THESIS

REGULARIZED LINEAR REGRESSION TO ESTIMATE THE SPATIAL SENSITIVITY GOVERNING THE PATTERN EFFECT, COMPARATIVE ANALYSIS TO CONTEMPORARY METHODS, AND OBSERVATIONAL APPLICATIONS

Submitted by

Leif Fredericks

Department of Atmospheric Science

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Spring 2025

Master's Committee:

Advisor: Maria Rugenstein
Co-Advisor: David W. J. Thompson

Daniel S. Cooley

ABSTRACT

REGULARIZED LINEAR REGRESSION TO ESTIMATE THE SPATIAL SENSITIVITY GOVERNING THE PATTERN EFFECT, COMPARATIVE ANALYSIS TO CONTEMPORARY METHODS, AND OBSERVATIONAL APPLICATIONS

How the spatially varying temperature field affects global radiation (i.e., the "pattern effect") is crucial to understanding how sensitive Earth's temperature is to anthropogenic forcing. We capture this phenomenon in a sensitivity map using regularized linear regression. When trained on 1,000 simulated years in a climate model, the resulting sensitivity maps are consistently able to explain over 75% of the variance in net top-of-atmosphere radiation in an out-of-sample internal variability test. However, when the training data are constricted to 24 years to mirror the length of available observations, that value ranges between 0% and 75% with a median of 50%. This implies that 24-year observational sensitivity maps produced by our method carry significant uncertainty. Tested against the forced climate response in an RCP 8.5 simulation, the ideal 1,000-year training case captures $\sim$75% of the forced response magnitude, while sensitivity maps derived from 24-year periods are unreliable for projecting the warming scenario. Acknowledging the implication that our results depend highly on the particular behavior of the last two decades, we present the first physically interpretable radiative feedback sensitivity maps derived entirely from observations. We then unify several alternative methods under a common training and testing procedure. These methods all generate predictive frameworks from internal variability, except for an included Green's function. The latter approach was the primary method used to generate pattern effect sensitivity maps prior to the methods discussed in this thesis, so it grounds our comparative analysis to the current state-of-the-field. All methods match or

improve upon the Green's function's ability to predict internal variability, but vary widely in their ability to predict a step forcing $4xCO_2$ warming simulation.

ACKNOWLEDGEMENTS

This material is based on work supported by NASA FINESST Fellowship no. 80NSSC24K0024 as well as by the NSF Climate and Large-Scale Dynamics program and the NASA Earth and Space Science Fellowship program.

I would like to thank my advisors, Dave Thompson and Maria Rugenstein, for their ready advice and lively discussions that prompted many of the ideas in this thesis. I would also like to thank the Rugenstein and Thompson groups along with all my other friends in the department for their camaraderie and support.

I would also like to acknowledge the contributions to the collaborative sensitivity map comparison effort for work not yet published. Thank you to Senne Van Loon for the CNN contribution, Fabrizio Falasca and Aurora Basinski-Ferris for the FDR contributions, Marc Alessi for the Green's function contribution, and Quran Wu for the PC regression contribution.

TABLE OF CONTENTS

Chapter 1: Introduction

The energy balance of the Earth system can be represented by the top-of-atmosphere (TOA) net radiative imbalance (N; Wm^{-2}) matching the balance between external radiative forcing (F; Wm^{-2}) and the Earth system radiative response (R; Wm^{-2}):

$$N = F + R \tag{1.1}$$

Following Gregory et al. (2004), we treat radiation as a linearized feedback to global surface temperature perturbation from the pre-industrial mean (ΔT; K), where λ_F is the global "forced climate feedback" parameter ($\mathrm{Wm}^{-2}\mathrm{K}^{-1}$; e.g., Rugenstein and Armour, 2021):

$$N = F + \lambda_F \Delta T \tag{1.2}$$

A positive external forcing F (e.g., increased atmospheric CO_2 concentrations) will create a positive radiative imbalance N, bringing the system out of an existing equilibrium. Global surface temperatures will rise from the imbalance, but in a stable climate λ_F is negative, so the system will regain equilibrium when the magnitude of $\lambda_F \Delta T$ matches the magnitude of the forcing. In practice, Equation 1.2 is complicated by the fact that the value of λ_F depends on the spatial distribution of the temperature perturbation, $\boldsymbol{\Delta T}$. This phenomenon, labeled the "pattern effect" (Andrews et al., 2015; Stevens et al., 2016), is responsible for nonlinearities in the global framework in simulated long-term warming (Rugenstein et al., 2020).

We will now introduce a distinction between the forced climate feedback (λ_F) and the "internal variability feedback," which we will identify as λ_I (e.g., Davis et al., 2024). This feedback acts on interannual timescales as a response to stochastic (as opposed to forced) temperature anomalies.

By definition, forcing is zero for internal variability, so Equation 1.1 reduces to $N = R$; i.e. the only sources of TOA imbalance are natural fluctuations in radiation. We relate internal fluctuations in R to internal fluctuations in surface temperature relative to the mean, δT:

$$R = \lambda_I \delta T \tag{1.3}$$

Just as the spatial distribution of forced temperature perturbations affects the forced feedback parameter, the pattern of temperature anomalies, $\boldsymbol{\delta T}$, affects the value of λ_I. Historical simulations, in which interannual variations are on a similar scale to the forced response, show that the internal variability feedback parameter fluctuates significantly with variable surface temperatures on multidecadal timescales (Andrews et al., 2022).

Research is ongoing into how internal variability feedbacks compare to forced climate feedbacks and how observational internal variability can be used to constrain model-simulated climate responses (e.g., Colman and Hanson, 2017; Dessler, 2013; Hall and Qu, 2006; He et al., 2021; Uribe et al., 2022; Zhou et al., 2015). However, recent work suggests that the respective patterns of internal variability temperature anomalies ($\boldsymbol{\delta T}$) and forced climate temperature departures ($\boldsymbol{\Delta T}$) lead to λ_I explaining between 12% and 26% of the variance in λ_F (Davis et al., 2024).

Most previous approaches connecting λ_I to λ_F rely on diagnostic comparisons of global feedbacks. Here, we build on the recently established objective to predict the forced climate radiative response from internal variability by leveraging spatial information in the temperature field (Rugenstein et al., 2025, in review). In this predictive framework, we represent the feedback parameter as a vector of the partial contribution of the temperature at each spatial location to global radiation. For example, Equation 1.3 becomes:

$$R \;=\; \boldsymbol{\lambda_I} \cdot \boldsymbol{\delta T} \;=\; \sum_i \frac{\partial R}{\partial T_i}\, \delta T_i, \tag{1.4}$$

where i indicates a particular location on Earth and T_i is the temperature at that location. The vector $\boldsymbol{\delta T}$ quantifies the relative influence of each spatial location, and we refer to the values $\partial R/\partial T_i$ as "spatial feedbacks." Mapping these spatial feedbacks globally creates a "sensitivity map," which illustrates how sensitive global radiation is to temperature changes at specific locations.

The most common example of a sensitivity map in the pattern effect literature is the Green's function (e.g., Barsugli and Sardeshmukh, 2002; Bloch-Johnson et al., 2023; Dong et al., 2019; Zhou et al., 2017). The Green's function combines equilibrated responses in atmospheric climate models to independently simulated patch perturbations, allowing one to create the sensitivity map of radiation forced by sea surface temperature (SST). These maps represent an equilibrated response to a sustained anomaly, and do not incorporate information from the ocean response.

One alternative method for predicting the radiative response R to a particular temperature pattern is to train a convolutional neural net (CNN) on many patterns of internal variability in fully-coupled atmosphere-ocean climate model simulations (Rugenstein et al., 2025, in review). Using existing output from fully-coupled models takes the interplay of ocean and atmosphere into account, and a CNN has the particular advantage of picking up on nonlinearities between the temperature field and R. However, their nonlinear, statistical fits limit physical interpretability, and they need hundreds of years for training. Falasca et al. (2024a) employ the fluctuation-dissipation relation (FDR) to predict R based on both the instantaneous SST field and those of previous time steps back to some defined limit using a convolution. As with the CNN, this can include the behavior of a coupled ocean-atmosphere system, and FDR can "embed" feedbacks at different timescales into the convolutional process. However, some spatial information is lost in the significant dimensionality reduction employed, and it has only been demonstrated training on control simulations of several hundred years.

Here, we explore linear methods to achieve the same goal. In Chapter 2, we propose an approach for developing sensitivity maps from observational internal variability. While several methods have generated sensitivity maps from simulated internal variability, they require more input than is available from observations (e.g., Bloch-Johnson et al., 2020; Falasca et al., 2024a; Rugenstein et al., 2025, in review). To address the sample size constraint, we explore the class of methods defined by regularized multilinear regression. This builds on the approach taken by Bloch-Johnson et al. (2020) who solve for the partial differentials $\partial R/\partial T_i$ — relating radiation R to temperature T at location i — by using ordinary least squares (OLS) multilinear regression (MLR). This approach requires a low-resolution discretization of the Earth's surface, 15°x15°, and over 1,000 simulated years. We find that by introducing regularization to MLR, we can achieve interpretable results from a nominal resolution of 480km x 480km (4.3° at the equator) and decades rather than centuries. Regularization places a penalty on the strength of individual feedbacks in the $\partial R/\partial T_i$ solutions, which removes some of the overfitting tendencies in the OLS solution. Ridge regression, one variation of such regularization, has previously been used as a step in approximating a Green's function sensitivity map (Kang et al., 2023) and in estimating climate responses from multiple spatial predictors (Ceppi et al., 2024; Ceppi and Nowack, 2021). Ours is the first effort to generate a radiative sensitivity map using regularized regression, and we present the first sensitivity maps derived from observations.

In Chapter 3, we position our method alongside other approaches that predict radiative imbalance based on internal variability training procedures. Along with the regularized regression methods we discuss in Chapter 2, these include maximum covariance analysis (as in Thompson et al., 2025, in review), ordinary least squares multilinear regression (as in Bloch-Johnson et al., 2020), principle component regression, partial least squares regression, fluctuation-dissipation relation (as in Falasca et al., 2024a), and convolutional neural net (as in Rugenstein et al., 2025, in review). We compare these methods by training all of them on the same climate model simulation of internal variability, then test their

predictions against an out-of-sample period of simulated internal variability. We also test them against a step forcing simulation in which CO_2 is instantaneously quadrupled (F in Equation 1.2). We also include a Green's function generated from the same climate model (Alessi and Rugenstein, 2023), and compare how well its sensitivity map performs in both tests. Our work represents the first systematic comparison of emerging predictive sensitivity methods, benchmarking them against the standard Green's function approach.

These three components — the attempt to predict internal variability with spatial feedbacks, the attempt to predict the response to a large climate forcing, and the sensitivity map — reflect the major attributes that we have identified as important in the relationship between temperature patterns and the Earth's global radiative response. Because we extract these spatial feedbacks from relationships in internal variability, their ability to predict the internal radiative response establishes a basic layer of trust that they capture the correct behavior. By then predicting the response in an abrupt CO_2 quadrupling, we test how the connections learned from internal variability apply to the forced climate response. This furthers the goal of connecting the internal variability feedback (λ_I) to the forced climate feedback (λ_F). A major motivation for this connection is the potential to constrain forced predictions using observable behavior. For this reason, we emphasize the applicability of each method to the observational record and directly apply our regularized regression method to the observations. Lastly, sensitivity maps connect these spatial feedbacks to physical interpretations. A definable negative or positive feedback in a given region may suggest causal explanations if it aligns with expectations for a particular process. A feature gains more credence when it appears across methods, and outlier sensitivity maps offer an opportunity to explore how methodological differences contribute to variations in learned behavior. The comparative information both across these three spatial feedback attributes and across methodologies serves as a starting point for explaining interannual dynamics, understanding how they connect to the forced response, identifying key regions responsible for predictions, and uncovering the underlying physical processes at play.

Chapter 2: Development of regularized regression methods culminating in observational sensitivity maps

2.1 Data and Methods

2.1.1 Data sources

The analyses performed in this investigation depend on model output from fully-coupled GCMs, satellite observational products, and reanalysis (also referred to as "observations" for concision).

NASA satellite top-of-atmosphere radiation measurements

Observationally-derived net TOA radiation comes from the NASA product CERES-EBAF4.2 (Clouds and Earth's Radiant Energy Systems - Energy Balanced and Filled v4.2; Loeb et al., 2018). Global means of upwelling shortwave radiation, downwelling shortwave radiation, and upwelling longwave radiation are combined into a full-spectrum net radiative imbalance. This corresponds to N (Wm^{-2}), so the forcing must be removed for an observationally-derived R. We assume the forcing is approximately linear over the observational period (e.g., Dessler and Forster, 2018). The absolute forcing magnitude does not matter for anomalies, so we can estimate R by linearly detrending N. The data span from 03/2000 to 02/2024. Monthly anomalies are calculated as departures from the climatological monthly means. Annual values are calculated as 12-month means of monthly anomalies.

ECMWF temperature reanalysis

For consistency with model output, surface air temperature is used for the observationally-derived temperature field. We use 2m atmospheric temperature from the ECMWF ERA5 reanalysis product (Hersbach et al., 2020). As with the CERES net TOA data, the forced re-

sponse is approximately removed from the temperature at each grid location by detrending the series. As with radiation observations, the data span from 03/2000 to 02/2024. Monthly anomalies are calculated as departures from the climatological monthly means. Annual values are calculated as 12-month means of monthly anomalies.

GCM piControl and RCP 8.5 warming scenario

For simulated internal variability, we use millennial-length pre-industrial control (piControl) model output from four models that participated in the LongRunMIP project (Rugenstein et al., 2019). This provides us with an idealized model testbed with access to a wealth of training data. Most of the analysis has been performed on MPI-ESM-1.2 (Max Planck Institute Earth System Model, v1.2) which has a piControl simulation length of 1,237 years (Mauritsen et al., 2019; Rohrschneider et al., 2019). This is the member of LongRunMIP for which we have access to a Green's function produced from its atmospheric component (Alessi and Rugenstein, 2023). The MPI-ESM piControl data are divided into a 1,000-year training set and a 237-year test set, with anomalies defined relative to the respective sets rather than the full piControl. In the absence of forcing, R is identical to N (Equation 1.1).

We use additional LongRunMIP member models with sufficiently long piControl simulations to test the robustness of our observational-length fitting analysis. We use 1,000 years of piControl from CESM-1.0.4 (Community Earth System Model, v1.0.4; Danabasoglu et al., 2012; Gent et al., 2011; Rugenstein et al., 2016), 2,000 years of piControl from CNRM-CM6-1 (Centre National de Recherches Météorologiques Climate Model v6-1; Saint-Martin et al., 2019; Voldoire et al., 2019), and 5,200 years of piControl from GFDL-CM3 (Geophysical Fluid Dynamics Laboratory Climate Model v3; Donner et al., 2011; Paynter et al., 2018). For consistency with MPI-ESM-1.2, the respective test sections of these three models are also 237 years in length, with the remainder making up the training sections. Anomalies

are again defined relative to respective sections rather than the full piControl. Similarly, for any subsets taken from these sections, anomalies are only with respect to that subset.

We also test against an RCP 8.5 warming scenario taken from the MPI-ESM-1.1 Grand Ensemble (Maher et al., 2019). This ensemble is performed on a slightly earlier version of the MPI-ESM (1.1 vs. 1.2), though very little changed between versions so it is a reasonable comparison. RCP 8.5 was chosen to transition from a state comparable to pre-industrial internal variability to a state with a large warming signal. This shows the transition from predicting internal variability to predicting a forced climate response. Anomalies for each ensemble member are defined relative to the first 30 years of that simulation (1871-1900). This is to stay consistent with defining anomalies relative to the particular data series under consideration (as with the training and testing sections in the piControl) and to avoid any coincidental offset between the starting point of the RCP 8.5 simulations and the particular set of training data used. Models only output N, so we need an estimate of total forcing (F) to estimate R in the warming scenario: $R = N - F$. We use an existing estimate of F for RCP 8.5 in MPI-ESM-1.1 (Alessi and Rugenstein, 2023), following the protocol established by Pincus et al. (2016).

2.1.2 Regularized regression methods

Our approach expands Equation 1.3 to predict the global radiative response to a spatial field of temperature anomalies. Assuming the TOA radiative imbalance can be approximated as the linear response to local temperature anomalies at each gridpoint i (T_i), as discretized on some reasonably fine grid scale, we treat these gridpoint temperatures as predictors. We approximate the internal variability system linearly as

$$R = \beta_1 T_1 + \beta_2 T_2 ... \beta_n T_n, \tag{2.1}$$

where β_i are coefficients on each of the predictor variables. The challenge with this assumption is that these local temperature predictors are highly correlated with one another and all may not be causally linked to global radiation.

The standard approach to solving a multilinear problem would be to estimate the coefficients β_i that minimize the error between the true system R and that predicted by Equation 2.1. This is the OLS solution implemented in Bloch-Johnson et al. (2020). Here we expand on that approach by choosing coefficients that satisfy additional constraints. The following equation defines the various regularizations we will consider by altering the minimization process:

$$\hat{\beta} = \arg\min_{\beta} \left(||y - X\beta||_2^2 + \alpha_1 ||\beta||_1 + \alpha_2 ||\beta||_2^2 \right) \tag{2.2}$$

In Equation 2.2, β is a candidate vector of predictor coefficients, $\hat{\beta}$ is the vector of coefficients chosen by the optimization, X is the matrix of local temperature time series $(T_i(t))$, y is the true time series of global radiation, α_1 is the regularization coefficient on the 1-norm of β, and α_2 is the regularization coefficient on the squared 2-norm of β. Setting α_1 and α_2 to zero in Equation 2.2 reduces the problem to the ordinary least squares (OLS) optimization. The equation states that the chosen β will be the vector that minimizes the combined sum of 1) the sum of squared error between predicted and actual radiation at each time step, 2) the sum of β_i magnitudes scaled by α_1, and 3) the sum of squared β_i values scaled by α_2.

Illustration

To conceptualize the effect of including the 1-norm and 2-norm of the coefficient vector in Equation 2.2, consider a 4-variable linear system: $y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$. For the sake of this argument, we have three candidate solutions for the vector $\beta = [\beta_1, \beta_2, \beta_3, \beta_4]$ that all recreate y reasonably well:

[-4, -14, 15, 3], [-5, 0, 0, 4], and [-3, -2, -2, 3].

Let us say that the error in the linear model's prediction, $||y - X\beta||_2^2$, is lowest for the first solution. This is the result from using OLS. However, we can see that the 1-norm of β, $||\beta||_1$, is lowest for the second solution (9). And similarly, the squared 2-norm of β, $||\beta||_2^2$, is lowest for the third solution (26). These are the three terms within the minimization of Equation 2.2, and the priorities set by α_1 and α_2 would determine which of these three solutions the optimization would select.

Ordinary least squares

To get the OLS solution, we can set α_1 and α_2 to zero in Equation 2.2. This approach has the advantage of prioritizing goodness of fit. However, this fit is specific to the range of X and y. Least-squares solutions, especially with a large number of predictors, tend to make use of mostly offsetting random fluctuations to fit noise in the training data, making them poor at out-of-sample prediction. We represent this behavior in the 4-variable example as large, offsetting coefficients on x_2 and x_3 for the first solution. This typically arises when the predictors x_2 and x_3 are highly correlated and have a small residual that happens to fit noise in y. This tendency towards overfitting explains some of the noise in the sensitivity maps of Bloch-Johnson et al. (2020).

LASSO

LASSO, (for Least Absolute Shrinkage and Selection Operator; Tibshirani, 1996), includes a nonzero α_1 penalty in Equation 2.2. α_1 is user-selected and sets how important "L1-regularization" is to the optimization. LASSO penalizes the 1-norm of β, the sum of magnitudes, which in practice promotes sparsity. This behavior is represented in the second solution to the 4-variable example. "Zero" often becomes the optimal solution for most indices of $\hat{\beta}$, a favorable property for isolating the predictors with the most predictive power. LASSO helps alleviate the weakness in the assumption that the predictors (T_i) are independent. If a group of predictors tend to move together, LASSO tends to zero out all but the most predictive of the group. This accounts for correlations among local

temperatures. LASSO also has a stronger argument for identifying causal connections. From Equation 2.2, a particular predictor must bring a large amount of predictive power to garner a nonzero coefficient when α_1 is high. The few local temperature variables offering the most predictive skill are likely candidates for true causal connections. As a caveat to these two advantages, zeroing out most members of a correlated group may overlook some true causal connections. Even if they are highly correlated, each could add a small amount of independent information. LASSO will tend to over-attribute the effect of the group (in our case, region) to the effect of the most predictive member of the group (in our case, specific location). In the context of this investigation, if two local temperatures causally linked to global radiation are correlated and spatially distant, it is possible LASSO would keep only one of them, affecting the visual interpretation of the resulting spatial feedback map.

Ridge regression

Setting α_1 to zero, but bringing in the third term in Equation 2.2 by selecting a nonzero α_2, defines ridge regression (Hoerl and Kennard, 1970). Ridge regression penalizes the squared β 2-norm, in other words, the sum of squares. Whereas LASSO tends towards sparse solutions, ridge tends towards evenness across $\hat{\beta}_i$ magnitudes, as is true in the third solution to the 4-variable example. For an intuitive geometric visualization on these behaviors, see Figure 2 in Tibshirani (1996). Even magnitudes help to prevent the overfitting tendencies of OLS multilinear regression. OLS can take two highly correlated predictors and assign one a large positive value and the other a large negative value, which amplifies the small residual between them to fit random noise. Ridge offsets OLS loss with the squared magnitude of $\hat{\beta}_i$ terms, so large values are particularly penalized. In a similar way, the tendency of LASSO to assign a large value to just one member of a correlated group would also be penalized heavily by ridge regularization. Ridge instead spreads a correlated group's influence among its members with relatively even-valued coefficients. This highlights

regions where the predictors tend to move together, and captures correlations better than OLS.

Elastic net

The regression loss function with nonzero values for both α_1 and α_2 defines an "elastic net" (Zou and Hastie, 2005), and trades all three terms in Equation 2.2. This approach aims to include the favorable properties of both LASSO and Ridge. Varying the relative magnitudes of α_1 and α_2 trades the sparsity promoted by LASSO with the evenness promoted by Ridge.

Maximum covariance analysis

Lastly, we consider another approach to selecting $\hat{\beta}$ that is relevant to the regularizations discussed. Maximum covariance analysis (MCA) identifies the spatial pattern in one field that explains the maximum possible amount of the covariance between that field and another (Bretherton et al., 1992). We are mapping the covariance between spatial temperature and global radiation, so it is reasonable to explore how well this pattern performs as a sensitivity map. In fact, we observe that the spatial pattern of MCA is the same pattern of relative magnitudes approached as α_2 approaches infinity in ridge regression, which is why we consider it in the same class of regression defined by Equation 2.2. MCA particularly highlights the structure of correlated predictors, and is examined in more detail in Thompson et al. (2025, in review).

2.1.3 Procedure to generate sensitivities

Equal-area MPAS grid

Both to reduce the number of predictors while preserving decent resolution in the tropics, and to create cleaner input data for the fitting algorithms by avoiding latitude-weighting, the spatial data from the GCMs and observational data sets were regridded

(using conservative remapping for consistency in spatial means) to an unstructured equal-area grid. The grid, taken from the Model for Prediction Across Scales (MPAS) Atmosphere mesh, has 2,562 horizontal grid cells, approximately consistent with a nominal resolution of 480km x 480km (Heinzeller et al., 2016).

Standardization of temperature and radiation variations

Certain regions, for example western boundary currents and extratropical land, have much greater temperature variance than regions like the West Pacific warm pool (WPWP) and Caribbean. These relatively large anomalies are more likely to be identified as "important" by the regularization process merely from the strength of the signal. To counter this, we divide time series by their temporal standard deviations prior to performing regression analysis, which we refer to as "standardization".

The solution $\hat{\beta}$ to Equation 2.2 is consequently in units of standard deviation, making it dependent on the statistics of the training data set. Rather than standardizing any out-of-sample test cases relative to the training data, a much more interpretable option is to rework the solution to be in the original units of temperature (K) and radiative imbalance anomalies (Wm^{-2}) as follows: To process $\hat{\beta}$ into a predictive sensitivity map for absolute temperature anomalies, we multiply each index of the solution vector by global radiation standard deviation in the training set over standard deviation in training temperature at location i to generate an intermediate prediction vector, β^*, in units of $\mathrm{Wm}^{-2}\mathrm{K}^{-1}$,

$$\beta_i^* = \hat{\beta}_i * \frac{\sigma_R}{\sigma_{T_i}}. \tag{2.3}$$

To illustrate the importance of this step, we perform an MCA regression on 1,000 years of MPI-ESM piControl internal variability both without (Figure 2.1a) and with (Figure 2.1b) standardizing the data and rescaling to physical units (Equation 2.3). These sensitivity maps both represent the partial contribution of local temperature at a given location to the global mean radiative imbalance, and the signs of the responses agree

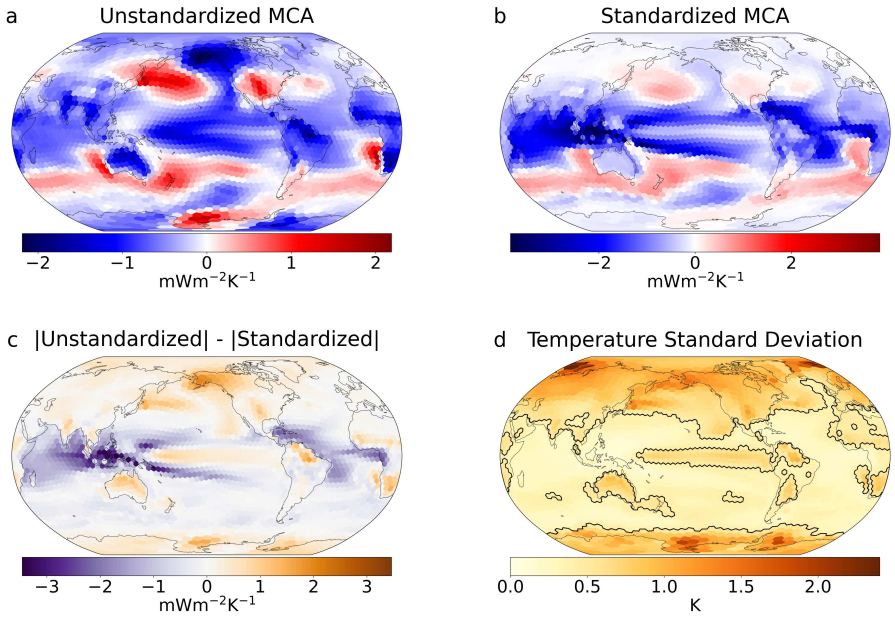

Figure 2.1: Maximum covariance analysis (MCA) solution for the MPI-ESM global climate model pre-industrial control showing the partial contribution of local temperature to global radiative imbalance (sensitivity map) for a) the original, unstandardized anomalies of temperature and radiation, and b) the standardized anomalies scaled to physical units as described by Equation 2.3. We plot spatial feedbacks in unit of milliwatts per square meter per Kelvin ($\mathrm{mWm^{-2}K^{-1}}$) for clarity. c) The difference in the magnitude of the response between (a) and (b). d) The local standard deviation in annual temperature in the control simulation. The zero-contour from (c) is superscribed onto (d) to highlight the similarity between the two.

between the two. However, the unstandardized procedure produces a much more uniform magnitude of regression coefficients across the globe, with land areas and higher latitudes in particular appearing relatively more important. Note that the scale has greater range for the standardized procedure (Figure 2.1b). Because the response is concentrated to fewer regions, they must have a stronger influence.

To contrast the strength of response between the two maps, we take the difference in magnitude between them (Figure 2.1c). This represents areas identified as more influential by the unstandardized process (positive values) and those found to be more influential by the standardized process (negative values). As expected, the regions highlighted by the former line up with the locations of highest temperature variance in the same period (Figure

2.1d). Standardizing the data before performing a regression eliminates the tendency to overvalue high variance areas, and we have found that it does improve out-of-sample predictive ability.

Regression dilution

We also find that, when used to recreate the training data, β^* (Equation 2.3) consistently underestimates the magnitude of the training radiative anomalies from the training temperature anomalies. This is analogous to "regression dilution", whereby sources of variance unaccounted for in the assumption that our predictor variables (T_i) fully define our predictand (R) tend to move the slope of the regression towards zero (Carroll and Ruppert, 1996; Frost and Thompson, 2000). This behavior is a known complication in estimating climate feedbacks (e.g., Gregory et al., 2020; Proistosescu et al., 2018). This is especially pronounced against large anomalies, so predictions using β^* would significantly and unrealistically depress the response in extreme years and in warming scenarios. To correct for the dilution bias, we calibrate β^* by enforcing that our prediction vector reproduces the variance of the training data.

To do so, we generate a predicted recreation of the training data radiation,

$$\hat{R} = X\beta^*, \tag{2.4}$$

then divide by the standard deviation of $\hat{R}$ and multiply by the standard deviation of the true training radiative time series:

$$J = \beta^* * \frac{\sigma_R}{\sigma_{\hat{R}}}. \tag{2.5}$$

J is the final sensitivity map in our procedure relating the partial contributions of local temperature anomalies to global TOA radiative anomalies. We assume J represents the system we describe in Equation 1.4, i.e.,

$$J = \left[\frac{\partial R}{\partial T_1}, \frac{\partial R}{\partial T_2}, \frac{\partial R}{\partial T_3} ... \right]. \tag{2.6}$$

The calibration in Equation 2.5 is analogous to orthogonal regression in single regression. This is a new approach for calibrating sensitivity maps that specifically does not rely on any information from the test cases and creates more realistic variance in out-of-sample applications. We acknowledge one potential risk in this approach is to over-attribute the magnitude of the radiative response to temperature.

Hyperparameter selection

In the case of LASSO, ridge regression, and elastic net, we use cross-validation to set the application specific hyperparameters α_1 and α_2. For the yearly MPI-ESM-1.2 piControl data, we have 1,000 years in the training set, which we split into a randomly selected training subset of 800 years and a validation subset of 200 years. Consecutive years are not necessarily kept together because our method treats each time step as an independent measurement. We use a gradient-descent optimization on the hyperparameter(s), finding a candidate sensitivity map from the training subset with candidate hyperparameter values. We define the loss function for these candidate maps as the root mean squared error (RMSE) between their predicted R in the validation subset and the actual validation subset R. The results are the values of α_1 and α_2 producing the sensitivity map that best predicts the validation subset when applied to the training subset. The optimized hyperparameters are then used with Equation 2.2 on the full 1,000-year training section to produce the cross-validated sensitivity map. Neither MCA nor OLS have hyperparameters, so we apply these methods directly to the 1,000-year training section.

2.2 Results

2.2.1 Sensitivity maps from ideal 1,000-year case

Using the methods outlined above, we create sensitivity maps for MCA, ridge regression, elastic net regression, LASSO, and OLS regression from the full training section. We then test them against the 237-year testing section annual-mean data, which functions as an out-of-sample internal variability test. Additionally, we include a null hypothesis where we calculate the global feedback parameter λ_I (see Equation 1.3) from an ordinary least squares regression of global radiation onto global temperature. The extent to which a given spatial feedback map improves upon the estimate from λ_I is the additional information that can be learned by considering the pattern effect.

The optimized regularized methods of LASSO, elastic net, and ridge regression produce sensitivity maps that explain 77% of the variance (R^2=0.77) in the out-of-sample internal variability test. This high level of prediictive skill suggests that these sensitivity maps effectively capture most of the effect of physical feedbacks in internal variability. OLS regression performs worse than regularized variants but still yields a sensitivity map explaining 61% of the test variance. There is a steeper drop in performance for MCA, which explains less than half of the variance at only 42%. The global feedback is the least predictive of the methods considered, explaining only 26% of the variance in the test data and confirming the importance of the spatial temperature pattern to the Earth's radiative response.

In the forced climate (Equation 1.2), we expect the response in forced radiation to be explainable primarily from forced change in surface temperature. The signature of anthropogenically forced climate change is much stronger than noise from other forcings (e.g., volcanoes). Conversely, the signal between interannual radiation and interannual temperature is of similar magnitude to the effects of noisy processes not captured by temperature alone (e.g., unforced variability in circulation). Therefore, we expect a limit to

out-of-sample variance explained somewhat less than 100%. Our method suggests this limit is around 80% for MPI-ESM.

Figure 2.2 shows the sensitivity maps from MCA, ridge regression, elastic net, and LASSO as spatial feedbacks. The value at each location is the unique, independent contribution from local temperature to global radiation. This will not necessarily agree with how temperature at that location correlates with broader, physical mechanisms. For example, in a hypothetical case where four locations correlate strongly with the El Nino-Southern Oscillation (ENSO), but one of them is also a slight indicator of a process dampening ENSO, the regression method ideally assigns the ENSO signal to the three and a countervailing signal to one.

The correlated structure of the temperature field dominates the MCA map (Figure 2.2a). The tropics have a nearly uniform negative feedback, with the exceptions of the Namibian and Western Australian coasts. This is in contrast to the negative/positive west/east divide we see across the tropical Pacific in Green's Function experiments (e.g., Bloch-Johnson et al., 2023), though Falasca et al. (2024a) did find similarly negative tropical feedbacks on timescales 1 year or longer in GFDL-CM4. However, the spatial clarity of this map comes with much less predictability than the other methods. So while it represents a relevant mode of the system, the implied connections from local temperature to global radiation are less representative of physical feedbacks than the other maps in Figure 2.2.

The cross-validated ridge regression sensitivity map (Figure 2.2b) is more heterogeneous than MCA, though correlated spatial locations still stand out. The WPWP dominates the negative feedback, and the equatorial eastern Pacific is a significant region of positive feedback, reminiscent of the Green's function spatial feedback pattern found by Alessi and Rugenstein (2023) in the same MPI-ESM model. There is also a relatively strong negative feedback contribution from the Caribbean and East coast of northern South America, consistent with the MPI-ESM Green's function, though more distributed along the coastline.

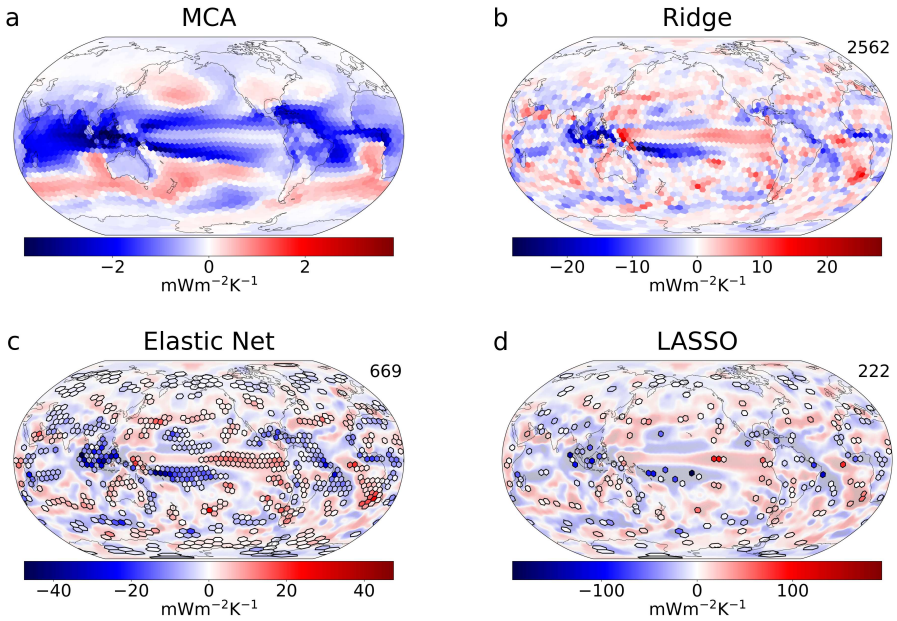

Figure 2.2: Estimating the spatial feedback from temperature at each gridpoint to global radiation using a) maximum covariance analysis (same as Figure 2.1b), b) ridge regression, c) elastic net regression, and d) LASSO regression from 1,000 years of MPI-ESM-1.2 piControl. For elastic net and LASSO (c,d, resp.), a blurred version of the ridge regression map is shown as a backdrop to emphasize the broader correlated regions. Only outlined grid locations are actually given nonzero values by these sparse regression methods. Numbers in the top right (b,c,d) indicate the number of nonzero coefficients.

However, there are also significant differences from the Green's function investigation. The equatorial band of positive feedback with a particularly strong signal over New Guinea is not present in prior Green's function investigations. This same feature does appear relevant in CNN approaches to predict global radiation from spatial temperature (Rugenstein et al., 2025, in review), and shows up in one formulation of the FDR map from Falasca et al. (2024a) on a monthly timescale, though is absent on the yearly timescale. This suggests the feature is more relevant to interannual variability than to an equilibrated response. To the north and south of this positive feature are two bands of negative feedbacks extending into the central Pacific from the west. These are roughly consistent with the climatological locations of the Inter-Tropical Convergence Zone (ITCZ) and the South Pacific Convergence Zone (SPCZ) in the MPI-ESM piControl. Though not as obvious as in the MCA map in Figure 2.2a, ridge regression in 2.2b shows generally positive extratropical feedbacks poleward of these two negative bands. Further poleward and eastward the feedbacks alternate back to negative, again more obviously in the MCA map. Beyond this alternating pattern, it is difficult to find structure in the extratropics.

While not fundamentally different in terms of regional significance, including sparsity in the elastic net and LASSO regressions offers an entirely new perspective on spatial feedbacks. Both Figures 2.2c and 2.2d imply no contribution from a majority of the Earth's surface, even in the tropics, without appreciable loss in predictive skill relative to the fully-leveraged ridge regression map of Figure 2.2b. A smoothed version of the ridge regression map is included as a backdrop for Figures 2.2c and 2.2d, while the locations with nonzero elastic net/LASSO coefficients are outlined in black. This highlights just how few local temperature predictors are needed, and many outlined points are close to zero, especially in the extratropics. The comparison to ridge regression shows that these isolated points correspond to larger regions, demonstrating how LASSO reduces correlated groups to their most predictive members. The largest difference between Figures 2.2c and 2.2d is the extent to which the elastic net map still maintains some structure in correlated regions.

Sparsity adds a previously unconsidered element to interpreting spatial feedbacks. The ability for elastic net and LASSO sensitivity maps to match the predictive ability of ridge regression indicates that the same feedback information can be gleaned from a few specific locations. However, we cannot guarantee these areas are uniquely predictive within wider correlated regions, or if they are the most representative of broadly predictive regions.

The transition from ridge regression to elastic net and ultimately LASSO gradually strips away superfluous information with respect to predictive skill, but at the cost of representing correlations in the temperature field. The relevant signs and locations of feedbacks present in the LASSO sensitivity map are the same as in the ridge regression map, but would be hard to interpret physically. Conversely, the most informative locations identified by LASSO are not simply the highest magnitude values in the ridge regression map. Because they highlight different information about the system, there is no objective best among these methods. Subjectively, we find the elastic net map of Figure 2.2c to be most interpretable, isolating the most important locations while still evoking a sense of the broader regional feedback structure. We recommend using all four methods to more fully characterize the system, but if one must be chosen we recommend elastic net.

2.2.2 Implications from 24-year sample size

The millennial-length piControl simulation is an ideal case to demonstrate the potential of these methods. However, we aim to extend this approach more generally, including to more data-constrained cases like the observational record within which we have CERES measurements. This limits the training data to just 24 years. The predictors greatly outnumber the observational instances, making multilinear regression particularly prone to overfitting.

We first address whether 24 years are sufficient to capture the physical processes characteristic of the system. We cannot perform an out-of-sample test on the observations, so we recreate the data limitation with global climate model output, and explore how repre-

sentative we can expect 24 years to be for a GCM. To do so, we resample with replacement 1,000 instances of consecutive 24-year stretches from the 1,000-year MPI-ESM-1.2 piControl simulation.

Though limited to 24 years of observations, we consider increasing our sample size by resolving them as monthly mean values rather than annual mean values. Monthly fluctuations around the seasonal climatology are higher in magnitude than annual fluctuations around the mean climatology, so a regression performed on a monthly timescale would not perform well at an annual internal variability test. However, there are connections between local temperature variance and global radiation variance that exist at both timescales. Our process standardizes the data before performing a regression, which captures these connections in a comparable way for both annual and monthly mean samples. Rather than translate the standardized regression back to monthly predictors, we can use Equations 2.3 and 2.5 to calibrate the result to the annual mean variance in the training data.

For each 24-year period in MPI-ESM-1.2, we use the cross-validated elastic net regression process to create a sensitivity map and subsequently test that map against the same out-of-sample test section. We do so for both $n = 24$ annually averaged data and $n = 288$ monthly averaged data over that period. In Figure 2.3 (far left column) we see a wide range of performance, with a median test variance explained of approximately 50% for both. In nearly all cases, the elastic net sensitivity map contains more information than the null hypothesis global feedback, which can only explain 26% of the variance in the test section R. We have suggested that the maximum out-of-sample variance we expect to explain in MPI-ESM-1.2 is around 80%, which is higher than we see with any 24-year period. Despite the increase in sample size, using monthly mean training data does not appreciably improve predictive skill. This suggests we are able to learn the relevant interactions from monthly mean data, but that no additional information is gained relative to the annual formulation.

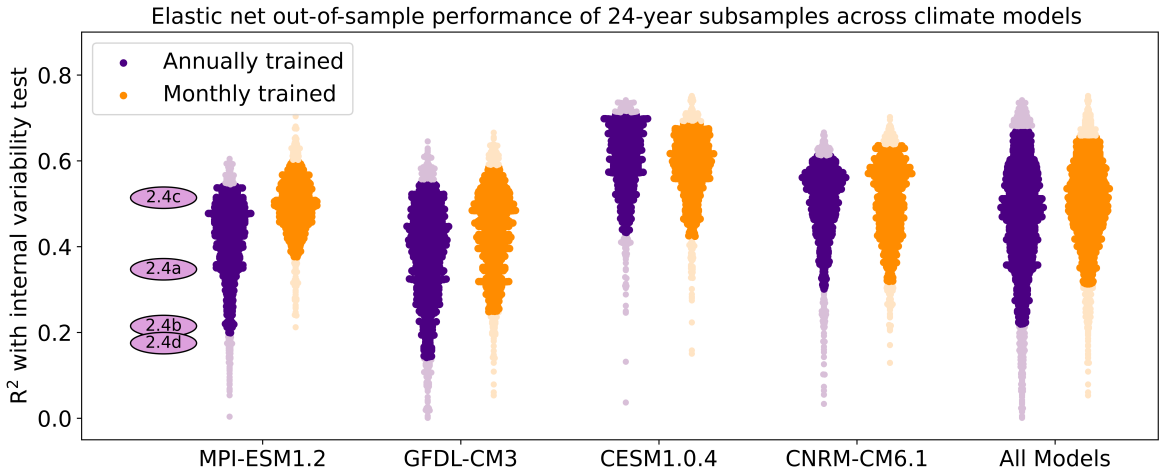

Figure 2.3: Scatter plot distributions showing how well sensitivity maps derived from elastic net regression on consecutive 24-year periods perform on out-of-sample internal variability. 1,000 instances are tested for each model using both annual and monthly mean data. The rightmost distribution includes all 4,000 scores across all 4 models. Darker shading indicates points between the 5th and 95th percentiles. Labeled circles correspond to scores of the sensitivity maps shown in Figures 2.4a, 2.4b, 2.4c, and 2.4d, resp.

To show that this behavior is not particular to a single GCM, we repeat the analysis for three additional GCMs that have contributed sufficiently long pre-industrial control simulations to LongRunMIP. As with MPI-ESM-1.2, we resample 1,000 periods from millennial-length piControl simulations in GFDL-CM3, CESM 1.0.4, and CNRM-CM6-1. For consistency with MPI-ESM-1.2, we reserve 237 years as a test section in each of these models. For each 24-year period in each model we use the same cross-validated elastic net regression process to create a sensitivity map, then compare the predicted to actual global radiation in the respective test sections (Figure 2.3). The distributions are similar for all four GCMs. Increasing the sample size with monthly means has even less effect in the other models. Considering all GCMs together, a given sensitivity map from 24 years could explain between 0% and 75% of the variance in out-of-sample internal variability. Maps derived from CESM-1.0.4 do perform better on average than in the other GCMs, which may imply this model has less variation in radiative feedback processes between decades. However, the CESM-1.0.4 distributions have shorter upper tails than the others, supporting

the idea that interannual temperature may not be able to explain more than approximately 80% of variance in interannual radiation.

The significant spread in sensitivity maps produced from such a small sample size emerges from two sources of uncertainty. One comes from the problem setup; the processes learned from a particular 24-year stretch of time may not apply well to other possible internal variability time periods. The other comes from methodological choice. In the ideal 1,000-year training period, we found no difference in the resulting sensitivity map regardless of how we divided it into the 800-year training subset and 200-year validation subset. However, the particular training/testing split for only 24 years does affect the outcome. In Figure 2.4 we show example sensitivity maps using elastic net regression trained on 24 years of annual-mean data from MPI-ESM-1.2, as done for the subsamples in Figure 2.3. The map in Figure 2.4a is derived from one 24-year period and explains 35% of the variance in the out-of-sample test (R^2=0.35). Figure 2.4b is a sensitivity map for a different 24-year period and only explains 21% of variance in the test. This demonstrates the first source of uncertainty; the two maps do not resemble one another and correspondingly have very different predictive ability.

To illustrate the second source of uncertainty, we then adjusted the training/validation split for the first and second time periods considered (Figures 2.4c and 2.4d, resp.). By coincidence, this happened to increase the number of nonzero predictors in both cases. In the first period, this improved the out-of-sample predictive ability from 35% to 51% variance explained. In the second, it slightly worsened the prediction from 21% to 17%. In both periods, spatial feedback signs and locations were qualitatively similar. This suggests that the qualitative features reflect real physical processes that exist in the training period. Figure 2.4c looks the most like the ideal case map in Figure 2.2c, in line with its leading performance. Both maps from period 1 (Figures 2.4a and 2.4c) are more predictive than those from period 2 (Figures 2.4b and 2.4d), supporting the idea that some time periods are more informative than others in capturing overall behavior in internal variability. However,

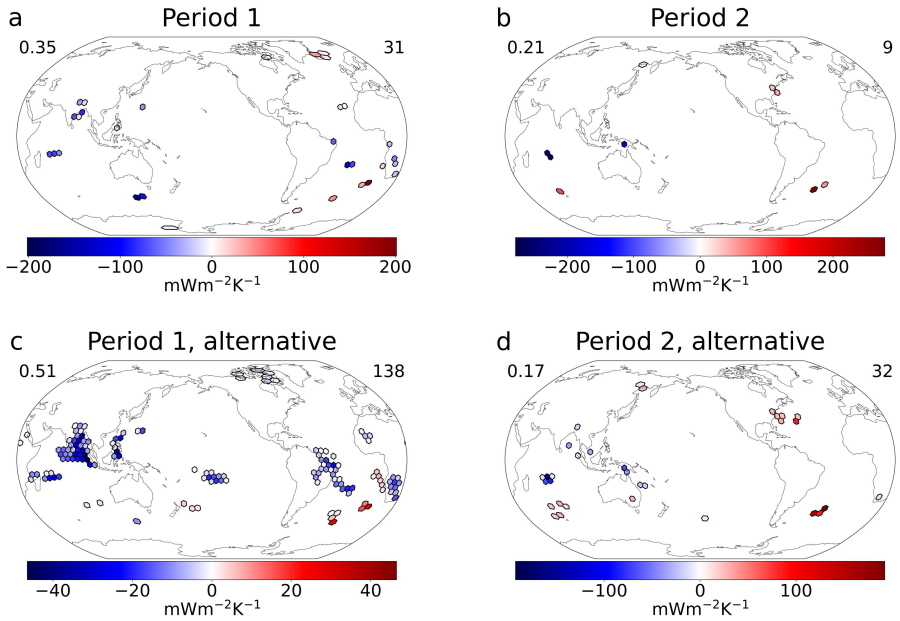

Figure 2.4: Elastic net sensitivity maps trained on 24 consecutive years of annual mean data from the MPI-ESM-1.2 piControl. a,c) maps trained on the same randomly selected 24-year period with different divisions of training and validation subsets in the fitting process. b,d) maps trained on a different 24-year period than (a) and (c), also with different training/validation divisions from one another. Numbers in the top right of a map indicate the number of nonzero predictors. Numbers in the top left indicate the R^2 value between the internal variability test and the sensitivity map prediction. Note that the color scale is different for each map.

the fact that the scored difference between training protocols in period 1 (51% vs. 35%) is larger than the difference between the two time periods in Figures 2.4a and 2.4b (35% vs. 21%) implies that a large amount of the spread in Figure 2.3 may come from methodological uncertainty. The performances of Figures 2.4a, 2.4b, 2.4c, and 2.4d are shown in Figure 2.3 where they would fall among the other 24-year sensitivity maps trained on annual mean data from MPI-ESM-1.2.

2.2.3 Application to a simulated warming scenario

We next show how well sensitivity maps learned from internal variability perform at predicting the radiative response in a forced climate change simulation. To do so, we use a 100-member ensmeble of the RCP 8.5 scenario in MPI-ESM-1.1, which begins at piControl conditions and transitions to a highly forced state over simulated years 1871-2099 (Figure 2.5).

In a single ensemble member, the ideal case elastic net sensitivity map (see Figure 2.2c) captures most of the forced response in global radiation (Figure 2.5a). Where internal variability dominates in the first half of the simulation, the prediction (red) matches the truth (black) well, then begins to diverge slightly as the warming signal starts to exceed the range of the pre-industrial internal variability. Even in the latter years of the simulation, the year-to-year variations are well-captured, and the absolute values are off by only about 1 Wm^{-2}. We also qualitatively demonstrate the uncertainty inherent to predictions using training data limited to 24 years. In Figure 2.5a we plot in cyan the predictions from 1,000 annually-trained elastic net sensitivity maps derived from 24 years of MPI-ESM-1.2 piControl (the same maps represented in the leftmost column of Figure 2.3). The spread of predictions overlaps both the ideal case prediction and the truth, but range from overestimating the response by 4 Wm^{-2} to showing a runaway climate with a positive overall feedback.

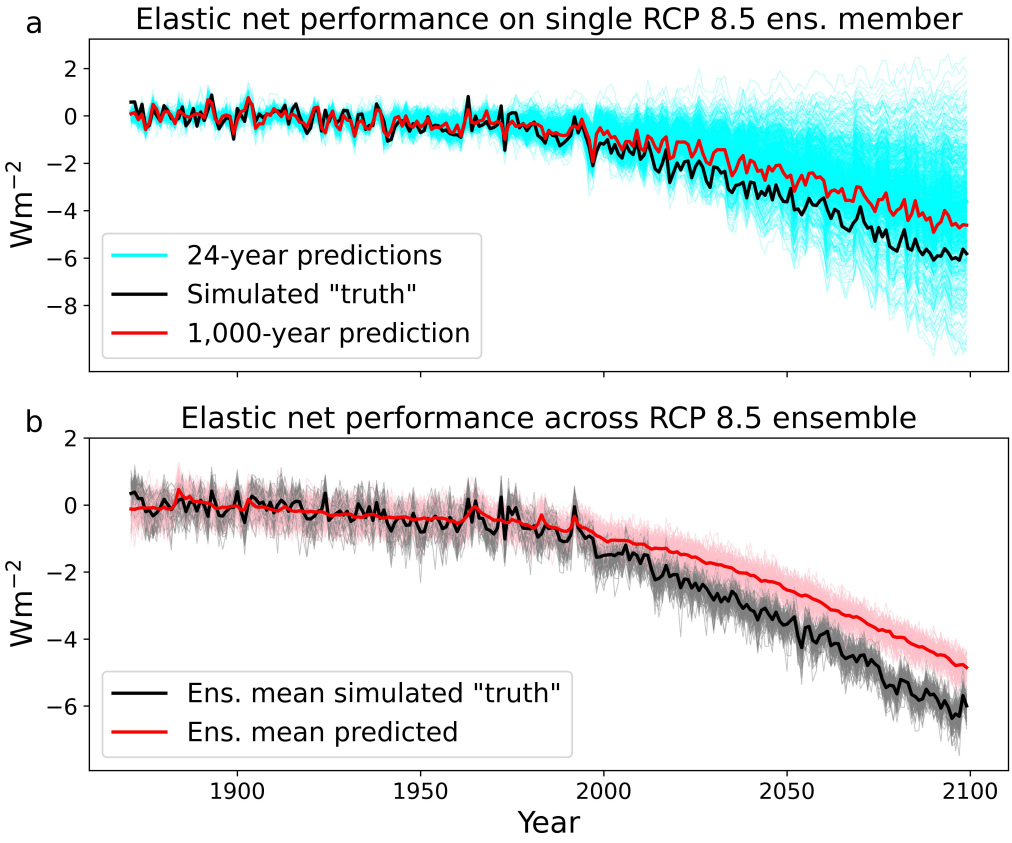

Figure 2.5: a) Predicted radiative response in one member of the 100-member MPI-ESM-1.1 RCP 8.5 ensemble using both the ideal 1,000-year case elastic net sensitivity map (red; from Figure 2.2c) and all 24-year annual mean elastic net sensitivity maps from MPI-ESM-1.2 represented in Figure 2.3 (cyan). b) predicted RCP 8.5 ensemble mean using the 1,000-year elastic net sensitivity map.

In Figure 2.5b we plot the forced climate response directly by taking the mean across all simulated RCP 8.5 ensemble members (black). This averages out stochastic internal variability leaving only the forced signal common to all members. We predict a corresponding ensemble using the ideal case elastic net map and also plot the analogous forced response predicted by our method (red). Thin lines show individual ensemble members for both simulated truth (grey) and predictions (pink). As with the single member explored in Figure 2.5a, the forced response reaches a difference of about 1 Wm^{-2} by the end of the simulation period, indicating the magnitude is consistently underestimated across ensemble members. Our sensitivity map representing the internal variability feedback λ_I does not capture everything about the forced response in radiation, so cannot be directly substituted as the forced climate feedback λ_F.

2.2.4 Application to observations

From our analysis in GCMs, we expect sensitivity map spatial structure to be informative from observational-length training data, despite the limits on predictive ability. As discussed in Methods, we estimate observational variations in spatial temperature and global radiation from locally detrended ERA5 2m-temperature anomalies and detrended global mean CERES-EBAF net TOA radiation anomalies, respectively. We apply MCA and our cross-validated ridge regression, elastic net regression, and LASSO regression methods to the observational data (Figure 2.6).

All four sensitivity maps in Figure 2.6 have more defined spatial structure than in the larger sample GCM fits, which is primarily because the system has experienced fewer configurations in the training data. This also means there is less information for ridge regression to capture relative to MCA, which explains why the MCA and ridge regression maps (Figures 2.6a,b, resp.) are essentially the same, unlike in Figure 2.2. In the tropics, the observations include strong positive feedbacks in the Southeast Pacific, North Atlantic, and Gulf of Africa. As in the GCM maps, we see an equatorial positive feedback flanked by

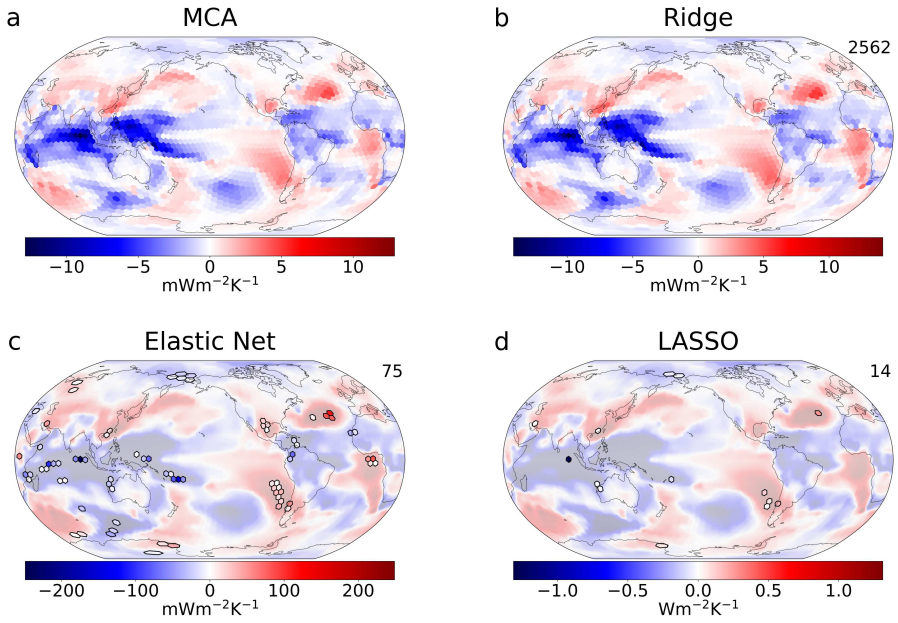

Figure 2.6: As in Figure 2.2, but using 24 years of locally detrended annual mean ERA5 2m-temperature data and detrended global CERES-EBAF net TOA radiation data.

negative feedbacks in the convergence zones. We also see an alternating positive/negative pattern extending poleward and eastward from the WPWP. The largest negative feedback extends from the WPWP into most of the Indian ocean.

The elastic net map (Figure 2.6c) is much more sparse than that of the GCM (Figure 2.2c). The strongest region of negative feedback including the Indian Ocean, WPWP, and SPCZ is captured, but the negative feedback near the ITCZ north of the equator is not. In fact, the only other included region in the Pacific is the positive feedback off the coast of South America. The equatorial positive band and wavelike alternating pattern in the Central Pacific are not included in the elastic net sensitivity map. The other two regions of strongest positive feedback seen in the ridge map (Figure 2.6b), the North Atlantic and Gulf of Africa, are included in the elastic net sensitivity map. Other locations do not appear to be as relevant. The sensitivity map from LASSO is extremely sparse, and many of the included points have values near zero. The only regions with notable representation are

the Indian Ocean, with a strong negative feedback, and the North Atlantic and Southeast Pacific, both with moderate positive feedbacks.

These feedback locations and signs likely reflect physical processes relevant to the last 24 years of the observational record. However, finer details like the specific locations included in sparse maps and their relative magnitudes are sensitive to the methodology, as implied Figure 2.4. This makes the sensitivity maps in Figure 2.6 qualitatively descriptive of the observations, but not necessarily effective at prediction outside the 24-year observational window. Figures 2.4 and 2.5 also imply that the spatial patterns in these specific 24 years may have little in common with the hypothetical equivalent of our 1,000-year example (i.e. the sensitivity map that would explain the most variance in any reachable pattern of internal variability). Taken together, we are comfortable asserting that the maps in Figure 2.6 characterize the feedbacks we have observed in the last 24 years, but not that they will be predictive of the internal variability in other possible time periods past or future.

2.3 Discussion

Predictive ability is a convenient way to establish credibility, but the sensitivity map framework also suggests physical processes that can explain the pattern effect. We contribute to GCM process understanding already inferred from previous Green's function experiments by interpreting these sensitivity maps, and we characterize the feedbacks specific to 2001-2024 using the same method.

Revisiting the Green's function approach, the major features consistent with the mechanistic understanding of Dong et al. (2019) show up in the MPI-ESM-1.2 sensitivity maps. The previously applied interpretation to the strongly negative WPWP feedback is that warm surface anomalies extend via deep convection and the weak temperature gradient (WTG) to set up conditions in regions of descent for strong inversions (e.g., Williams et al., 2023), and thus favorable conditions for globally-cooling low marine clouds (e.g., Andrews and Webb, 2018). This links a globally net negative feedback to warming in

regions of deep convection, and we find that the major tropical regions of negative feedback roughly correspond to areas of deep convection in Figure 2.2. In line with that interpretation, warming the surface in regions of descent decreases the local inversion strength, weakening the conditions for low marine clouds and decreasing global albedo (e.g., Wood and Bretherton, 2006). We would therefore expect those regions of marine low clouds to contribute positively to the global feedback. The western South American coast, the western coast of Australia, and the Namibian coast are all regions of marine low clouds that are consistently positive in Figure 2.2. The California coast is another such region, but has a less pronounced or absent positive feedback in the MPI-ESM sensitivity maps.

Broadly speaking, the observationally-derived sensitivity maps echo these features that seem to agree with previous Green's function studies. Deep-convective regions have strongly negative feedbacks and subtropical highs have positive feedbacks, which aligns with the first-order understanding gained from causal studies (e.g., Bloch-Johnson et al., 2023; Dong et al., 2019). The WPWP is one predominant region of negative feedback, with branches extending into the convection regions of the ITCZ and SPCZ. The observationally-driven regressions are also consistently negative over equatorial land areas, with coherent regions of negative feedback over the Amazonian and Congo basins. As expected, the subtropical highs in the observations are areas of positive feedback, with particular strength in the Southeast Pacific.

We also find features that challenge expectations from Green's function experiments. The equatorial Pacific band of positive feedback — found across MPI-ESM sensitivity maps and seen in some observational maps (Figures 2.6a and 2.6b) — is strongly reminiscent of the ENSO anomaly in the equatorial cold tongue. This is consistent with findings from Ceppi and Fueglistaler (2021), which suggest that ENSO and the TOA radiative imbalance interact in a two-way coupling. They find evidence in observations and models for a positive cloud radiative effect through shortwave effects leading the warm phase of ENSO, along with a negative planck response during and following the warm peak.

In the subtropics and extratropics, we note evidence for an alternating pattern of positive and negative feedbacks emanating poleward and eastward from the west-central Pacific. This is best seen in MCA maps (Figures 2.2a, and 2.6a), but is also recognizable amidst the noisier ridge regression maps (Figures 2.2b and 2.6b). The equatorial positive feedback band is flanked by the negative feedbacks we have postulated to emerge from deep convection in the ITCZ and SPCZ. Taken collectively, the pattern across the tropical Pacific visually resembles the expected circulation response to a warming in the center of the basin (Gill, 1980; Matsuno, 1966), which in this case may be driven by warm ENSO anomalies in the Central Pacific. From the center of the Matsuno-Gill pattern, we can interpret the alternating pattern to be wavelike in nature, in line with an extratropical temperature response via the propagation of Rossby waves (e.g., Branstator, 1985; Hoskins and Karoly, 1981). These patterns resemble the temperature response to ENSO in the wave-mediated atmospheric bridge (Alexander et al., 2002). We cannot definitively assert whether these features are merely correlated with an underlying driver of global radiation like ENSO, or if temperature forcings at these locations have a causal effect to amplify and dampen the global response to temperature fluctuations in these regions.

Chapter 3: Comparative analysis between regularization methods and the contemporary state-of-the-field methods

3.1 Introduction

In this Chapter, we bring the methods developed in Chapter 2 into context with the current state of the field. A number of methods are actively being pursued that both predict TOA radiation from the spatial temperature field and produce physically interpretable sensitivity maps. Through a collaborative effort, we are able to ensure the fairest possible comparison by training all methods on the same set of data and performing identical predictive tests. This is the first such effort to unify pattern effect sensitivity methods.

3.2 Data and Methods

3.2.1 Data sources

As in Chapter 2, we use a millennial-length pre-industrial control (piControl) from the LongRunMIP project (Rugenstein et al., 2019) to provide an ample supply of training data. All data come from MPI-ESM-1.2 (Max Planck Institute Earth System Model, v1.2; Mauritsen et al., 2019; Rohrschneider et al., 2019), which is the same model as used in Chapter 2. We choose this model for the same advantages previously discussed; the piControl simulation is relatively long at 1,237 years, and we have access to a Green's function produced from its atmospheric component, which is especially relevant for comparative analysis (Alessi and Rugenstein, 2023).

Unlike in Chapter 2, we only reserve 100 years of piControl to serve as an out-of-sample internal variability test. This is to accommodate methods that improve sensitively to marginal increases in training data while preserving a long enough testing set to capture any multidecadal oscillations. We also change how we divide the data. Some methods use information from lagged relationships in their projections, so we can no longer treat

each year as an independent measurement. Because both the training and test sets must be continuous, we choose the first 100 years to be the test set.

In addition to the internal variability test, we also include a warming scenario test set in which methods are given the temporally evolving temperature field for a 4xCO_2 step forcing simulation. This is a 1,000-year simulation in MPI-ESM-1.2 where CO_2 is instantaneously quadrupled from pre-industrial conditions, also performed under the auspices of LongRunMIP.

3.2.2 Green's function

We include the Green's function in this comparison to show how the emerging methods we consider compare to the relatively more established protocol. As discussed in Section 2.1, the Green's function combines the equilibrated response in atmospheric climate models to patch perturbations, allowing one to create the sensitivity map of radiation forced by sea surface temperature (SST; e.g., Barsugli and Sardeshmukh, 2002; Bloch-Johnson et al., 2023; Dong et al., 2019; Zhou et al., 2017). A simulation with a particular spatial time series of SSTs acts as a control, then a number of experiments superimpose single patches of anomalous warming or cooling for ten years or longer. The difference between the forced simulations and the control simulation represents the partial contribution from a given local SST perturbation to an atmospheric model's radiative response to that perturbation. The resulting sensitivity map indicates whether a warm anomaly at a given sea surface location will have a positive or negative net effect on global radiation. While Green's functions causally link radiation to SSTs by construction, the isolated patch perturbations result in temperature fields that do not naturally emerge. Because there is no way to enforce these SST patterns outside of a model, there is also no way to create an analogous function for the observable Earth system. Additionally, Green's functions represent an equilibrated response to a sustained patch anomaly, so they are primarily relevant to the forced pattern of warming.

3.2.3 Maximum covariance analysis

The first emerging method we consider in the comparison is maximum covariance analysis (MCA). We perform MCA slightly differently in this Chapter relative to how it was presented in Chapter 2. For the comparative analysis, we follow the procedure employed by Thompson et al. (2025, in review). We do not regrid the data from the horizontal resolution as retrieved, 1.875°x1.875°. We also do not standardize the data, which leads to stronger magnitudes in high-variance temperature regions (see Figure 2.1). This is a more direct application of the method as presented in Bretherton et al. (1992). MCA predictions in the test sections are performed as is Chapter 2, by projecting the sensitivity map onto the spatial temperature fields.

3.2.4 Ordinary least squares multilinear regression

Following Bloch-Johnson et al. (2020), we employ ordinary least squares (OLS) multilinear regression (MLR) as one of the considered methods. We first regrid (with conservative remapping to preserve spatial means) the temperature data in both the training and test sets to an extremely low horizontal resolution 15°x15° grid. Because this grid is rectilinear, we then apply latitude cosine weighting to the predictors, i.e. the temperatures at each gridpoint.

The ordinary least squares solution as defined in Section 2.1, reproduced here, minimizes the squared error between predicted and true R:

$$\hat{\beta} = \arg\min_{\beta} \left(||y - X\beta||_2^2 \right). \tag{3.1}$$

In Equation 3.1, y is the true R in the training data set, X is the matrix of local temperature time series ($T_i(t)$), β is a candidate vector of predictor coefficients, and $\hat{\beta}$ is the vector of coefficients chosen by the optimization. The least squares solution can also be found as:

$$\hat{\beta} = \left(X^T X \right)^{-1} X^T y, \tag{3.2}$$

which is how we calculate the OLS MLR sensitivity map. Our test predictions from this method are also found by projecting the sensitivity map onto the spatial temperature fields in the test data.

3.2.5 Regularized regression

We apply the regularized regression methods (ridge regression, LASSO regression, and elastic net regression) exactly as described in Chapter 2. Both because the number of years in the training set has changed, and because the division between train and test is different, the regularized regression results do not perfectly match the sensitivity maps and correlations in Section 2.2, though they are nearly identical.

3.2.6 Principle component regression

Principle component (PC) regression addresses the dimensionality problem by computing the empirical orthogonal functions (EOFs) of the spatial temperature field and treating these, rather than individual gridpoint locations, as predictors. Analogous to Equation 2.1, this linearization can be written:

$$R = \beta_1 PC_1 + \beta_2 PC_2 ... \beta_n PC_n, \tag{3.3}$$

where PC_i is the principle component corresponding to the ith EOF, and β_i is the regression coefficient. These coefficients are selected using ordinary least squares. The EOFs here are constructed from the concatenated spatiotemporal temperature fields from the training set, internal variability test set, and 4xCO_2 test set. This is done because the spatial behavior of temperature in the test cases must be included for the EOFs to have meaning relative to them. Once the EOFs have been constructed from temperature data alone, the coefficients β_i in Equation 3.3 are fit using R only from the training set. This maintains the requirement that the training method not have any knowledge of the truth in the test sets; it merely incorporates information from the test into the dimensionality reduction step. Because

the temperature anomalies in the 4xCO_2 are much larger than the anomalies in internal variability, these are reduced by a factor of 10 for the EOF determination step.

The advantage to using EOFs rather than individual gridpoints is that the first EOF will explain the most variance in the temperature field of any possible map, the second the second most, etc. In a system with m gridpoint locations, it would take m EOFs to completely recreate the original data. However, the vast majority of the variance in the original data can be recreated with far fewer EOFs. In this case, 100 EOFs explain approximately 90% of the variance, so we chose values near 100.

We also smooth the data by introducing a running mean, which means our system has two free hyperparameters: 1) the number of EOFs included and 2) the number of years included in the running mean. Rather than optimize these parameters, we implement the model with 4 variations in each and consider this the methodological uncertainty in PC regression. We consider including 80, 90, 100, and 110 EOFs, and we fit each of those options against running means of 40, 80, 120, and 160 years. The mean prediction across all 16 combinations defines this method's nominal prediction for each test. We calculate spatial sensitivity maps as the linear combination of the β_i coefficients and the EOFs, and we use the mean map as the nominal sensitivity map. Because the system is fully linear, one could use either this mean sensitivity map projected onto spatial temperature or the full PC regression formulation of Equation 3.3 to calculate the predicted R.

3.2.7 Partial least squares

Another method frequently applied to ill-posed multilinear regression models is partial least squares (PLS). This approach can be seen as an extension of MCA with results that closely resemble ridge regression. As with MCA, it relates the covariance of two fields to one another. Using the notation of Equation 3.1, our case relates the $(space\ x\ time)$ matrix of temperature anomalies (X) and the $(1\ x\ time)$ matrix of global radiation anomalies (y).

In MCA, we identify the pattern in X that explains the most variance in y, which becomes our sensitivity map.

In PLS, we then remove the influence of that pattern from the data X and y by subtracting the regressed projections. MCA can then be performed on the residuals to find a second component pattern representing the covariance not captured in the first MCA pattern. This process can be repeated any number of times. Because of this connection, MCA is sometimes referred to as "PLS1" regression, i.e. PLS regression keeping only the first component pattern. This is a broad description of the process, and multiple algorithms exist to remove the influences of the covariance patterns from the data (Wegelin, 2000). For simplicity and reproducibility, we use the default implementation of PLS regression in the scikit-learn python package (Pedregosa et al., 2011).

As with the EOF analysis, we must choose how many components to include. We randomly select 20% of the training set to serve as a validation subset and perform PLS from 1 to 200 components on the remaining 80%. We find a clear minimum error from the validation subset at 12 components retained. We then apply PLS regression with 12 components to the full training set. The sensitivity map represents the combined procedure from all 12 components, and R predictions are found by projecting this sensitivity map onto spatial temperature anomalies.

3.2.8 Fluctuation-dissipation relation

The methods to this point have all been various ways to identify a single sensitivity map which can be projected onto spatial temperature anomalies to predict R. This is identical to how predictions are made with the Green's function sensitivity map, making it the most comparable interpretation. However, some methods incorporate additional information from the temperature field, so do not fit cleanly into the sensitivity map framework as we have defined it. Nonetheless, these methods fall within our comparison parameters; they can be trained entirely from internal variability in surface temperature and TOA radiation,

and they can predict global radiation in out-of-sample cases given only the temperature field.

FDR-coupled

The first such method we consider is the fully-coupled fluctuation-dissipation relation (FDR) approach presented in (Falasca et al., 2024a). This approach defines a state matrix ($S(t)$) that includes both the spatiotemporal fields of surface temperature (S_{TS}) and TOA radiation (S_{TOA}). I.e.,

$$S(t) = \big[S_{TS},\ S_{TOA}\big](t). \tag{3.4}$$

The fluctuation-dissipation framework specifically predicts the average response in $S(t)$ across an ensemble ($\delta\langle S(t)\rangle$) to an external perturbation ($\delta f(t)$). Because we have a long piControl, we can divide it into independent realizations of shorter time periods to define our "ensemble" for training. When we apply FDR to a test section, we are predicting the mean response to a hypothetical ensemble of the test section. The forcing $f(t)$ has the same shape as the state matrix $S(t)$, so we use temperature anomalies for perturbations in S_{TS} and zeros for anomalies in S_{TOA}. The response is then calculated using the convolution:

$$\delta\langle S(t)\rangle = \int_0^t G(\tau)\delta f(t-\tau)d\tau. \tag{3.5}$$

In Equation 3.5, $G(t)$ is a response function derived from the training set of the piControl. For details on how it has been approximated for this application, see Falasca et al. (2024b). The state vector $S(t)$ evolves according to Equation 3.5 and the supplied temperature anomalies ($f(t)$). In the fully-coupled formulation, this has the unique property among our methods to allow surface temperatures to respond to the forcing. Temperature anomalies are treated as external forcings rather than prescribed surface temperature in the test cases. Spatial TOA radiation also evolves in the state matrix, which can then be globally averaged

to predict R. These predictions depend on lagged temperature relationships to predict R, though we expect more recent temperatures to have a stronger influence. We limit the influence of prior time periods by defining a critical timescale τ_∞ at which the convolution goes to zero. In the coupled formulation, we use a critical timescale of 10 years. We use monthly mean anomalies rather than yearly anomalies when training $G(t)$, but we use yearly temperatures for $f(t)$ because this more closely represents an external forcing.

Because predictions emerge from an ever-developing state matrix, the concept of a sensitivity map is harder to define. We apply a similar logic to the construction of the Green's function to represent the cumulative influence with $\tau_\infty = 10$ years from a given forcing. We do so by applying a 1 K forcing to each spatial location separately and noting the equilibrated response in global radiation once $S(t)$ has stabilized. This gives us the causal change in global radiation due to a change in local temperature, $\partial R/\partial T_i$. Though not used like the Green's function map for prediction, this is the closest process to determine a sensitivity map to that used in the Green's function method. These sensitivity maps are also less spatially resolved than the others considered here. To reduce the dimensionality of the spatial field, a community detection method (see Falasca et al., 2024a,b) is used to group gridpoint locations into large contiguous regions. Each of these regions makes up one spatial location.

FDR-atmospheric

Though it does not predict radiation in the same manner as the Green's function, the coupled FDR method does provide the nearest analogy to how the Green's function is produced. This raises the possibility to use FDR as an emulator for Green's function simulations and make a more directly comparable sensitivity map. We refer to this as the atmospheric version of FDR because the method is only used to generate a Green's function-style sensitivity map, which can then be projected onto temperature anomalies. There is

no evolution of the state variable $S(t)$ involved in the prediction. Surface temperature is not able to respond to anomalies, hence the distinction from the coupled version.

Because this approach predicts R from the concurrent temperature field, we develop the atmospheric FDR sensitivity map with $\tau_\infty = 1$ month to capture fast-response feedbacks. We create the sensitivity map as with the coupled version, using the equilibrated response to 1 K perturbations at each location. We then predict R by projecting this map onto temperature anomalies in both test cases.

3.2.9 Convolutional neural net

The convolutional neural net (CNN; see LeCun et al., 2015) is another method that does not generate a conventional sensitivity map. This is because CNNs are capable of using nonlinearities in the spatial field of predictors to improve their predictions. CNNs are particularly useful in identifying spatial patterns, making them appropriate for this task. We train the CNN by feeding it the spatial patterns of temperature in the piControl training set, and we optimize for ability to predict R in the training set. For further details on the particular procedure used for this problem, see Rugenstein et al. (2025, in review).

The concept of a sensitivity map does not apply well to this method because the connections identified by the CNN are nonlinear. The sensitivity of R to temperature at a given location also depends on the rest of the temperature field, so there is effectively a different sensitivity map for every unique spatial pattern of temperature anomalies. It is possible to identify the sensitivity of a single prediction to a single predictor using eXplainable artificial intelligence (XAI; e.g., Mamalakis et al., 2022), so it is possible to find the sensitivity map that corresponds to each temperature pattern $T_i(t)$. We can approximate the mean behavior of the CNN by averaging all single-instance sensitivity maps across the testing period. CNN predictions employ the weights and nonlinearities learned from training, so cannot be approximated by projecting this sensitivity map onto temperature anomalies.

3.3 Results

In Figures 3.1 and 3.2 we plot the sensitivity map for each method along with the time series predictions for both the internal variability test and the step forcing 4xCO_2 test. We score performance in the internal variability test by how much variance can be explained (i.e., R^2) in the true simulated R by the model-predicted R. As we did in Chapter 2, we also calculate the global internal variability feedback (λ_I) by regressing global radiation onto global temperature in the training set to serve as a null hypothesis for the pattern effect. The particular division of training and testing data in this Section 3.2 result in global λ_I explaining 42% of the variance in the out-of-sample test. This suggests that the test set in this case is more representative of the full piControl than in Chapter2.

In order from lowest to highest variance explained, the methods are:

- the null hypothesis (R^2=0.42),
- the Green's function (R^2=0.53),
- MCA (R^2=0.55),
- atmosphere-only FDR (R^2=0.57),
- fully-coupled FDR (R^2=0.69),
- PC regression (R^2=0.71),
- OLS multilinear regression (R^2=0.76),
- CNN (R^2=0.80),
- PLS (R^2=0.80),
- LASSO regression (R^2=0.81),
- elastic net regression (R^2=0.82), and
- ridge regression (R^2=0.82).

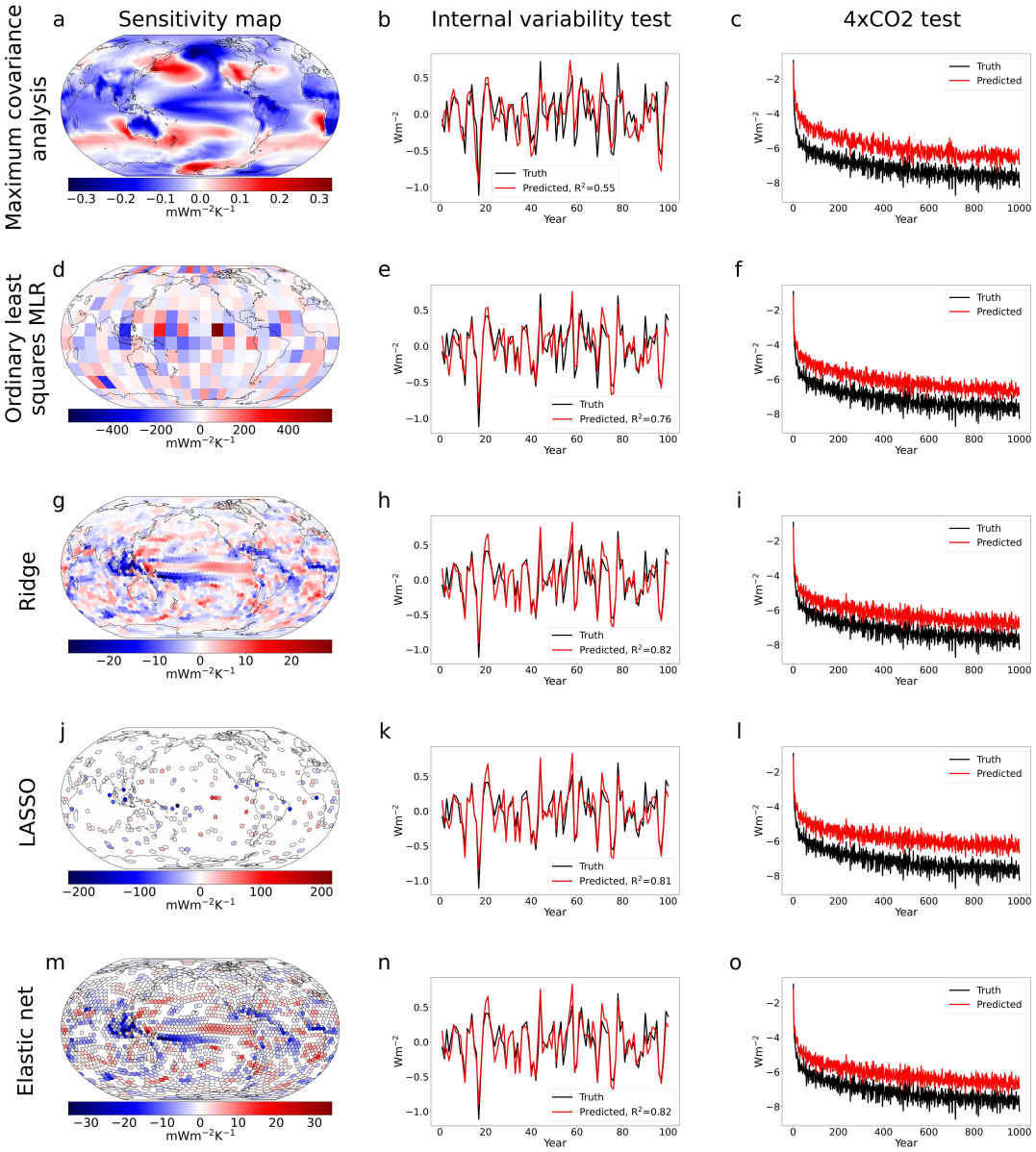

Figure 3.1: Comparison, part 1. Sensitivity map, internal variability prediction results, and 4xCO_2 prediction results, respectively, for a,b,c) maximum covariance analysis, d,e,f) ordinary least squares multilinear regression, g,h,i) ridge regression, j,k,l) LASSO regression, and m,n,o) elastic net regression. Red lines indicate predicted radiative time series, black lines represent the simulated truth. Note that colorbars are not consistent across sensitivity maps.

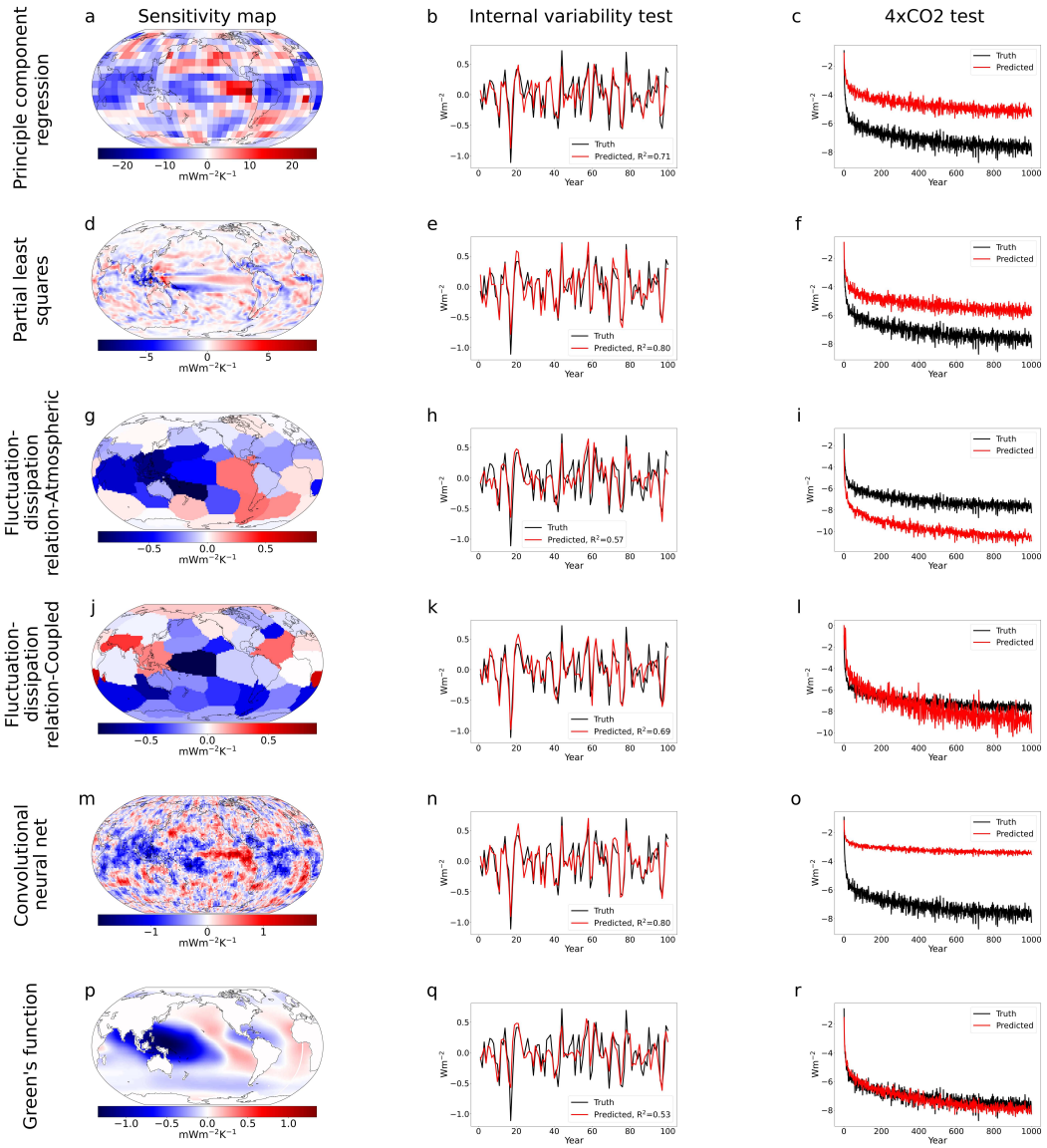

Figure 3.2: Comparison, part 2. Sensitivity map, internal variability prediction results, and 4xCO_2 prediction results, respectively, for a,b,c) principle component regression, d,e,f) partial least squares regression, g,h,i) atmosphere-only fluctuation-dissipation relation, j,k,l) fully-coupled fluctuation-dissipation relation, m,n,o) convolutional neural net, and p,q,r) Green's function. Red lines indicate predicted radiative time series, black lines represent the simulated truth. Note that colorbars are not consistent across sensitivity maps. The sensitivity maps for fully-coupled FDR (g) and the CNN (m) do not directly project onto temperature anomalies to predict radiation and are instead approximately representative of the full prediction method, see text.

We see an improvement in the regularized regression methods relative to Chapter 2, further evidence that the test case may better represent the full piControl. These performances are still in line with the idea that the maximum amount of variance in global radiation explainable by spatial temperature is approximately 80%.

Variance explained would not be a good metric for the step forcing test because variance in the truth is overwhelmingly dominated by the long-trend negative curve. Any prediction that also swings in a similar manner, regardless of magnitude, would explain nearly all the variance in R. This raises the question of what metric is most appropriate to evaluate predictive performance. For internal variability, we prefer R^2 over error measurements because on interannual timescales we expect the true causal connections to be represented in the signs and relative magnitudes of anomalous swings in radiation. We care less that the absolute magnitude for some anomalies may be wrong because we know that a decent portion (perhaps 20% as our results suggest) of interannual variations in R come from sources other than surface temperature. In the forced scenario, however, we do expect much more of the forced response in radiation to be captured in the forced temperature response. Both are strong signals driven directly by the same known external perturbation. Unlike in internal variability, the magnitude of the predicted response is of primary interest, and an underestimate or overestimate would significantly change our estimate of the Earth's sensitivity to climate change. To quantify this feature, we report the percentage of overall response captured, as measured from the last 30 years of the predicted and simulated radiative time series.

From underprediction to overprediction:

- CNN (44%),
- PC regression (67%),
- PLS (74%),
- LASSO regression (82%),

- MCA (85%),
- elastic net regression (87%),
- OLS multilinear regression (87%),
- ridge regression (88%),
- the Green's function (104%),
- the null hypothesis (110%),
- fully-coupled FDR (115%), and
- atmosphere-only FDR (137%).

We note that the Green's function employed here is not the raw Green's function from simulation outputs, but rather a version that has been "calibrated" in anticipation of predicting the forced response. We therefore are not surprised that the Green's function prediction very nearly matches the forced response. The null hypothesis performs particularly well at predicting the forced response, though from Figure 4 in Bloch-Johnson et al. (2020) we see that this is a model-dependent coincidence specific to MPI-ESM-1.2.

3.4 Discussion

3.4.1 Green's function

The Green's function sensitivity map (Figure 3.2p) is broadly consistent with Green's function maps from other models (Bloch-Johnson et al., 2023), and from this aligns with the proposed physical connection we discuss in Section 2.3. Namely, deep convection and the weak temperature gradient (WTG) set up conditions in regions of descent for strong inversions, which in turn leads to conditions favorable for low maritime clouds. These clouds have a strongly negative cloud radiative effect, and thus contribute to an overall negative radiative anomaly. Similarly, local warming in areas of descent would disrupt

these conditions, contributing to a positive overall feedback. The signature of this possible explanation, negative feedbacks in convective regions and positive feedbacks in subsidence regions, is clearly present in the Green's function sensitivity map. The Green's function does the best at predicting the forced response, in part by construction, but does the worst among the spatial methods at predicting internal variability.

The primary advantage of the Green's function is that it is inherently causal. By perturbing SST directly and calculating the response, we can be more certain that the formulation $\partial R/\partial T_i$ applies. We also have the ability to trace potential physical explanations by exploring individual patch perturbation simulations. We can also apply Green's functions to other variables that may be forced by temperature (e.g., precipitation; Alessi and Rugenstein, 2023).

These advantages emerge from the process of performing experiments in atmospheric climate models, which explicitly model atmospheric processes. However, this experimental setup also limits the Green's function's applicability. To identify the forced signal, large magnitude patch perturbations are used that create unrealistic temperature patterns. How the system responds to these patches in isolation may not have relevance to real spatial patterns of warming. Green's functions are also computationally expensive to run, and are currently limited to SST forcing rather than globally resolved surface temperature. There is also no way to create a Green's function from the observational record because there is not way to isolate a single sustained patch anomaly. Their explanatory power is therefore limited to the accuracy of climate model physics.

3.4.2 Maximum covariance analysis

Though on a different grid, the MCA sensitivity map (Figure 3.1a) is qualitatively the same as the unstandardized MCA map in Figure 2.1a. The magnitudes in high variance temperature regions are more pronounced in this map than in the standardized MCA sensitivity map discussed in detail in Chapter 2, but the locations of the positive and

negative feedbacks are essentially identical. In this comparison, MCA explains 55% of out-of-sample variance in R, which is higher than the 42% from the MCA map in Figure 2.2a, but because we found that MCA performed better after standardization we believe this is due to the different division between training and testing rather than in procedure.

The advantage to MCA is its simplicity both in application and interpretation. It is closely related to pointwise linear regression (see Thompson et al., 2025, in review), and sensitivity maps derived from MCA tend to have a high degree of spatial cohesiveness. We subjectively organize Figures 3.1 and 3.2 from the simplest method to the most complex, and in the same sense that the largest advantages and disadvantages to the Green's function both come from its complexity, the main disadvantage to MCA is also its simplicity. MCA does not utilize much of the information available in the training data. Because MCA is the same pattern as both ridge regression with $\alpha_2 >> 1$ and PLS with only one component retained, we know from cross-validation in both those methods that the MCA pattern is not the most informative. If it were, our cross-validation would have yielded the same results for MCA, PLS, and ridge. In fact, we see this in the observational sensitivity maps (Figure 2.6). The cross-validated ridge map is very similar to the MCA map because there is much less information to be gained from the 24-year training data set relative to the 1,000-year training data set.

3.4.3 Ordinary least squares multilinear regression

Ordinary least squares MLR is also a relatively simple implementation, as shown in Section 3.2. The objective of OLS MLR is to prioritize goodness of fit in the training data. As discussed in Chapter 2, this comes with a tendency to overfit to the training data at the cost of out-of-sample applicability. OLS still performs reasonably well at out-of-sample prediction, performing in the upper half of methods. It also predicts 87% of the forced response, in line with the other regression-based methods. However, the disadvantage from fitting noise affects not only predictive skill, but also interpretability. The sensitivity

map (Figure 3.1d) is incredibly noisy, with a number of locations where a strong positive and negative feedback are situated adjacent to one another. Some feedbacks are in line with expectations, for example a generally negative WPWP feedback and a positive feedback off the Peruvian coast. But these features would not stand out without prior expectations, and any given gridpoint feedback is difficult to trust in isolation. The predictive skill tells us that this map is identifying a real signal, but the noise makes it nearly impossible to interpret.

3.4.4 Regularized regression

The regularized regression results are perfectly in line with the results from Chapter 2. As noted, the LASSO sensitivity map (Figure 3.1j), elastic net sensitivity map (Figure 3.1m), and ridge sensitivity map (Figure 3.1g) are slightly different from the maps shown in Figure 2.2 due to the division of training and testing sets. However, the feedbacks implied by each map are essentially identical, demonstrating that these methods characterize the full piControl system well when given millennial-length training data. All three methods perform similarly. They have average performance in the 4xCO_2 test, predicting about 87% of the response, but perform the best at the out-of-sample internal variability test, explaining about 82% of the variance in R.

Relative to the other methods considered, regularized regression is still a simple approach. The complications introduced are largely via the hyperparameters α_1 and α_2, which makes the solution more dependent on protocol than MCA or OLS. These methods can be performed relatively quickly, though do take more time than MCA and OLS on large training sets. The relative merits of ridge, LASSO, and elastic net are discussed in detail in Chapter 2.

3.4.5 Principle component regression

The PC regression sensitivity map (Figure 3.2a) is consistent with the Green's function results in general form, with negative feedback in regions of tropical ascent and positivie

feedbacks in the extratropics and regions of subsidence. The map also includes some of the equatorial positive feedback band flanked by negative feedbacks that we identify in the regularized regression maps. While qualitatively similar to many of the other methods, this approach underperforms in predictive skill at both tests.

The advantages of this method are the relatively fast implementation and dimensionality reduction. By using EOFs, the number of predictors is greatly reduced while maintaining most of the information in the temperature field. However, this approach is limited by the necessity to define EOFs that are relevant to both the training and testing data. In this case, we address the problem by including the test case temperature data in the EOF calculation, but this has the potential to skew the results based on the relative length of training and testing data sets. It also somewhat complicates the sensitivity map's general applicability because it contains information specific to the particular intended test set. This implementation also uses a relatively low horizontal resolution grid which may limit its ability to identify relevant small-scale features. These complications may explain why this method has less predictive skill than methods finding qualitatively similar spatial feedbacks.

3.4.6 Partial least squares

The PLS sensitivity map (Figure 3.2d) is qualitatively nearly identical to that of ridge regression (Figure 3.1g). It performs comparably to regularized regression against internal variability, though does not capture as much of the forced response. One unique aspect of our ridge regression implementation is the calibration step (Equation 2.5), which ensures the ridge sensitivity map does not under- or overestimate the magnitude of large anomalies in the training set. This may make our method better at approaching the significant anomalies in the forced response.

PLS also has similar advantages to ridge regression. The method can tolerate a large number of predictors relative to observations, which allows for much more fine-scale

detail than several of the lower resolution sensitivity maps. Cross-validation is easier than with ridge regression because the PLS hyperparameter requires checking relatively few integers while the optimal α_2 parameter in ridge regression can vary by orders of magnitude depending on the dimensions of the training data.

The iterative application of MCA makes this method marginally more complicated than those prior to it in Figure 3.1 and 3.2, though it is still relatively easy to conceptualize. As with ridge, spatial heterogeneity in the sensitivity map complicates the physical interpretation.

3.4.7 Fluctuation-dissipation relation

The sensitivity map for the atmospheric version of FDR (Figure 3.2g) is broadly in line with most of the other maps. The WPWP and surrounding regions are strongly negative, subsidence regions in the tropical southern hemisphere are positive, and there is even a hint at an equatorial positive feedback. The clustered presentation smooths out the finer-scale details, making it difficult to identify or contrast with features that appear contiguous and correlated in other maps. This map, which is applied as a projection onto temperature anomalies, does not perform particularly well at either test. The main advantage is that, unlike regression- or covariance-based methods, this approach creates a sensitivity map in the same manner as the Green's function. We apply patch perturbations and allow the FDR state matrix, essentially a simple climate model emulator, to equilibrate. The connections are therefore causal, making it more appropriate to assume the sensitivity map represents $\partial R/\partial T_i$. This method is much faster and computationally cheaper than the Green's function, making it much more flexible to apply.

The fully-coupled version of FDR produces a sensitivity map that stands alone among all methods considered (Figure 3.2j). The spatial feedbacks in this map not only differ from all other maps, but also directly disagree in sign for the most critical locations we have identified. In this map, the WPWP contributes a positive feedback, and regions of subtropical

descent have negative feedbacks. Nonetheless, this method performs reasonably well in both tests. The caveat to this sensitivity map is that it is one of two (along with the CNN) that is not directly projected onto the spatial temperature anomalies. The Green's function style patch perturbation used to produce this map shows how sustained anomalies might affect the equilibrated response, but the predictions in Figures 3.2k and 3.2l are generated within the FDR protocol, forced by the temporally evolving temperature field.

This complexity makes the method harder to interpret, but also allows the system to respond as a whole to the forcing. Falasca et al. (2024a) propose that these fully-coupled sensitivity maps integrate complex feedbacks on longer timescales stretching back to the limit set by τ_∞. At these timescales, the role of the ocean becomes important, which cannot be captured by any of the other methods, including the Green's function. The relatively mediocre performance in predictive skill is balanced by the unique ability to integrate feedbacks across timescales. The assumption that the piControl can be treated as an ensemble also requires a fairly large training data set, which is another disadvantage to this method. It is unlikely that the 24 years available from the observations would be sufficient.

3.4.8 Convolutional neural net

CNNs are the most complicated of the emerging methods we compare to the Green's function, and the sensitivity map in Figure 3.2m is correspondingly difficult to interpret. CNNs introduce a second level to the pattern effect because spatial feedbacks are themselves dependent on the pattern. The sensitivity map shown is the average approximation across the internal variability test, but the map for a given year or across a different test would look different. To illustrate this, we also show the average sensitivity map across the 4xCO_2 test in Figure 3.3.

Given enough training data, CNNs can be extremely effective at prediction. The CNN performs about as well as any other method at the internal variability test. However, the

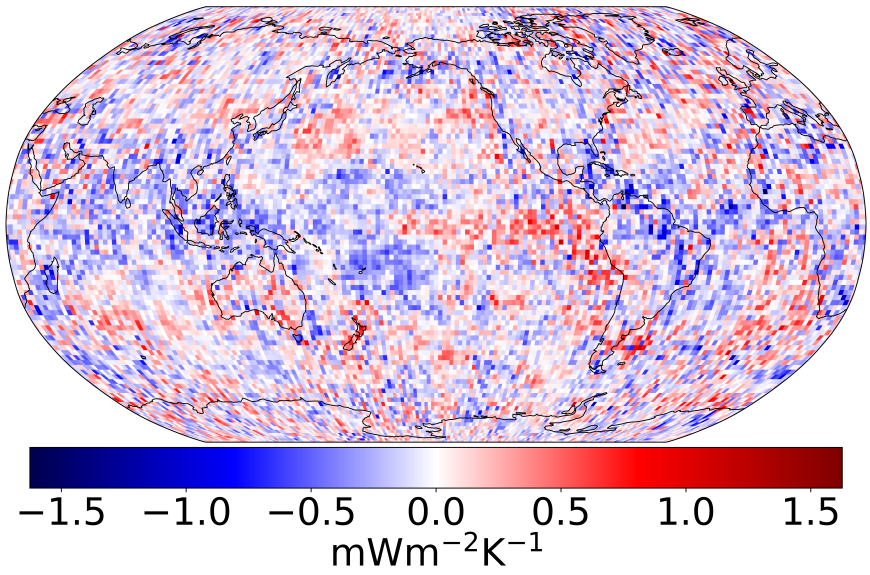

Figure 3.3: Convolutional neural net sensitivity map formed from the average XAI sensitivity at each time step in the 4xCO_2 test time series. This is an equally valid alternative to the CNN sensitivity map representation in Figure 3.2m.

CNN is the worst by far at predicting the forced response in the 4xCO_2 test. This is contrary to the results in Rugenstein et al. (2025, in review) who find that a CNN trained on internal variability does well at predicting the forced response. There are a number of reasons to explain this disconnect, which speaks to the complexity and sensitivity behind this method. The biggest difference is an order of magnitude in training data. Whereas the CNN in this comparison trained on 1,000 years, the CNN in Rugenstein et al. (2025, in review) trained on 18,400 years. Neural networks are very sensitive to data set size, so this could explain much of the difference. Our CNN also trained on piControl internal variability, while theirs trained on detrended warming scenario ensembles. It is possible certain patterns of internal variability do not occur in our piControl that do appear in the 4xCO_2 simulation, which would make the CNN's extrapolation task much harder. Lastly, architecture and randomness also play a role in CNN skill, and slightly different design choices may have affected this CNN's ability to predict forced anomalies.

As with FDR, the CNN needs a large training data set to effectively predict out-of-sample, so it is also a poor candidate for observational applications. These data-heavy, relatively

complex approaches are more suited to emulating or replacing the extensive simulations required by the Green's function. Though dependent on climate models, they contribute significantly to our understanding of the pattern effect and the connection between internal variability feedbacks and forced climate feedbacks. In a similar vein, (Kang et al., 2023) also approximate a Green's function from existing climate model simulations. We did not include that method in this comparison because it requires a forcing simulation rather than training exclusively on internal variability.

Chapter 4: Conclusion

In Chapter 2, we have presented a physically-interpretable sensitivity map that characterizes the feedback from the spatial surface temperature field to the global radiative response, both in an idealized climate model case and entirely based on observational data. In the MPI-ESM GCM, maps produced by this method can explain a large majority of the variance in out-of-sample tests if given millennium-length training data. From only 24 years, we do not find sensitivity maps to perform consistently at our predictive tests, but we believe they do capture the qualitative feedbacks relevant to the particular training time period. In that context, observational sensitivity maps derived from our methods can contribute to our process understanding of the pattern effect as observed.

We are left with the following key takeaways from Chapter 2:

- Regularized regression methods are effective at generating predictive internal variability sensitivity maps relating spatial temperature anomalies to global radiation anomalies when trained on a millenium-length GCM piControl simulation. This is seen in the 77% variance explained in out-of-sample R.

- Our GCM sensitivity maps are much more predictive than using the global feedback parameter to estimate global radiation from global temperature, which can only explain 26% of the variance out-of-sample.

- Our GCM sensitivity maps are much more predictive than using the global feedback parameter to estimate global radiation from global temperature, which can only explain 26% of the variance out-of-sample.

- Across several GCMs, these methods also result in a median predictive ability around 50% of out-of-sample variance explained even when the sample size is reduced by two orders of magnitude from 1,000 years to 24 years.

- However, the limited sample size results in maps that vary significantly both from training time period and training methodology. While they may not reliably perform well at predictive tests, they likely do reflect real physical processes that exist in the training data

- While the sensitivity maps derived from millennium-length internal variability do well at predicting interannual variations in an RCP 8.5 warming scenario, they consistently underestimate the magnitude of the forced response. Sensitivity maps derived from only 24 years overlap with the simulated truth, but with extreme uncertainty.

- Sensitivity maps derived from GCM piControl have the following major features: 1) a dominant negative feedback centered in the WPWP, 2) positive feedbacks in regions of subsidence/low marine clouds, 3) an equatorial band of positive feedback echoing the ENSO signature in the cold tongue, 4) bands of negative feedback flanking the positive equatorial band, roughly colocated with convergence zones, 5) an alternating pattern of positive and negative feedbacks extending eastward and poleward from the WPWP, hypothesized to be connected to the atmospheric bridge, and 6) negative feedbacks along the Atlantic coastlines of South America and the Gulf of Africa.

- Sensitivity maps derived from observations have the following major features: 1) a dominant negative feedback extending from the WPWP well into the Indian Ocean, 2) positive feedbacks in regions of subsidence/low marine clouds, especially strong in the Southeast Pacific, 3) a slight band of positive feedback in the equatorial Pacific, 4) bands of negative feedback flanking the equatorial Pacific, roughly colocated with convergence zones, 5) an alternating pattern of positive and negative feedbacks extending eastward and poleward from the WPWP, hypothesized to be connected to the atmospheric bridge, 6) negative feedback along the Atlantic coast of South America and over land areas of the Amazonian and Congo basins, and 7) a strong positive feedback in the North Atlantic.

In Chapter 3, we have presented our regularized regression methods alongside several other new methods to predict global radiative anomalies from spatial temperature anomalies. We also compare to the Green's function to show how these methods perform at both an internal variability test and a 4xCO_2 step forcing test relative to the previously dominant method.

We are left with the following key takeaways from Chapter 3:

- All spatial feedback approaches, including the Green's function, improve upon the global feedback null hypothesis in predicting internal variability. All new methods considered improve upon the Green's function at this test.

- The null hypothesis and the Green's function perform the best at predicting the forced response in the 4xCO_2 test. However, we explain the former as a model-dependent coincidence and the latter as being reached somewhat by construction. The new models both over- and underestimate the magnitude of the forced response. The CNN performs the worst with a significant underestimation.

- All sensitivity maps, with the notable exception of fully-coupled FDR, generally show the canonical features expected from prior Green's function analysis. This includes negative feedbacks in tropical regions of ascent and positive feedbacks in tropical regions of subsidence. Most methods also show some hint of an equatorial postivie feedback band not found in the Green's function map.

- The fully-coupled FDR sensitivity map looks nothing like that of any other included method. We explain this as a feature from integrating complex feedbacks across longer timescales than any other method, with the unexpected pattern signaling the importance of the ocean-atmosphere coupling.

- Performance in the internal variability test does not predict performance in the 4xCO_2 test. We see a large avenue for future work relating the internal variability feedback

to the forced climate feedback via the pattern effect, motivated in a large part by this discrepancy.

Bibliography

Alessi, M. J. and Rugenstein, M. A. A. (2023). Surface temperature pattern scenarios suggest higher warming rates than current projections. *Geophys. Res. Lett.*, 50(23).

Alexander, M. A., Bladé, I., Newman, M., Lanzante, J. R., Lau, N.-C., and Scott, J. D. (2002). The atmospheric bridge: The influence of ENSO teleconnections on air–sea interaction over the global oceans. *J. Clim.*, 15(16):2205–2231.

Andrews, T., Bodas-Salcedo, A., Gregory, J. M., Dong, Y., Armour, K. C., Paynter, D., Lin, P., Modak, A., Mauritsen, T., Cole, J. N. S., Medeiros, B., Benedict, J. J., Douville, H., Roehrig, R., Koshiro, T., Kawai, H., Ogura, T., Dufresne, J.-L., Allan, R. P., and Liu, C. (2022). On the effect of historical SST patterns on radiative feedback. *J. Geophys. Res. D: Atmos.*, 127(18):e2022JD036675.

Andrews, T., Gregory, J. M., and Webb, M. J. (2015). The dependence of radiative forcing and feedback on evolving patterns of surface temperature change in climate models. *J. Clim.*, 28(4):1630–1648.

Andrews, T. and Webb, M. J. (2018). The dependence of global cloud and lapse rate feedbacks on the spatial structure of tropical pacific warming. *J. Clim.*, 31(2):641–654.

Barsugli, J. J. and Sardeshmukh, P. D. (2002). Global atmospheric sensitivity to tropical SST anomalies throughout the indo-pacific basin. *J. Clim.*, 15(23):3427–3442.

Bloch-Johnson, J., Rugenstein, M., and Abbot, D. S. (2020). Spatial radiative feedbacks from internal variability using multiple regression. *J. Clim.*, 33(10):4121–4140.

Bloch-Johnson, J., Rugenstein, M. A. A., Alessi, M. J., and others (2023). The green's function model intercomparison project (GFMIP) protocol. *Authorea*.

Branstator, G. (1985). Analysis of general circulation model sea-surface temperature anomaly simulations using a linear model. part I: Forced solutions. *J. Atmos. Sci.*, 42(21):2225–2241.

Bretherton, C. S., Smith, C., and Wallace, J. M. (1992). An intercomparison of methods for finding coupled patterns in climate data. *J. Clim.*, 5(6):541–560.

Carroll, R. J. and Ruppert, D. (1996). The use and misuse of orthogonal regression in linear errors-in-variables models. *Am. Stat.*, 50(1):1–6.

Ceppi, P. and Fueglistaler, S. (2021). The el niño–southern oscillation pattern effect. *Geophys. Res. Lett.*, 48(21).

Ceppi, P., Myers, T. A., Nowack, P., Wall, C. J., and Zelinka, M. D. (2024). Implications of a pervasive climate model bias for low-cloud feedback. *Geophys. Res. Lett.*, 51(20).

Ceppi, P. and Nowack, P. (2021). Observational evidence that cloud feedback amplifies global warming. *Proc. Natl. Acad. Sci. U. S. A.*, 118(30):e2026290118.

Colman, R. and Hanson, L. (2017). On the relative strength of radiative feedbacks under climate variability and change. *Clim. Dyn.*, 49(5-6):2115–2129.

Danabasoglu, G., Yeager, S. G., Kwon, Y.-O., Tribbia, J. J., Phillips, A. S., and Hurrell, J. W. (2012). Variability of the atlantic meridional overturning circulation in CCSM4. *J. Clim.*, 25(15):5153–5172.

Davis, L. L. B., Thompson, D. W. J., Rugenstein, M., and Birner, T. (2024). Links between internal variability and forced climate feedbacks: The importance of patterns of temperature variability and change. *Geophys. Res. Lett.*, 51(24):e2024GL112774.

Dessler, A. E. (2013). Observations of climate feedbacks over 2000–10 and comparisons to climate models. *J. Clim.*, 26(1):333–342.

Dessler, A. E. and Forster, P. M. (2018). An estimate of equilibrium climate sensitivity from interannual variability. *Journal of Geophysical Research: Atmospheres*, 123(16):8634–8645.

Dong, Y., Proistosescu, C., Armour, K. C., and Battisti, D. S. (2019). Attributing historical and future evolution of radiative feedbacks to regional warming patterns using a green's function approach: The preeminence of the western pacific. *J. Clim.*, 32(17):5471–5491.

Donner, L. J., Wyman, B. L., Hemler, R. S., Horowitz, L. W., Ming, Y., Zhao, M., Golaz, J.-C., Ginoux, P., Lin, S.-J., Schwarzkopf, M. D., Austin, J., Alaka, G., Cooke, W. F., Delworth, T. L., Freidenreich, S. M., Gordon, C. T., Griffies, S. M., Held, I. M., Hurlin, W. J., Klein, S. A., Knutson, T. R., Langenhorst, A. R., Lee, H.-C., Lin, Y., Magi, B. I., Malyshev, S. L., Milly, P. C. D., Naik, V., Nath, M. J., Pincus, R., Ploshay, J. J., Ramaswamy, V., Seman, C. J., Shevliakova, E., Sirutis, J. J., Stern, W. F., Stouffer, R. J., Wilson, R. J., Winton, M., Wittenberg, A. T., and Zeng, F. (2011). The dynamical core, physical parameterizations, and basic simulation characteristics of the atmospheric component AM3 of the GFDL global coupled model CM3. *J. Clim.*, 24(13):3484–3519.

Falasca, F., Basinski-Ferris, A., Zanna, L., and Zhao, M. (2024a). Diagnosing the pattern effect in the atmosphere-ocean coupled system through linear response theory. *arXiv* [*physics.ao-ph*].

Falasca, F., Perezhogin, P., and Zanna, L. (2024b). Data-driven dimensionality reduction and causal inference for spatiotemporal climate fields. *Phys Rev E*, 109(4-1):044202.

Frost, C. and Thompson, S. G. (2000). Correcting for regression dilution bias: comparison of methods for a single predictor variable. *J. R. Stat. Soc. Ser. A Stat. Soc.*, 163(2):173–189.

Gent, P. R., Danabasoglu, G., Donner, L. J., Holland, M. M., Hunke, E. C., Jayne, S. R., Lawrence, D. M., Neale, R. B., Rasch, P. J., Vertenstein, M., Worley, P. H., Yang, Z.-L., and Zhang, M. (2011). The community climate system model version 4. *J. Clim.*, 24(19):4973–4991.

Gill, A. E. (1980). Some simple solutions for heat-induced tropical circulation. *Q. J. R. Meteorol. Soc.*, 106(449):447–462.

Gregory, J. M., Andrews, T., Ceppi, P., Mauritsen, T., and Webb, M. J. (2020). How accurately can the climate sensitivity to co2 be estimated from historical climate change? *Clim. Dyn.*, 54(1-2):129–157.

Gregory, J. M., Ingram, W. J., Palmer, M. A., Jones, G. S., Stott, P. A., Thorpe, R. B., Lowe, J. A., Johns, T. C., and Williams, K. D. (2004). A new method for diagnosing radiative forcing and climate sensitivity. *Geophys. Res. Lett.*, 31(3).

Hall, A. and Qu, X. (2006). Using the current seasonal cycle to constrain snow albedo feedback in future climate change. *Geophysical Research Letters*, 33(3).

He, H., Kramer, R. J., and Soden, B. J. (2021). Evaluating observational constraints on intermodel spread in cloud, temperature, and humidity feedbacks. *Geophys. Res. Lett.*, 48(17).

Heinzeller, D., Duda, M. G., and Kunstmann, H. (2016). Towards convection-resolving, global atmospheric simulations with the model for prediction across scales (MPAS) v3.1: an extreme scaling experiment. *Geosci. Model Dev.*, 9(1):77–110.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Jean-Noël Thépaut (2020). The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, 146(730):1999–2049.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.

Hoskins, B. J. and Karoly, D. J. (1981). The steady linear response of a spherical atmosphere to thermal and orographic forcing. *J. Atmos. Sci.*, 38(6):1179–1196.

Kang, S. M., Ceppi, P., Yu, Y., and Kang, I.-S. (2023). Recent global climate feedback controlled by southern ocean cooling. *Nat. Geosci.*, 16(9):775–780.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.

Loeb, N. G., Doelling, D. R., Wang, H., Su, W., Nguyen, C., Corbett, J. G., Liang, L., Mitrescu, C., Rose, F. G., and Kato, S. (2018). Clouds and the earth's radiant energy system (CERES) energy balanced and filled (EBAF) top-of-atmosphere (TOA) edition-4.0 data product. *J. Clim.*, 31(2):895–918.

Maher, N., Milinski, S., Suarez-Gutierrez, L., Botzet, M., Dobrynin, M., Kornblueh, L., Kröger, J., Takano, Y., Ghosh, R., Hedemann, C., Li, C., Li, H., Manzini, E., Notz, D., Putrasahan, D., Boysen, L., Claussen, M., Ilyina, T., Olonscheck, D., Raddatz, T., Stevens, B., and Marotzke, J. (2019). The max planck institute grand ensemble: Enabling the exploration of climate system variability. *Journal of Advances in Modeling Earth Systems*, 11(7):2050–2069.

Mamalakis, A., Ebert-Uphoff, I., and Barnes, E. A. (2022). Explainable artificial intelligence in meteorology and climate science: Model fine-tuning, calibrating trust and learning new science. In *xxAI - Beyond Explainable AI*, Lecture notes in computer science, pages 315–339. Springer International Publishing, Cham.

Matsuno, T. (1966). Quasi-geostrophic motions in the equatorial area. *J. Meteorol. Soc. Japan*, 44(1):25–43.

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., Brovkin, V., Claussen, M., Crueger, T., Esch, M., Fast, I., Fiedler, S., Fläschner, D., Gayler, V., Giorgetta, M.,

Goll, D. S., Haak, H., Hagemann, S., Hedemann, C., Hohenegger, C., Ilyina, T., Jahns, T., Jimenéz-de-la Cuesta, D., Jungclaus, J., Kleinen, T., Kloster, S., Kracher, D., Kinne, S., Kleberg, D., Lasslop, G., Kornblueh, L., Marotzke, J., Matei, D., Meraner, K., Mikolajewicz, U., Modali, K., Möbis, B., Müller, W. A., Nabel, J. E. M., Nam, C. C. W., Notz, D., Nyawira, S.-S., Paulsen, H., Peters, K., Pincus, R., Pohlmann, H., Pongratz, J., Popp, M., Raddatz, T. J., Rast, S., Redler, R., Reick, C. H., Rohrschneider, T., Schemann, V., Schmidt, H., Schnur, R., Schulzweida, U., Six, K. D., Stein, L., Stemmler, I., Stevens, B., von Storch, J.-S., Tian, F., Voigt, A., Vrese, P., Wieners, K.-H., Wilkenskjeld, S., Winkler, A., and Roeckner, E. (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.

Paynter, D., Frölicher, T. L., Horowitz, L. W., and Silvers, L. G. (2018). Equilibrium climate sensitivity obtained from multimillennial runs of two GFDL climate models. *Journal of Geophysical Research: Atmospheres*, 123(4):1921–1941.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Pincus, R., Forster, P. M., and Stevens, B. (2016). The radiative forcing model intercomparison project (RFMIP): experimental protocol for CMIP6. *Geosci. Model Dev.*, 9(9):3447–3460.

Proistosescu, C., Donohoe, A., Armour, K. C., Roe, G. H., Stuecker, M. F., and Bitz, C. M. (2018). Radiative feedbacks from stochastic variability in surface temperature and radiative imbalance. *Geophys. Res. Lett.*, 45(10):5082–5094.

Rohrschneider, T., Stevens, B., and Mauritsen, T. (2019). On simple representations of the climate response to external radiative forcing. *Clim. Dyn.*, 53(5-6):3131–3145.

Rugenstein, M., Bloch-Johnson, J., Abe-Ouchi, A., Andrews, T., Beyerle, U., Cao, L., Chadha, T., Danabasoglu, G., Dufresne, J.-L., Duan, L., Foujols, M.-A., Frölicher, T., Geoffroy, O., Gregory, J., Knutti, R., Li, C., Marzocchi, A., Mauritsen, T., Menary, M., Moyer, E., Nazarenko, L., Paynter, D., Saint-Martin, D., Schmidt, G. A., Yamamoto, A., and Yang, S. (2019). LongRunMIP: Motivation and design for a large collection of millennial-length AOGCM simulations. *Bull. Am. Meteorol. Soc.*, 100(12):2551–2570.

Rugenstein, M., Bloch-Johnson, J., Gregory, J., Andrews, T., Mauritsen, T., Li, C., Frölicher, T. L., Paynter, D., Danabasoglu, G., Yang, S., Dufresne, J.-L., Cao, L., Schmidt, G. A., Abe-Ouchi, A., Geoffroy, O., and Knutti, R. (2020). Equilibrium climate sensitivity estimated by equilibrating climate models. *Geophys. Res. Lett.*, 47(4).

Rugenstein, M., Van Loon, S., and Barnes, E. A. (2025). Convolutional neural networks trained on internal variability predict forced response of toa radiation by learning the pattern effect. in review.

Rugenstein, M. A. A. and Armour, K. C. (2021). Three flavors of radiative feedbacks and their implications for estimating equilibrium climate sensitivity. *Geophys. Res. Lett.*, 48(15).

Rugenstein, M. A. A., Sedláček, J., and Knutti, R. (2016). Nonlinearities in patterns of long-term ocean warming. *Geophysical Research Letters*, 43(7):3380–3388.

Saint-Martin, D., Geoffroy, O., Watson, L., Douville, H., Bellon, G., Voldoire, A., Cattiaux, J., Decharme, B., and Ribes, A. (2019). Fast-Forward to perturbed equilibrium climate. *Geophys. Res. Lett.*, 46(15):8969–8975.

Stevens, B., Sherwood, S. C., Bony, S., and Webb, M. J. (2016). Prospects for narrowing bounds on earth's equilibrium climate sensitivity: EARTH'S EQUILIBRIUM CLIMATE SENSITIVITY. *Earths Future*, 4(11):512–522.

Thompson, D. W. J., Rugenstein, M., Forster, P. M., and Fredericks, L. (2025). An observational estimate of the pattern effect on climate sensitivity: The importance of the eastern tropical pacific and land areas. in review.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Series B Stat. Methodol.*, 58(1):267–288.

Uribe, A., Bender, F. A.-M., and Mauritsen, T. (2022). Observed and CMIP6 modeled internal variability feedbacks and their relation to forced climate feedbacks. *Geophys. Res. Lett.*, 49(24).

Voldoire, A., Saint-Martin, D., Sénési, S., Decharme, B., Alias, A., Chevallier, M., Colin, J., Guérémy, J.-F., Michou, M., Moine, M.-P., Nabat, P., Roehrig, R., Salas y Mélia, D., Séférian, R., Valcke, S., Beau, I., Belamari, S., Berthet, S., Cassou, C., Cattiaux, J., Deshayes, J., Douville, H., Ethé, C., Franchistéguy, L., Geoffroy, O., Lévy, C., Madec, G., Meurdesoif, Y., Msadek, R., Ribes, A., Sanchez-Gomez, E., Terray, L., and Waldman, R. (2019). Evaluation of CMIP6 DECK experiments with CNRM-CM6-1. *J. Adv. Model. Earth Syst.*, 11(7):2177–2213.

Wegelin, J. (2000). A survey of partial least squares (PLS) methods, with emphasis on the two-block case. *University of Washington, Tech. Rep*.

Williams, A. I. L., Jeevanjee, N., and Bloch-Johnson, J. (2023). Circus tents, convective thresholds, and the non-linear climate response to tropical SSTs. *Geophys. Res. Lett.*, 50(6):e2022GL101499.

Wood, R. and Bretherton, C. S. (2006). On the relationship between stratiform low cloud cover and lower-tropospheric stability. *J. Clim.*, 19(24):6425–6432.

Zhou, C., Zelinka, M. D., Dessler, A. E., and Klein, S. A. (2015). The relationship between interannual and long-term cloud feedbacks. *Geophys. Res. Lett.*, 42(23).

Zhou, C., Zelinka, M. D., and Klein, S. A. (2017). Analyzing the dependence of global cloud feedback on the spatial pattern of sea surface temperature change with a G reen's function approach. *J. Adv. Model. Earth Syst.*, 9(5):2174–2189.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Series B Stat. Methodol.*, 67(2):301–320.