029/2022MS003527

Taylor, M., Caldwell, P., Bertagna, L., Clevenger, C., Donahue, A., Foucar, J., … Wu, D. (2023). 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.* New York, NY, USA: Association for Computing Machinery. doi: 10.1145/3581784.3627044

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

Tebaldi, C., Debeire, K., Eyring, V., Fischer, E., Fyfe, J., Friedlingstein, P., … others (2021). Climate model projections from the scenario model intercomparison project (scenariomip) of cmip6. *Earth System Dynamics*, *12*(1), 253–293.

Tebaldi, C., Dorheim, K., Wehner, M., & Leung, R. (2021). Extreme metrics from large ensembles: investigating the effects of ensemble size on their estimates. *Earth System Dynamics*, *12*(4), 1427–1501.

Tebaldi, C., & Knutti, R. (2007). The use of the multi-model ensemble in probabilistic climate projections. *Philosophical transactions of the royal society A: mathematical, physical and engineering sciences*, *365*(1857), 2053–2075.

Tebaldi, C., & Knutti, R. (2018). Evaluating the accuracy of climate change pattern emulation for low warming targets. *Environmental Research Letters*, *13*(5), 055006.

Trefethen, L. N., & Bau III, D. (1997). *Numerical linear algebra.* SIAM.

Waidelich, P., Batibeniz, F., Rising, J., Kikstra, J. S., & Seneviratne, S. I. (2024). Climate damage projections beyond annual temperature. *Nature Climate Change*, 1–8.

Watson-Parris, D., Rao, Y., Olivié, D., Seland, Ø., Nowack, P., Camps-Valls, G., ... others (2022). Climatebench v1. 0: A benchmark for data-driven climate projections. *Journal of Advances in Modeling Earth Systems*, *14*(10), e2021MS002954.

Williams, C., & Rasmussen, C. (1995). Gaussian processes for regression. In D. Touretzky, M. Mozer, & M. Hasselmo (Eds.), *Advances in neural information processing systems* (Vol. 8). MIT Press. Retrieved from `https://proceedings.neurips.cc/paper_files/paper/1995/file/7cce53cf90577442771720a370c3c723-Paper.pdf`

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

Yu, S., Hannah, W., Peng, L., Lin, J., Bhouri, M. A., Gupta, R., ... others (2024). Climsim: A large multi-scale dataset for hybrid physics-ml climate emulation. *Advances in Neural Information Processing Systems*, *36*.

Yuan, X.-C., Zhang, N., Wang, W.-Z., & Wei, Y.-M. (2021). Large-scale emulation of spatio-temporal variation in temperature under climate change. *Environmental Research Letters*, *16*(1), 014041.

Zelazowski, P., Huntingford, C., Mercado, L. M., & Schaller, N. (2018). Climate pattern-scaling set for an ensemble of 22 gcms–adding uncertainty to the imogen version 2.0 impact system. *Geoscientific Model Development*, *11*(2), 541–560.

Open Research Section

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

Acknowledgments

This work acknowledges support by Schmidt Sciences, LLC, through the Bringing Computation to the Climate Challenge, 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.