CAMBRIDGE
UNIVERSITY PRESS

APPLICATION PAPER

Analyzing climate scenarios using dynamic mode decomposition with control

Nathan Mankovich[1] (iD), Shahine Bouabid[2] (iD), Peer Nowack[3] (iD), Deborah Bassotto[1] (iD) and Gustau Camps-Valls[1] (iD)

[1]Image Processing Laboratory, Universitat de València, València, Spain
[2]Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA
[3]Institute of Theoretical Informatics and Institute of Meteorology and Climate Research Karlsruhe Institute of Technology, Karlsruhe, Germany
Corresponding author: Nathan Mankovich; Email: Nathan.mankovich@gmail.com

Abstract

Understanding the complex dynamics of climate patterns under different anthropogenic emissions scenarios is crucial for predicting future environmental conditions and formulating sustainable policies. Using Dynamic Mode Decomposition with control (DMDc), we analyze surface air temperature patterns from climate simulations to elucidate the effects of various climate-forcing agents. This improves upon previous DMD-based methods by including forcing information as a control variable. Our study identifies both common climate patterns, like the North Atlantic Oscillation and El Niño Southern Oscillation, and distinct impacts of aerosol and carbon emissions. We show that these emissions' effects vary with climate scenarios, particularly under conditions of higher radiative forcing. Our findings confirm DMDc's utility in climate analysis, highlighting its role in extracting modes of variability from surface air temperature while controlling for emissions contributions and exposing trends in these spatial patterns as forcing scenarios change.

Impact Statement

This paper presents an application of dynamic mode decomposition with control to surface air temperature to extract modes of variability and compare different emissions scenarios. It highlights the opportunities of control-based modeling to separately inspect the effect of climate dynamics and forcing from emissions on projections of future climate.

1. Introduction

Modeling and understanding climate *modes of variability*, like the El Niño-Southern Oscillation (ENSO) and the North Atlantic Oscillation (NAO), are essential for predicting climate impacts on global ecosystems and human activities. Modes of variability can be affected by human activity, significantly influence global weather conditions, and are linked to fundamental physical processes in the atmosphere and ocean. Thus, they are crucial for devising effective climate mitigation and adaptation strategies. For example, modes of variability used to understand the interconnectedness of global weather patterns and

their far-reaching effects are often referred to as *teleconnections* (Diaz et al., 2001; Tsonis et al., 2008; Bridgman and Oliver, 2014).

Dimensionality reduction methods are fundamental in extracting modes of variability for analyzing climate dynamics and teleconnections. Principal Component Analysis (PCA), a.k.a. Empirical Orthogonal Functions (EOF), is traditionally used to extract spatiotemporal features (Bauer-Marschallinger et al., 2013; Volkov, 2014, Foroontan et al., 2016). PCA is, however, a linear model, which limits its ability to capture nonlinear processes. Nonlinear formulations, such as kernel PCA, are better suited for nonlinear feature extraction (Schölkopf et al., 1998). Many other PCA extensions have been proposed to address specific climate studies like extended EOF, Multivariate EOF, Principal Oscillation Patterns, and Nonlinear PCA Gehne et al. (2014), Wa et al. (2008), Hannachi et al. (2007), Monahan (2001). Although PCA/EOF and its variants effectively reduce dimensionality, they fall short of capturing the dynamic behavior of climate systems (e.g., the oscillation frequency of modes of variability) due to their inability to process temporal information. This limitation spurred the development of methods like Linear Inverse Modeling (LIM) (Penland and Sardeshmukh, 1995) and Dynamic Mode Decomposition (DMD) (Kutz et al., 2016), which are designed to identify dynamically significant spatiotemporal patterns. LIM continues to find applications in climate science (Wills et al., 2020), though it is not the only dynamical systems-based model used to extract modes of variability in this field. Other methods, such as singular spectrum analysis, are reviewed by Ghil et al. (2002), with subsequent developments including Average Predictability Time (DelSole and Tippett, 2009a, b) and, more recently, Bayesian Linear Dynamical Mode Decomposition (Gavrilov et al., 2020). In contrast, DMD has primarily been applied in fields like fluid dynamics, with a few notable exceptions in climate science, such as emulating sea surface temperature data (Erichson et al., 2019; Navarra et al., 2021), detecting transitions in the North Atlantic Oscillation (Gottwald and Gugole, 2020), and analyzing large-scale climate datasets (Xiong et al., 2023).

Here we use Dynamic Mode Decomposition with control (DMDc) (Proctor et al., 2016), which extends DMD/LIM methods and extracts modes of variability driven by a control variable (Figure 1). Although modes of variability like ENSO indices are usually derived from Sea Surface Temperature (SST) datasets, SSTs undergo trends under climate forcings, which makes it hard to cleanly identify changes in such modes on top of the underlying global warming trend. Our approach applies DMDc to Surface Air Temperature (SAT) data and allows us to separate potential changes in climate modes—such as the ENSO or the NAO—from simultaneous trends in the mean climate state.
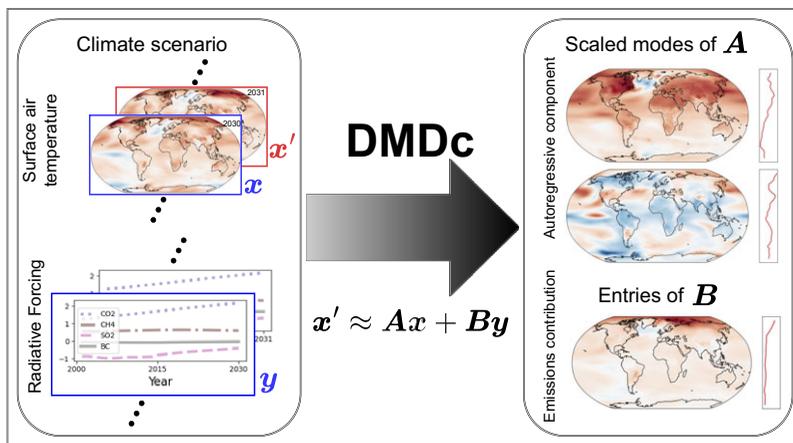

Figure 1. *The DMDc model for estimating the autoregressive component and emissions contributions to global Surface Air Temperature (SAT). SAT at time t (in years) is x, the last 30 years $(t - 29, t - 28, \ldots, t)$ of radiative forcing are in y, and SAT at time $t + 1$ is x'. We take $t = 2030$ in the figure above. The autoregressive component contains the scaled modes of A, which are scaled eigenvectors of a low-rank estimate of A. The forcing contribution patterns are entries of a low-rank estimate of B scaled by entries of y.*

We highlight the utility of DMDc on SATs under different future emissions scenarios. Specifically, we let x denote SAT at time t and x' be SAT at time $t+1$. DMD assumes the model $x' = Ax$ whereas DMDc includes a linear contribution of the past 30 years of radiative forcing from emissions in the control variable y. DMDc results in the model

$$x' = Ax + By. \tag{1}$$

This novel application discerns distinct climate responses under multiple emissions scenarios. In particular, DMDc extracts modes of variability that separate into autoregressive modes (eigenvectors of A) derived from the year-to-year relationship of SAT with itself and "control" contributions associated with the radiative forcing (using B) induced by emissions over the past 30 years. We find that the autoregressive modes include warming trend and ENSO modes, which can be attributed to thermo-dynamically and dynamically forced trends under different forcing scenarios. We discovered that the emissions contribution from DMDc extracts the variable local impact of radiative forcing for different emissions agents over time, including warming from carbon-based emissions and some cooling from aerosols.

2. Data and methods

This section describes the ClimateBench dataset and summarizes how DMDc is used in our experiments.

2.1. Dataset: ClimateBench

We use the local annual SAT variable from the ClimateBench dataset (Watson-Parris et al., 2022) at approximately 2 degree spatial resolution (96 latitude pixels and 144 longitude pixels), paired with annual emissions for four of the main anthropogenic forcing agents: carbon dioxide (CO_2), methane (CH_4), sulfur dioxide (SO_2) and black carbon (BC). The SAT data is derived from the most recent version of the Norwegian Earth System Model (NorESM2) (Seland et al., 2020) while the emission data is sourced from the Community Emissions Data System (Hoesly et al., 2018).

The data spans the years 1850 to 2050. These years are split into historical simulation from the Coupled Model Intercomparison Project Phase 6 (CMIP6) (Eyring et al., 2016) and different future simulations corresponding to emissions scenarios from ScenarioMIP protocol (O'Neill et al., 2016). The historical data spans the years 1850 to 2014, and the scenario data covers the years 2015 to 2100. We focus on four possible realistic future trajectories called Shared Socioeconomic Pathways (SSPs) (Riahi et al., 2017). From highest to lowest forcing, the selected scenarios are SSP585, SSP370, SSP245, and SSP126 (Figure 2).

The following analysis uses the SAT from the historical period along with emissions data for each scenario. When these analyses are run on the raw emissions data, forcing agents have a minuscule and inconsistent contribution across emissions scenarios. These contributions do not align with known interactions between forcing agents and surface air temperature. Therefore, we use the procedure outlined in (Leach et al., 2021; Bouabid et al., 2024) to compute estimates of the radiative forcing level induced by each forcing agent from their annual emissions data (Figure 2(b)). Specifically, radiative forcing is computed using the gas cycle model and forcing model of FaIRv2.0.0 calibrated for NorESM2-LM to convert (1) emission data into concentration and (2) concentration into levels of effective radiative forcing (Leach et al., 2021).

2.2. Applying DMDc to ClimateBench

To analyze these data, we use a Python implementation of the DMDc model (Demo et al., 2018; Ichinaga et al., 2024) and specify the radiative forcing contribution from emissions as a control variable. Mirroring the notation by Proctor et al. (2016), we assume the linear model for local SAT in Eq. (1). In this model, x is the flattened local SAT at year t, y is the estimated radiative forcing for each agent from the last 30 years:
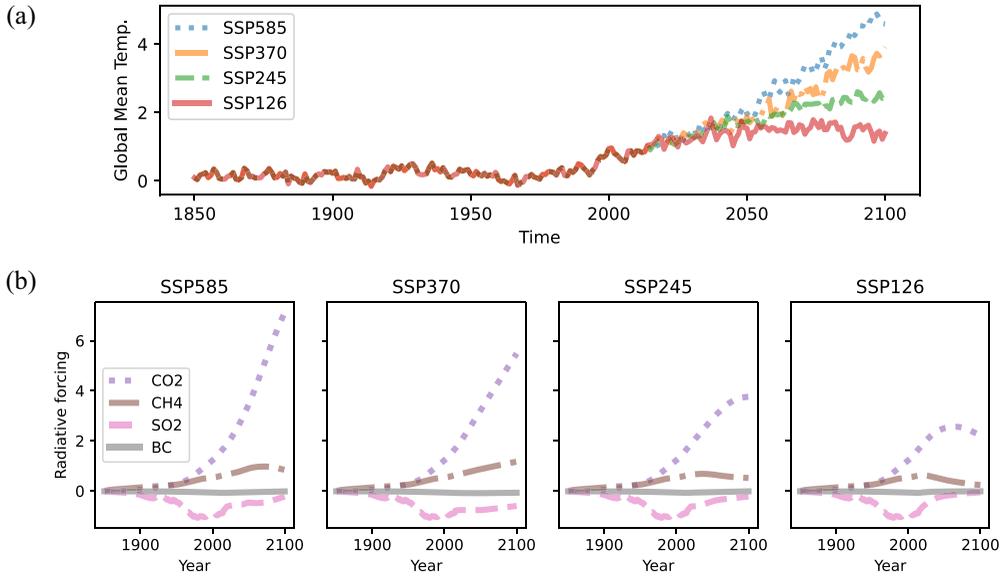
(a)

(b)

Figure 2. *The ClimateBench dataset and the derived radiative forcing. This dataset contains four different emissions scenarios: SSP585, SSP370, SSP245, and SSP126 (ordered from highest forcing to lowest forcing). (a) Spatially averaged yearly global temperature from the ClimateBench dataset. (b) Radiative forcing derived from the ClimateBench dataset.*

$t, t-1, \ldots, t-29$, and x' is the SAT for the next year, $t+1$. Neither the SAT nor the forcing time series are mean-centered before running our analysis. In the DMDc model, A corresponds to the autoregressive component and B to the emissions (in this case, radiative forcing) contribution. To simplify the detection of dynamics within the autoregressive component of the system, we assume A is rank 5. The definitions and symbols for this applying DMDc to ClimateBench appear in Table 1.

The modes of variability of the autoregressive component, along with their contribution over time, come from the scaled modes, ξ_m, and corresponding eigenvalues, λ_m, of A. The contribution of radiative forcing can be found using the product of entries of B and y corresponding to different forcing agents and times. We now present further details on the DMDc algorithm along with extraction of the autoregressive component and emissions contributions.

Table 1. *Definitions of symbols for applying DMDc to SAT and emissions from the ClimateBench dataset*

Symbol	Definition
$x(t)$	Flattened global temperature at time t
$y(t)$	Radiative forcing at time t
N	Number of samples
R	Number of features
K	Number of control features
$x_1, \ldots, x_N$	Input sample sequence
$y_1, \ldots, y_N$	Input control sequence
$x'_1, \ldots, x'_N$	Output sample sequence
$X = [x_1, \ldots, x_N]$	$R \times N$ input sample matrix
$Y = [y_1, \ldots, y_N]$	$K \times N$ input control matrix
$X' = [x'_1, \ldots, x'_N]$	$R \times N$ output sample matrix

2.2.1. Time delayed dataset structure

We first organize the SAT and emissions data from ClimateBench into variables (Table 1).

Our data consists of a flattened local SAT time series

$$\{\boldsymbol{x}(0), \boldsymbol{x}(1), \ldots, \boldsymbol{x}(T)\} \subset \mathbb{R}^R, \tag{2}$$

where $R = 96 \times 144$ is the dimensionality of the flattened local mean surface temperature grid and a radiative forcing time series (estimated from the ClimateBench emission data following the procedure outlined in (Leach et al., 2021; Bouabid et al., 2024))

$$\{\boldsymbol{y}(0), \boldsymbol{y}(1), \ldots, \boldsymbol{y}(T)\} \subset \mathbb{R}^S, \tag{3}$$

where $S = 4$ for the 4 forcing agents considered (CO_2, CH_4, SO_2, and BC).

Let $\tau = 1$ year be the lag in the autoregressive component (see supplementary material for a detailed explanation), $J = 30$ years be the delay for the time delay embedding (Takens, 2006) of emissions, and define $N = T - J - \tau + 1$. We construct the input and output sample matrices as

$$\boldsymbol{X} = [\boldsymbol{x}(J)\, \boldsymbol{x}(J+1)\ldots\boldsymbol{x}(T-\tau)] \in \mathbb{R}^{R \times N}, \tag{4}$$

$$\boldsymbol{X}' = [\boldsymbol{x}(J+\tau)\, \boldsymbol{x}(J+\tau+1)\ldots\boldsymbol{x}(T)] \in \mathbb{R}^{R \times N}. \tag{5}$$

This results in a dataset of N samples, each with R sample features. Further, using $K = SJ$, we define the input control signal following

$$\boldsymbol{y}_n = \begin{bmatrix} \boldsymbol{y}(n) \\ \boldsymbol{y}(n-1) \\ \vdots \\ \boldsymbol{y}(n-J+1) \end{bmatrix} \in \mathbb{R}^K, \tag{6}$$

which we concatenate into the control input matrix

$$\boldsymbol{Y} = [\boldsymbol{y}_J \ldots \boldsymbol{y}_{T-\tau}] \in \mathbb{R}^{K \times N}. \tag{7}$$

Thus, we have a dataset of N control inputs, each with K control features.

2.2.2. DMDc procedure

Dynamic Mode Decomposition with Control (DMDc) (Proctor et al., 2016) assumes the linear state space model with control

$$\boldsymbol{x}'_n = \boldsymbol{A}\boldsymbol{x}_n + \boldsymbol{B}\boldsymbol{y}_n. \tag{8}$$

In matrix form, this model can be block-factorized following

$$\boldsymbol{X}' = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{B}\boldsymbol{Y} = [\boldsymbol{A}\,\boldsymbol{B}]\begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{bmatrix} = [\boldsymbol{A}\,\boldsymbol{B}]\boldsymbol{\Omega}. \tag{9}$$

$$\underbrace{\phantom{\begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{bmatrix}}}_{\boldsymbol{\Omega}}$$

Using this, DMDc aims to accomplish two tasks.

1. Estimate $\boldsymbol{A}$ and $\boldsymbol{B}$ with least-squares regression using the data $\boldsymbol{X}, \boldsymbol{X}', \boldsymbol{Y}$.
2. Compute the eigendecomposition of $\boldsymbol{A}$, which allows the study of the modes of the autoregressive dynamics.

This is classically achieved using Singular Value Decomposition (SVD) based methods (Tu, 2013), which compute a reduced-rank approximation of $\boldsymbol{A}$ and $\boldsymbol{B}$. The procedure we follow is outlined in Algorithm

1 and outputs eigenvalues along the diagonal of $\mathbf{\Lambda}$, dynamic modes in the columns of $\mathbf{\Phi}$, and a reduced rank approximation of $\boldsymbol{B}$ in $\check{\boldsymbol{B}}$. DMDc generalizes DMD by including control (App. A for details).

Algorithm 1 SVD-based DMDc procedure (Proctor et al., 2016).

1: **procedure** DMDc($\mathbf{\Omega}, M, M_\Omega$)
2:     Compute M_Ω-truncated SVD $\mathbf{\Omega} \approx \begin{bmatrix} \boldsymbol{U}_A \\ \boldsymbol{U}_B \end{bmatrix} \mathbf{\Sigma} \boldsymbol{V}^*$
3:     Compute M-truncated SVD $\boldsymbol{X}' \approx \widehat{\boldsymbol{U}} \widehat{\mathbf{\Sigma}} \widehat{\boldsymbol{V}}^*$
4:     Approximate A as $\overline{\boldsymbol{A}} = \boldsymbol{X}' \boldsymbol{V} \mathbf{\Sigma}^{-1} \boldsymbol{U}_A^*$
5:     Approximate B as $\overline{\boldsymbol{B}} = \boldsymbol{X}' \boldsymbol{V} \mathbf{\Sigma}^{-1} \boldsymbol{U}_B^*$
6:     Compute $\tilde{\boldsymbol{A}} = \widehat{\boldsymbol{U}}^* \overline{\boldsymbol{A}} \widehat{\boldsymbol{U}}$ and $\tilde{\boldsymbol{B}} = \widehat{\boldsymbol{U}}^* \overline{\boldsymbol{B}}$
7:     Compute the eigendecomposition $(\mathbf{\Lambda}, \boldsymbol{W})$ of $\tilde{\boldsymbol{A}}$
8:     Compute the dynamic modes $\mathbf{\Phi} = \boldsymbol{X}' \boldsymbol{V} \mathbf{\Sigma}^{-1} \boldsymbol{U}_A^* \widehat{\boldsymbol{U}} \boldsymbol{W}$
9:     Rank-M approximation $\check{\boldsymbol{B}} = \widehat{\boldsymbol{U}} \widehat{\boldsymbol{U}}^* \overline{\boldsymbol{B}}$
10:     **return** $(\mathbf{\Lambda}, \mathbf{\Phi}, \check{\boldsymbol{B}})$.
11: **end procedure**

In experiments, we use a full-rank SVD of $\mathbf{\Omega}$, e.g., $M_\Omega = \text{rank}(\mathbf{\Omega})$. We compute a reduced-rank approximation of $\boldsymbol{B}$ by restricting the rank of the SVD of $\boldsymbol{X}'$. This chosen rank M will determine the number of dynamic modes. We choose $M = 5$ as we experimentally observe it captures the dominant singular values (see supplementary material). Instead of using the least squares approximations ($\overline{A}$ and $\overline{B}$) to emulate or reconstruct the SAT series, we use the first $M = 5$ dynamic modes from Algorithm 1 and $\check{B}$ to extract modes of variability.

2.2.3. Autoregressive component

We use DMDc to analyze the dynamics from the matrix A by first computing the eigenvalue decomposition $\tilde{A} \boldsymbol{W} = \boldsymbol{W} \mathbf{\Lambda}$. This eigenvalue decomposition gives rise to the eigenvalues, dynamic modes, and amplitudes (Table 2). The spatial pattern is captured in the dynamic mode, the amplitude captures the fixed contribution of the dynamic mode, and the eigenvalue summarizes the oscillation and trend over time.

Spatial patterns. The vector of amplitudes, $\boldsymbol{\alpha} = [\alpha_1 \cdots \alpha_M]^T$, is defined as the minimizer of the least squares problem $\mathbf{\Phi} \boldsymbol{\alpha} \approx \boldsymbol{x}_1$. The product of a dynamic mode and its corresponding, fixed amplitude is the *scaled mode* (Krake et al., 2021) denoted

$$\boldsymbol{\xi}_m = \alpha_m \boldsymbol{\varphi}_m. \tag{10}$$

We visualize the scaled modes in our experiments to provide a realistic scaling (e.g., correcting for sign flips) of the spatial patterns of the dynamic modes (Figures 4, 5, and 6).

Table 2. *Decomposition of $\tilde{A}$*

Symbol	Definition
$\boldsymbol{w}_m$	Eigenvectors of $\tilde{A}$ (columns of $\boldsymbol{W}$)
λ_m	Eigenvalues of $\tilde{A}$ (diagonal of)
$\boldsymbol{\varphi}_m$	Dynamic modes (columns of)
α_m	Amplitudes (mth row of $\boldsymbol{\alpha}$)
$\boldsymbol{\xi}_m = \alpha_m \boldsymbol{\varphi}_m$	Scaled (dynamic) modes

Temporal patterns. Inspired by (Krake et al., 2021), we determine the evolution of the scaled modes over time with the help of the following propositions. See supplementary material for the proofs.

Proposition 1. *The dynamic modes, $\boldsymbol{\varphi}_m$, are the eigenvectors of $\overline{A}\widehat{U}\widehat{U}^*$ with eigenvalues λ_m.*

Proposition 2. *$\lambda_m\boldsymbol{\xi}_m$ is the evolution of a scaled mode, $\boldsymbol{\xi}_m$, over τ time steps in the subspace spanned by the columns of $\widehat{U}$. Specifically,*

$$\lambda_m\boldsymbol{\xi}_m = \overline{A}\widehat{U}\widehat{U}^*\boldsymbol{\xi}_m \in \mathbb{R}^R. \tag{11}$$

By linearity, the evolution of two scaled modes over τ time steps is $\lambda_j\boldsymbol{\xi}_j + \lambda_k\boldsymbol{\xi}_k$. We use this fact to visualize the evolution of two scaled modes with complex conjugate eigenvalues (Figure 6b).

The *trend and oscillation frequency* of a scaled mode can be extracted from its associated eigenvalue. When written in exponential form, the mth eigenvalue is $\lambda_m = r_m e^{i\omega_m}$. The magnitude $r_m = |\lambda_m|$ is the magnitude of the contribution of the scaled mode over time. A magnitude of $r_m = 1$ is stable, $r_m < 1$ is decaying, and $r_m > 1$ is a diverging contribution. The complex portion $e^{i\omega_m}$ controls the frequency of oscillation

$$f_m = \frac{\omega_m}{2\pi} = -i\log\left(\frac{\lambda_m}{|\lambda_m|}\right)/(2\pi). \tag{12}$$

We combine the previous two propositions and the definition of the dynamic mode amplitudes. Using $\boldsymbol{\Xi}$ to represent the matrix whose columns are scaled dynamic modes ($\boldsymbol{\xi}_m$), this results in the approximation of the autoregressive part as.

$$A\boldsymbol{x}_n \approx \boldsymbol{\Xi}\boldsymbol{\Lambda}\boldsymbol{\Xi}^\dagger\boldsymbol{x}_n.$$

The details of this approximation can be found in the supplementary material.

2.2.4. Emissions contribution

The columns of the matrix $\check{\boldsymbol{B}} \in \mathbb{R}^{N \times K}$ are fixed spatial patterns of the linear forcing contribution from emissions (Figure 3). The scale of each of these contributions to SAT is exactly the corresponding entry in $\boldsymbol{y}$.

The evolution of the contribution of radiative forcing over times $(1, 2, \ldots, N)$ is $(\check{\boldsymbol{B}}\boldsymbol{y}_1, \check{\boldsymbol{B}}\boldsymbol{y}_2, \ldots, \check{\boldsymbol{B}}\boldsymbol{y}_N)$. The mean of each vector $\check{\boldsymbol{B}}\boldsymbol{y}_n$ is the spatial mean contribution of emissions to temperature over time (Figure 7 top left).

Time-lagged radiative forcing. We use the time-lagged structure of $\boldsymbol{y}_n$ to separate the radiative forcing signal into years (Eq. (6)). Each entry of $\boldsymbol{y}(j) \in \mathbb{R}^S$ (denoted $\boldsymbol{y}(j)_s$) corresponds to a different forcing agent (e.g., CO_2). In our experiments, we have $S = 4$ forcing agents.

Structure of $\check{\boldsymbol{B}}$. The spatial contribution of radiative forcing over the last J years is determined by decomposing $\check{\boldsymbol{B}}$ using the structure of $\boldsymbol{y}_n$. First, $\check{\boldsymbol{B}}$ is decomposed into blocks of size $N \times S$ corresponding to each year in $\boldsymbol{y}_n$
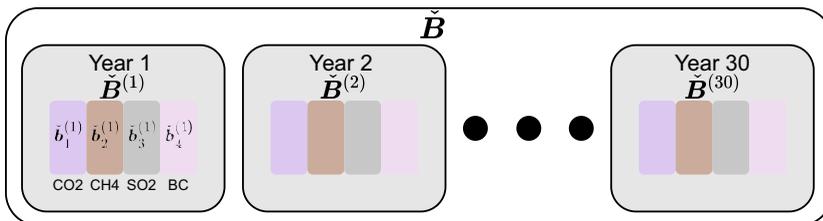
Figure 3. The structure of the forcing contribution matrix $\check{\boldsymbol{B}}$.

$$\check{\boldsymbol{B}} = \begin{bmatrix} \overset{(1)}{\check{\boldsymbol{B}}} & \overset{(2)}{\check{\boldsymbol{B}}} & \cdots \overset{(J)}{\check{\boldsymbol{B}}} \end{bmatrix}. \tag{13}$$

Then the spatial contribution of the input radiative forcing agent s (e.g., CO_2) over $(1, 2, \ldots, J)$ time steps in the past are grouped into the sth columns of each of $\left(\overset{(1)}{\check{\boldsymbol{B}}}, \overset{(2)}{\check{\boldsymbol{B}}}, \ldots, \overset{(J)}{\check{\boldsymbol{B}}} \right)$, and is denoted $\left(\overset{(1)}{\check{\boldsymbol{b}}_s}, \overset{(2)}{\check{\boldsymbol{b}}_s}, \ldots, \overset{(J)}{\check{\boldsymbol{b}}_s} \right)$. This structure is summarized in Figure 3.

This process unravels the matrix multiplication into pieces corresponding to time lag and forcing agent as

$$\check{\boldsymbol{B}} \boldsymbol{y}_n = \sum_{j=1}^{J} \overset{(j)}{\check{\boldsymbol{B}}} \boldsymbol{y}(j) = \sum_{j=1}^{J} \sum_{s=1}^{S} \overset{(j)}{\check{\boldsymbol{b}}_s} \, \boldsymbol{y}(n-j+1)_s. \tag{14}$$

The spatial pattern corresponding to the linear contribution of a forcing agent s at year j before the predicted year $n+1$ is found in the entries of $\overset{(j)}{\check{\boldsymbol{b}}_s} \, \boldsymbol{y}(n-j+1)_s \in \mathbb{R}^R$. Using this method, we determine the forcing contribution for each agent as the products between entries of $\check{\boldsymbol{B}}$ and $\boldsymbol{y}_n$ (Figures 7–9). See Appendix B for details of the usage.

Limitations. DMDc does not explicitly address the possible correlation between the time series dimensions and, as a result, can suffer from the same pitfalls as linear regression (e.g., difficulty in interpreting the coefficients because they may not linearly separate). If we are using a form of regularization in fitting the A and B matrices, that may help mitigate this effect by ensuring we do not "mix up" factors as much. We leave this as an interesting avenue for future study. We acknowledge that this potential for correlation limits the extent to which we state that one particular factor is responsible for the obtained pattern. For example, a correlation between CO_2 and CH_4 forcings can be due to a shared latent variable (human activity) that drives them, and interpreting the coefficients of a linear model such as DMDc as the causal effect of a forcing (CO_2 or CH_4) has on the pattern does not correct for this confounding.

Summary. Our application of DMDc to SAT results in the decomposition of SAT at year $n+1$ as

$$\boldsymbol{x}_{n+1} \approx \boldsymbol{\Xi} \boldsymbol{\Lambda} \boldsymbol{\Xi}^{\dagger} \boldsymbol{x}_n + \sum_{j=1}^{J} \sum_{s=1}^{S} \overset{(j)}{\check{\boldsymbol{b}}_s} \, \boldsymbol{y}(n-j+1)_s. \tag{15}$$

In this decomposition, the mth column of $\boldsymbol{\Xi}$ is a scaled mode ($\boldsymbol{\xi}_m$) and the mth diagonal entry of $\boldsymbol{\Lambda}$ is its associated eigenvalue (λ_m). The term $\overset{(j)}{\check{\boldsymbol{b}}_s} \, \boldsymbol{y}(n-j+1)_s$ is the spatial pattern of the forcing contribution of emissions agent s at j years before year $n+1$.

3. Results

For our analysis, we use DMDc on the ClimateBench dataset (Watson-Parris et al., 2022) to detect modes of variability from local, gridded annual mean SAT at $\approx 2°$ horizontal resolution while including the forcing information from various emissions modalities (Figure 1 and Section 2). We compare these representations across four emissions scenarios: the Shared Socio-economic Pathways (SSPs) 585, 370, 245, and 126. This analysis results in modes of variability from (1) *the autoregressive component*: year-to-year autoregression of the SAT time series (from A), and (2) *the emissions contribution*: the effect of the last 30 years of radiative forcing from different emissions types on SAT (from B).

3.1. Autoregressive component

The scaled modes of the autoregressive component of DMDc detect the year-to-year autoregressive contribution of SAT. We use *scaled* modes to remove sign ambiguity for spatial plots (Section 2). This component contains the effect of the previous year's temperature on the next year and the forcing contributions that are not captured by emissions contribution. Scaled modes are fixed spatial patterns that oscillate over time and are extracted from the data fields using the eigenvectors of a low-rank estimate of A. The temporal evolution of a scaled mode is determined by its associated eigenvalue (Section 2).

The scaled modes extracted from SAT display a clear warming trend and other common climate patterns, effectively capturing the year-to-year impact of previous temperatures and external forcings (Figures 4–6) (Hasselmann, 1976). DMD run on the same data extracts similar spatial patterns, trends, and oscillations (see comparison to DMD in supplementary material). However, the DMD model does not incorporate emissions information and thus cannot extract the essential emissions contribution patterns detected by DMDc.

The global warming trend is captured in two scaled modes related to warming and cooling with associated eigenvalues. We plot the sum of these two scaled modes in Figure 4. Higher emissions scenarios have a higher magnitude warming trend. Higher emission scenarios have larger eigenvalues for the warming mode (0.98 for SSP585 and 0.93 for SSP126), indicating increased stability over time.

Global warming modes reveal spatial patterns such as intensified land warming in the Northern Hemisphere, especially in Russia (Figure 5), and the North Atlantic Warming Hole (NAWH) (Collins et al., 2013) (Figure 4). There are various mechanisms under discussion to contribute to the existence of the NAWH, from a weakened Atlantic Meridional Overturning Circulation (AMOC) to changes in atmospheric wind patterns (Drijfhout et al., 2012; Bellomo et al., 2021; He et al., 2022) as a combination of the effects of aerosol and greenhouse gas forcing (Qasmi, 2023). North Atlantic temperature variability is generally also associated with the North Atlantic Oscillation (NAO) due to two-way interactions between the atmosphere and ocean (Wang et al., 2004). Since the warming pattern does not oscillate, this suggests that the NAWH can be interpreted as a thermodynamically forced trend rather than a feature of long-term internal variability in the NAO. Additionally, all scenarios feature a South Atlantic warming hole that exhibits no discernible trend with rising emissions. Moreover, all scenarios show a mild El Niño-
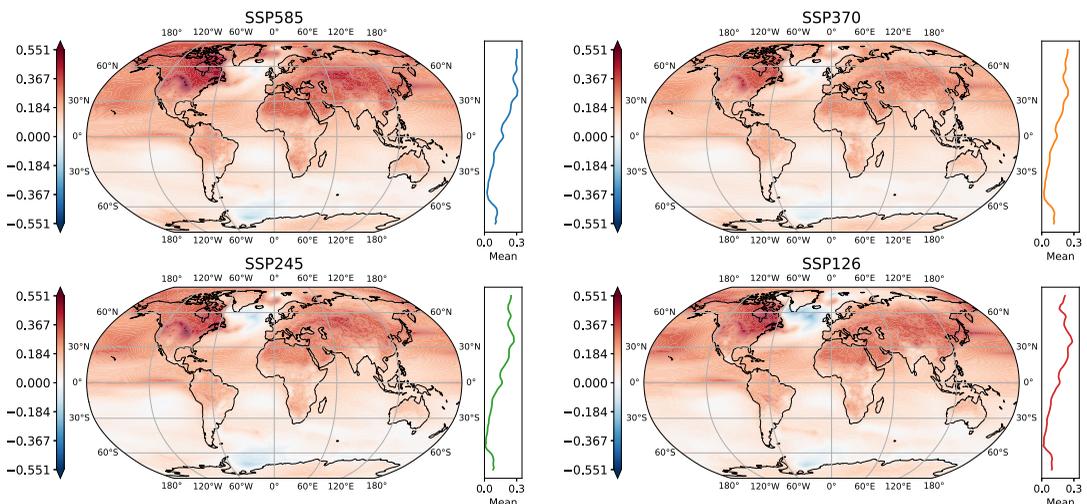

Figure 4. *The global warming trend is associated with the two largest real eigenvalues. The scenarios ordered from high to low emissions are SSP585 (blue), SSP370 (orange), SSP245 (green), and SSP126 (red). Each spatial pattern is the sum of the two scaled dynamic modes associated with the two largest real eigenvalues. The plot on the right is the mean latitudinal temperature profile. The color bars and latitudinal profiles are in °C.*
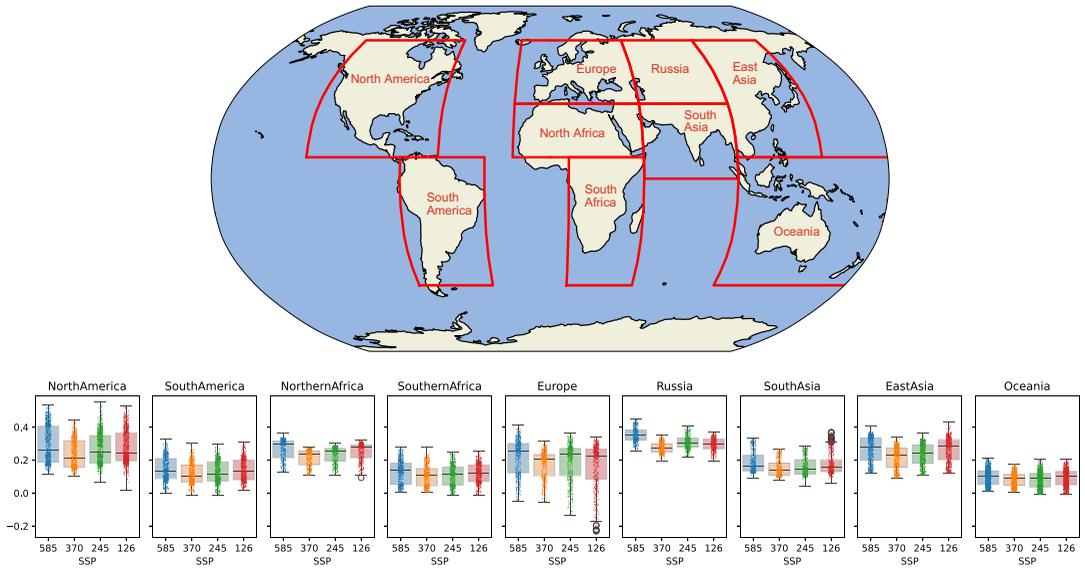
Figure 5. *The SAT warming trend by land mass region over the entire scenario. The scenarios ordered from high to low emissions are SSP585, SSP370, SSP245, and SSP126. Each boxplot summarizes the distribution of the global warming scaled mode values in each region. The whiskers cover the entire distribution of the temperature values. We see larger warming relative to the global mean for the higher forcing scenarios in the Northern Hemisphere, especially Russia.*

like signal. This pattern is often connected to a weakening of the Pacific zonal Walker circulation under CO_2 forcing in climate models, which enforces our discovery that this pattern is weakest in the high aerosol forcing scenario, SSP370 (Watanabe et al., 2024).

Two other scaled modes are associated with complex eigenvalues indicating oscillation over time (Figure 6). The tropical Pacific patterns we find in these modes have their peak positive amplitude in the Central Pacific—somewhat akin to an average Central Pacific El Niño event. While our approach to characterizing climate mode behavior is different, our pattern resembles those found in the ENSO analysis in the original NorESM2 model documentation paper by Seland et al., (2020, Figure 23). There, the El Niño composite signal of NorESM2 showed a larger amplitude than observed, with a clear peak amplitude in the Central Pacific (rather than a more elongated East-to-Central Pacific warm anomaly). Therefore, we call these scaled modes the "ENSO modes." These modes oscillate and have a clear pattern that resembles central Pacific El Niño events for all scenarios except SSP585, which is a La Niña pattern that evolves into a central Pacific El Niño event one year into the future using the associated eigenvalues. These ENSO patterns oscillate at somewhat different frequencies for different scenarios (from 0.15/year for SSP126 to 0.18/year for SSP585). Except for SSP245, the frequency of ENSO increases in higher emissions scenarios (see Autoregressive component in supplementary material). There is evidence of this pattern of increased emissions in conjunction with increasing ENSO variability in many climate model ensembles (Timmermann et al., 1999; Malik et al., 2020). However, these oscillation differences could be estimation uncertainty given the application of DMDc to single ensemble members.

3.2. Emissions contribution

The emissions contribution contains the linear impact of the past 30 years of radiative forcing from the four of the main anthropogenic forcing agents: carbon dioxide (CO_2), methane (CH_4), sulfur dioxide (SO_2) and black carbon (BC) on the current year SAT. Recall that these patterns cannot be detected using
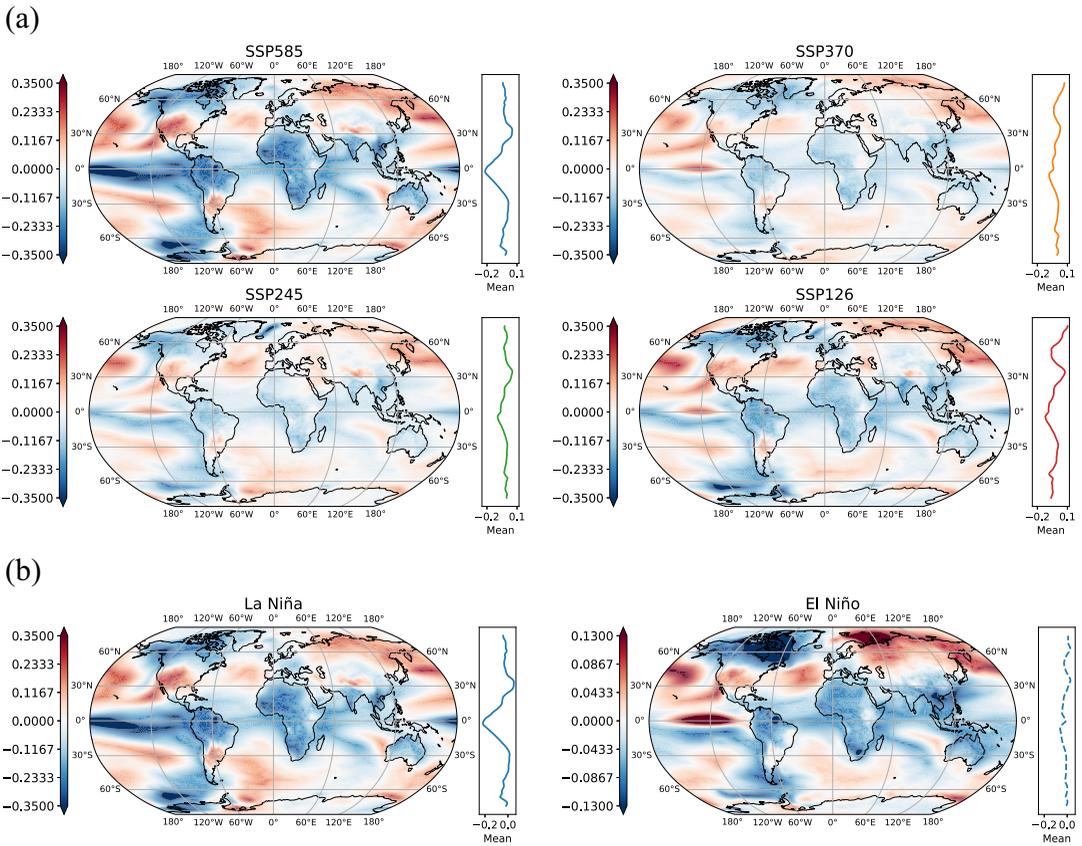
(a)

(b)

Figure 6. *Analysis of ENSO oscillation with DMDc. The scenarios ordered from high to low emissions are SSP585(blue), SSP370 (orange), SSP245 (green), and SSP126 (red). The plot on the right of each map is the mean latitudinal temperature profile. The color bars and latitudinal profiles are in °C. (a) ENSO-like spatial patterns appear in all emissions scenarios, specifically, a central Pacific El Niño event in SSP126, SSP245, and SSP370, whereas La Niña in SSP585. Each spatial pattern is the sum of the two scaled dynamic modes associated with complex conjugate eigenvalues. (b) We observe a Niña-to-Niño phase transition in SSP585 as we evolve the spatial pattern one year into the future using the associated eigenvalues (Section 2). Therefore, a pattern akin to a central Pacific El Niño event is visible in all scenarios. (Left) The spatial pattern is the sum of the two scaled dynamic modes associated with complex conjugate eigenvalues. (Right) The spatial pattern is the sum of the two scaled dynamic modes associated with complex conjugate eigenvalues, multiplied by their associated eigenvalue.*

methods like DMD because they do not include emissions information. The emissions contribution includes part of the linear thermodynamic response of SAT to different emissions modalities because, in DMDc, they linearly shift the next year's temperature.

The spatial mean of the emissions contribution increases for higher emissions scenarios (Figure 7, top left). The impact of each of the last 30 years of forcing on a current year increases as we move further into the past (Figure 7, top right). Carbon radiative forcing dominates the spread of the forcing contribution (Figure 7, bottom row) and is known to have a delayed effect on temperature (Zickfeld and Herrington, 2015). This explains the increased impact of radiative forcing on SAT as we look further into the past. Additionally, we compute the grid point standard deviation over all forcing agents from 2051 to 2100. We see an increase in standard deviation from 0.14 °C in SSP126 to 0.20 °C in SSP585 which indicates an overall higher variability in forcing impact in higher forcing scenarios.
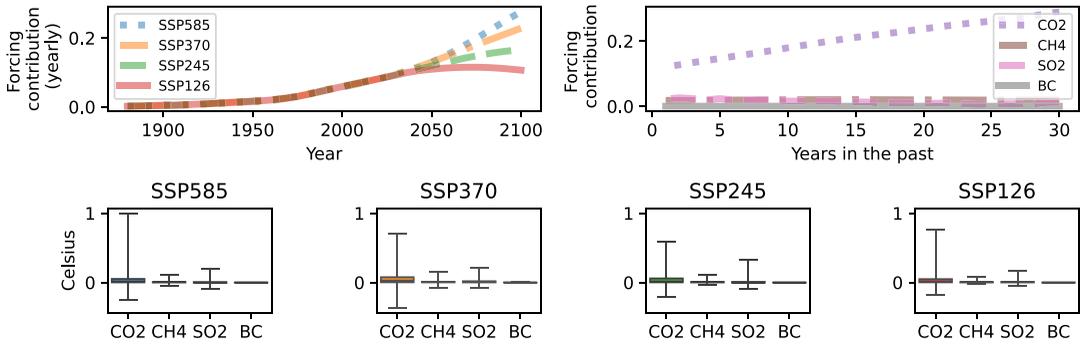
Figure 7. *Analysis of emissions contribution with DMDc. The scenarios ordered from high to low emissions are SSP585, SSP370, SSP245, and SSP126. Units are °C for the vertical axis in all plots. (Top left) Each annual value represents the spatial mean of the past 30 years of radiative forcing contributions from emissions to global SAT. (Top right) The spatial mean of the cumulative forcing from years 2051–2100 as we look further back from the predicted year averaged over across all four scenarios. (Bottom) The distribution of SAT contribution for each forcing agent for each scenario. The whiskers include each point in the data distribution.*

Although DMDc properly isolates the effects of radiative forcing from carbon on SAT, it does not fully capture a mean global cooling from aerosols (see forcing contribution in supplementary material). However, the forcing contribution does capture cooling from SO_2 on northern landmasses with high human populations under some scenarios (SSP585 in Figure 8). The warming in other scenarios could be driven by relative change due to decreasing SO_2 emissions over time (Figure 2(b) in Section 2 for the decreasing magnitude of SO_2).
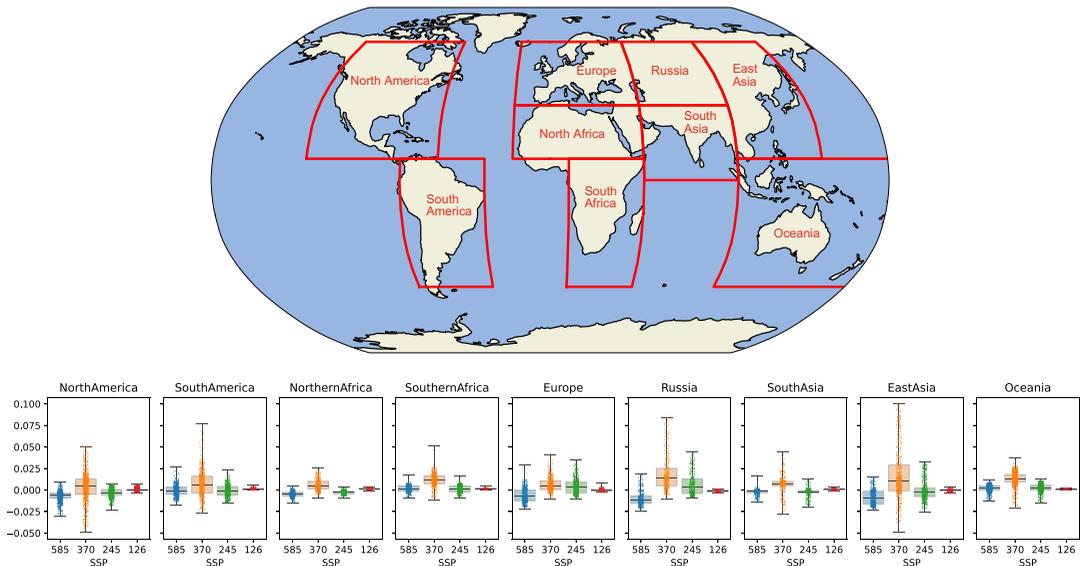
Figure 8. *The forcing contribution to SAT from SO_2 by land mass region looking 30 years backwards from the year 2101. Each boxplot is the distribution of SAT contribution for the forcing contribution of SO_2. The whiskers cover each point in the data distribution. The scenarios, ordered from high to low emissions, are SSP585, SSP370 SSP245 and SSP126 Although SSP585 exhibits cooling from SO_2, we see warming from SO_2 in other scenarios.*
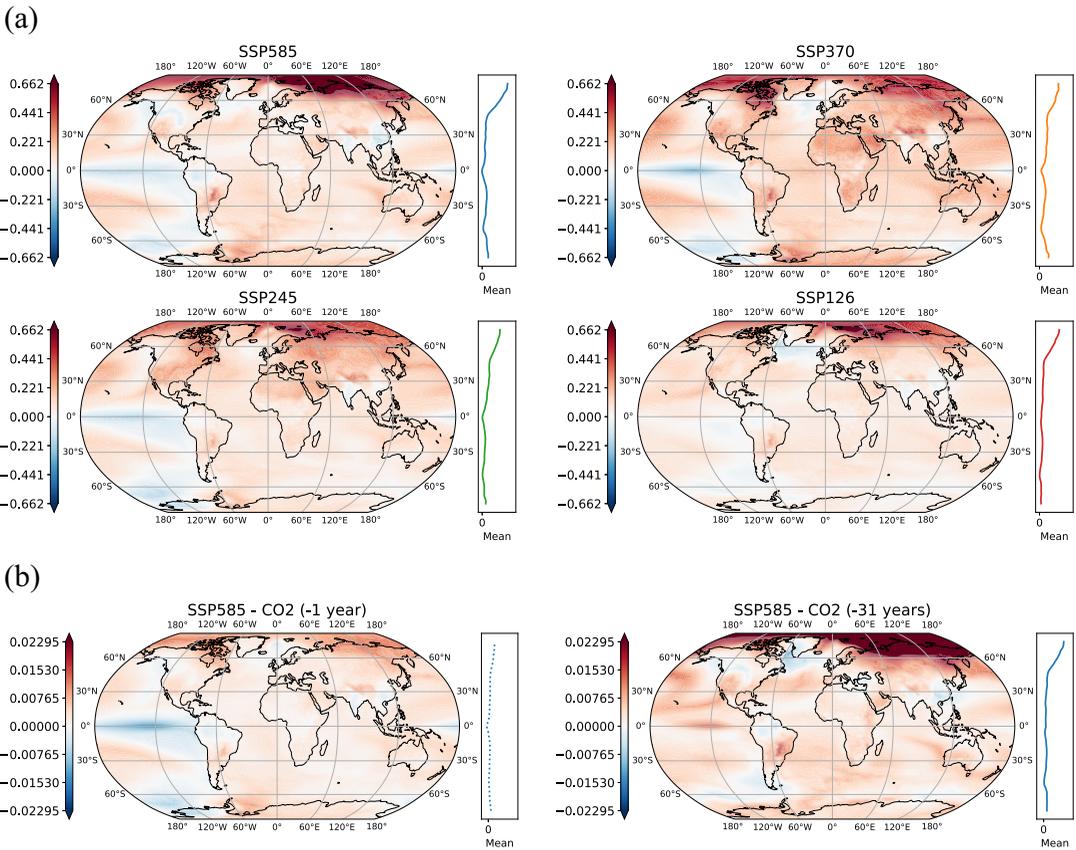
(a)

(b)

Figure 9. *Visualization of the linear impact of radiative forcing on SAT. The scenarios ordered from high to low emissions are SSP585 (blue), SSP370 (orange), SSP245 (green), and SSP126 (red). The plot on the right of each map is the mean latitudinal temperature profile. The color bars and latitudinal profiles are in °C. (a) The cumulative effect of the past 30 years of radiative forcing on SAT (averaged over the output years 2050 to 2100). We see an increased contribution to SAT for higher forcing scenarios and polar amplification. Due to the DMDc model, this only contains the linear contribution of radiative forcing on SAT. (b) The changing pattern of the effect of radiative forcing from carbon on SAT for different years before the predicted year in SSP585 averaged over predicted years* 2050 *to* 2100*). We see a higher contribution from emissions when we look further into the past from the predicted year.*

The local impacts of the total forcing contribution include increased warming in high latitudes of the Northern Hemisphere (Figure 9). This polar warming pattern is consistent with Arctic amplification, a well-documented phenomenon in climate science literature: polar regions warm faster than the rest due principally to ice-albedo feedback mechanisms (Pithan and Mauritsen, 2014; Rantanen et al., 2022). This pattern is especially expected under the high greenhouse gas emissions (Meehl et al., 2020).

4. Conclusion

We used DMDc to extract modes of variability and emissions contributions under different climate scenarios. To do this, we leveraged ClimateBench, a novel, rich dataset that provides local, gridded yearly mean surface air temperatures while including crucial emissions information. Previous methods (e.g., LIM/DMD) are limited to only considering an autoregressive model for surface air temperature while ignoring essential control information like radiative forcing. DMDc addresses this issue by incorporating

radiative forcing as a control variable. When applied to surface air temperatures, DMDc determined distinct (e.g., dynamic versus thermodynamic) climate responses under different emissions scenarios, producing insights into climate change dynamics.

Spatially explicit patterns and temporally resolved trends of impacts were quantified through running DMDc on these data. Specifically, we detected a global warming trend and ENSO (an important modulator of climate extremes) in the autoregressive component and also quantified their increase in higher emissions scenarios. We uncovered the known North Atlantic warming hole and saw its strength decrease as we increased the emissions scenario. Unique to DMDc, the forcing component extracted spatial contributions of four radiative forcing agents: CO_2, CH_4, SO_2, and BC over the last 30 years. As expected, the forcing contribution is higher in higher emissions scenarios. We also saw warming from carbon across all scenarios and some landmass cooling from aerosol in specific scenarios. Other, more nuanced patterns identified by DMDc include changes in emissions contributions as we analyze the effect of emissions over the past 30 years. Finally, we identified spatial patterns of the forcing contributions, which included Arctic amplification.

Our future work will include algorithm refinement and application of DMDc to other climatological variables. For example, our method for determining the emissions contribution from DMDc does not account for potential partial correlations between variables. Exploration of non-linear forcing contribution and other DMD variants (e.g., adding constraints) may address partial correlations, increase forcing contribution to SAT (Deng et al., 2020), improve extraction of the aerosol-based cooling signal, and result in more temporally stable (e.g., non-decaying) autoregressive modes. Finally, further analysis of the control signal with larger datasets could lead to a novel detection attribution strategy. Overall, our study demonstrates the opportunities of DMDc for climate science, paving the way for more novel, data-driven analysis of climate data.

References

Bauer-Marschallinger B, **Dorigo WA**, **Wagner W and Van Dijk AIJM** (2013) How oceanic oscillation drives soil moisture variations over mainland Australia: An analysis of 32 years of satellite observations. *Journal of Climate* 26(24).

Bellomo K, **Angeloni M**, **Corti S and von Hardenberg J** (2021) Future climate change shaped by inter-model differences in Atlantic meridional overturning circulation response. *Nature Communications* 12(1), 1–10.

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

Bridgman HA and Oliver JE (2014) The Global Climate System: Patterns, Processes, and Teleconnections. New York: Cambridge University Press.

Collins M, **Knutti R**, **Arblaster J**, **Dufresne J-L**, **Fichefet T**, **Friedlingstein P**, **Gao X**, **Gutowski WJ**, **Johns T**, **Krinner G**, et al. (2013) Long-Term Climate Change: Projections, Commitments and Irreversibility. IPCC.

DelSole T and Tippett MK (2009a) Average predictability time. Part I: Theory. *Journal of the Atmospheric Sciences 66*(5), 1172–1187.

DelSole T and Tippett MK (2009b) Average predictability time. Part II: Seamless diagnoses of predictability on multiple time scales. *Journal of the Atmospheric Sciences 66*(5), 1188–1204.

Demo N, **Tezzele M and Rozza G** (2018) PyDMD: Python dynamic mode decomposition. *Journal of Open Source Software 3*(22), 530.

Deng J, **Dai A and Xu H** (2020) Nonlinear climate responses to increasing CO2 and anthropogenic aerosols simulated by CESM1. *Journal of Climate 33*(1), 281–301.

Diaz HF, **Hoerling MP and Eischeid JK** (2001) ENSO variability, teleconnections and climate change. *International Journal of Climatology: A Journal of the Royal Meteorological Society 21*(15), 1845–1862.

Drijfhout S, **Van Oldenborgh GJ and Cimatoribus A** (2012) Is a decline of AMOC causing the warming hole above the North Atlantic in observed and modeled warming patterns? *Journal of Climate 25*(24), 8373–8379.

Erichson NB, **Mathelin L**, **Kutz JN and Brunton SL** (2019) Randomized dynamic mode decomposition. *SIAM Journal on Applied Dynamical Systems 18*(4), 1867–1891.

Eyring V, **Bony S**, **Meehl GA**, **Senior CA**, **Stevens B**, **Stouffer RJ and Taylor KE** (2016) Overview of the Coupled Model Intercomparison Project phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development 9*(5), 1937–1958.

Forootan E, **Khandu K**, **Awange Jl**, **Schumacher M**, **Anyah RO**, **Van Dijk AIJM and Kusche J** (2016) Quantifying the impacts of ENSO and IOD on rain gauge and remotely sensed precipitation products over Australia. *Remote Sensing of Environment 172*.

Gavrilov A, **Kravtsov S and Mukhin D** (2020) Analysis of 20th century surface air temperature using linear dynamical modes *Chaos: An Interdisciplinary Journal of Nonlinear Science 30*(12).

Gehne M, **Kleeman R and Trenberth KE** (2014). Irregularity and decadal variation in ENSO: A simplified model based on principal oscillation patterns. *Climate Dynamics*, *43*(12):3327–3350.

Ghil M, **Allen M**, **Dettinger M**, **Ide K**, **Kondrashov D**, **Mann ME**, **Robertson AW**, **Saunders A**, **Tian Y**, **Varadi F**, et al. (2002) Advanced spectral methods for climatic time series. *Reviews of Geophysics 40*(1), 3–10.

Gottwald GA and Gugole F (2020) Detecting regime transitions in time series using dynamic mode decomposition. *Journal of Statistical Physics 179*(5–6), 1028–1045.

Hannachi A, **Jolliffe IT and Stephenson DB** (2007) Empirical orthogonal functions and related techniques in atmospheric science: A review. *International Journal of Climatology 27*(9), 1119–1152.

Hasselmann K (1976) Stochastic climate models part I. Theory. *Tellus 28*(6), 473–485. https://doi.org/10.1111/j.2153-3490.1976.tb00696.x.

He C, **Clement AC**, **Cane MA**, **Murphy LN**, **Klavans JM and Fenske TM** (2022) A North Atlantic warming hole without ocean circulation. *Geophysical Research Letters 49*(19), e2022GL100420.

Hoesly RM, **Smith SJ**, **Feng L**, **Klimont Z**, **Janssens-Maenhout G**, **Pitkanen T**, **Seibert JJ**, **Vu L**, **Andres RJ**, **Bolt RM**, et al. (2018) Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the community emissions data system (CEDS). *Geoscientific Model Development 11*(1), 369–408.

Ichinaga SM, **Andreuzzi F**, **Demo N**, **Tezzele M**, **Lapo K**, **Rozza G**, **Brunton SL and Kutz JN** (2024) PyDMD: A Python package for robust dynamic mode decomposition. *arXiv preprint arXiv:2402.07463*.

Krake T, **Reinhardt S**, **Hlawatsch M**, **Eberhardt B and Weiskopf D** (2021) Visualization and selection of dynamic mode decomposition components for unsteady flow. *Visual Informatics 5*(3), 15–27.

Kutz JN, **Brunton SL**, **Brunton BW and Proctor JL** (2016) Dynamic mode decomposition: Data-driven modeling of complex systems. *SIAM*.

Leach NJ, **Jenkins S**, **Nicholls Z**, **Smith CJ**, **Lynch J**, **Cain M**, **Walsh T**, **Wu B**, **Tsutsui J and Allen MR** (2021) FaIRv2. 0.0: A generalized impulse response model for climate uncertainty and future scenario exploration. *Geoscientific Model Development 14*(5), 3007–3036.

Malik A, **Nowack PJ**, **Haigh JD**, **Cao L**, **Atique L and Plancherel Y** (2020) Tropical Pacific climate variability under solar geoengineering: Impacts on ENSO extremes. *Atmospheric Chemistry and Physics 20*(23), 15461–15485.

Meehl GA, **Senior CA**, **Eyring V**, **Flato G**, **Lamarque J-F**, **Stouffer RJ**, **Taylor KE and Schlund M** (2020) Context for interpreting equilibrium climate sensitivity and transient climate response from the CMIP6 earth system models. *Science Advances 6*(26), eaba1981.

Monahan AH (2001) Nonlinear principal component analysis: Tropical Indo–Pacific Sea surface temperature and sea level pressure. *Journal of Climate 14*(2), 219–233.

Navarra A, **Tribbia J and Klus S** (2021) Estimation of Koopman transfer operators for the equatorial Pacific SST. *Journal of the Atmospheric Sciences 78*(4), 1227–1244.

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

Penland C and Sardeshmukh PD (1995) The optimal growth of tropical sea surface temperature anomalies. *Journal of Climate 8*(8), 1999–2024.

Pithan F and Mauritsen T (2014) Arctic amplification dominated by temperature feedbacks in contemporary climate models. *Nature Geoscience 7*(3), 181–184.

Proctor JL, **Brunton SL and Kutz JN** (2016) Dynamic mode decomposition with control. *SIAM Journal on Applied Dynamical Systems 15*(1), 142–161.

Qasmi S (2023) Past and future response of the North Atlantic warming hole to anthropogenic forcing. *Earth System Dynamics 14*(3), 685–695.

Rantanen M, **Karpechko AY**, **Lipponen A**, **Nordling K**, **Hyvärinen O**, **Ruosteenoja K**, **Vihma T and Laaksonen A** (2022) The Arctic has warmed nearly four times faster than the globe since 1979. *Communications Earth & Environment 3*(1), 168.

Riahi K, **Van Vuuren DP**, **Kriegler E**, **Edmonds J**, **O'neill BC**, **Fujimori S**, **Bauer N**, **Calvin K**, **Dellink R**, **Fricko O**, et al. (2017) The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change 42*, 153–168.

Schölkopf B, **Smola A and Müller K-R** (1998) Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation 10*(5), 1299–1319.

Seland Ø, **Bentsen M**, **Olivié D**, **Toniazzo T**, **Gjermundsen A**, **Graff LS**, **Debernard JB**, **Gupta AK**, **He Y-C**, **Kirkevåg A**, et al. (2020) Overview of the norwegian earth system model (NorESM2) and key climate response of CMIP6 deck, historical, and scenario simulations. *Geoscientific Model Development 13*(12), 6165–6200.

Solomon A and Newman M (2012) Reconciling disparate twentieth-century Indo-Pacific ocean temperature trends in the instrumental record. *Nature Climate Change 2*(9), 691–699.

Takens F (2006) Detecting strange attractors in turbulence. *Dynamical Systems and Turbulence, Warwick 1980: Proceedings of a Symposium Held at the University of Warwick 1979/80*. Springer, pp. 366–381.

Timmermann A, **Oberhuber J**, **Bacher A**, **Esch M**, **Latif M and Roeckner E** (1999) Increased El Niño frequency in a climate model forced by future greenhouse warming. *Nature 398*(6729), 694–697.

Tsonis AA, **Swanson KL and Wang G** (2008) On the role of atmospheric teleconnections in climate. *Journal of Climate 21*(12), 2990–3001.

Tu JH (2013) Dynamic Mode Decomposition: Theory and Applications. *PhD thesis*, Princeton University.

Volkov DL (2014) Do the North Atlantic winds drive the nonseasonal variability of the Arctic Ocean sea level? *Geophysical Research Letters 41*(6).

Wa B, **Wu Z**, **Li J**, **Liu J**, **Chang C-P**, **Ding Y and Wu G** (2008) How to measure the strength of the East Asian summer monsoon. *Journal of Climate 21*(17), 4449–4463.

Wang W, **Anderson BT**, **Kaufmann RK and Myneni RB** (2004) The relation between the North Atlantic oscillation and SSTs in the North Atlantic basin. *Journal of Climate 17*(24), 4752–4759.

Watanabe M, **Kang SM**, **Collins M**, **Hwang Y-T**, **McGregor S and Stuecker MF** (2024) Possible shift in controls of the tropical Pacific surface warming pattern. *Nature 630*(8016), 315–324.

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

Wills RCJ, **Sippel S and Barnes EA** (2020) Separating forced and unforced components of climate change: The utility of pattern recognition methods in large ensembles and observations. *Variations 18*(2), 1–10.

Xiong W, **Ma M**, **Sun P and Tian Y** (2023) Koopmanlab: A PyTorch module of Koopman neural operator family for solving partial differential equations. *arXiv preprint* arXiv:2301.01104.

Zickfeld K and Herrington T (2015) The time lag between a carbon dioxide emission and maximum warming increases with the size of the emission. *Environmental Research Letters 10*(3), 031001.

A. From DMD/LIM to DMDc

In DMD (Tu, 2013) (a.k.a. LIM (Penland and Sardeshmukh, 1995)), the model is $dx/dt = Cx + \varepsilon$ where ε is noise. Without the noise term and a time lag $\tau = 1$, this translates to

$$x(t+1) = Ax(t) = \exp(C)x(t). \tag{A.1}$$

Generally, LIMs use this model where the features of $x(t)$ are the principal components of the data rather than the raw data. LIMs are used as naive estimators of the forced response from climate variables via the "least damped mode" of A (Solomon and Newman, 2012). This corresponds to the mode of A associated with the largest, entirely real eigenvalue. If the forced response is in the least damped mode, the internal variability is thus contained in the other modes of A.

We will now re-capitulate (projected) DMD (Tu, 2013). Define the snapshot matrices as before, and take the rank M SVD: $X \approx U\Sigma V^*$. DMD assumes that the data satisfy $U^* x_{n+1} = \tilde{A} U^* x_n$ where $\tilde{A} = U^* \tilde{A} = U^* X' V \Sigma^{-1}$ with eigenvectors $\tilde{A} W = W \Lambda$. We use the (projected) dynamic modes, $\Phi = UW$. Define the amplitudes as the least squares approximation $\Phi \alpha \approx x_1$ and the scaled modes as $\xi_m = \varphi_m \alpha_m$. As with DMDc, the scaled modes make up the autoregressive contribution. Multiplication of a scaled mode by its associated eigenvalue, λ_m evolves the scaled mode one-time step into the future. The DMDc model generalizes LIM/DMD by including a linear control term instead of noise. The DMDc model with a time lag of $\tau = 1$ is $\hat{U}^* x_{n+1} = \tilde{A} \hat{U}^* x_n + \tilde{B} y_n$. A summary of these differences is in Table A1.

Table A1. *The spatial and temporal patterns from DMD and DMDc for climate analysis. Both methods have spatial and temporal patterns for the autoregressive component (scaled modes of A). DMDc adds spatial and temporal patterns for the emissions contribution by analyzing the matrix $\check{B} \approx B$ (Figure 3) using the time-lagged radiative forcing stored in y*

	Model for $x(t+1)$	Parameters	Spatial patterns	Temporal patterns
DMD	$Ax(t)$	A	ξ_m	λ
DMDc	$Ax(t) + By(t)$	A & B	ξ_m & $\check{b}_s^{(j)} \, \check{y}(j)_s$	λ_m & $y(1),\ldots,y(J)$

B. Emissions contribution

In the upper left panel of Figure 7, we plot the mean entry of $\check{B}y_n$. Let $\mathcal{N}$ be the set of all n corresponding to years 2051 to 2100. In the upper right panel, we plot the spatial mean entry of

$$\sum_{n \in \mathcal{N}} \check{b}_s^{(j)} \, y(n-j+1)_s \tag{B.1}$$

for $j = 1, 2, \ldots, 30$ years in the past and emissions agent s. In the bottom panel, we visualize the distribution of the entries of $\check{b}_s^{(j)} \, y(n-j+1)_s$ over $j = 1, 2, \ldots, 30$ years in the past and a fixed emissions agent s. In Figure 8, we plot the distribution of the entries of $\check{b}_s^{(j)} \, y(n-j+1)_s$ for s corresponding to SO_2 and certain landmass regions. In Figure 9(a) we plot

$$\sum_{j=1}^{30} \frac{1}{50} \sum_{n \in \mathcal{N}} \check{b}_s^{(j)} \, y(n-j+1)_s. \tag{B.2}$$

In Figure 9(b) we plot

$$\frac{1}{50} \sum_{n \in \mathcal{N}} \check{b}_s^{(j)} \, y(n-j+1)_s \tag{B.3}$$

for $j = 1$ and $j = 30$.

Cite this article: Mankovich N, Bouabid S, Nowack P, Bassotto D and Camps-Valls G (2025). Analyzing climate scenarios using dynamic mode decomposition with control. *Environmental Data Science*, 4: e16. doi:10.1017/eds.2025.8