SANDIA REPORT

SAND2024-xxxx
Printed September 2024

Sandia National Laboratories

CLimate Impact: Determining Etiology thRough pAthways (CLDERA)

Diana Bull, Kara Peterson, Lyndsay Shand, Laura Swiler, Irina Tezaur

LEADERSHIP TEAM: Diana Bull, Kara Peterson, Lyndsay Shand, Laura Swiler, Irina Tezaur, Ben K. Cook, Andrew Salinger, Clare Amann, Bernadette Watts, Rob Leland

CONTRIBUTORS: Luca Bertagna, Hunter Brown, Meredith Brown, Mauricio Campos, Max Carlson, Kenny Chowdhary, Joseph Crockett, Warren Davis, Thomas Ehrmann, Robert Garrett, Katherine Goode, Mamikon Gulian, Carole Hall, Graham Harper, Joseph Hart, James Hickey, Benjamin Hillman, Brent Houchens, Jose Gabriel Huerta, Daniel Krofcheck, Justin Li, Indu Manickam, Kellie McClernon, Audrey McCombs, Jeffrey Nichol, Matthew Peterson, Daniel Ries, Mark A. Smith, Andrea Staid, Andrew Steyer, James Derek Tucker, Benjamin Wagman, Jerry Watkins, Christopher Wentland, Everett Wenzel, Robert Michael Weylandt, Andrew Yarger

UNIVERSITY PARTNERS: Christiane Jablonowski and Joseph Hollowed (University of Michigan), Xiaohong Liu and Allen Hu (Texas A&M University), Bo Li and Sam Shi-Jun (University of Illinois), Konstantinos Tsigaridis and Ram Singh and Kate Marvel (Columbia University)

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185
Livermore, California 94550

Issued by Sandia National Laboratories, operated for the United States Department of Energy by National Technology & Engineering Solutions of Sandia, LLC.

Available to DOE and DOE contractors from

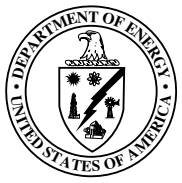
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@osti.gov
Online ordering: http://www.osti.gov/scitech

Available to the public from

U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Road
Alexandria, VA 22312

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.gov
Online order: https://classic.ntis.gov/help/order-methods



ABSTRACT

Climate impacts have broad economic, health, political, and national security ramifications. Societally relevant impacts are typically farther downstream, are the product of multiple interacting processes, and can arise over small regions and timeframes because their sources are short-term and localized. Short-term forcings (as can be seen in volcanic eruptions, climatic tipping points (e.g., the collapse of rainforests or the disappearance of sea ice), or in increasingly plausible climate interventions) fundamentally possess low signal-to-noise and could benefit from accounting for the multiple conditional processes through which a downstream impact arises. Under the Grand Challenge LDRD CLDERA (CLimate impacts: Discovering Etiology thRough pAthways), we have developed tools to enable downstream impact attribution from geographically and temporally localized source forcings in the climate. CLDERA developed methods that can distinguish *how* a localized source drives the climate system to respond with particular impacts. The *how* is embodied in pathways – the spatio-temporally evolving chain of physical processes that connects a source to a series of increasingly distant impacts. Novel analytic methods in pursuit of downstream impact attribution were developed and demonstrated on simulations and observations of the 1991 eruption of Mt. Pinatubo in the Philippines. As described within this report we have

- developed stratospheric expertise and aerosol modeling capabilities in E3SM,
- created original methods to detect and model pathways from source-to-impact, and
- advanced climate attribution through novel methods, cases, and approaches.

Further, CLDERA developed a tiered verification process consisting of controlled datasets to prototype, verify, and refine the original method development. CLDERA increased Sandia's footprint in the climate analytics community and developed new climate collaborations whilst also creating a cadre of climate analysts at Sandia. The products from CLDERA have been extensive with a total of 9 journal articles published, 12 articles submitted and under review, and an additional 8 articles in preparation[1]. We have produced 1750 simulated years[2] and developed 9 code-bases. This report details these accomplishments and serves as a summary of the work completed during the CLDERA Grand Challenge.

Recommended Citation:

Diana Bull, Kara Peterson, Lyndsay Shand, Laura Swiler, Irina Tezaur, Ben K. Cook, Andrew Salinger, Clare Amann, Bernadette Watts, Rob Leland, Luca Bertagna, Hunter Brown, Meredith Brown, Mauricio Campos, Max Carlson, Kenny Chowdhary, Joseph Crockett, Warren Davis, Thomas Ehrmann, Robert Garrett, Katherine Goode, Mamikon Gulian, Carole Hall, Graham Harper, Joseph Hart, James Hickey, Benjamin Hillman, Brent Houchens, Jose Gabriel Huerta, Daniel Krofcheck, Justin Li, Indu Manickam, Kellie McClernon, Audrey McCombs, Jeffrey Nichol, Matthew Peterson, Daniel Ries, Mark A. Smith, Andrea Staid, Andrew Steyer, James Derek Tucker, Benjamin Wagman, Jerry Watkins, Christopher Wentland, Everett Wenzel, Robert Michael Weylandt, Andrew Yarger, Christiane Jablonowski, Joseph Hollowed, Xiaohong Liu, Allen Hu, Bo Li, Sam Shi-Jun, Konstantinos Tsigaridis, Ram Singh, and Kate Marvel. "CLimate impact: Determining Etiology through Pathways (CLDERA)." Technical Report. SAND2024-XXX. September 2024, Sandia National Laboratories, Albuquerque NM.

[1] Our website `https://www.sandia.gov/cldera/` will continue to be updated with publication information

[2] Data from the full E3SMv2-SPA simulation campaign including pre-industrial control, historical, and Mt. Pinatubo ensembles will be hosted at Sandia National Laboratories with location and download instructions announced on `https://www.sandia.gov/cldera/e3sm-simulations-data/` when available.

This page intentionally left blank.

ACKNOWLEDGEMENT

The opportunity for this Grand Challenge Research was made possible by the LDRD Office and the Chief Research Office. They provided additional, and much appreciated, support from Jeff Dorman and Perspectives (Carol Jacobus, Whitney Wegner, and Greg Wong).

We received guidance, feedback, and encouragement from an external advisory board who graciously met with the team three times. Dave Bader was the chair, and members included: Chris Bretherton, Omar Ghattas, Jordan Kazemi, Jean-Francois Lamarque, Doug Nychka, Karen Rosenloft, Steve Scharre, and Alex Wong.

We received additional guidance and encouragement from an internal advisory board over the course of three meetings. Members included: Mark Taylor, Amir Mohagheghi, Kathy Simonson, Habib Najm, Peter Davies, and Tim Trucano.

We were also given the opportunity to interact with and receive support from Sandia's Earth Science Review Board and the Computing & Information Sciences Review Board.

Additional support from Jim Foucar for help in building E3SM on our HPC, Steve Monk for reserved nodes on Boca and for help in building Flight, and the HPC resource team writ large for all of their support. Diana Hackenburg helped to build and maintain our website as well as craft communicable material.

Each of the leadership teams managers were instrumental in helping: Tom Lowry, Tim Wildey, Aubrey Eckert, Jen Gaudioso, and Jerry McNeish. As well as the administrative support staff with their support with team meetings, maintaining contracts and much more (Cookie Santamaria, DeVonna Flanery, Laura Bravo, Catherine McGuire, Michelle Vaughn).

CLDERA was envisioned with the support of many other individuals including: Jim Stewart, Lori Parrott, Erik Webb, Dave Sandison, S. Scott Collis, Susan Seestrom, Carol Adkins, Amy Powell, Erika Roesler, Lauren Wheeler, Elizabeth Keller, Oksana Guba, Pete Bosler, Andrew Bradley, and many others who we hope will forgive us for not adding them here.

This research used resources of the National Energy Research Scientific Computing Center (NERSC), a Department of Energy Office of Science User Facility using NERSC award BER-ERCAP0026535. Data from the full E3SMv2-SPA simulation campaign including pre-industrial control, historical, and Mt. Pinatubo ensembles will be hosted at Sandia National Laboratories with location and download instructions announced on `https://www.sandia.gov/cldera/e3sm-simulations-data/` when available.

This page intentionally left blank.

CONTENTS

This page intentionally left blank.

LIST OF FIGURES

This page intentionally left blank.

LIST OF TABLES

This page intentionally left blank.

ACRONYMS & DEFINITIONS

ACC Anthropogenic Climate Change

AOA Age of Air tracer

AOD Aerosol Optical Depth

BAE Bayesian Approximation Error approach

BDC Brewer-Dobson circulation

BURDENS04 Vertically integrated SO4 concentration (sulfate aerosols)

CaStLe Causal Space-Time Stencil Learning

CLDERA CLimate impacts: Discovering Etiology thRough pAthways Grand Challenge LDRD

CT CLDERA-Tools, available at `https://github.com/sandialabs/cldera-tools`

DAG Directed Acyclic Graph

DLM Dynamic Linear Model

EAST Explicit Aerosol Source Tagging

EOF Empirical Orthogonal Function

E3SM Energy Exascale Earth System Model

E3SMv2-SPA E3SM version 2, with Stratospheric Prognostic Aerosol implementation

ENSO El Niño Southern Oscillation

ESN Echo State Network

FLNT Net longwave flux as top of the model

FSDS Downwelling solar flux at surface

FSDSC Clearsky downwelling solar flux at surface

HSW-V Held-Suarez-Williamson-volcano configuration of E3SM

GHG Greenhouse Gas

INLA Integrated nested Laplace approximation, a data fusion method

IPCC Intergovernmental Panel on Climate Change

ISML In-Situ Machine Learning for Intelligent Data Capture

ITCZ Inter-Tropical Convergence Zone

MERRA-2 Reanalysis data: Modern-Era Retrospective Analysis for Research and Applications, Version 2

NAO North Atlantic Oscillation

NCAR National Center for Atmospheric Research

NLP Natural Language Processing

NOAA National Oceanic and Atmospheric Administration

PFI Permutation Feature Importance

QBO Quasi Biennial Oscillation

QOI Quantity of Interest

RBF Radial Basis Function

RFR Random Forest Regression

SAI Stratospheric aerosol injection

SHAP SHapley Additive exPlanation feature importance method

SO2 Sulfur dioxide

SPV Stratospheric Polar Vortex

stZFI spatio-temporal Zero Feature Importance

T50 Temperature at 50 hPa (in stratosphere)

TREFHT Reference Height Temperature (near surface)

VAR Vector Autoregressive Model

ZFI Zeroed Feature Importance

1. INTRODUCTION

Climate impacts have broad economic [166, 20], health [104], political [200], and national security ramifications [120]. Beyond advancing science, discovery and quantification of the impact's source are needed to guide policy decisions that may ameliorate or establish liability for undesirable impacts, as well as guide U.S. national security posture for impacts that may require response. However, in highly coupled complex nonlinear systems, like the Earth's climate, traditional understanding of causal relationships does not apply and many drivers may contribute to a detected change or impact.

Detection establishes that a change has occurred while attribution ascribes a source as the cause of that change [70]. Starting with assessment report 1 in 1990 [1], detection and attribution (DA) has been a 30-year focus for the International Panel on Climate Change (IPCC). Much progress has been made in climate change attribution that strives to identify impacts of long-term anthropogenic climate change (ACC) through greenhouse gas emissions. The methods have addressed three classes of attribution [18]: (i) attribution of changes in climate state variables to anthropogenic activity (e.g., global mean temperature); (ii) assigning relative responsibility to different sectors, activities, and entities that contribute emissions; and (iii) evaluation of the relative frequency and severity of extreme events due to climate change. These attribution methods have two common limiting characteristics. First, they require long-term records to identify a signal strength above natural variability; in effect they assume the *only* causal factors are independent and include only internal variability and the external forcing (e.g. ACC or solar radiation modification (SRM)). Second, attribution requires positing source-impact pairs (as done in (i) and (ii)) that can only notionally, not quantitatively, link a source through a series of cascading impacts (as done in (iii) and sometimes (i)). By limiting evaluation to a single source-impact pair, these studies necessarily exclude other important quantities which could increase certainty in attribution.

These characteristics make current attribution methods unsuitable for signals that are mired with high variability like societally relevant impacts which are typically farther downstream or impacts arising over small regions and timeframes because their sources are short-term and localized. Downstream impacts like droughts, flooding, or crop yields are the product of multiple interacting processes, and thus may require conditional attribution techniques to be successful (i.e. violating both of the characteristics highlighted above). Short-term forcings (as can be seen in volcanic eruptions, climatic tipping points (e.g., the collapse of rainforests or the disappearance of sea ice), or in increasingly plausible climate interventions like SRM) fundamentally possess low signal-to-noise and could benefit from accounting for the multiple conditional processes through which a downstream impact arises.

The goal of CLDERA is to enable conditional attribution in the climate by developing quantitative relationships (through multiple variables over space and time) between a climate forcing and its downstream impacts.

1.1. Approach

A primary focus of CLDERA is to develop methods that can distinguish *how* a localized source drives the climate system to respond with particular impacts. The *how* is embodied in pathways – the spatio-temporally evolving chain of physical processes that connects a source to a series of increasingly distant impacts. Pathways combine evidence from multiple processes (or quantities of interest) to strengthen connection between source and impact. The causal nature of these relationships can be explored by employing the epidemiological framing present in the Bradford Hill Criteria [74]: evaluating the sensitivity of response to source size, establishing the consistency of the impact across scale and subject to mediators, and through demonstration that the impact is specific to the source (i.e. through evaluation of other variables). With knowledge of pathways, specific methods to then attribute the magnitude of source forcing to the detected impacts were possible.

Novel analytic tools were developed and demonstrated on simulations and observations of the 1991 eruption of Mt. Pinatubo in the Philippines. Considered the 20th century's second-largest volcanic eruption [122], this event caused global climatic impacts, including decreases in near-surface temperatures [132, 167], lower stratospheric temperature increases [93], precipitation [49, 179, 60, 7], and global sea-level [25], as well as increases in cirrus cloud cover [100, 197] and diffusivity of incoming radiation [147, 140]. CLDERA focused on the surface and stratospheric temperature impacts as well as on impacts to agricultural productivity.

CLDERA tailored the US DOE's flagship Earth System Model (ESM), the Energy Exascale Earth System Model (E3SM) [55], to simulate the chemical and microphysical evolution of the volcanically erupted sulfur-dioxide gas into sulfate aerosols in the stratosphere [13]. Ensemble analyses were used to evaluate the susceptibility of a pathway to initial conditions. Results from the simulation campaign are presented in Section 3. Much research was performed under CLDERA to advance the scientific understanding of stratospheric processes within E3SM and to evaluate the effects of Mt. Pinatubo debated and further downstream impacts as presented in Chapter 7.

This simulation output and observations from the eruption were used to develop computational analysis tools specifically designed to elucidate and represent pathways between the Mt. Pinatubo forcing and a series of increasingly distant, but well documented, impacts in the Simulated Pathways and Observed Pathways Thrusts. This is a source-to-impact forward mapping (light blue or red cones) as seen in Figure 1-1. A key part of these approaches focused in dimension-reduced approximations of the system dynamics used to deepen understanding of causal relationships. The Simulated Pathway results are presented in Chapter 4 and the Observed Pathway results are presented in Chapter 5.

The Attribution Thrust is retrospective by design, looking to ascribe the experienced impact to a source magnitude. This is the impact-to-source inverse direction (green cone) as seen in Figure 1-1. The Attribution Thrust results are presented in Chapter 6.

Table 1-1 provides a summary description of the various methods developed in CLDERA for pathway identification and verification, detection of significant changes, and attribution of those changes. These methods are described in more detail throughout this report, especially in Chapters 4, 5, and 6.

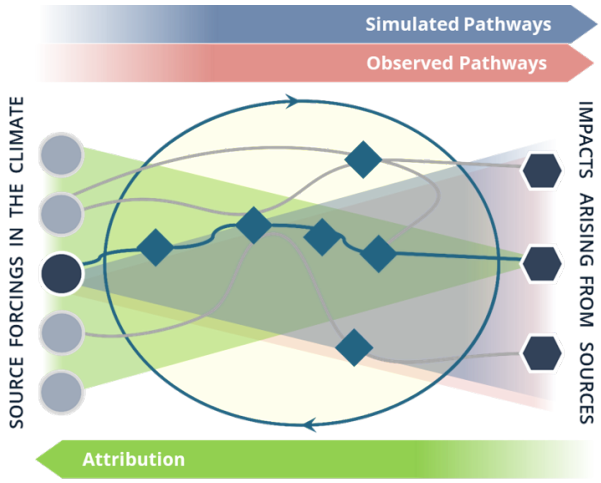


Figure 1-1. Representation of pathways (connective lines) from source to impacts with process-nodes shown as diamonds. Simulated and observed pathways will be traced to impacts (left to right); resulting data will then be used to attribute impacts back to the source (right to left).

Method in CLDERA	Brief description	Thrust
Random Forest Regression (RFR)	Generate feature pathway networks using RFR feature importance analysis	Simulated Pathways
Clustering	Detect changes by finding & tracking non-stationary variable clusters	Simulated Pathways
Multivariate Mining	Identify pathways by mining frequent multivariate cluster evolutions and extracting meaningful sequences of events.	Simulated Pathways
Profiling	Dynamically detect and trace pathways through the E3SM software as the software executes (*in situ*)	Simulated Pathways
Echo State Network (ESN)	Quantify variable relationships and their evolution in time and space, as captured by a black-box machine learning model (ESN)	Observed Pathways
Dynamic Linear Model (DLN)	Characterize multivariate relationships in terms of correlations which are allowed to vary across space and time.	Observed Pathways
Space-Time Changepoint Detection	Detect changepoint in climate data over time and space as a function of distance from an event.	Observed Pathways
Functional Changepoint Detection	Detect changepoint in climate data from functional measurements over time and space.	Observed Pathways
Data Fusion	Fuse data taken at different spatial and temporal resolutions.	Observed Pathways
Inverse Optimization	Identify source characteristics given observations of downstream impacts	Attribution
Causal Discovery	build a network that represents a system's spatiotemporal causal dynamics	Attribution
Multistep Conditional Attribution	Identify fingerprints (features) and a multi-step approach which can detect and attribute downstream (noisy) impacts from localized source forcings	Attribution

Table 1-1. Overview of methods developed and investigated under the CLDERA project.

Method robustness is central to CLDERA's design and at the urging of our advisory boards (Internal and External), CLDERA implemented a tiered verification approach. We generated verification datasets with known and controlled patterns to quantitatively evaluate the efficacy of the data analytic solutions. Synthetic datasets such as these can be used as a basis for quantitatively evaluating methods which attempt to extract spatial and temporal patterns. Tiered verification, which uses benchmark datasets of increasing complexity, was used by a variety of CLDERA methods to identify and address weaknesses where appropriate, and provide quantitative evaluation of the method accuracy before being applied on fully coupled ESM simulations. Synthetic and idealized formulations have been constructed to test the accuracy and robustness of the tools

before graduating and proving their capabilities on fully complex Mt. Pinatubo pathways. Tiered verification will be discussed in Chapter 2.

1.2. Outcomes

We highlight important outcomes from CLDERA:

- CLDERA enabled multi-step attribution in the climate through development of new methods that quantitatively assess pathways between a climate forcing and its downstream impacts.
- Under CLDERA, we developed a stratospheric prognostic aerosol capability in E3SM (v2-SPA) which allows for evolution of volcanic aerosols and tagging and diagnostics of various types of aerosols (volcanic, anthropogenic, etc.)
- We performed a massive E3SMv2-SPA simulation campaign, generating 400+ TB of data encompassing source magnitude varying ensembles representing 1750 simulated years[1].
- 7 of 9 code-bases have been copyrighted and released with the remaining two under copyright assertion.
- CLDERA increased Sandia's footprint in the climate analytics community and developed new climate collaborations.
- CLDERA created a cadre of climate analysts, with a team of more than 40 Sandians working across three divisions and four universities. This represented significant workforce development, as many of the CLDERA team had not worked in climate science previously.
- CLDERA developed a tiered verification process that can be utilized in other domains. Simplified and controlled data sets were developed with key characteristics of the multi-step pathways. These datasets were used to prototype, verify, and refine the CLDERA methods.
- As of the time of publication of this report, we have published 9 journal articles, submitted 12 articles which are under review, and are preparing an additional 8 articles[2]. Additionally, CLDERA staff have given 79 conference presentations and posters and 17 invited talks.

CLDERA has advanced climate attribution science by identifying impacts from localized sources and demonstrating current limits of attribution with respect to noisy downstream impacts. Further, this work has developed state-of-the-art machine learning, reduced order modeling, and statistical tools to elucidate and represent pathways, which are a foundational means for understanding dependent causal-like relationships in many complex nonlinear systems.

New methods and tools developed through CLDERA offer a framework that can be translated to other localized or episodic sources such as large wildfires, changes to ocean currents (e.g., Atlantic Meridional Overturning Circulation), or climate interventions (e.g., stratospheric aerosol injection).

[1] Data from the full E3SMv2-SPA simulation campaign including pre-industrial control, historical, and Mt. Pinatubo ensembles will be hosted at Sandia National Laboratories with location and download instructions announced on `https://www.sandia.gov/cldera/e3sm-simulations-data/` when available.

[2] Our website `https://www.sandia.gov/cldera/` will continue to be updated with publication information

The CLDERA methods could also be extended to understand and attribute impacts from tipping points in the climate system.

With these new capabilities, Sandia will be able to:

- Advance climate science. This includes expanding the work from volcanic eruption impacts to new problem sets (e.g., impacts from tipping points). It also includes improving the predictive capabilities of ESMs using machine learning tools to formalize pathways and cross-validate observational and simulation pathways.
- Analyze climate impacts. CLDERA methods support detection and attribution of impacts. These tools can help advance the knowledge of the magnitude, timing, and location of ensuing impacts, including an understanding of the minimum source magnitude necessary to identify impacts.
- Motivate sound climate actions. CLDERA methods can provide decision support using ESMs, potentially helping inform policies, agreements, regulations, and treaties. They could also be used to help develop requirements for monitoring and measurement systems.

This page intentionally left blank.

2. TIERED VERIFICATION

The novel methods investigated in CLDERA are at the interface of statistics, machine learning, and data science. These data analytic methods are a new area for much of the climate modeling community with interpretability and explainability challenges. The large-scale spatio-temporal climate data will challenge the capabilities of these data analytic methods.

A key framework to quantitatively assess performance of Earth System Models, ESMs, is the Model Intercomparison Projects, MIPs, for which simulation and output design are specified ensuring that spread in outcome between models is due to the characteristics of the models themselves. The MIPs enable quantitative inter-model assessment of performance by standardizing simulation inputs/design and outputs. They also serve to direct and focus refinements of models assumptions, forcing response, parameterizations, and biases to achieve better model performance.

Just as ESMs embed key assumptions, have tunable parameterizations, and possess biases, so too do many of the advanced tools used on climate datasets. MIPs play an important role for climate models: they help climate modelers understand and expose issues relating to the key assumptions, tunable parameterizations, and model biases. In a similar fashion, we argue that benchmark datasets be developed to test the advanced spatiotemporal statistical and ML methods being developed for climate datasets. For example, the most common technique used to identify natural modes in the climate, PCA, has been shown to only be effective for one dominant mode [39]. [107] showed the most common explainability technique for neural networks (gradient or smooth gradient methods) failed when applied to an attribution benchmark dataset. [154] created a causality benchmark platform in which the causal structure is known with high confidence to deepen trust in method performance.

Advanced analytic techniques are increasingly applied to extract spatial and temporal patterns from climate datasets. Traditional pattern identification techniques, like principal component analysis mentioned above, also struggle with climate datasets, requiring preprocessing to obtain interpretable patterns. Data analytic methods require significant expertise to produce interpretable results, in part because there is a lack of canonical datasets with known structure and dynamics of increasing realism and fidelity to test their efficacy.

In the first year of CLDERA, we recognized the need to develop a set of verification datasets for which patterns can be controlled such that the efficacy of the methods can be quantitatively evaluated. We set up a process to test the CLDERA methods on benchmark datasets of increasing complexity, starting with relatively simple datasets with known relationships between variables and working up the verification hierarchy to datasets of increasing complexity and realism. This process is called "tiered verification." The tiers for CLDERA are explained in Figure 2-1.

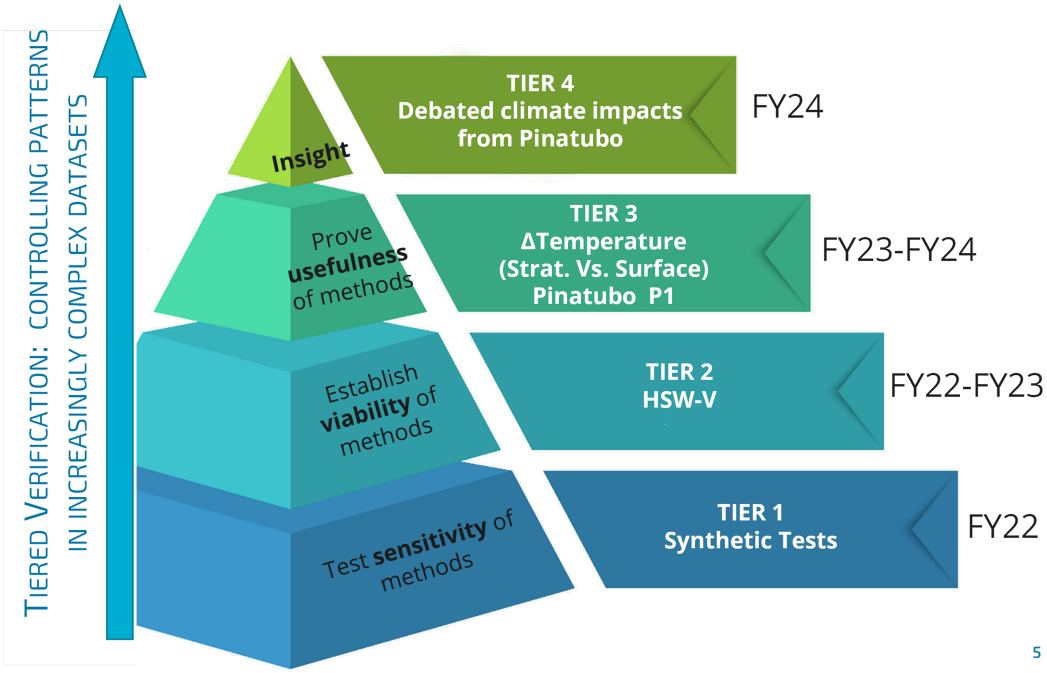


Figure 2-1. Representation of the levels of increasing complexity on which CLDERA methods were demonstrated. The base tier, Tier 1, was designed to test the sensitivity of the methods and enable proof-of-concept in operation. Tier 2 looked to establish the viability of the methods through an idealized yet realistic representation of the eruption. Tier 3 was designed to prove the usefulness of the methods through application on the fully coupled system. Finally insight can be achieved in Tier 4 with application of the methods on more noisy, downstream, debated impacts.

CLDERA developed simplified and controlled data sets to support tiered verification of the CLDERA methods, as shown in Figure 2-1, with key characteristics of the multi-step attribution problem to refine methods, explore sensitivities, establish viability, and prove usefulness of advanced methods. One advantage of a simpler problem is to understand why a solution may be wrong or right. Each tier offers something new towards the goal. Verification is designed to demonstrate capability and lower the risk of using advanced analytic methodologies (which have not usually been tested extensively on geospatial data) for more complex problems. Our efforts toward producing verification benchmarks are a first step. As the climate community increasingly relies on advanced statistical inference and machine learning approaches, it is critical to develop common testbeds with large numbers of benchmarks and synthetic datasets which can be widely shared and used to understand the accuracy of the inferences about spatial patterns and relationships.

A brief description of each Tier level is given below.

2.1. Tier 1: Synthetic

The purpose of Tier 1 is to establish proof-of-concept for the various methods developed under CLDERA. Synthetic datasets were used to: build methodological components, identify performance characteristics of methods, test sensitivity of the methods, compare and contrast the CLDERA approaches with state-of-the-art approaches, and highlight tradeoffs between methods.

In particular, for the inverse optimization and functional changepoint methods, CLDERA needed to build the methdological components and ensure that they performed properly. The dynamic linear model used synthetic data to test dimension reduction performance, and random forest regression used Tier 1 to verify the role of intermediate steps. The CaStLe approach used synthetic datasets for comparison with a state-of-the-art causal algorithm, and echo state neworks used Tier 1 dataset to select between explainable/interpretable approaches and more general approaches which are not as explainable. Overall, the Tier 1 synthetic datasets provided a helpful learning environment for the CLDERA team to test out and prove feasibility of the methods. Examples of synthetic datasets are outlined below.

2.1.1. Coupled Equations

A simple example of a verification problem to exhibit causal dynamics is a set of coupled equations that show variables related to other variables at previous time points. Such a process is called an autoregressive model; one with multiple variables as shown in Equation 2.1 is called a Vector AutoRegressive model (VAR) in statistical terminology. The equations also define a structural causal model in the terminology of causal dynamics. As a simple example, consider the following VAR:

$$\begin{aligned} W_t &:= 0.9W_{t-1} + \eta_t^W \\ X_t &:= 0.8X_{t-1} + 0.4W_{t-1} + 0.2Z_{t-3} + \eta_t^X \\ Y_t &:= 0.5Y_{t-1} + 0.2X_{t-2} + \eta_t^Y \\ Z_t &:= 0.6Z_{t-1} + 0.3Y_{t-1} + \eta_t^Z \end{aligned} \tag{2.1}$$

where each $\eta \sim \mathcal{N}(0, 1)$ is IID Gaussian noise. These relations form a causal model for simulated realizations of this process.

The model in Equation 2.1 can easily be extended to include space: the variables W, X, Y, and Z can have subscripts for both space and time. In this way, stochastic spatial processes with a known trend and noise structure and known autoregressive relationships can be generated, representing multivariate series with known relationships over space and time.

Several of the methods developed under CLDERA used VARS equations for the first tier of verification. A nice example of developing a robust 2-D spatio-temporal data generator is described in Section 2.1 of [125]. One challenge when creating spatio-temporal data is that it often becomes non-stationary, but [125] describes a method to generate Gaussian neighborhood dependence matrices which result in stable spatio-temporal dynamics.

2.1.2. Causal Relationships

Moving up the verification hierarchy, we describe another type of verification problem: canonical partial differential equations (PDEs). VARs are discrete, linear models, whereas PDEs model continuous, nonlinear dynamics. One example that was used to verify the CaStLe method is Burgers' equation [19], a PDE model of advective and diffusive transport.

We define Burgers' equation in a 2D space as:

$$\frac{\partial u}{\partial t} + u(\alpha \frac{\partial u}{\partial x} + \beta \frac{\partial u}{\partial y}) = c(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}) + f \tag{2.2}$$

where α and β are advection coefficients for the x and y directions, respectively, c is the diffusion coefficient, and f is a forcing term. This formulation was used to instantiate a Gaussian *plume* in the plane, which advects in a specified angle and diffuses at a controlled rate.

Burgers' equation is a useful benchmark environment because it is used to model a broad class of physical phenomena, such as fluid mechanics, nonlinear acoustics, gas dynamics, and traffic flow. Any physical phenomena in which a quantity transports and diffuses in space can be modeled with Burgers' equation. Demonstrating performance with the model illustrates its applicability to a wide array of application areas. In CLDERA, CaStLe was used to find the continuous advection angle from gridded Burgers' equation simulation data.

2.1.3. Gaussian Plume

Another PDE is the Gaussian plume model, which is commonly used to model advection of a plume. The Gaussian plume model was developed primarily for verification of the inverse optimization approach (see Section 6.1 and references therein), but other CLDERA methods used these datasets for verification as well.

We briefly describe the plume model, to show how it was used to implement key features of the Mt. Pinatubo plume advection while still remaining somewhat stylized and abstracted.

The SO_2 Plume problem is as follows: we consider a model for atmospheric transport of sulfur dioxide SO_2 in a rectangular region in space parametrized by longitudinal position x_1 and altitude x_2. We denote the concentration by $u(x_1, x_2, t)$ at time t. We take $L_{x_1} = 200$ (km), $L_{x_2} = 20$ (km), and generate u as the solution to the equations

$$\begin{aligned} &\frac{\partial u}{\partial t} - \kappa \Delta u + \mathbf{v} \cdot \nabla u - S\mathbf{e}_2 \cdot \nabla u = R(u) + f && \text{on } \Omega \times [0, T] \\ &\nabla u \cdot \mathbf{n} = 0 && \text{on } \partial\Omega \times [0, T] \\ &u = 0 && \text{on } \Omega \times \{0\} \end{aligned}$$

where κ is the diffusion coefficient,

$$\mathbf{v} = (v_1(x_1, x_2, t), 0)$$

describes the wind, with

$$\begin{aligned} v_1(x_1, x_2, t) = &\left(1 + 0.1 \cos\left(\frac{2\pi t}{60}\right)\right) \\ &\times \left(1 + 0.2 \cos\left(\frac{6\pi x_1}{40}\right) - 0.1 \sin\left(\frac{4\pi x_1}{40}\right)\right)\left(0.25 + 3.75 \sin\left(\frac{\pi x_2}{20}\right)\right), \end{aligned}$$

the term

$$S\mathbf{e}_y = (0, S) = \left(0, \sqrt{\frac{8}{3}\frac{\rho_{SO_2}}{\rho_{\text{atmo}}}\frac{g}{C_s}r}\right)$$

describes the effect of the particles falling due to gravity with a terminal speed S, and

$$R(u) = -\gamma u$$

is the reaction function modeling chemistry with γ being the e-folding time. We generate data using the forcing term

$$f(t, x_1, x_2) = z(t) \exp\left(-100(x_1 - 5)^2\right) \exp\left(-\frac{(x_2 - 5)^2}{16}\right)$$

with forcing amplitude

$$z(t) = \lambda_1 \exp\left(-\lambda_2 t\right)$$

to model SO_2 injection.

Note that the Gaussian plume model specifically includes terms such as the forcing magnitude (size of the eruption), the atmospheric winds, and the reaction rate of SO_2 decay. These are key features we need to identify, model, and account for in the various pathway and attribution methods. This is why the Gaussian plume model was an important verification problem for CLDERA.

2.2. Tier 2: HSW-V

In the second tier, the CLDERA methods were tested and demonstrated on a simplified climate model (HSW-V, which is described below). Tier 2 involved a different phase of learning: the HSW-V was used to understand the sensitivity of CLDERA methods to data preparation, identify issues with windowing and normalization (e.g. for fingerprinting), and explore the use of spatial dimension reduction methods. Tier 2 led the inverse optimization team to identify that PCA was insufficient and they pivoted to using radial basis functions (RBFs) for spatial dimension reduction. For fingerprinting, Tier 2 also led to an understanding of the deficiencies of PCA and exploration of other features to be used in attribution. For RFR and profiling, Tier 2 highlighted that graph visualization with regional specificity presented significant visualization barriers.

The Tier 2 datasets also led to methodological developments. For changepoint methods, the HSW-V datasets pushed the team to develop epidemic changepoint methods (where both the start and end of a change are detected, not just the start). The HSW-V data was heavily used by the inverse optimization team to explore advection dynamics with radial basis functions whose locations progressed in time. The CaStLe method also explored spatial progression with HSW-V, and algorithms from natural language processing (NLP) [3] were employed within cluster multivariate mining, via association rule mining, on this data.

Since the HSW-V datasets were still somewhat abstracted, the lack of corresponding observational data encouraged novel "in silico" model experiments and generation of ensemble methodologies, especially for echo state networks and fingerprinting. Finally, the lack of a background signal (confounding background aerosols) was helpful in some methods (e.g. inverse optimization) but problematic for other methods (multivariate DLM, clustering and multivariate mining). Overall, Tier 2 provided a necessary and valuable step in the verification hierarchy for method testing and development.

2.2.1. Held-Suarez-Williamson-volcano

This section describes the Held-Suarez-Williamson-volcano (HSW-V) model configuration developed and utilized in CLDERA. The HSW-V configuration was run within the fully-coupled E3SM simulation but is an idealized configuration that is more akin to an "atmosphere only model" with no land topology and simplified boundary conditions from other components of E3SM. HSW-V was a critical element of the tiered verification which allowed testing of the statistical and ML tools for the detection of physical pathways in E3SM. Here we present a summary. The HSW-V model development is described in detail in [77]:

- Joseph P. Hollowed, Christiane Jablonowski, Hunter Y. Brown, Benjamin R. Hillman, Diana L. Bull, and Joseph L. Hart. "HSW-V v1.0: localized injections of interactive volcanic aerosols and their climate impacts in a simple general circulation model." Geoscientific Model Development, 17, 5913–5938, https://doi.org/10.5194/gmd-17-5913-2024, 2024.

These idealized climate simulations were built upon the so-called Held-Suarez-Williamson [194] model configuration with added volcanic forcings (called HSW-V, see [77]). Such an idealized

E3SM configuration with reduced complexity provided a testbed for the CLDERA analysis tools that were developed by the 'Simulated Pathways', 'Observed Pathways', and 'Attribution' thrusts. The idea was to supply a controlled environment for the development and testing phases of some CLDERA tools before they were applied to complex climate data from coupled E3SM simulations with added aerosol chemistry.

Our goal was to design simpler E3SMv2 configurations with analytic, tunable, Mt. Pinatubo-like volcanic eruptions and embedded prescribed physical pathways. This simpler HSW-V setup only uses the atmospheric component of E3SM at a coarse horizontal resolution with approximately 200 km grid spacing and E3SMv2's 72 vertical levels, utilizes a flat and dry earth without topography, does not incorporate an annual cycle, and replaces E3SM's complex physical parameterization package with two simple forcing functions. The latter are Rayleigh friction below 700 hPa and near the model top, as well as a Newtonian temperature relaxation towards a prescribed equilibrium temperature. These two processes mimic the turbulence in the planetary boundary layer and the solar radiation, respectively. Applying these forcings enables the model to run over multi-year time scales while exhibiting a realistic atmospheric circulation [194]. For example, the HSW-V climate has a realistic temperature distribution, exhibits the typical tropospheric zonal jets in the midlatitudes, and has polar jets in the stratosphere. However, some aspects of the stratospheric circulation are not represented, such as the annual cycle, the Quasi-Biennial Oscillation (QBO) in the tropical stratosphere or Sudden Stratospheric Warmings (SSWs) in the polar regions. While these processes could be added to the idealized setup, they were not the focus of the idealized HSW-V configuration.

Our focus was on the inclusion of newly-developed and standalone parameterizations into E3SMv2 with the Held-Suarez-Williamson (HSW) forcing to simulate an idealized injection, evolution, and radiative forcing by stratospheric volcanic aerosols. These new parameterizations link aerosol concentrations to the temperature field directly, and do not depend on the presence of other radiation or chemistry packages to mediate the forcing effect. Specifically, our implementation injects sulfur dioxide (SO_2) and ash with a specified vertical profile into the stratosphere, and makes use of simple exponential decay functions for their removal. In particular, the e-folding decay times of 25 days (for SO_2) and 24 hours (for ash) were used. In addition, SO_2 underwent a chemical decay. This chemical decay of SO_2 is modeled as a perfect conversion to a long-living sulfate aerosol which persists in the stratosphere. All three species (SO_2, ash and sulfate) are implemented as tracers in E3SMv2's model framework and are transported by the dynamical core's advection algorithm. The aerosols contribute simultaneously to local heating of the stratosphere and cooling of the surface by a simple plane-parallel Beer–Lambert law applied on two zonally symmetric radiation broadbands in the longwave and shortwave ranges. The implementation parameters can be tuned to produce realistic temperature anomaly signatures of large volcanic events.

The particular implementation for the CLDERA team is shown in [77]. The chosen configuration demonstrates that the E3SMv2 HSW-V runs mimic the volcanic eruption of Mt. Pinatubo in 1991. The idealized simulations contain a single isolated volcanic event against a statistically uniform climate, where no background aerosols or other sources of externally forced variability are present. In addition, ensemble simulations without a Mt. Pinatubo-like volcanic eruption were conducted. These volcanic-eruption simulations and their no-eruption counterfactuals thereby provided insight into the impact of the volcanic aerosols on the atmospheric circulation in a

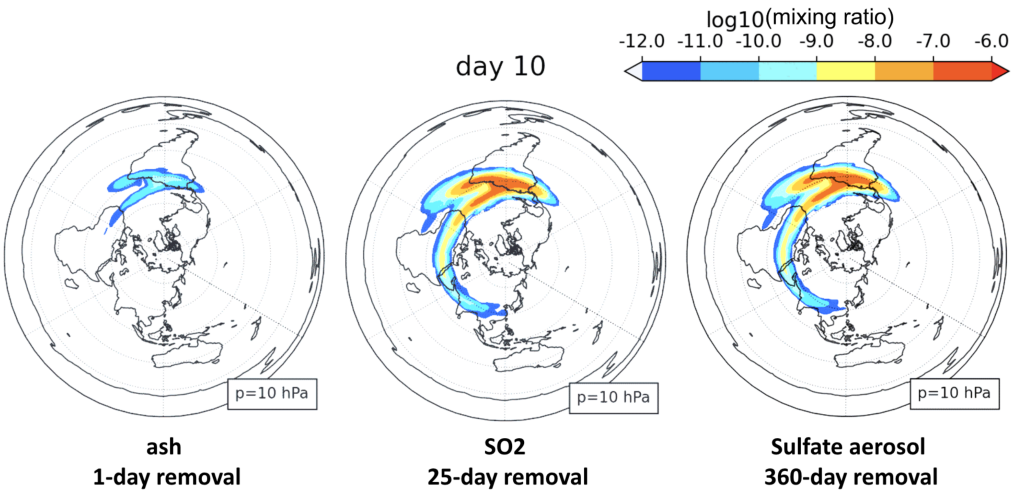


Figure 2-2. Tracer distributions at the 10 hPa pressure level, 10 days post-eruption. From left to right are the log-scaled mixing ratios (dimensionless ratio of tracer mass to air mass, in kg/kg) for volcanic ash, sulfur dioxide, and sulfate aerosol, respectively. Notes below each tracer label is that tracer's corresponding removal timescale.

controlled and therefore simpler environment. It made the HSW-V configuration a simpler-to-understand tool for the development of CLDERA's climate source-to-impact attribution methods, and now has been suggested as general modeling strategy for other science teams in [77].

Figures 2-2 and 2-3 demonstrate a single simulation from the HSW-V model. Figure 2-2 shows the evolution of the volcanic plume over the 10 days following a Pinatubo-like eruption near 15°N and 120°E. Over this timescale, the volcanic ash has almost entirely been removed, and the SO_2 is gradually being converted to a long-living sulfate aerosol. The plume has been transported halfway around the globe, but has not yet established a zonally-symmetric distribution. Figure 2-3 shows the radiative forcing of the sulfate aerosols over 5 months post-eruption, as well as the resulting temperature anomalies. The global-mean temperature forcing peaks at approximately 0.35 K/day near 30 hPa during the second month, which drives temperature anomalies of up to 3 K throughout much of the stratosphere by the fifth month. A more detailed display and discussion of the tuned temperature response in this model is shown in [77]. Finally, HSW-V runs were generated with different volcanic forcing magnitudes (e.g. different *sizes* of Mt. Pinatubo). The ensembles of varying magnitudes played a critical role in the development of the climate source-to-impact pathways and the attribution methods in CLDERA.

2.3. Tier 3: Temperature changes from Mt. Pinatubo

Tier 3 involved fully-coupled modeling of the volcano using E3SM's prognostic aerosol capability, which is documented in Section 3.1. Tier 3 moved beyond the idealized volcano simulated in HSW-V to a large-scale, realistic set of simulation ensembles that are described in Chapter 3. The goal of Tier 3 was to both identify pathways from source (Mt. Pinatubo) to impact (surface temperature changes or stratosphere temperature changes) and to attribute the temperature changes to the

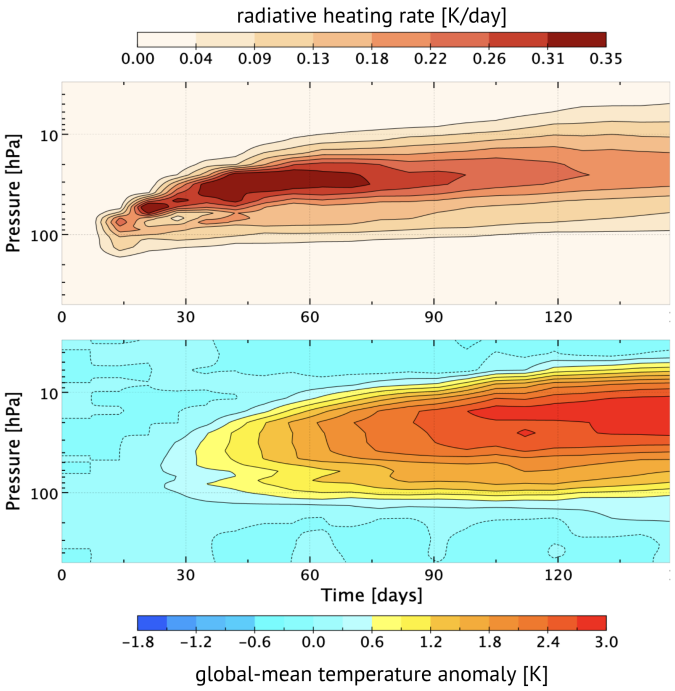


Figure 2-3. The global-mean forcing of temperature by volcanic aerosols in HSW-V from 3 hPa to 500 hPa over the first 5 months of evolution. The eruption occurs at day 15. Variables shown are (top) the aerosol radiative heating rate and (bottom) the temperature anomaly, defined as the difference between the volcanic simulation and the time-average of a 10-year HSW simulation with no volcanic forcing.

magnitude of the volcanic source forcing. Pathways from source to either surface or stratospheric temperature changes are referred to as *Pathway 1* by CLDERA.

Mt. Pinatubo's impacts to temperature were documented in following SAND report:

- Thomas Ehrmann, Benjamin Wagman, Diana Bull, Benjamin Hillman, Joseph Hollowed, Hunter Brown, Kara Peterson, Laura Swiler, Jerry Watkins, and Joseph Hart. "Identifying Northern Hemisphere Stratospheric and Surface Temperature Responses to the Mt. Pinatubo Eruption within E3SMv2-SPA." Technical Report. SAND2024-12730. September 2024, Sandia National Laboratories, Albuquerque NM. [35]

Tier 3 covers much of the method development and demonstrations covered in this report, Chapters 4-6. Several lessons were learned in CLDERA based on the Tier 3 datasets. We learned that having a range of source magnitudes (e.g. hypothetical Pinatubos of different sizes) was critical for our methods. This range supports the dose-response formalism in the Bradford-Hill causality framework (e.g. larger dose leads to a larger response). Tier 3 also enabled methodologies for handling intra-ensemble variability (especially for fingerprinting and inverse optimization). The certainty of statements about temperature impacts, even in the case of a hammer like Mt. Pinatubo, are difficult to make. Bayesian approaches, which result in probabilistic statements, help. Tier 3 datasets facilitated Bayesian approaches in inverse optimization, multi-step conditional attribution, and multivariate dynamic linear models. Finally, the realism of the Tier 3 datasets showed how autocorrelations can dominate interpretations (e.g., from RFR, multivariate DLM) and necessitated addressing autocorrelation.

2.4. Tier 4: Debated impacts from Mt. Pinatubo

The final tier of CLDERA's tiered verification process is the identification of pathways, detection of changes, and attribution of impacts that are either further downstream than the P1 pathway temperature impacts and/or noisier with more uncertainty (e.g. lower signal-to-noise ratio). These noisier, debated impacts include Mt. Pinatubo's impact on agricultural productivity, its effect on the south Asian monsoon, and its effect on cirrus cloud formation. The agricultural productivity impact is discussed in Section 7.8. This impact was addressed through standard climate scientist approaches and also with conditional multi-step attribution approaches. The south Asian monsoon impact is discussed in Section 7.7 and was addressed by the Echo State Network approach. The final impact of Mt. Pinatubo on cirrus cloud changes is discussed in Section 7.6. Overall, the Tier 4 datasets and impact analysis did prove more challenging. There is still much work to be done in multivariate analysis of subtle changes in climate impacts in the face of significant noise in climate dynamics.

3. SIMULATION CAMPAIGN

A major component of the CLDERA project was a simulation campaign generating coupled climate model ensembles for the years around the Mt. Pinatubo eruption. These ensembles were designed for use in developing novel methods to detect, model, and attribute impacts from the eruption. For this effort we have used the U.S. Department of Energy's Energy Exascale Earth System Model (E3SM) version 2 [55]. E3SMv2 is a fully coupled global model that includes atmosphere, ocean, ice, and land components. The E3SMv2 atmosphere model resolves the stratosphere and contains process models for clouds and aerosols, which are important components for modeling large volcanic eruptions. However, the released version of E3SMv2 prescribes the radiative impacts of volcanic aerosols from historical eruptions, like Mt. Pinatubo, rather than evolving the volcanic stratospheric aerosols directly. For CLDERA, we required the capability to modify the mass of the Mt. Pinatubo eruption for our analyses, which necessitated implementing a prognostic volcanic aerosol capability in the code to evolve SO_2 gas from the eruption into sulfate aerosols. The new implementation, denoted E3SMv2-SPA, modifies aerosol microphysics to accurately simulate stratospheric volcanic aerosols and was validated against observational data as described in [13].

With E3SMv2-SPA developed, longer-term coupled climate simulations were performed that confirmed that the changes to stratospheric aerosols did not change the average climate in E3SMv2. This verification and the procedure for the prognostic aerosol implementation are described in Section 3.1. After validation and verification of E3SMv2-SPA, a series of historical simulations that include years surrounding the Mt. Pinatubo eruption in June of 1991 were completed. The simulations can be divided into three simulation sets that differ in terms of the length of the simulations, the initialization procedure, and the addition of new diagnostic capabilities. The first set, full variability, was initialized in 1985, which is long enough before the eruption to result in ensemble members that are independent in the context of climate mode variability at the time of the eruption in June of 1991. This set is described in Section 3.2 and each ensemble includes both a Mt. Pinatubo-sized eruption of 10 Teragrams (Tg) of SO_2 and a corresponding counterfactual (CF) with no eruption. The second set, limited variability, was initialized in June 1991 with a run that most closely resembled the actual climate conditions at the time of the Pinatubo eruption and is described in Section 3.3. The final set, limited variability with source tagging, is described in Section 3.4 and uses the same initialization procedure as the limited variability set, but incorporates additional diagnostic capabilities to distinguish between different aerosol sources in the simulations. The new diagnostic capabilities are are described in Section 3.5. This ensemble set also includes members with varying eruption magnitudes from 0 Tg to 15 Tg. An overview of the simulation sets are provided in Table 3-1 and more complete descriptions are included in the following sections. In all cases the simulations were run with the standard E3SMv2 low-resolution configuration that combines 110 km atmosphere, 165 km land, and 30-60 km ocean and sea ice.

The E3SMv2-SPA simulation ensembles (full variability, limited variability, and limited variability

Simulation Set	Years	# Ensembles (Total)	Description
Pre-industrial Control	100, perpetual 1850	1 (1)	• Initialized from E3SMv2 piControl simulation. • For verification that E3SMv2-SPA maintains climate of E3SMv2.
Historical	1850-2014	2 (2)	• Initialized from E3SMv2-SPA piControl. • For verification of transient simulation with E3SMv2-SPA.
Full Variability	1985-1998 (1991-1998 CF)	15 (30)	• Initialized January 1985 from E3SMv2-SPA historical. Each ensemble member has perturbed initial condition. • Includes Pinatubo 10Tg SO_2 and counterfactual 0Tg SO_2.
Limited Variability	June 1991 – 1998	6 (12)	• Initialized in June 1991 from E3SM-SPA ensemble with best fit to climate mode observations at time of Pinatubo. • Includes Pinatubo 10Tg SO_2 and counterfactual 0Tg SO_2.
Limited Variability, Mass varying with Tagging	June 1991 – 1998	15 (120)	• Initialized in June 1991 from E3SM-SPA ensemble with best fit to climate mode observations at time of Pinatubo. • Mass variation: 0, 1, 3, 5, 7, 10, 13, 15Tg SO_2. • Includes tagged aerosols and radiation diagnostics.
Limited Variability, Altitude varying with Tagging	June 1991 – 1998	2 (6)	• Initialized from E3SMv2-PA historical June 1988. • Altitude variation: 16-18, 20-22, 22-24 km. • Includes tagged aerosols and radiation diagnostics.

Table 3-1. Overview of simulation ensemble sets from E3SMv2-SPA generated under the CLDERA project.

with tagging) include default output for the ocean and sea ice for the transient historical water cycle case (WCYCL20TR). For the atmosphere and land we created catered lists with additional variables for daily and higher temporal resolutions that were designed for analysis of Mt. Pinatubo impacts. The outputs were selected based on comparison with the Arise-SAI output fields (Appendix A in [144]) along with atmospheric and land process expertise. The outputs for this simulation campaign are described in Appendix A.

3.1. Prognostic Volcanic Aerosol Implementation in E3SMv2

Contributing Authors: Hunter Brown (8931), Benjamin Wagman (8931), Xiaohong Liu (Texas A & M University), Allen Hu (Texas A & M University), Kara Peterson (1442)

In historical simulations with the standard released E3SMv2 light extinction from stratospheric volcanic eruptions is prescribed from the GloSSAC reanalysis data set [176] based on satellite measurements. Under this configuration volcanic aerosols are not explicitly modeled and there is no ability to modify characteristics of an historical eruption in a simulation. The main goal of the CLDERA project has been to develop novel detection and attribution methods using the Mt. Pinatubo eruption as an exemplar. To more fully understand the impacts of Mt. Pinatubo and how the impacts change as a function of eruption strength it was necessary to implement a prognostic volcanic aerosol formulation in E3SMv2. In the prognostic formulation, the sulfate aerosol precursor, SO_2, is emitted in the stratosphere and undergoes chemical reactions to form sulfate aerosol. This implementation was critical for constructing impact pathways that include the chemical and microphysical evolution of volcanic aerosols.

In E3SMv2, aerosol evolution in the atmosphere is modeled with the four-mode Modal Aerosol Model (MAM4) microphysics package [99]. In this package all aerosol species are represented by

three size modes along with a forth mode representing freshly emitted black carbon and organic carbon from combustion. The modal distributions of aerosols evolve based on nucleation (aerosol formation), evaporation (size reduction), condensation and coagulation (size growth), and dry/wet deposition (aerosol removal). MAM4 had been designed to accurately model tropospheric aerosol evolution, but is not configured to accurately model the evolution of massive injections of SO_2 in the stratosphere from volcanic eruptions.

The modified code developed under the CLDERA project, E3SMv2-SPA (Stratospheric Prognostic Aerosol), includes enhancements to MAM4 that improve the stratospheric aerosol representation without degrading the representation of tropospheric aerosols or the modeled climate [13]. The major modifications to the code include (1) allowing aerosol growth from the accumulation to coarse mode in the stratosphere to increase aerosol size following volcanic eruptions and (2) adjustment of the coarse-mode and accumulation mode size distribution to increase aerosol lifetime.

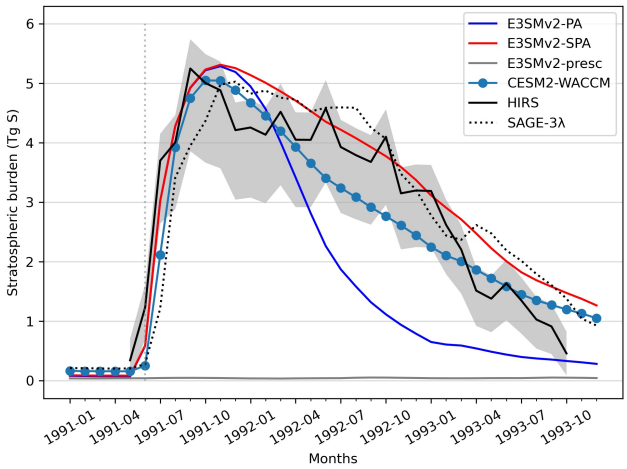


Figure 3-1. Stratospheric sulfate burden (Tg of the sulfur mass contribution) for E3SMv2-SPA, the standard E3SMv2 with prescribed stratospheric aerosol forcing (E3SMv2-presc), a version of E3SMv2 with prognostic aerosols before the SPA modifications (E3SM-PA), a full-chemistry climate model with prognostic volcanic aerosols (CESM2-WACCM), observational data from SAGE-λ and the High Resolution Infrared Radiation Sounder (HIRS).

To validate the implementation, atmospheric simulations with E3SMv2-SPA were run from 1990 to 1993 covering the Mt. Pinatubo eruption. The prescribed extinction from GloSSAC data was turned off and volcanic SO_2 emissions from the VolcanEESMv3.11 data set [121] including modifications from [115] were used as input. This data set provides emissions from historical eruptions and in the 1990-1993 period this includes eruptions from Mt. Pinatubo, Cerro Hudson, Spurr, and Lascar volcanoes. Results from E3SMv2-SPA were compared with observations of sulfate burden, aerosol optical depth, top-of-atmosphere radiative flux, and aerosol size distributions. Results were also compared with version 2 of the Community Earth System Model (CESM2) [28] and the Whole Atmosphere Community Climate Model version 6 (WACCM6) [47], which uses a similar microphysical treatment to E3SMv2 for prognostic aerosols, but includes full atmospheric chemistry.

Results for aerosol burden and aerosol optical depth are shown in Figures 3-1 and 3-2. In both cases the E3SMv2-SPA implementation matches well with both observational data and the full-chemistry

prognostic aerosol model (CESM2-WACCM).

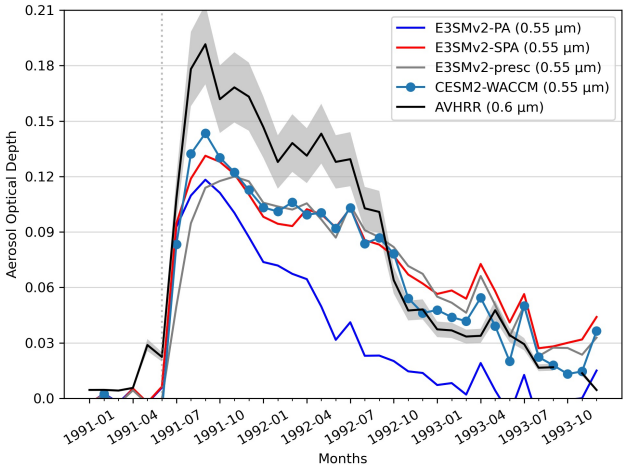


Figure 3-2. Stratospheric aerosol optical depth over the ocean and across latitudes 60S - 60N from E3SMv2-SPA, the standard E3SMv2 with prescribed stratospheric aerosol forcing (E3SMv2-presc), a version of E3SMv2 with prognostic aerosols before the SPA modifications (E3SM-PA), a full-chemistry climate model with prognostic volcanic aerosols (CESM2-WACCM), and observational data from AVHRR.

After validating the E3SMv2-SPA code modifications, fully coupled simulations were performed to investigate the effect of E3SMv2-SPA modifications on the modeled climate in E3SMv2. The first simulation was a one-hundred year long pre-industrial control simulation, which uses constant reference year 1850 forcing for inputs that include greenhouse gas emissions, solar irradiance, and aerosols and is designed to investigate the equilibrium climate. The E3SMv2-SPA model was initialized from year 101 of the E3SMv2 500-year pre-industrial control simulation [55] and used averaged volcanic SO_2 emissions from the VolcanEESMv3.11 data set over the years 1850-2014 following [115]. Several quantities including global average top-of-atmosphere (TOA) net radiation and surface temperature were evaluated to ensure that E3SMv2-SPA maintains a stable climate under this forcing. Results demonstrated that E3SMv2-SPA has a near-zero net TOA radiation (shown in Figure 3-3 in comparison with E3SMv2) and near-constant surface temperature with average value matching E3SMv2.

In addition to the pre-industrial control simulation, two historical simulations from 1850 to 2014 were run with E3SMv2-SPA that used transient historical forcings for greenhouse gas emissions, solar irradiance, and other inputs along with SO_2 emissions from historical volcanic eruptions from the VolcanEESMv3.11 data set. The first historical simulation was initialized from year 61 of the E3SMv2-SPA pre-industrial control simulation. A second historical simulation was run that included additional atmospheric diagnostic tracers. In order to spin up the tracers before the beginning of the historical simulation, ten extra years of the pre-industrial control simulation were run with the tracers and the historical was initialized from year 111 of the extended simulation. Results from both historical simulations compared well with the five released historical ensembles from E3SMv2 as shown in Figure 3-4.

More details on the prognostic volcanic aerosol implementation in E3SMv2-SPA and validation against observational data sets are documented in [13]:

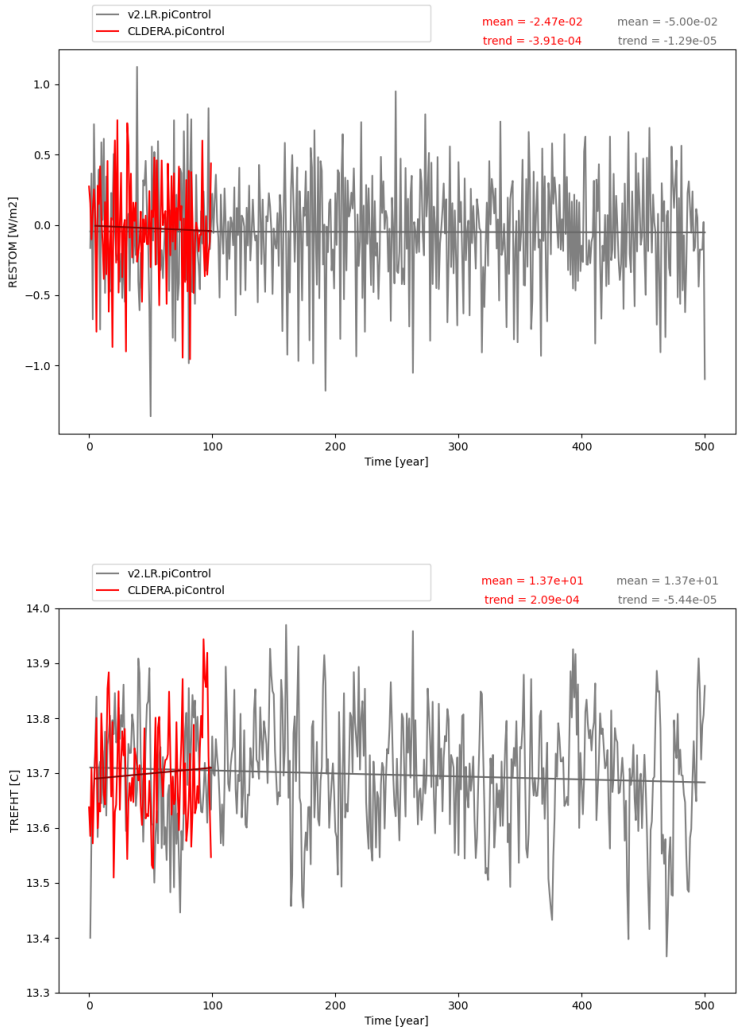


Figure 3-3. Global average top-of-model net radiation (W/m^2) (left) and reference height (surface) temperature ($^\circ$ C) for the E3SMv2 pre-industrial control simulation in grey and the E3smv2-SPA pre-industrial control simulation in red. Results demonstrate that the modifications in E3SMv2-SPA maintain a stable climate with pre-industrial control forcing with an average surface temperature the same as E3SMv2 and near-zero net top-of-model radiation.

- Brown, H. Y., Wagman, B., Bull, D., Peterson, K., Hillman, B., Liu, X., Ke, Z., and Lin, L.: Validating a microphysical prognostic stratospheric aerosol implementation in E3SMv2 using observations after the Mount Pinatubo eruption, Geosci. Model Dev., 17, 5087–5121, https://doi.org/10.5194/gmd-17-5087-2024, 2024 .

The E3SMv2-SPA code is available to download from `https://github.com/sandialabs/CLDERA-E3SM`.

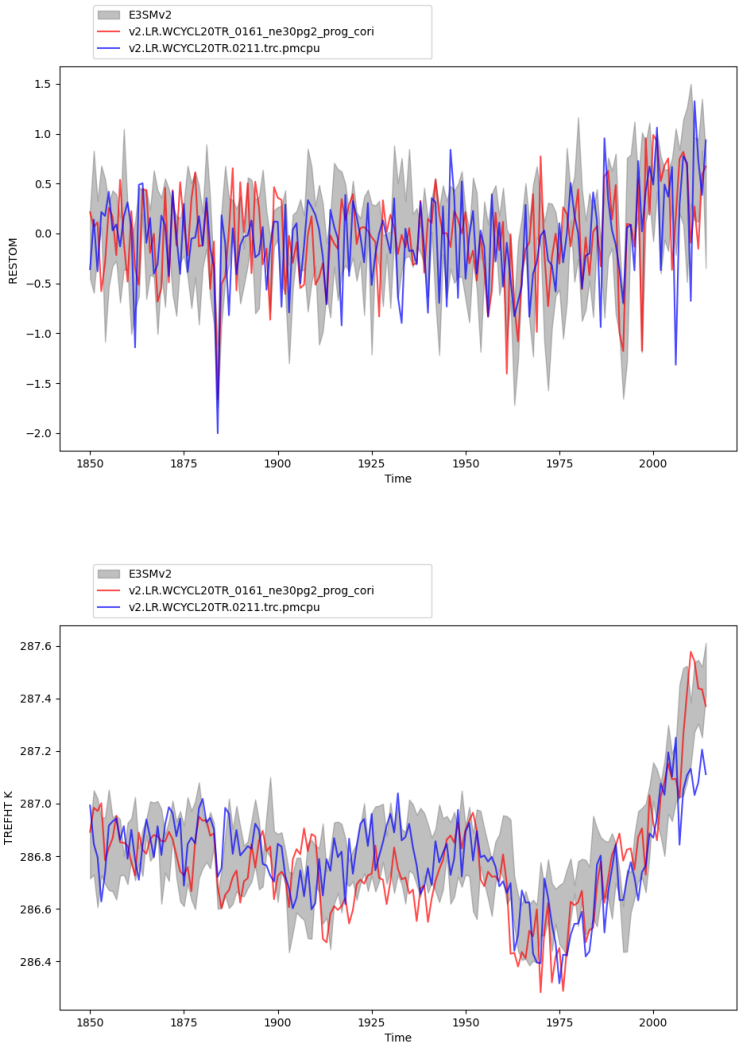


Figure 3-4. Global average top-of-model net radiation (W/m^2) and reference height (surface) temperature ($^\circ$ C) for the five E3SMv2 historical ensembles in grey and the E3smv2-SPA historical simulations in red and blue. The E3SMv2-SPA historical simulations match the spread of the E3SMv2 historical simulations well.

3.2. Full Variability Simulations

Contributing Authors: Benjamin Wagman (8931), Hunter Brown (8931), Jerry Watkins (8734), Tom Ehrmann (8931), Ben Hillman (8931), Kara Peterson (1442), Diana Bull (8931)

The initial set of Mt. Pinatubo ensembles, denoted full variability or fullvar, were designed to maximize the natural variability at the time of the Mt. Pinatubo eruption in June of 1991. The simulations were initialized from January 1985 of the second historical ensemble with each ensemble member having a small (10^{-14}) random atmospheric temperature perturbation. The initialization procedure allowed time for the ocean and atmosphere modes to become independent

by the time of the eruption. A total of 15 ensemble members were generated in this configuration running from 1985 to 1998.

In addition to the historical Mt. Pinatubo ensemble members, 15 corresponding counterfactual ensembles with no eruption were generated. Each member of the counterfactual set was initialized in January of 1991 from the corresponding Mt. Pinatubo ensemble with no perturbation so that the coupled simulations are bit-for-bit until the time of the eruption. The counterfactual ensemble provided a means to isolate impacts from the Mt. Pinatubo eruption by subtracting simulated fields in the counterfactual ensemble member from the field in the corresponding ensemble member with the Mt. Pinatubo eruption. This process is illustrated in Figure 3-5 where the difference in global averaged fields for the Mt. Pinatubo surface and stratospheric temperature pathways are shown in relation to the corresponding counterfactual. The fields include aerosol optical depth, net shortwave flux at top of model, net longwave flux at top of model, clear sky downwelling surface flux, and stratospheric temperature at 50 hPa, and reference height temperature. The time of the Mt. Pinatubo eruption is designated with the red dashed line and the figures demonstrate that the simulations and counterfactuals match before the eruption and after the eruption occurs the Mt. Pinatubo simulation and corresponding counterfactual begin to diverge. Strong signals of the eruption are seen in the aerosol optical depth, clear sky downwelling shortwave radiation, and stratospheric temperature where the impact anomaly is far from zero.

Fullvar simulations were utilized in the following research:

- Daniel Ries, Katherine Goode, Kellie McClernon, and Ben Hillman. "Using feature importance as exploratory data analysis tool on earth system models." Submitted to *GMD Special Issue: Theoretical and computational aspects of ensemble design, implementation, and interpretation in climate science*, July 2024. The review preprint is avialable at: `https://gmd.copernicus.org/preprints/gmd-2024-133/` [145]
- Thomas Ehrmann, Benjamin Wagman, Diana Bull, Benjamin Hillman, Joseph Hollowed, Hunter Brown, Kara Peterson, Laura Swiler, Jerry Watkins, and Joseph Hart. "Identifying Northern Hemisphere Stratospheric and Surface Temperature Responses to the Mt. Pinatubo Eruption within E3SMv2-SPA." Technical Report. SAND2024-12730. September 2024, Sandia National Laboratories, Albuquerque NM. [35]
- Benjamin M. Wagman, Hunter Brown, Joe Hollowed, Diana Bull, Tom Ehrmann, Ben Hillman, Laura Swiler, Kara Peterson. "Does prognostic volcanic aerosol enhance climate variability in Earth System Model simulations of volcanic eruptions?" SAND2024-12725o. September 2024, Sandia National Laboratories, Albuquerque NM. [183]

3.3. Limited Variability Simulations

Contributing Authors: Benjamin Wagman (8931), Hunter Brown (8931), Jerry Watkins (8734), Tom Ehrmann (8931), Ben Hillman (8931), Kara Peterson (1442), Diana Bull (8931)

In contrast to the full variability simulation set, the limited variability ensembles were designed to provide a set of simulations that have less internal variability and to match as closely as possible the

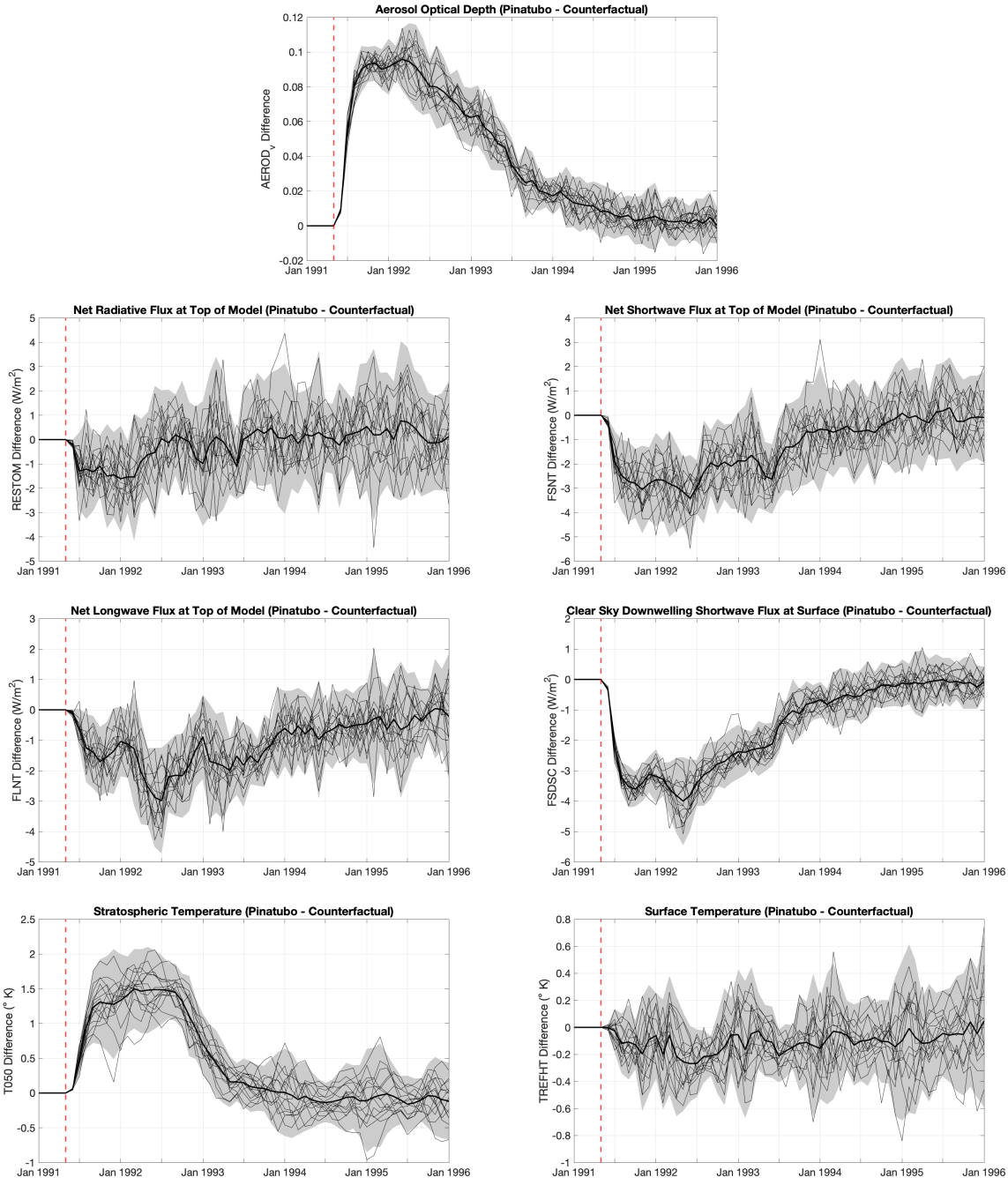


Figure 3-5. Selected global averaged monthly fields for the full variability ensemble set. For each variable the counterfactual is subtracted to define an impact anomaly. The dark black line is the average of the first ten ensemble members with the 2 σ spread indicated by the gray shading. The time of the eruption is indicated with the red dashed line.

observed climate at the time of the Mt. Pinatubo eruption. For this ensemble set we initialized the runs on June 1, 1991, approximately two weeks before the eruption, which enables the simulations to diverge at the synoptic scale while still preserving the more slowly evolving climate modes by the time of the eruption. To identify a suitable initial condition we considered June 1 values

of select climate modes from years 1986-1996 from the second historical run and four ensemble members from the full variability ensemble, which provided 55 possible initial conditions. Two main climate modes were identified as being of particular interest in determining the appropriate initial conditions: the El-Niño Southern Oscillation (ENSO) and the Quasi-Biennial Oscillation (QBO). For ENSO we used the NINO3.4 Index to determine the phase of ENSO and the monthly-averaged mean temperature within the NINO3.4 domain to determine the trend in ENSO. For the QBO we used the daily-averaged zonal mean zonal wind between 5°N and 5°S at 50 hPa and 25 hPa. We used two levels for the QBO to capture the vertical structure of the oscillation and ensure that the full phase of the QBO is represented. For the ocean data we used the historical archives of sea-surface temperature used as forcing in E3SM, and for the atmospheric data we used the Modern-Era Retrospective analysis for Research and Applications (MERRA-2) reanalysis. Plots of the monthly averaged NINO3.4 Index and the daily QBO at 50 hPA over time for the full variability and limited variability ensembles are shown in Figure 3-6. The limited variability modes from both ENSO and QBO are generally well-aligned for the first year after the eruption.

Quantity of Interest	**Observed June 1, 1991 Value**	**Data Source**	**Simulated June 1, 1991 Value**
Nino3.4	2.7	E3SM Ocean Forcing Data	1.82
Nino 3.4 Trend	0.185/month	E3SM Ocean Forcing	-0.585/month
QBO at 50 HPa	10.01 m/s	MERRA-2	3.21 m/s
QBO at 25 HPa	-22.33 m/s	MERRA-2	-7.09 m/s

Table 3-2. Selection criteria from observational data sources for Nino3.4 and QBO climate modes in limvar initialization compared with simulation values used for initialization.

In Figure 3-7 the difference in monthly global averaged fields for the Mt. Pinatubo surface and stratospheric temperature pathways are shown in relation to the corresponding counterfactual for the limited variability ensemble. The time of the eruption is again indicated by a red dashed line. In this case, the ensemble members start just 15 days after the eruption.

Limvar simulations were utilized in the following research:

- Meredith G. L. Brown, Matt Peterson, Irina Tezaur, Kara Peterson, and Diana Bull. "Random Forest Regression Feature Importance for Climate Impact Pathway Detection." Submitted to *Journal of Computational and Applied Mathematics*, September 2024. [16]
- Warren L. Davis, Max Carlson, Irina Tezaur, and Diana Bull. "Spatio-temporal Multivariate Cluster Evolution Analysis for Detecting and Tracking Climate Impacts." Intended submission to *Journal of Computational and Applied Mathematics*, September 2024. [31]
- Thomas Ehrmann, Benjamin Wagman, Diana Bull, Benjamin Hillman, Joseph Hollowed, Hunter Brown, Kara Peterson, Laura Swiler, Jerry Watkins, and Joseph Hart. "Identifying Northern Hemisphere Temperature Responses to the Mt. Pinatubo Eruption through Limited Variability Ensembles." Submitted to *Climate Dynamics*, September 2024. [36]
- Thomas Ehrmann, Benjamin Wagman, Diana Bull, Benjamin Hillman, Joseph Hollowed, Hunter Brown, Kara Peterson, Laura Swiler, Jerry Watkins, and Joseph Hart. "Identifying

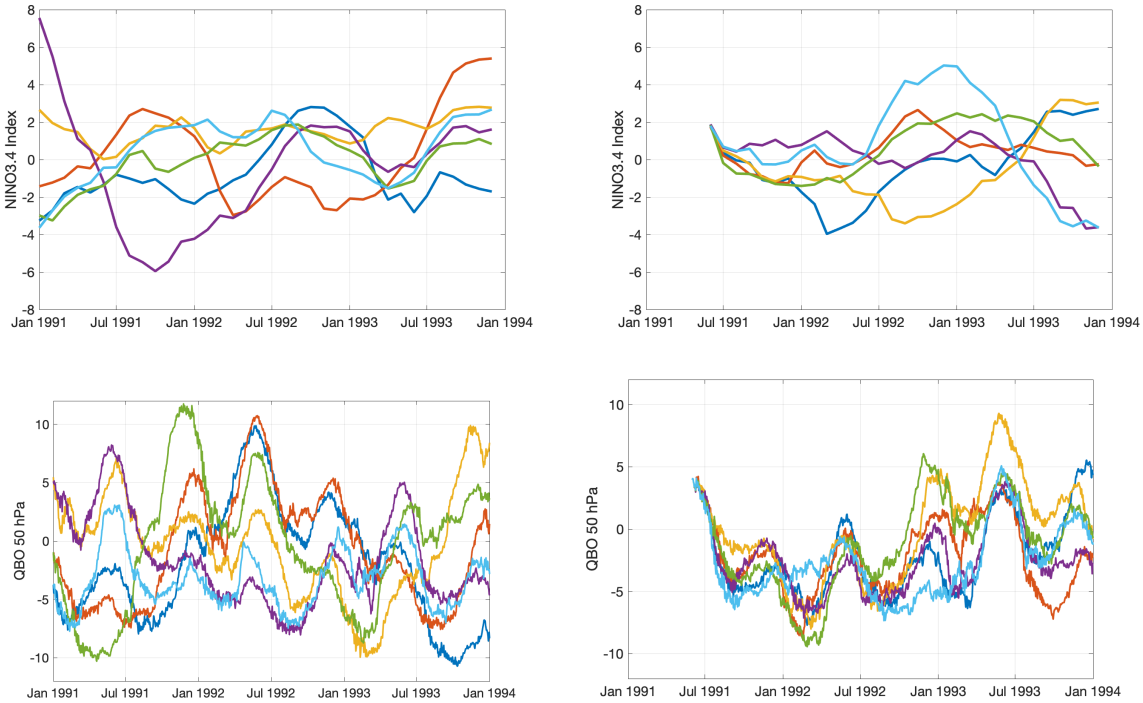


Figure 3-6. Timeseries of NINO3.4 index and QBO at 50 hPA for six full variability (left) and limited variability (right) ensembles. The limited variability initialization results in smaller spread in the climate modes over the first year.

Northern Hemisphere Stratospheric and Surface Temperature Responses to the Mt. Pinatubo Eruption within E3SMv2-SPA." Technical Report. SAND2024-12730. September 2024, Sandia National Laboratories, Albuquerque NM. [35]

- Daniel Ries, Kellie McClernon, Thomas Ehrmann, and Diana Bull. "Using Spatio-Temporal Feature Importance to Identify Drivers of Indian Monsoon after the Mount Pinatubo Eruption." SAND2024-12581O. September 2024, Sandia National Laboratories, Albuquerque NM. [146]

- Hollowed, J., Jablonowski, C., Ehrmann, T., Hillman, B., ... "Volcanic Aerosol Modification of the Stratospheric Circulation in E3SMv2 Part I: Wave-Mean Flow Interaction." To be submitted to *Climatic Dynamics* by October 2024. [76]

- Krofcheck, D., Crockett, J., Peterson, K., Li, J., Bull, D. "Volcanic stratospheric aerosol injection may buffer against global ecohydrological impacts of hotter and drier climates." To be submitted December 2024. [89]

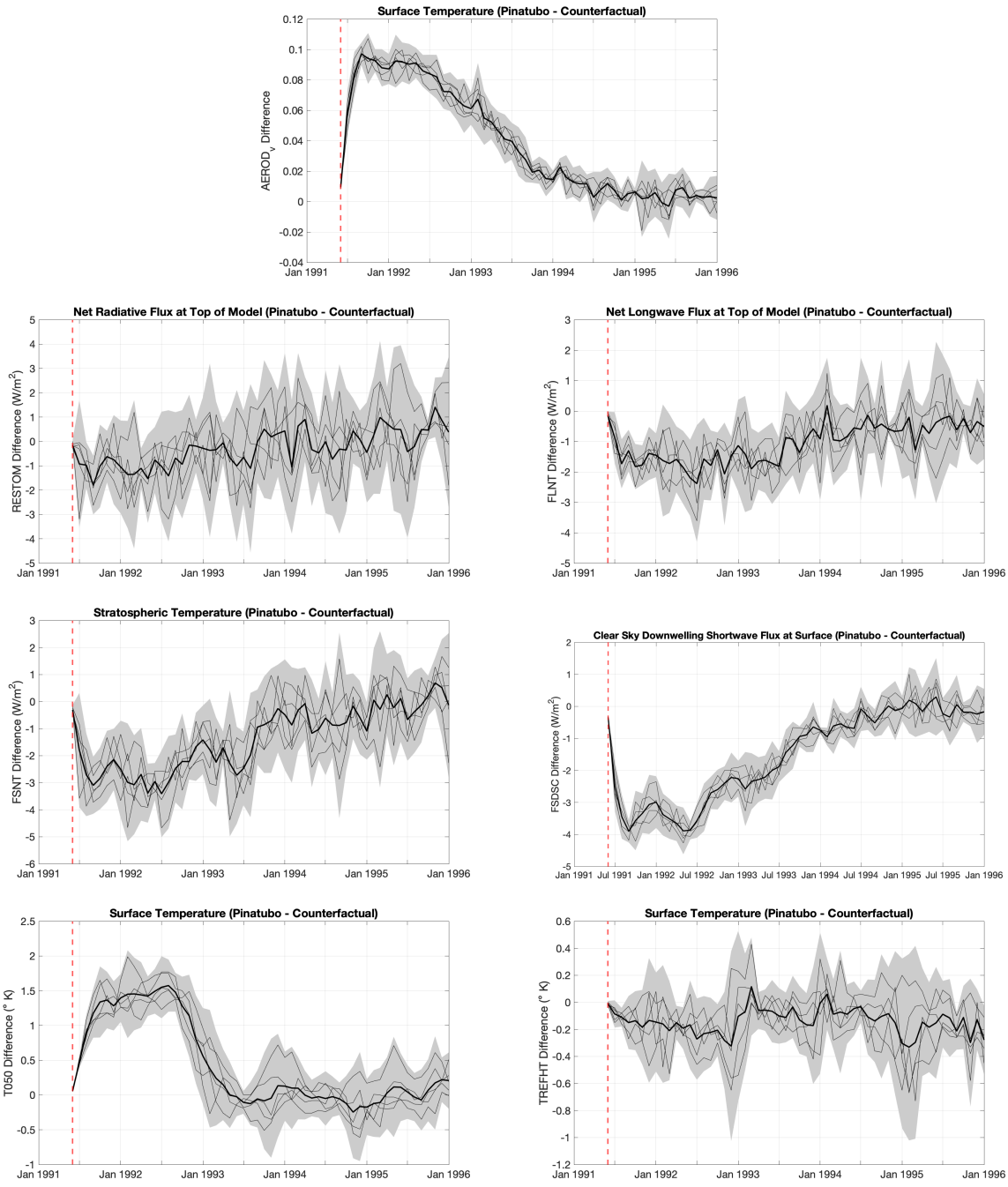


Figure 3-7. Selected global averaged monthly fields for the limited variability ensemble set. For each variable the counterfactual is subtracted to define an impact anomaly. The dark black line is the average of the six ensemble members with the 2 σ spread indicated by the gray shading. The time of the eruption is indicated with the red dashed line.

3.4. Limited Variability Simulations with Aerosol Tagging and Eruption Mass Variation

Contributing Authors: Benjamin Wagman (8931), Hunter Brown (8931), Jerry Watkins (8734), Tom Ehrmann (8931), Ben Hillman (8931), Kara Peterson (1442), Diana Bull (8931)

The final set of simulation ensembles uses the same initialization procedure as the Limited Variability ensemble and included varying mass eruptions along with additional diagnostic tools. In order to understand the limits of detection and attribution a set of 15 ensembles with eruption magnitudes of 1, 3, 5, 7, 13, and 15 Tg in addition to the 0 Tg counterfactual and 10 Tg standard Mt. Pinatubo eruption were completed. For each of these ensembles additional diagnostics including tagging of the Mt. Pinatubo aerosols to distinguish from background aerosols and radiation calls that separate out the Mt. Pinatubo contributions to radiative fluxes were performed as described in Sections 3.5.1 and 3.5.2.

In figure 3-8, plots of monthly global averaged fields for the first 12 ensemble members are shown. The effects of the mass variation are apparent in the figures where each color represents an ensemble of mass values with the bold line the ensemble average difference. The aerosol optical depth, clear sky downwelling radiation, and stratospheric temperature show relatively clear delineations between the eruption masses. In the other fields, however, there is much overlap between the different eruption magnitudes illustrating the challenges in attributing a source to a particular impact in the climate.

As for the full variability and original limited variability ensemble set, this varying magnitude ensemble set was designed to have the paired counterfactual and eruption simulations be bit-for-bit the same before the eruption. To achieve this, care was taken to run corresponding ensembles on the same computing cluster, either Perlmutter at NERSC, Boca at SNL, or Flight at SNL. Due to changes in the operating system of Boca, ensemble 2 15Tg and ensemble 3 5Tg and 15Tg are non-BFB after timestep 6 when compared with other simulations in their respective ensembles. These simulations were run after the TOSS4 RHEL-8 system upgrade on the Boca computing cluster which upgraded compilers and other E3SM third party libraries. This difference impacts the analysis at early times, but after the first year the variability between ensemble and counterfactual is in line with other ensemble members.

An additional issue was found in ensemble 10, where the simulations were run on a new cluster, Flight, and included modified initial conditions to add the ST80 tracer (discussed in Section 7.2. This was due to an error when modifying the initial condition to add a revised ST80 tracer that also caused changes to the initial surface air pressure (PS). A revised simulation was performed that fixed the problem for the 10Tg simulation (it was run without the revised ST80 tracer). However, the majority of the ensemble 10 simulation suite run on flight does possess this issue. Again, like ensemble 2 and 3 over long time frames, the ensemble 10 simulation suite does not look out of sample.

Lastly, comparisons of the 10 Tg eruptions at early times showed that ensemble members 11 and 12 (especially 11) seem to be outliers in the first month for daily aerosol distributions when compared with the first nine ensemble members. These were also simulated on Flight, but careful inspection of

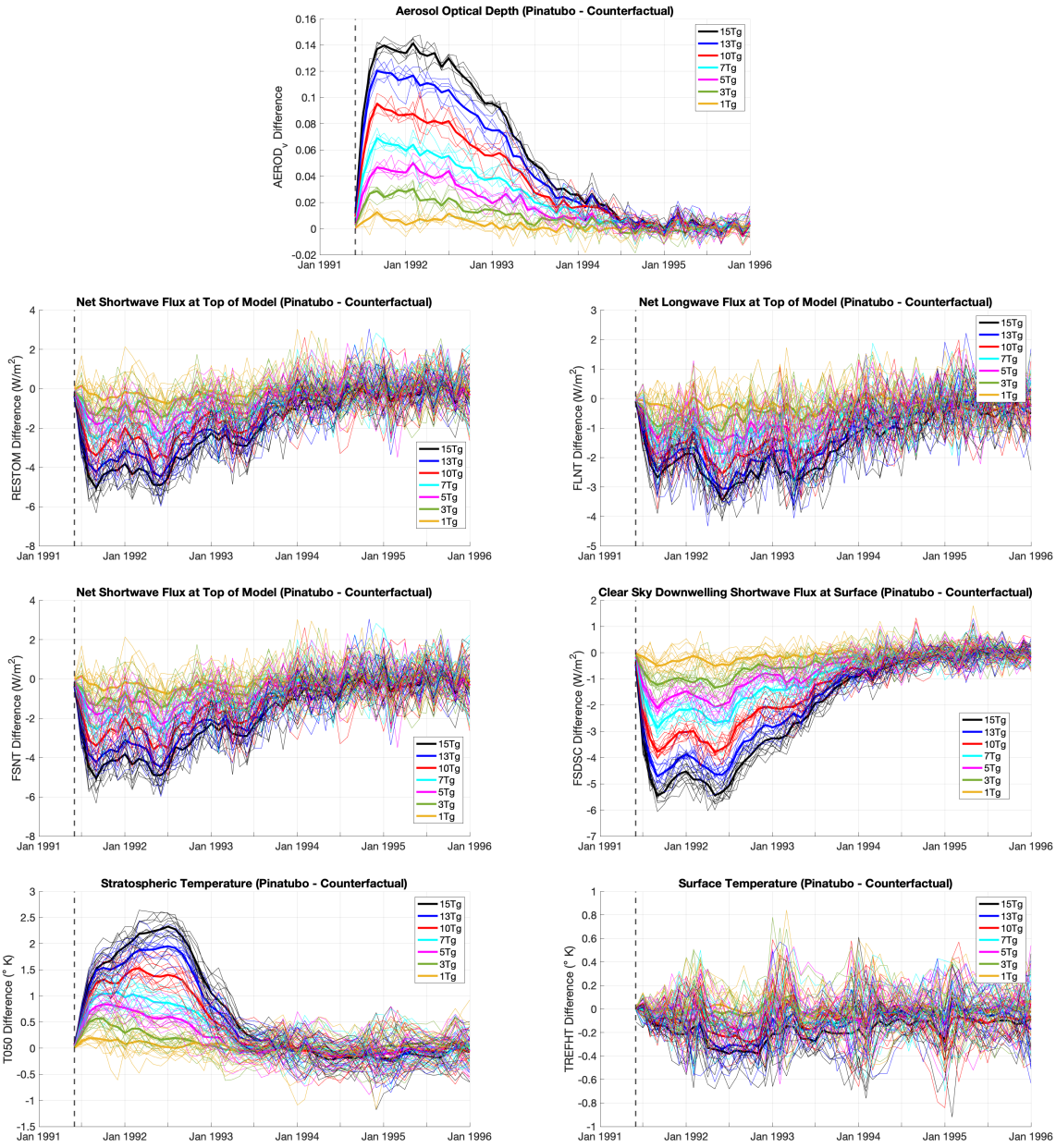


Figure 3-8. Selected global averaged monthly fields for the mass varying limited variability ensemble set. For each variable the counterfactual is subtracted to define an impact anomaly. The bold lines indicate the ensemble mean for a given mass variation.

initialization files showed no cause. At later times, all ensemble members show similar variability in monthly global averaged fields (Figure 3-8).

Limvar simulations and the variable source magnitudes were utilized in the following research:

- Christopher R. Wentland, Michael Weylandt, Laura P. Swiler, Thomas S. Ehrmann, Diana Bull. Probabilistic multi-step attribution for short-term climate forcings. Submitted to *Journal of Climate*, Aug. 2024. This manuscript is available on arXiv: `https://doi.org/`

10.48550/arXiv.2409.01396. [191]

- Joseph Hart, Indu Manickam, Mamikon Gulian, Laura Swiler, Diana Bull, Thomas Ehrmann, Hunter Brown, Benj Wagman, and Jerry Watkins. Stratospheric aerosol source inversion: Noise, variability, and uncertainty quantification, submitted to *Journal of Machine Learning for Modeling and Computing*. This manuscript is available on arXiv, https://doi.org/10.48550/arXiv.2409.06846. [66]
- Jerry Watkins, Luca Bertagna, Graham Harper,Andrew Steyer, Irina Tezaur, and Diana Bull. "Entropy-based feature selection for capturing impacts in Earth system models with extreme forcing." Submitted to *Journal of Computational and Applied Mathematics*, September 2024. [187]
- Li, J., Wentland, C., Swiler, L, Peterson, K., Weylandt, M., Krofcheck, D., Wagman, B., Bull, D. "Attributing Mt. Pinatubo in spatial patterns of vegetation response." To be submitted November 2024. [94]
- Hollowed, J., Jablonowski, C., Ehrmann, T., Hillman, B., ... "Volcanic Aerosol Modification of the Stratospheric Circulation in E3SMv2 Part II: Tracer Sensitivity." To be submitted by December 2024. [75]

3.5. Strategies for Isolating Mt. Pinatubo Impacts

Contributing Authors: Hunter Brown (8931), Benjamin Wagman (8931), Ben Hillman (8931)

3.5.1. Aerosol Tagging

A variety of atmospheric aerosol and gas-phase species are tracked within E3SM from their emission/formation to their removal from the atmosphere. These species are typically an amalgamation of all sources in the atmosphere (e.g., the sulfate aerosol tracer represents contributions from anthropogenic and natural sources). The addition of a source specific tracer typically requires modifications to the underlying source code, the species namelists, and the chemical formulas - an editing of hundreds of lines of code. The 'aerosol tagging' method adapts the source code to allow for a less invasive method for doing this, whereby the user can specify new source-specific aerosol species via the model runscript with minimal changes to the namelist files. This method applies to the aerosol species sulfate (SO_4), black carbon (BC; i.e., soot), primary (i.e., directly emitted) organic aerosol/matter (POA or POM), and secondary (i.e., formation from gas phase precursors) organic aerosol (SOA). The different tagged species are determined by the user-specified emission files, which should sum to the total global emissions of that species. These tags are active tracers, meaning that they are affected by, and effect changes in, atmospheric radiation and dynamics.

The tagging code, referred to as the Explicit Aerosol Source Tagging (EAST) system in the literature, was originally incorporated in E3SMv1 [201] and was ported into E3SMv2 as a part of CLDERA. EAST provides a method for exploring climate sensitivity to different emission source regions and emission sectors within a single, physically consistent model simulation. In EAST,

the model applies the same physical, chemical, and dynamical processes to each additional aerosol source species. Contrast this with traditional emission sensitivity approach, whereby a single source emission is removed in a counterfactual simulation and the difference is taken between this simulation and the default to infer differences in aerosol contributions. The counterfactual method becomes computationally expensive with multiple source regions (analogous to tagged species in our case), requiring additional model simulations for each separate sources. It also assumes a linear response to emission perturbation [185, 201], which will introduce uncertainty due to internal variability across model simulations which can lead to nonlinear dynamical responses. The improved computational efficiency of E3SMv2 compared to E3SMv1 [55] makes EAST ideal for our version of E3SM as the addition of multiple tracer species will have a smaller impact on computational cost than in CESM or E3SMv1.

EAST has been used extensively to explore aerosol source contributions to pollution and radiative impacts over a variety of receptor regions. Early work with this code looked at the source contribution of BC burden and direct radiative forcing in the Arctic using the Community Earth System Model (CESM) [185]. A sulfate tagging capability was later added to CESM to explore source-apportionment of global forcing of SO_4 [202]. This was followed by the inclusion of SOA and POA tags in CESM [203], allowing for the inclusion of tagged brown carbon (i.e., light absorbing POA) and its impacts on snow albedo [14].

In CLDERA we design our tagging experiments to separate atmospheric sulfate contributions into three sources: Pinatubo, Cerro Hudson, and all other contributions (i.e., anthropogenic, off-gassing surface volcanic emissions, and ocean contributions from dimethyl sulfide (DMS)). The other taggable species BC, POM, and SOA are treated as a single global source (same as the default model). Tagging tracks all sulfate mass diagnostics as well as the sulfate gas precursor emissions of sulfur dioxide (SO_2) and the intermediary gas-phase sulfuric acid (H_2SO_4). It also tracks sulfate contribution to visible aerosol optical depth - a measure of aerosol opacity to visible wavelengths of light in the atmosphere which depends on sulfate aerosol number concentration, incident wavelength of light, and particle size. To isolate the source specific (e.g., Pinatubo or Hudson) impact on atmospheric radiation balance, tagging must be used in conjunction with the model radiation diagnostic capability.

Tagged simulation output was utilized in the following research:

- Joseph Hart, Indu Manickam, Mamikon Gulian, Laura Swiler, Diana Bull, Thomas Ehrmann, Hunter Brown, Benj Wagman, and Jerry Watkins. Stratospheric aerosol source inversion: Noise, variability, and uncertainty quantification, submitted to *Journal of Machine Learning for Modeling and Computing*. This manuscript is available on arXiv, `https://doi.org/10.48550/arXiv.2409.06846`. [66]

3.5.2. Radiation Diagnostics

E3SM has the capability to make separate, parallel radiation calculations that can be customized to include or exclude certain aerosol tracers. This process is illustrated in Figure 3-9. All of these radiation calls include the impacts of clouds and greenhouse gases (GHGs). The removal of all aerosols gives an aerosol-free background (i.e., impacts only from GHGs and clouds) which can

be used to calculate an aerosol only forcing (i.e., direct aerosol forcing) when it is subtracted from the default radiation call (contributions from GHGs, all aerosol, and clouds) ([48] and references therein). Since the radiation calls are run in parallel, the aerosol forcing is the true forcing in the model because the atmospheric dynamics and microphysics are identical across radiation calls. For the same reasons that tagging is beneficial to the calculation of aerosol tracer contributions, these radiation diagnostics are more accurate than the traditional method for calculating an aerosol forcing which relies on a difference between a default simulation and a counterfactual simulation (see Figure 3-10. Separate simulations can exhibit different atmospheric dynamics, which introduces uncertainty in the results, and this uncertainty is not present in radiation diagnostic differences resulting in a cleaner signal.

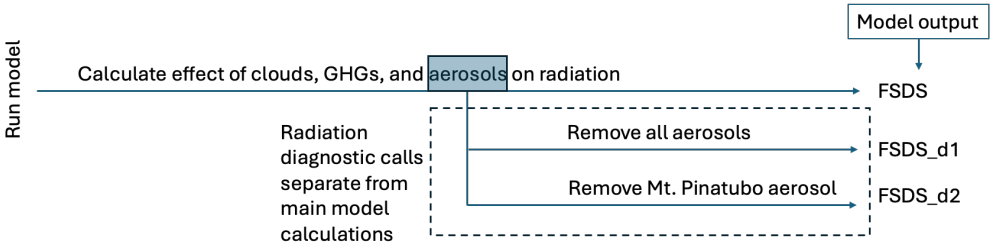


Figure 3-9. The radiation diagnostics procedure in E3SM where additional calls to the radiation model are performed with specific aerosol sources removed.

Aerosol tagging [201] allows for preferential removal of specific species and source aerosol contributions to the radiation calls, allowing for a source specific understanding of aerosol contributions. For example, removal of the tagged Pinatubo aerosol from the radiation call means that the differences between the default radiation call and this results in the Pinatubo-only impacts on atmospheric radiation. The Pinatubo flux (F_{Pin}) is derived from the difference between the atmosphere radiation flux diagnostic from the default radiation call (F) and the radiation call neglecting Pinatubo (F_{noPin}):

$$F_{Pin} = F - F_{noPin} \tag{3.1}$$

A note on the difference between the [48] method and our own: they have an additional counterfactual simulation run to determine change in flux due to anthropogenic aerosol, denoted by Δ. So their calculations are

$$F_{anthro} = \Delta(F - F_{noaerosol}), \tag{3.2}$$

where the delta is a difference between the default and no anthropogenic emission counterfactual simulations. This is necessary for their simulations because they are unable to isolate the anthropogenic sources from other natural sources of aerosol. We can isolate sources with tagging, so we don't require the Δ term.

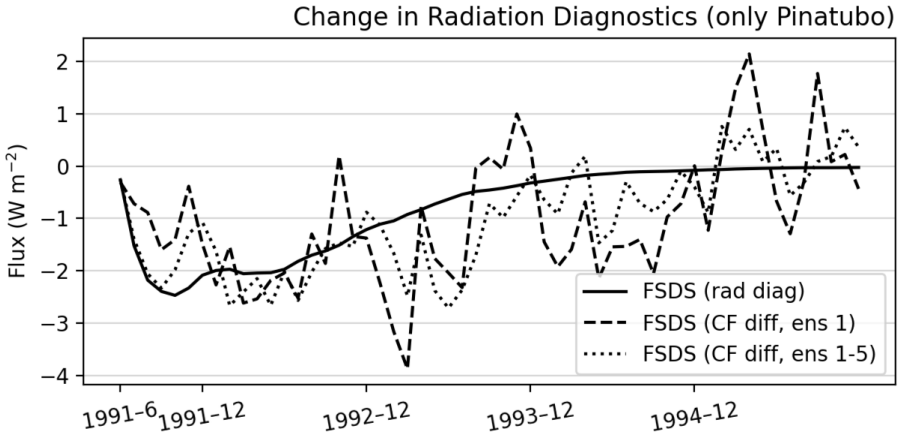


Figure 3-10. The radiation diagnostics provide a capability to isolate the effects of the Mt. Pinatubo eruption on radiation fields like Flux of Shortwave Downwelling radiation at the Surface (FSDS) and provides a clearer signal than what is produced subtracting a counterfactual simulation.

This page intentionally left blank.

4. SIMULATED PATHWAYS THRUST

The simulated pathways thrust under CLDERA has focussed on developing a fundamentally new and blended computational and information science approach to discovering, confirming and explaining connective relationships within the climate system, referred to herein as source-impact pathways. The thrust has produced three complementary approaches for source-impact pathway detection, each described in more detail in the following subsections: (i) Random Forest Regression (RFR), (ii) Profiling, and (iii) Clustering/Multi-Variate Data Mining. Whereas subthrusts (i) and (iii) are purely data-driven in that they operate on a set of pre-generated data, subthrust (ii) performs pathways-tracking analyses in-situ while running E3SM, and is the only in-situ method developed under CLDERA. As depicted in Figure 4-1, the three simulated pathways subthrusts are synergistic: both methods developed and data generated under one subthrust may be utilized by the other subthrusts. Pathways identified by the simulated pathways thrust have several use cases. Since pathways provide insight into the possible downstream impacts of a climate source, they can be helpful in detection. They can additionally be helpful in attribution by employing the epidemiological framing present in the Bradford Hill Criteria [74] which may be investigated in future work.

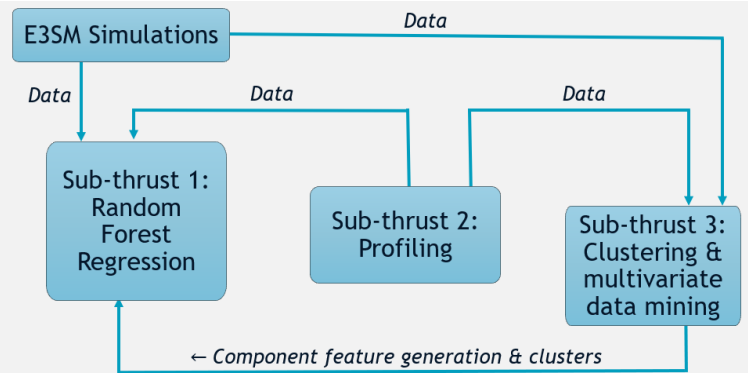


Figure 4-1. Illustration showing organization of the simulated pathways thrust

4.1. Random Forest Regression (RFR)

Contributing Authors: Meredith Brown (8931), Matt Peterson (1441), Irina Tezaur (8734), Kara Peterson (1442), Diana Bull (8931).

The goal of the RFR subthrust is to use time-series data to create a directed graph comprised of nodes and edges, in which the nodes represent features of interest and the edges represent relationships between these features. These relationships are directional, weighted and have a time lag associated

with them. Once a graph is fully constructed, it can be used to trace out source-impact "pathways", defined as the interactions of a set of variables in space-time due to an external forcing, within the provided time-series data (in our case, climate data).

Rather than utilizing RFR for classification or regression tasks (the most common use case for RFR) [159], we propose a fundamentally new pathway identification RFR-based workflow, summarized in Figure 4-2 [15]. This workflow consists of three key steps: (i) train random forest (RF) regressors on a set of spatio-temporal features of interest, (ii) calculate their pairwise feature importances, and (iii) translate these feature importances into a weighted pathway network (i.e., a weighted directed graph), which can be used to trace out and rank interdependencies between climate features and/or modalities. One unique capability of our approach is that, once a pathway network graph has been constructed using RFR and feature importance, it can be queried to answer questions such as "what is the most direct source-impact path?" and "what is relative strength of a given pathway?". Other strengths of the method include its low computational cost and its ability to discover pathways from (either simulated or observed) data.

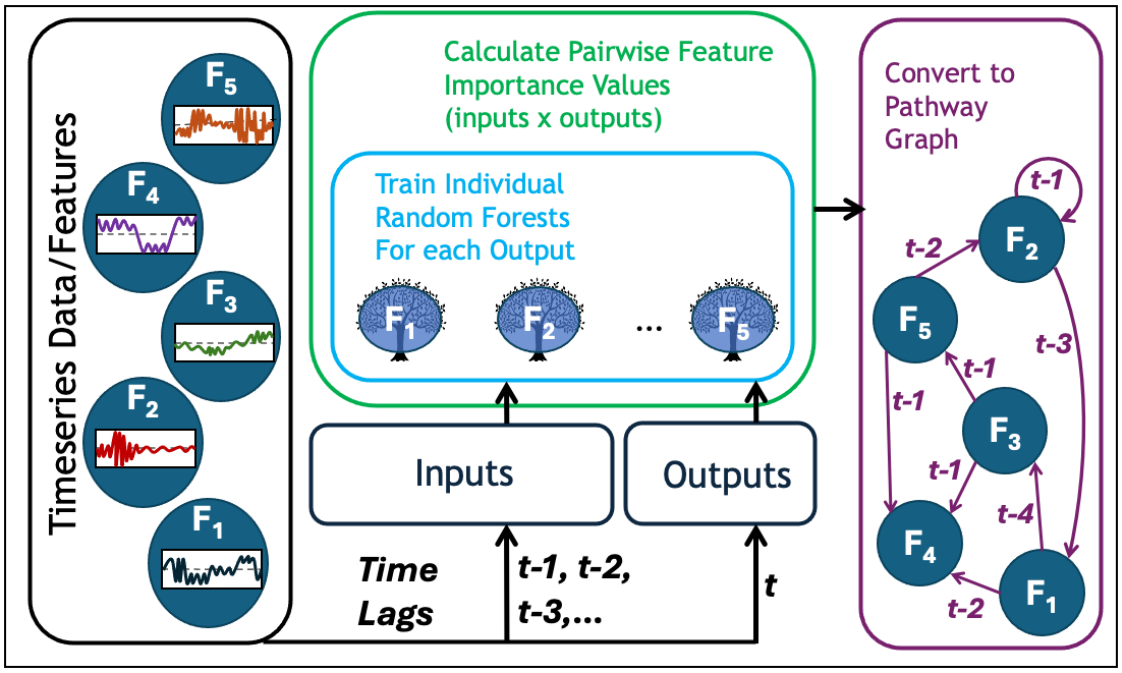


Figure 4-2. Visual depiction of the RFR-based pathway construction approach.

In the first step of our workflow, for each spatio-temporal feature of interest, we train an individual RF with the outputs given by the feature of interest at time t and the inputs are all the features at previous time steps/lags, e.g., $[t-1, t-2, t-3, ...]$. This provides insight into relationships between different features of interest as well as the time lags at which these relationships occur.

In step two of our approach, pairwise relationships between features of interest are quantified using feature importances, obtained via a post-processing step applied to the RF model. We chose to use the SHapley Additive exPlanation (SHAP) method [160] for calculating feature importances, as several recent studies have shown that SHAP values can achieve better results for assessing feature importance in machine learning models than other metrics for various problems of interest

[109, 186, 33]. Additionally and importantly, SHAP proved to be a computationally efficient metric for our analysis.

The final step of our workflow is to use feature importance information data to create a directed graph between the various inputs/outputs. Each node in the graph represents a feature of interest, and each edge represents a directed relationship between any two features of interest (see the right panel of Figure 4-2). The edge will contain two pieces of information: the SHAP value and the associated time lag. To identify the strongest relationships, we prune out edges having a SHAP value below a predetermined threshold; see [15] for more details.

The new RFR workflow for pathway identification described above was verified using three increasingly complex benchmark problems: (i) a set of synthetic coupled equations, (ii) the Held Suarez Williamson-Volcano (HSW-V) test case [77], an idealized atmosphere-only configuration of version 2 of the U.S. Department of Energy's Energy Exascale Earth System Model (E3SMv2) with simplified physics/topography and a point source aerosol injection representing a volcanic eruption, and (iii) a fully coupled E3SMv2 simulation of the 1991 eruption of Mount Pinatubo in the Philippines.

It was demonstrated in [15] that our RFR-based workflow is able to correctly identify the surface cooling and stratospheric pathways for the Mount Pinatubo exemplar (iii). In the surface cooling pathway, which we focus on here for the sake of brevity, an increase in aerosol optical depth (AEROD_v) leads to a decrease in the amount of shortwave radiation reaching the surface (FSDSC), followed by a lowering of the temperature at the surface (TREFHT). Our RFR-based analysis was consistent with these known processes. The strongest connections are those that go into the Subtropical North FLNTC from AEROD_v in both latitudinal bands and from Tropical FLNTC. In addition to intra-zonal band relationships (e.g., AEROD_v$\rightarrow$ FLNTC), we also see some cross-band connections, many of which show a dependence of a Subtropical band variable on a Tropical band variable. These relationships are expected, as they demonstrates a general northward flowing direction from the Tropics to the Subtropical North, which follows the evolution of the volcanic plume that was observed after the Mount Pinatubo eruption. The time lags associated with many of the relationships uncovered are on the order of 21-36 days, which is consistent with the time it took for the aerosols from Mount Pinatubo to encircle the globe. More results, including for the surface cooling pathway, can be found in [15]:

- Meredith G. L. Brown, Matt Peterson, Irina Tezaur, Kara Peterson, and Diana Bull. "Random Forest Regression Feature Importance for Climate Impact Pathway Detection." Submitted to *Journal of Computational and Applied Mathematics*, September 2024. `https://arxiv.org/html/2409.16609v1`

4.1.1. Lessons learned

Applying our new RFR-based workflow to the CLDERA tiered verification problems led to many refinements in the underlying algorithm, including insights into how to window the data around the eruption and normalize each input globally per feature to be between 0 and 1 or -1 and 1. We additionally learned that feature importance-based pruning is essential, as, without pruning, pathway graphs have too many edges, making them uninterpretable.

In applying our approach to actual climate data from HSW-V and Mount Pinatubo simulations, we discovered that the most dominant edges invariably corresponded to "self lags", i.e., the analysis revealed that most variables were primarily dependent on themselves at the previous time. These relationships make physical sense, but can make the identification of non-trivial pathways more difficult. The domination of self lags between variables was particularly evident when we applied our method to the HSW-V data, where we were looking for pathways containing just two nodes: aerosol optical depth and temperature. We hypothesize that this result can be circumvented by including more intermediary variables to our analysis.

Lastly, when applying our method to realistic climate data (HSW-V and Pinatubo), we commonly observed edges appearing in both directions between two features where we only expected an edge in a single direction. At first, we believed that these relationships were erroneous; however, upon further inspection and discussion with climate experts, we concluded that our approach is likely picking up on some hidden feature dependencies between variables. For example, a "back edge" from TREFHT to AEROD_v may be explained by the fact that surface temperature can potentially impact winds (a variable not included in our analysis), which can in turn impact AEROD_v and aerosol spread. Again, we conjecture that adding more intermediary variables to our analysis can help us reconcile our results with the underlying dynamics.

4.1.2. Future research directions

The following research tasks would be of interest towards continuing to refine our RFR-based pathway approach.

- Instead of performing a fixed windowing of the data [15], one could consider employing sliding time windows, which could provide better insight into the temporal system-wide dynamics and allow us to avoid the algorithm picking out primarily self lags.
- Thus far, we have applied our method to data averaged globally as well as by latitudinal bands. Future work should look at more regional analyses, e.g., by performing averaging over IPCC regions [82].
- The relationships identified by our RFR-based approach are correlative. As such, they are not guaranteed to be causal. A very useful future research endeavor would be to apply some kind of causal analysis to the pathway graphs produced by our approach, so as to pick out causal vs. correlative edges.

4.2. Profiling

Contributing Authors: Andrew Steyer (1442), Luca Bertagna (1446), Graham Harper (1442), Jerry Watkins (8734), Irina Tezaur (8734), Diana Bull (8931).

The profiling subthrust under CLDERA performs in-situ coanalysis of E3SM model variables for source-impact pathway identification. The main contributions of profiling are: (i) the development of a lightweight and efficient open-source software library known as CLDERA-Tools (`https:`

//github.com/sandialabs/cldera-tools), and (ii) a mathematical framework and associated algorithm for identifying pathways, represented as time-dependent directed acyclic graphs (DAGs), from high-frequency time-series data available in-situ while running E3SM.

CLDERA-Tools (CT) is a C++ library designed to perform in-situ coanalysis in a host application (in this case, E3SM), with the goal of minimizing the amount of code that needs to be added to the application. CT works by calculating and tracking a set of user-specified quantities of interest (QOIs), e.g., global max/min/average, zonal means, regional/masked integrals (possibly weighted), vertical integrals (possibly weighted), horizontal integrals and more. The user must additionally provide a set of baseline data (e.g., data from a previously-run counterfactual simulation). Anomalies are flagged by comparing QOIs computed within the code using CT to this baseline using bounds tests and Z-scores (see [172] for more detail). Importantly, these comparisons are done in-situ while running an E3SM simulation. All inputs to CT are conveniently specified in a `yaml` file created by the user.

During model initialization, CT interfaces are called to "register" the model variables that will be needed during the QOIs calculation. Besides the metadata (name, layout, data type, etc.), the user can also provide a pointer to the area of memory where the variable is stored, allowing CT to access the model data at run time without any additional memory movement. If the variable is computed at runtime in temporary arrays, data will have to be copied over to CT at every time iteration. At runtime, CT will compute all the QOIs at once, and store them in an internal database, which can also be dumped to a file via PnetCDF (`https://parallel-netcdf.github.io`).

As mentioned earlier, profiling represents pathways as time-dependent DAGs where nodes in the DAG correspond to QOIs and edges in the graph represent relationships between QOIs (informed by e.g., subject matter expertise and/or problem dynamics), with the direction of the edges denoting the flow of impacts. Our approach starts with a so-called base-DAG $\mathcal{G}_B = (V_B, E_B)$ consisting of a set of r-vertices $V = \{v_l\}_{l=1}^{r}$ with node v_l corresponding to the QOI Q_l and a set of edges $E = \{e_j\}_{j=1}^{s}$ determined by assumed relationships between QOIs. These hypothesized relationships are informed by subject matter expertise (e.g., increases in tracer concentrations causes increases in aerosol optical depth). Associated to each node v_l is a (potentially time-dependent) bounds test $\tau_{l,m}$, taking Boolean values in $\{0, 1\}$, that determines when a node will be active ($\tau_{l,m} = 1$) or inactive ($\tau_{l,m} = 0$) with m representing the model time-step. A pathway DAG $\mathcal{G} = \{\mathcal{G}_m\}_{m=0}^{M}$ is time-dependent DAG where at each time-step m the DAG $\mathcal{G}_m = (V_m, E_m)$ is comprised of a set of vertices $V_m \subseteq V$ and set of edges $E_m \subseteq E$ i.e. is a subgraph of the base-DAG. For more details on this algorithm, the reader is referred to [172].

4.2.1. Verification on HSW-V test case

We demonstrated the efficacy and efficiency of CLDERA-Tools using the idealized HSW-V test case [77] in the following article [172] and that is summarized below:

- Andrew Steyer, Luca Bertagna, Graham Harper, Jerry Watkins, Irina Tezaur, and Diana Bull. "In-situ data extraction for pathway analysis in an idealized atmosphere configuration of E3SM." Submitted to *Computing in Science & Engineering*, August 2024. `https://arxiv.org/abs/2408.04099`

In [172] we show the added cost of running HSW-V with CLDERA-Tools in-situ analysis. In this study we take mean additional run-time of a 10 member ensemble running E3SM with in-situ analysis enabled tracking 7, 35, 175, or 875 QOIs corresponding to integrated value of a model field in the stratosphere over a zonal region and compare to the mean value of a 10 member ensemble not tracking any QOIs to measure the percent added run-time. Percent added cost was minimal (less than 2%) for runs tracking 175 QOIs or fewer and only about 12.5% when tracking 875 QOIs. See Section 5.1 of [172] for more details.

In our pathway analysis, we considered five model fields: stratospheric SO2 concentration (SO2), stratospheric SO4 concentration (SUL), aerosol optical depth (AOD), stratospheric temperature (T), and surface temperature (ST) on four zonal regions (equatorial (e), subtropical north (s), temperature north (t), and polar north (p)). From these fields and regions, we defined the following QOIs, denoted by $Q(\mathrm{x})$: $Q \in \{\mathrm{SO2}, \mathrm{SUL}, \mathrm{AOD}, \mathrm{T}, \mathrm{ST}\}$. Here, $x \in \{\mathrm{e}, \mathrm{t}, \mathrm{s}, \mathrm{p}\}$, so that $Q(\mathrm{x})$ is the integrated value of a given QOI over the horizontal region x (and over the vertical region defined by the stratosphere for SO2, SUL, and T). To try to answer the question of "how long can you go?" – that is, how small can the eruption be to be detectable? – we performed our analysis not only on a simulation of the Mt. Pinatubo eruption, but also on two variants of the Mt. Pinatubo eruption, in which 5 Tg and 20 Tg of sulfur were released into the stratosphere.

We constructed pathway DAGs based on the above QOIs using Algorithm 1 from [172]. The bounds tests used to add/remove nodes from the pathway DAG in our analysis (see Algorithm 1 in [172]) were defined in terms of the expected maximum concentration for SO2(x), SUL(x), and AOD(x) (since these variable have zero concentration in the eruption-free counterfactual), as well as in terms of a Z-score computed for the variables T(x) and ST(x). The mean and standard deviation of this Z-score were calculated from the standard deviation of a counterfactual ensemble. We considered four values for the Z-score, 0.75, 1.0, 1.5, and 2.0, corresponding to experiments Ex1, Ex2, Ex3 and Ex4, respectively. For more details regarding these experiments, as well as visuals of the key results the reader is referred to Section 5 of [172]. QOIs are typically active sooner and for a longer time-interval in runs with larger eruption mass, in line with expectations. There is an amount of variability, in particular for the first activation time in the 5 Tg eruption ensemble for AOD(e), T and ST. Experiments which are more sensitive (i.e. smaller Z-score values) correspond to T(x) and ST(x) (x=e,s,t,p) becoming active sooner and for longer time-intervals in line with physical expectations. SO2(p) does not typically become active since most SO2 reacts into SO4 before it has been transported to the polar regions.

4.2.2. Verification on Pinatubo surface heating pathway in E3SMv2-SPA

Next, we apply our method to a fully-coupled simulation of Mt. Pinatubo performed using E3SMv2-SPA, and two perturbed mass variants of this eruption. We consider five model fields: vertically integrated SO4 concentration (BURDSO4), net long-wave flux at the top of the model (FLNT), clearsky downwelling solar flux at surface (FSDSC), vertically integrated stratospheric temperature (T), and near-surface temperature (TREFHT). To define QOIs we take the zonally integrated values of these fields in the four northerly zones: (equatorial (e), subtropical north (s), temperature north (t), and polar north (p)). We denote by $Q(x)$ the integrated value of field Q over the zonal region x where Q =BURDSO4,FLNT,FSDSC,TREFHT,T and x =e,s,t,p. This leads to the P1 pathway

(see Figure 4-3) where SO4 impacts both FLNT and FSDSC which then impact T and TREFHT, respectively. For bounds testing, we use the Z-score test for each QOI where a variable is set to be active if its Z-score is above 1 and then set to be inactive if its Z-score subsequently falls under 0.5. We consider 9 member limited variability ensembles of eruption runs with eruption masses of 5 Tg, 10 Tg, and 15 Tg. As for the HSW-V analysis in the previous section, we look at the mean and standard error for both the first time and total time a given variable is active.

Results for our pathway analysis are in Figures 4-4 through 4-7. For the equatorial and subtropical north regions, QOIs become active sooner and for longer time-intervals. In the temperate and polar north regions, this is not always the case (e.g. in Figure 4-4 the first time TREFHT(t) becomes active is always sooner for the 10 Tg eruption runs than for the 15 Tg eruption runs). We suspect that increased internal variability relative to the HSW-V simulations results in this non-ideal behavior.

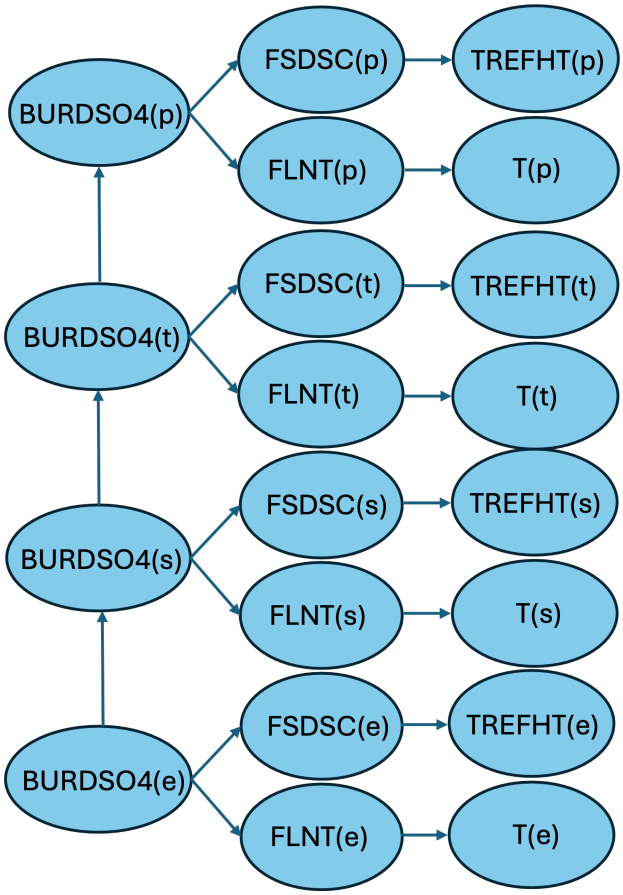


Figure 4-3. Visualization of the Pinatubo stratospheric and surface temperature pathways hypothesized in E3SMv2-SPA.

4.2.3. Entropy-based feature selection

The high frequency data generated by CT suffers from two problems: high variability and high dimensionality. In particular, pathway DAGs created by our algorithm can be extremely large and complex, making them difficult to analyze/interpret. In an effort to mitigate this problem, we

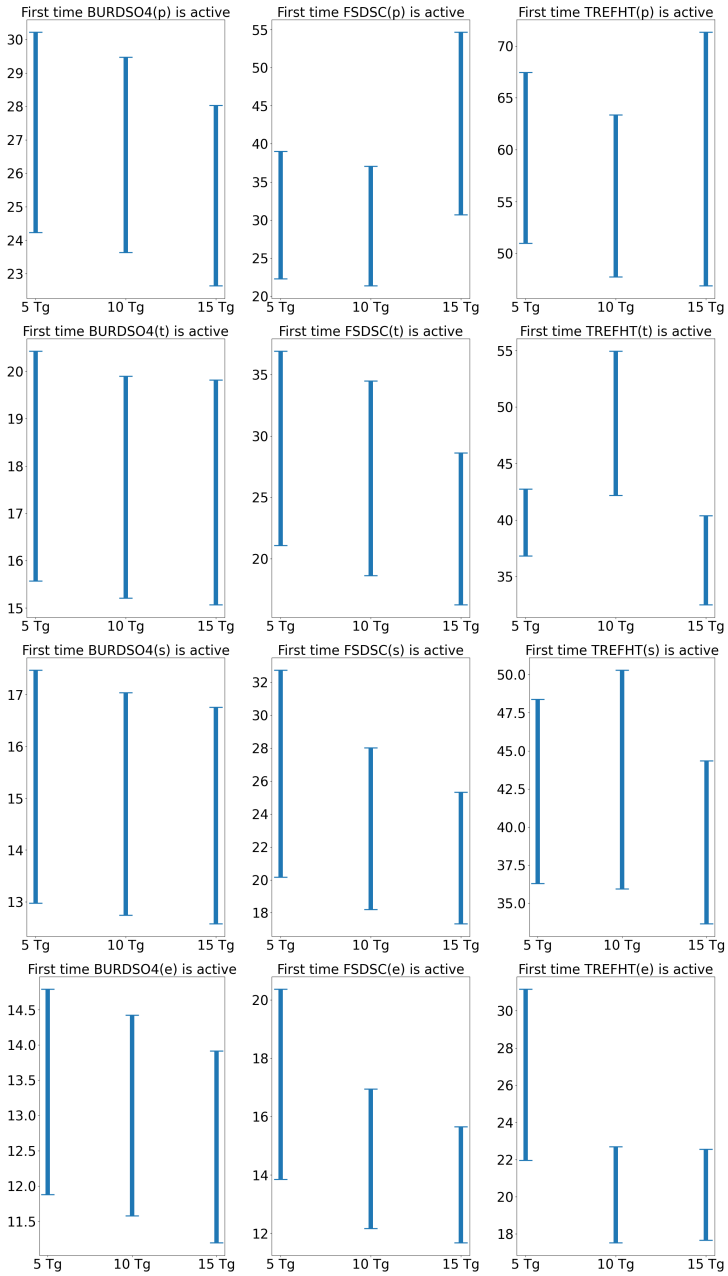


Figure 4-4. Pinatubo temperature pathways: first activation times with error bars denoting the standard error (mean and standard error taken with respect to the 9 member ensemble corresponding to an eruption mass of 5,10,15 Tg) for BURDSO2, FSDSC, TREFHT.

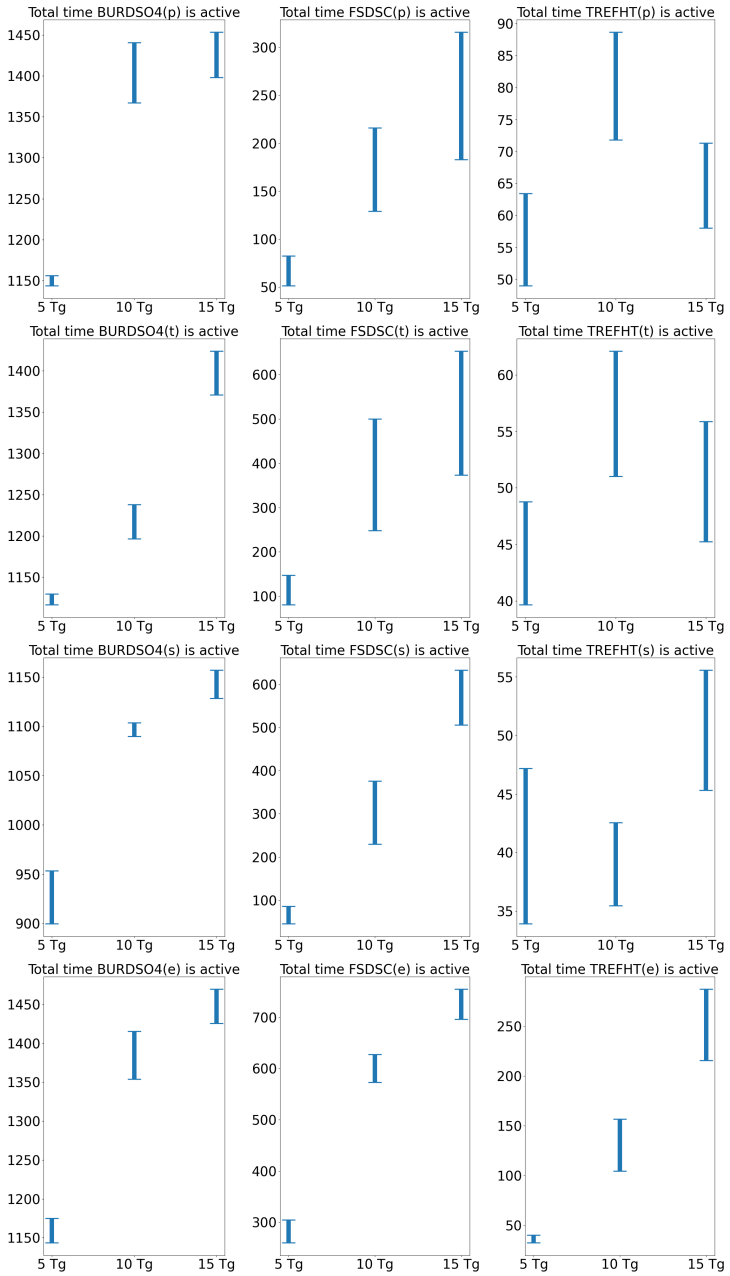


Figure 4-5. Pinatubo temperature pathways: total activation times with error bars denoting the standard error (mean and standard error taken with respect to the 9 member ensemble corresponding to an eruption mass of 5,10,15 Tg) for BURDSO2, FSDSC, TREFHT in all four zones.

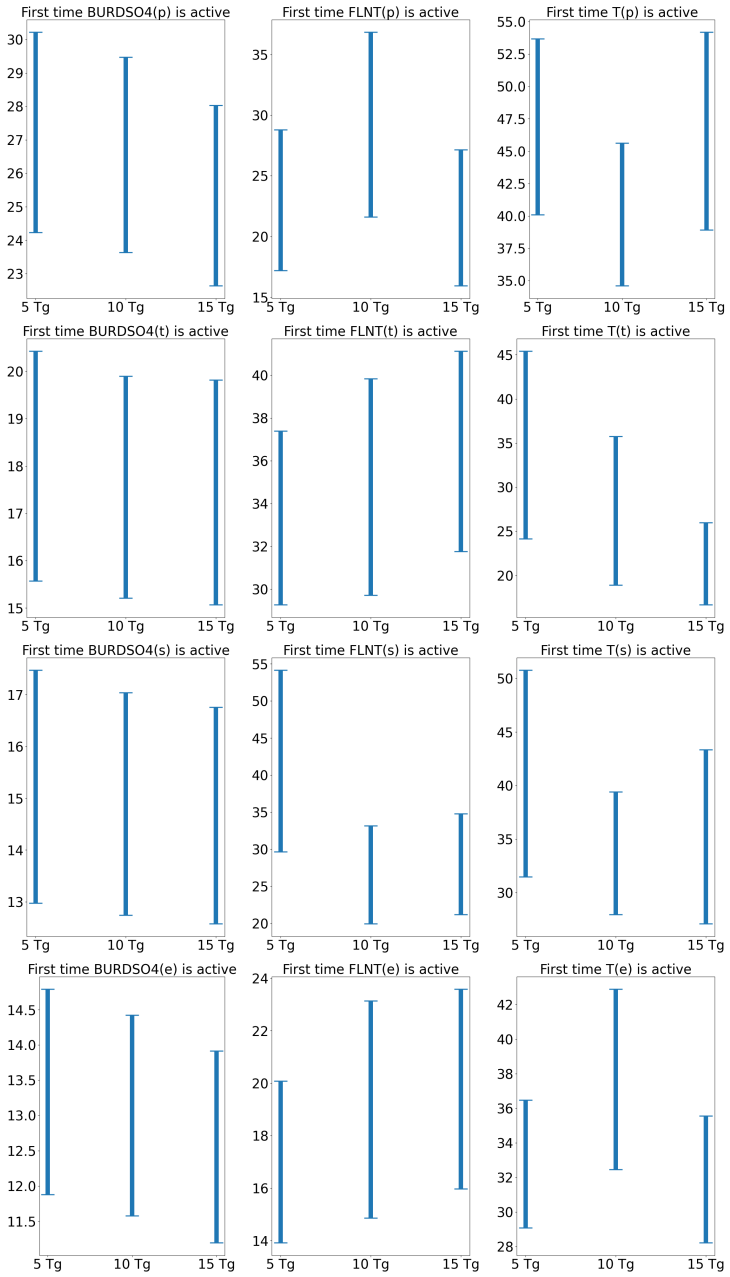


Figure 4-6. Pinatubo temperature pathways: first activation times with error bars denoting the standard error (mean and standard error taken with respect to the 9 member ensemble corresponding to an eruption mass of 5,10,15 Tg) for BURDSO2, FLNT, T.

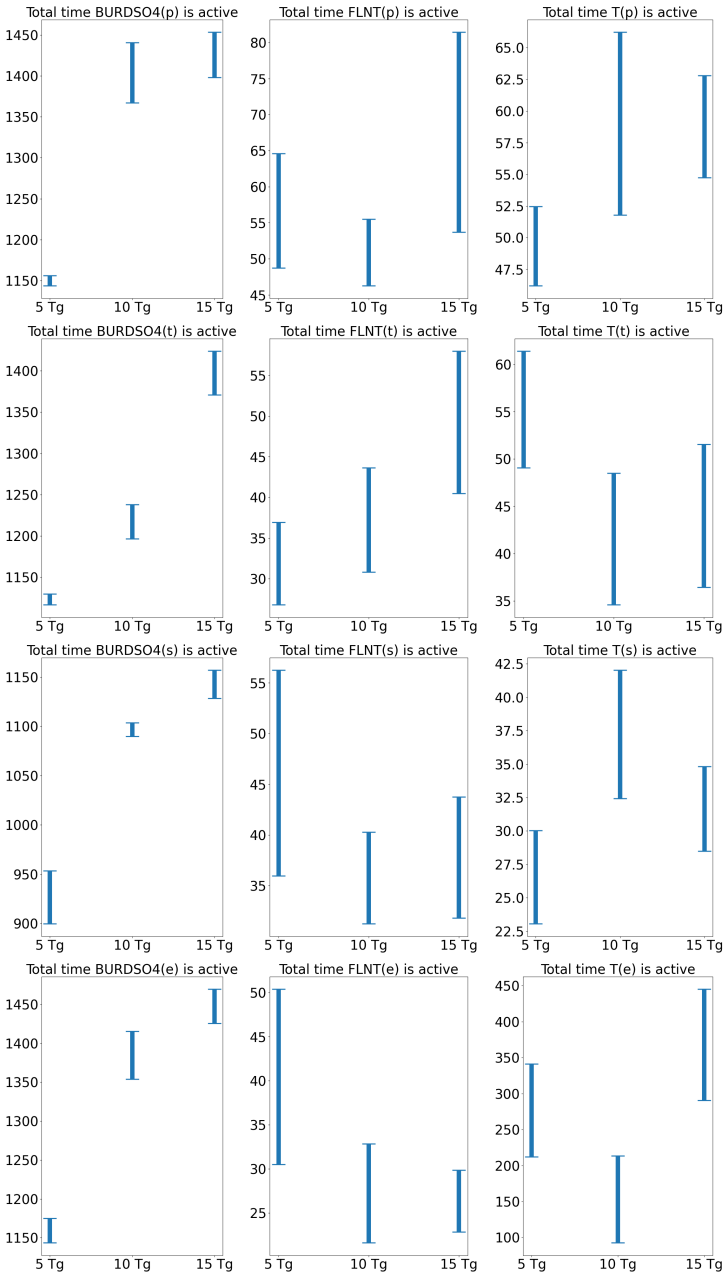


Figure 4-7. Pinatubo temperature pathways: total activation times with error bars denoting the standard error (mean and standard error taken with respect to the 9 member ensemble corresponding to an eruption mass of 5,10,15 Tg) for BURDSO2, FLNT, T.

developed an entropy-based feature selection method that attempts to focus on the most important features related to a given source-to-impact pathway. The novelty in the method is in its use of cross-fuzzy entropy [199] and changepoint detection to isolate regions of constant entropy when comparing data from an ensemble with the source (in our case, the eruption) to a counterfactual ensemble (an ensemble without the eruption). Here, entropy refers to a measure of regularity and synchronicity between the two signals. A publication on our new entropy-based approach is currently in preparation [187]:

- Jerry Watkins, Luca Bertagna, Graham Harper,Andrew Steyer, Irina Tezaur, and Diana Bull. "Entropy-based feature selection for capturing impacts in Earth system models with extreme forcing." Submitted to *Journal of Computational and Applied Mathematics*, September 2024.

In this paper, we were able to detect a statistically significant decrease in temperate north near surface air temperature between April and November of 1992 and construct a dependency graph with direct connections to aerosol optical depth and solar flux at the surface from the eruption.

4.2.4. Lessons learned

The primary challenge encountered by the profiling subthrust stemmed from the fact that the pathway DAGs produced by our methods can be incredibly complex, making them very difficult to interpret. Our efforts in developing the entropy-based feature selection approaches described in Section 4.2.3 and [187] are a promising mitigation that has the potential to simplify pathway analysis and interpretation.

A positive lesson learned from this work was our discovery that profiling using CT does not introduce significant overhead to an E3SM simulation being run. It was unknown that this would be the case a priori.

4.2.5. Future research directions

We are currently extending CT to the E3SM Land Model (ELM), towards using this tool to study the impact of volcanic eruptions on agriculture and crop production. This work is in support of a new FY25 LDRD project, which will use profiling and CT to develop optimal SAI-based controls and study their impact on the temperature.

We envision other future applications of profiling and CT, including:

- in-situ monitoring and real-time manipulation of internal model variables within the E3SM on the native model grid;
- studying the impacts of climate intervention strategies by using CT to modify internal model variables;
- applying causal analysis approaches to DAGs produced by profiling, so as to distinguish correlative relationships from causal ones.

4.3. Clustering and multivariate data-mining

Contributing Authors: Max Carlson (8734), Warren Davis (1441), Irina Tezaur (8734), Diana Bull (8931)

The clustering and multivariate data mining subthrust under CLDERA developed a new clustering and data miniming-based methodology for detecting impacts and identifying source-impact pathways using (simulated or observed) climate data. Our work in this area leveraged algorithms developed in a synergistic ASCR-funded project known as In-Situ Machine Learning for Intelligent Data Capture (ISML) [161].

Our approach begins by breaking up the spatial grid on which a dataset is defined into so-called "analysis partitions", and calculating reduced representations of the data known as "signatures" within each partition (see Figure 1 in [31]). The simplest example of a signature is a variable average. Whereas the primary goal of ISML was to identify anomalies in the data by comparing signatures with each other, here the objective was to track the spatio-temporal evolution of non-stationary variable clusters. Toward this effect, we performed a k-means clustering over each variable's signature independently using data over all spatial locations, timesteps and, when applicable, ensemble members. To ensure rigor and explainability, we clustered the data over several values of k, and determined the most stable value using metrics such as intra-cluster distance and the adjusted rand index; for details, please see [31]. After a stable number of clusters k was determined, each variable was clustered independently, and multivariate groupings of clusters were analyzed. We demonstrated that the prevalence of certain cluster patterns can provide insight into impacts within the underlying climate data. For instance, a dataset from a simulation in which the Earth is generally warmer (e.g., under the ICC's RCP-8.5 emissions scenario) is likely to have a larger number of clusters corresponding to higher temperatures. By tracking the number of hotter vs. cooler clusters in time and performing statistical t-tests to compare these cluster counts/histories with those in an analogous counterfactual simulation, detection is possible using our approach.

Our next step following detection is actual pathway identification. In [31], we demonstrated that this is possible by: (i) mining frequent multivariate cluster evolutions in space and time, and (ii) comparing cluster evolution patterns in an ensemble of Pinatubo simulations to their counterfactual analogs to identify statistically significant differences. The first step can be done efficiently by utilizing algorithms from natural language processing (NLP) [3]. While our approach has the potential to discover pathways "from scratch", we found that such pathways are difficult to verify given the sheer number of cluster evolution patterns that are typically identified. To circumvent this challenge, we worked with climate subject matter experts to: (i) postulate several pathways expected to be present in the data, (ii) translate these pathways into rules that can be mapped to specific cluster evolution patterns, and (iii) use our approach to confirm the presence of these pathways in the relevant data. In addition to finding pathways, our approach enables us to investigate the precise locations of a particular pattern evolution, and visualize it on a geospatial map.

More details on our approach, including results for the Pinatubo stratospheric heating pathway, can be found in [31]:

- Warren L. Davis, Max Carlson, Irina Tezaur, and Diana Bull. "Spatio-temporal Multivariate Cluster Evolution Analysis for Detecting and Tracking Climate Impacts." Intended submission to *Journal of Computational and Applied Mathematics*, September 2024.

4.3.1. Lessons learned

Most of the methodologies developed under the clustering and multivariate data-mining subthrust are not computationally expensive. Calculating the stability scores can be expensive, but these calculations only have to be done once per variable at the beginning, and do not need to be recalculated for the inference steps. The most expensive part of our workflow is the extraction/mining of the cluster evolution patterns. Not only are the patterns themselves potentially an exhaustive enumeration of cluster IDs per state variable, but the evolutions of the patterns is also a combinatorial explosion. While we were able to make the data mining part of our approach more efficient through the use of the `NLTk` library, the sheer volume of patterns identified within our data was intractable to interpret/analyze. As a result, we focused our analysis in [31] on using our approach in a confirmatory rather than an exploratory manner.

4.3.2. Future research directions

This research presents a significant advancement in our capability to find statistically significant differences in pattern dynamics within climate simulations. However, as implemented in this project, the approach is currently used primarily for confirmatory analyses. There is a great opportunity to use these capabilities for *discovery*, where we mine the pattern evolutions and then extract those that are statistically significant for subject matter expert (SME) review. If the evolution patterns discovered are unknown to climate SMEs but physically plausible, they have the potential to inform climate science. If the evolution patterns are not physically plausible, our approach can be used to identify biases or bugs within the climate model used to generate the data used in the analysis. This latter case might be particularly effective when observational data is available and can be included in our analysis. The feasibility of using our methods for discovery would depend on having the time and resources to combat the curse of dimensionality mentioned above, and the availability of possible alternative approaches mining the pattern evolutions (e.g., different conceptual representations, different data structures, distributed processing, etc.).

5. OBSERVED PATHWAYS THRUST

Similar to the Simulated Pathways Thrust, the Observed Pathways Thrust focused on developing novel statistical approaches to discovering, confirming and explaining relationships within the climate system with the specific focus of confirming the temperature source-impact pathways for most methods. The main difference under this thrust was that the focus was on observational datasets where only one “replicate" of the system is available. This requires us to leverage methods that strongly rely on the spatial and temporal dependency structures. The goal of the developed methods was to identify meaningful quantitative relationships between the multiple variables along the pathway which included aerosol optical depth, radiative flux and temperature, described further in the next section. Again, similar to the simulated pathways thrust, the method developed under the observational thrust can also be applied to simulated data and have uses beyond those which are highlighted here. Two primary methods resulted from this work: (1) Feature Importance applied to Echo State Networks (ESN-FI), a multi-variable method which considered a single predictor and multiple independent variables within a nonlinear machine learning model called an echo state network and (2) a Multivariate Space-Time Dynamic Model (MV-STDM), a multivariate approach which considered multiple predictors while assuming a simplified space-time evolution of the processes. Both approaches successfully identified known relationships and provide unique ways of interpreting the relationships between the multiple climate processes as they evolve over space and time.

In addition to these pathway motivated methods, two novel changepoint detection methods were developed under this thrust. As opposed to anomalies, a change point refers to a change in the underlying data generating mechanism. In practice, this usually means detecting a statistically significant change in the mean or variance of the observed processes. Prior to this work, state of the art change point methods have assumed change points in spatial data occur at the same time rather than sequentially as is the case with many atmospheric events such as a volcanic eruption. The methods developed under CLDERA are the first to allow for spatially-varying change points.

Lastly, due the multimodal nature of true observational data, e.g. from ground sensors or space-based monitoring, a data fusion effort was also a part of this thrust. Multiple methods for data fusion were compared including a multivariate tensor approach, an integrated nested Laplace approximation (INLA) approach [21], a multiresolution Gaussian process model [LatticeKrig, 129] and SSDF[123]. Further, this thrust demonstrated the importance of accounting for differences in the support of each data type, e.g. the different spatial resolutions that each data set is collected over.

The remainder of this chapter summarize the data used under this thrust, followed by an overview of each novel method developed along with outcome highlights and summaries of published, under review or in progress manuscripts that resulted from this work.

5.1. Observational Data

Methods under this thrust were applied to at least two data tiers of the tiered verification method: either synthetic or HSW-V dataset and at least one observational dataset. The echo state network approach described in section 5.2 was additionally applied the E3SM ensemble to compare results with observational data. For the temperature pathway, reanalysis data from Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2) was used as the “observational" dataset to test all methods. Specifically, the following MERRA-2 variables were considered: aerosol optical depth [TOTEXTTAU, 51], shortwave radiative flux [SWTNT, 52], longwave radiative flux [LWTUP, 52], stratospheric temperature at 50mb [T050, 50] and surface temperature measured 2m above the surface [T2M, 53]. Most methods used coarsened data at 7.5 degree resolution working with a 48×24 longitude-latitude grid at monthly time steps. The functional changepoint method detailed in section 5.5 was the only method to consider daily time steps. Only the data fusion methods considered true observations such as temperature data collected by the TOVS instrument aboard NOAA satellites in orbit during the time of the eruption.

5.2. Echo State Network

Contributing Authors: Daniel Ries (5574), Katherine Goode (5573), Kellie McClernon (5573)

The development of algorithmic techniques for quantifying relationships between source and impact variables related to a climate event (i.e., a climate pathway) could help inform policy decisions. Machine learning may provide an approach to understanding relationships between climate variables associated with such a climate event. We explore the use of echo state networks (ESN) to characterize a climate pathway associated with the 1991 volcanic eruption of Mount Pinatubo in the Philippines, which acts as a natural occurrence of an stratospheric aerosol injections (SAI). In particular, we consider the relationships between the variables of sulfur dioxide, aerosol optical depth (AOD), and stratospheric temperature associated with the eruption of Mount Pinatubo. This climate variable pathway has been well-studied, so we can compare our results to the relationships previously described in the literature. See [57] for the full details associated with these methods and analyses.

ESNs are a machine learning algorithm intended for temporal and spatio-temporal data that provide improved computational efficiency compared to recurrent and convolutional neural networks. However, ESNs lack interpretability, which is crucial for understanding how model inputs relate to outputs. We propose two methods for computing feature importance (FI) with ESNs for spatio-temporal data: spatio-temporal permutation feature importance (stPFI) and spatio-temporal zeroed feature importance (stZFI). Both methods “adjust" (permute or set to 0) input variables of interest over a block of times and quantify the effect on the predictive performance of the model. The FI values provide insight into how much an input variable contributes to the predictions over time.

Simulated datasets are used to evaluate the two FI methods. The datasets each contain one input variable with a known relationship to the response variable and a suprious input variable. The study builds confidence in the FI approach, showing that stZFI effectively identifies the importance of across varying noise levels and block sizes, while stPFI is more prone to detecting spurious

importance. Future work could explore data with more complex non-linear relationships to further validate these findings.

The analysis of the Mount Pinatubo climate variables is implemented using Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2) data. We train ESNs to forecast stratospheric temperature one month ahead given lagged values of AOD and stratospheric temperature. All variables contain monthly values on a 24×48 equally spaced latitude and longitude lattice from 1980-1995. By starting with data from 1980, the eruption of a second volcano is included: the 1982 volcanic eruption of El Chichón in southeast Mexico. A total of 25 ESNs are trained to account for random variability in the models. stPFI and stZFI are applied separately for the input variables of AOD and stratospheric temperature with a block size of 6 (i.e., how important are the previous six months for forecasting one month ahead?). Figure 5-1 shows the feature importance values. The results show that the importance of AOD for forecasting stratospheric temperature increases after the eruptions of both El Chichón and Mount Pinatubo, which agree with previous climate science research and supports the capabilities of the approach.

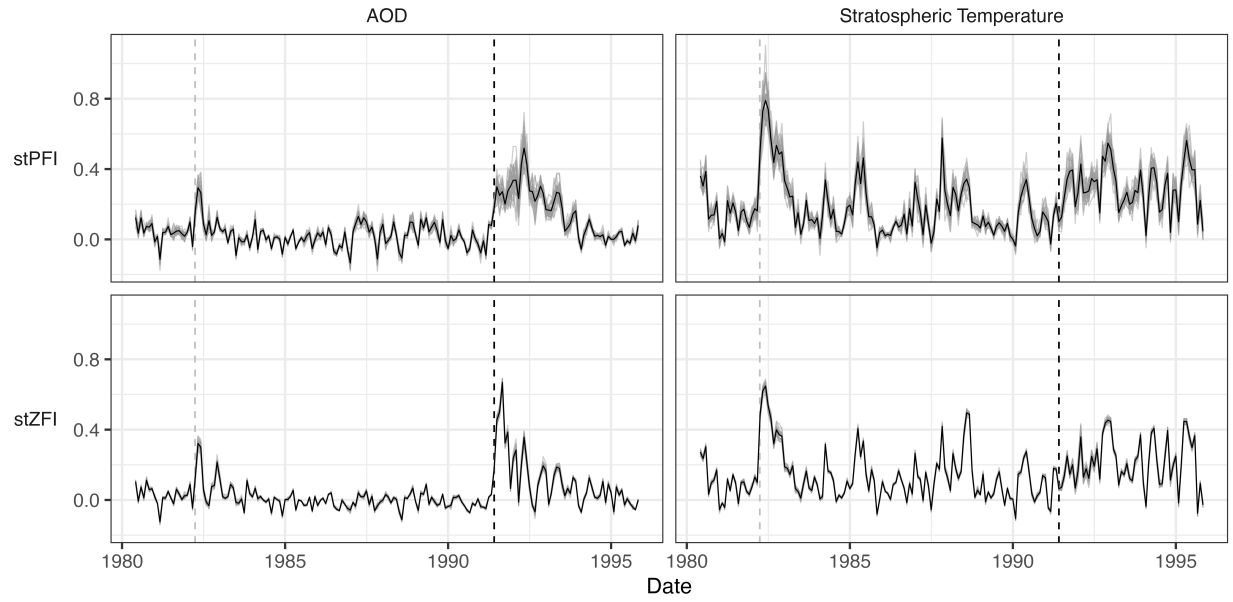


Figure 5-1. Feature importances on MERRA-2 data with a block size of 6. The gray vertical dashed line indicates the eruption of El Chichón, and the black vertical dashed line indicates the eruption of Mount Pinatubo.

5.2.1. Model Validation

The usefulness of conclusions drawn from stZFI [57] are dependent on the inherent usefulness of the ESN model on which they are predicated. Ensuring a good model fit is crucial before using stZFI. For example, [57] used the *repeated hold-out* approach when fitting the ESN to MERRA-2 data of the Mount Pinatubo eruption. Repeated hold-out assumes the objective is forecasting and creates sequentially smaller testing sets from the end of the time series. However, when the analysis goal is to capture variable relationships rather than forecast, an ideal testing set would contain the

event of interest - here an eruption. With observational data, we cannot obtain such a test set since we only have one instance. Climate model simulations, because they can be replicated, offer a way to create ideal training and testing sets for such analyses. [112] compares a new model validation metric, *replicate cross-validation*, which uses replicates from climate models, with repeated hold-out. This paper applies the two metrics under consideration to the idealized climate model *HSW-V* [77] to explore the potential of using climate replicates to inform the development and assessment of statistical techniques for observational data.

We consider the pathway of using aerosol optical depth, stratospheric temperature (50mb), and surface temperature to predict future surface temperatures. Hyperparameters for the ESN, as detailed in [57], are tuned using replicate cross-validation across the 5 HSW-V replicates. Performance of the replicate cross-validation (CV) metric and the repeated hold-out method are assessed for HSW-V replicate 1 with 100 ensembled ESNs. Figure 5-2 shows that the RMSE on the hold-out test set provides a conservative approximation to the replicate CV in the absence of replicates, i.e. observational data. The replicate CV is still preferable since it provides an measure of performance on a new data set. [112] also applied stZFI across all HSW-V ensembles, but results are omitted here as they are similar to the conclusions drawn from MERRA-2 data in [57]. See [145] for stZFI on HSW-V presented by latitude instead of globally.

Further details on model development and validation can be found in [57, 112]:

- K. Goode, D. Ries, and K. McClernon, "Characterizing climate pathways using feature importance on echo state networks", Stat. Anal. Data Min.: ASA Data Sci. J. 17 (2024), e11706. https://doi.org/10.1002/sam.11706
- McClernon, Kellie. Goode, Katherine. Ries, Daniel. "A comparison of model validation approaches for echo state networks using climate model replicates". Spatial Statistics. 2024. https://doi.org/10.1016/j.spasta.2024.100813

5.2.2. Feature Importance & ESMs

Earth system models (ESMs) provide a wealth of information for climate scenarios which have not been observed. This includes the generation of counterfactuals which omits the occurrence of an event. This is particularly useful for a major event like Mount Pinatubo's eruption, because it serves as an exemplar for a stratospheric aerosol injection (SAI). ESMs also generate ensembles to capture variability across the climate system. From a statistical perspective, these can be thought of as independent replicates. However, ESMs generate large quantities of data which can be difficult to understand and make conclusions. The stZFI method developed by [57] can not only be used as an inferential tool, but also as an exploratory tool. As an exploratory data analysis (EDA) tool, stZFI can provide much richer insights into the interrelationships of variables the ESM outputs compared to simple statistics like means and correlations. Although machine learning models have become popular in the climate science community, they are often used as predictive models rather than explanatory or exploratory tools. As [178] points out, "the ultimate objective of using a neural network can also be the interpretation of what the network has learned rather than the output itself". This helps motivate the development and application of an machine learning EDA tool whose purpose is discovery rather than prediction.

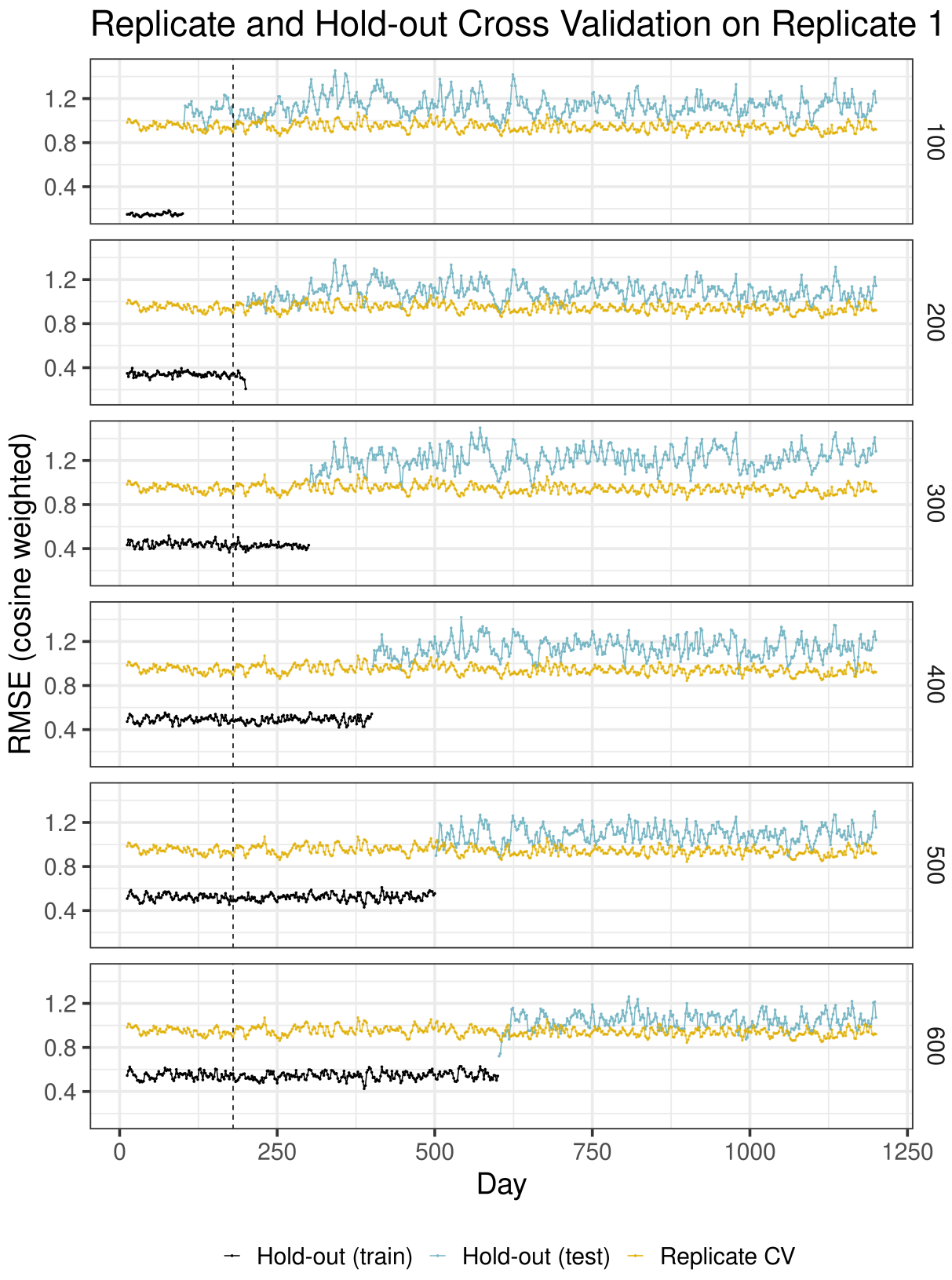


Figure 5-2. Comparison of *repeated hold-out* train and test RMSEs to *replicate cross-validation* RMSEs. All RMSEs cosine-weighted on normalized anomaly scale for HSW-V replicate 1. For repeated hold-out, ensemble of ESNs trained through day shown on row label on the right of plots. Vertical dashed line indicates aerosol injection, day 179.

To illustrate the potential of stZFI as an EDA tool we use the 1991 Mount Pinatubo eruption as an exemplar for a SAI event. ESM data from E3SM "fullvar" and MERRA-2 reanalysis data are used to showcase the method. We explore two temperature pathways: using aerosol optical depth, long-wave upwelling radiation, and stratospheric temperature at 50mb to forecast future stratospheric temperature at 50mb, and aerosol optical depth, short-wave downwelling cloud free radiation, and surface temperature at 2m to forecast future surface temperature at 2m. Data is monthly from 1991-1998. Additionally, we explore using stZFI on a simpler ESM that has a single climate driver: a synthetic injection of aerosols. We refer to these data as HSW-V, and they serve as a simple test case. Echo state networks as described in [57] are used, with hyperparameters selected and predictive performance assessed using replicate cross-validation [112]. Global feature importance for the input variables is calculated over time to see the importance of each input on the model's ability to predict the output on a global scale. [145] also introduced a new *regional* stZFI metric that shows the regional impacts of the input variables on the output variable. This new metric allows scientists working with ESMs to quickly understand the regional effects of global events.

Summarizing results across the three data sets, the main takeaways were:

- *Aerosols* had the most consistent FI results. In all cases, there was a clear increase in stZFI for predicting temperatures immediately after the SAI, which decreases over time.
- *Radiative flux* variables associated with E3SM and MERRA2 had relatively similar FI trends. For long-wave radiative flux, there was no clear trend in FI with values close to 0 across all times when predicting stratospheric temperatures. For short-wave radiative flux, there was a slight increase in FI after the SAI when predicting surface temperatures.
- *Temperature* FI values agreed between HSW-V and E3SM but differed from MERRA2 results. With stratospheric temperatures, the HSW-V and E3SM results showed a clear increase in FI after the SAI, but the MERRA2 results showed a noisy possible increase in FI. With surface temperatures, the HSW-V and E3SM results showed no FI trends, but the MERRA2 results showed a steadily increasing trend in FI.

The stZFI results can be used to point to new hypotheses and research directions. For example, the upward trend in stZFI for T2M is unlikely due to Mount Pinatubo alone, and could lead to additional research. Another example suggested by the latitudinal contribution plots is the question of how the latitude of an SAI event will affect its impacts. It also could help find areas where climate models do not match observational data. stZFI shows the variables a model is using, and when, in order to predict. Therefore, discrepancies between a climate model and observational data could point modelers to relationships a ESM is not currently capturing.

A full explanation of the data, model and results can be found in [145]:

- Daniel Ries, Katherine Goode, Kellie McClernon, and Ben Hillman. "Using feature importance as exploratory data analysis tool on earth system models." Submitted to *GMD Special Issue: Theoretical and computational aspects of ensemble design, implementation, and interpretation in climate science*, July 2024. The review preprint is avialable at: `https://gmd.copernicus.org/preprints/gmd-2024-133/`

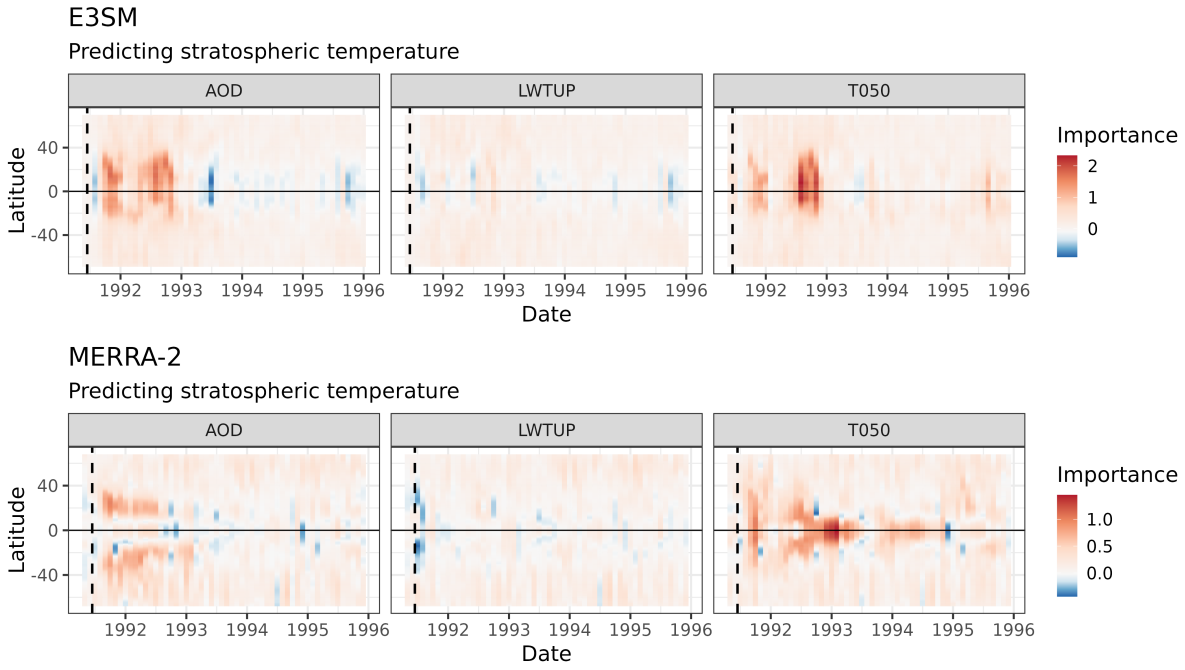


Figure 5-3. Latitudinal contributions to stZFI for E3SM and MERRA-2 for models predicting T050. Note importance scales are different for E3SM and MERRA-2. Vertical dashed lines denote the June 15, 1991 Mount Pinatubo eruption.

5.2.3. Regional Feature Importance

While the global climate variable pathways of the Mount Pinatubo eruption are well-known, many regionally specific impacts are unknown or under debate, prompting the need to develop methods that can identify regional changes in climate variable relationships. [57] developed an explainability metric, stZFI, that quantifies the importance of globally defined variables on predicting global stratospheric temperature. [145] presented the metric in terms of the location-specific contributions to stZFI and showed stZFI average by latitude. We extend the ESN methodology to allow for *regional stZFI* that quantifies the importance of regionally defined variables on the response and aggregate the stZFI location-specific contributions by region [113]. Thus we are able to make inferences about the relative importance of a regionally-specific covariate process on the response at a specified location. We apply the idea of *tiered verification*, using increasingly complex data sets, to validate our method.

Verification data sets assessed in [113] were simulated spatio-temporal Gaussian processes (GP) where the response GP was linearly related to one covariate GP with random noise, HSW-V [77], and E3SMv2 “fullvar" [13]. The `listenr` package [58] was used to fit the ESN and calculate stZFI. The key results on each data set were:

- *Linear model* stZFI correctly assigned zero importance to extraneous covariate not related to the response and illustrated the effect of the signal-to-noise ratio on the noisiness of the importance function.

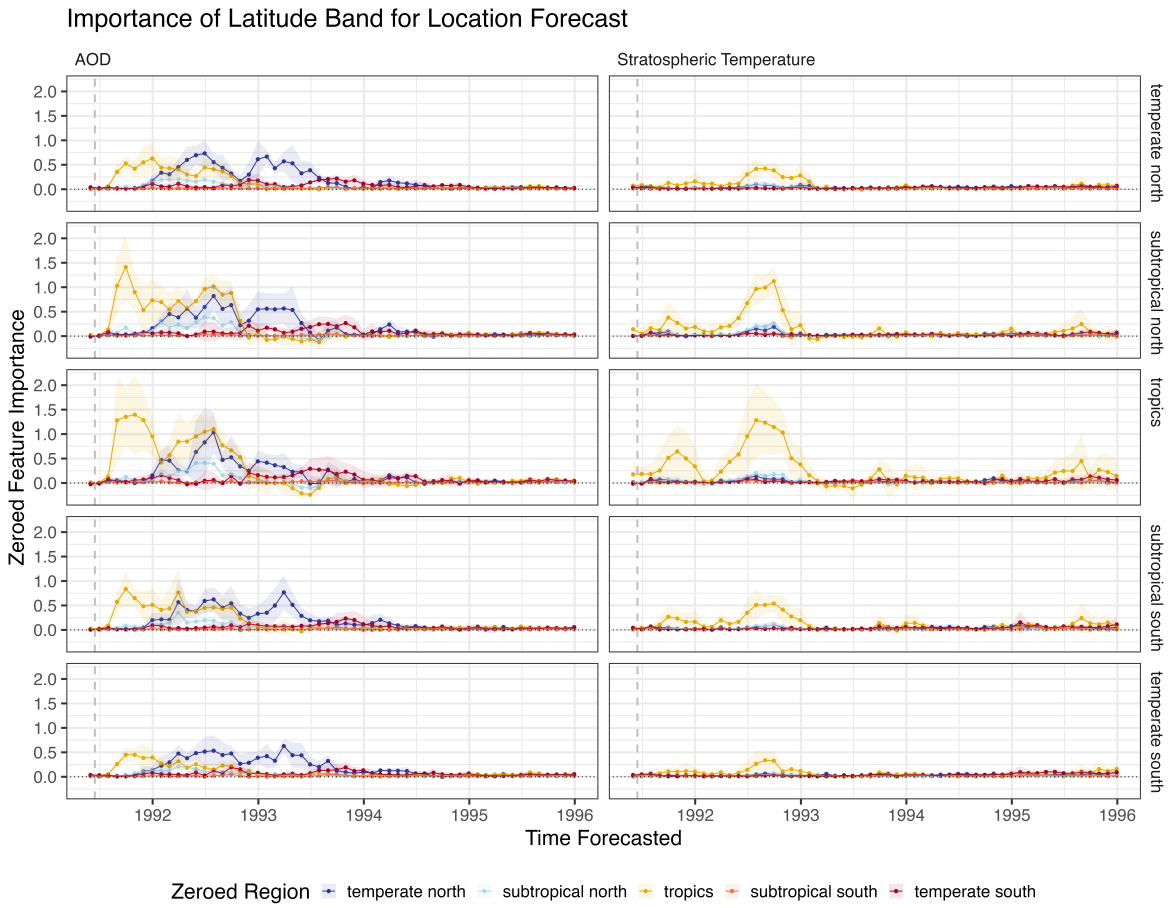


Figure 5-4. Contributions to stZFI shown by response latitude band along the vertical axis for regional features from E3SM predicting stratospheric temperature. Shading represents stZFI ± 1 standard deviation among E3SM ensembles. Vertical dashed lines denote the June 15, 1991 Mount Pinatubo eruption. (Long wave radiation (FLNT) excluded for clarity of presentation.)

- *HSW-V* feature importance results showed the method is able to identify the timing of increases in AOD by latitude and reflect the relative strength of tropical temperature on predicting temperature in other latitudes.

- *E3SMv2* captured the relative importance of AOD and stratospheric temperature, where AOD as the driver of temperature increases is most important. Temperate north AOD importance strengthened in accord with the spread of aerosols around September 1991 out from the origin of the injection, Figure 5-4.

Overall, the importance metric over time agreed with the temporal evolution of the underlying data generating mechanism and was able to identify the relative importance of regional processes on the response. The method showed its ability to faithfully capture changes over time in the regional variable relationships suggesting its usefulness for assessing changing dynamics resulting from either natural climate instigators (e.g. volcanic eruptions or potentially ENSO) or artificial climate mitigation strategies, such as SAI or cloud seeding.

Further details on this research can be found in [113]:

- McClernon, Kellie, Ries, Daniel, and Goode, Katherine. "Verifying regional feature importance on echo state networks for climate pathways." SAND2024-12404O. September 2024, Sandia National Laboratories, Albuquerque NM.

5.3. Dynamic Linear Model

Contributing Authors: Lyndsay Shand (5573), J. Gabriel Huerta (5573), Robert Garrett (5573)

A Multivariate Space-Time Dynamic Linear Model (MV-STDM) was developed to characterize the multivariate and the dynamic nature of climate impacts following the Mt. Pinatubo eruption, a natural analog for stratospheric aerosol injection. Full details of the framework with results are presented in [45] and here we provide a summary of the key contributions of this work to the CLDERA-LDRD project. The goal was to formulate a statistical spatio-temporal model that performs well for interpolation/prediction of spatio-temporal fields with multiple variables while capturing multivariate relationships. Additionally, the model can quantify the contribution of increased AOD levels to changes in stratospheric and surface temperature with associated uncertainty quantification. The model was implemented in a Bayesian framework, and use efficient radial basis function representations as in [27]. The basis function coefficients vary spatially over a spatial grid domain and temporally via Dynamic Linear Models as in [192] with a transition matrix $\boldsymbol{A}$. The variability between state transitions (basis function coefficients) is modeled via a spatial autoregressive matrix. Implementation of the MV-STDM was implemented in R and C++ and is available at `github.com/garretrc/MV-STDM`. Please see [45] for full details of the method and results:

- Garrett, R., Shand, L. and Huerta, J. G.. A Multivariate Space-Time Dynamic Model for Characterizing the Atmospheric Impacts Following the Mt. Pinatubo Eruption. Submitted to *Environmetrics* August 2024.

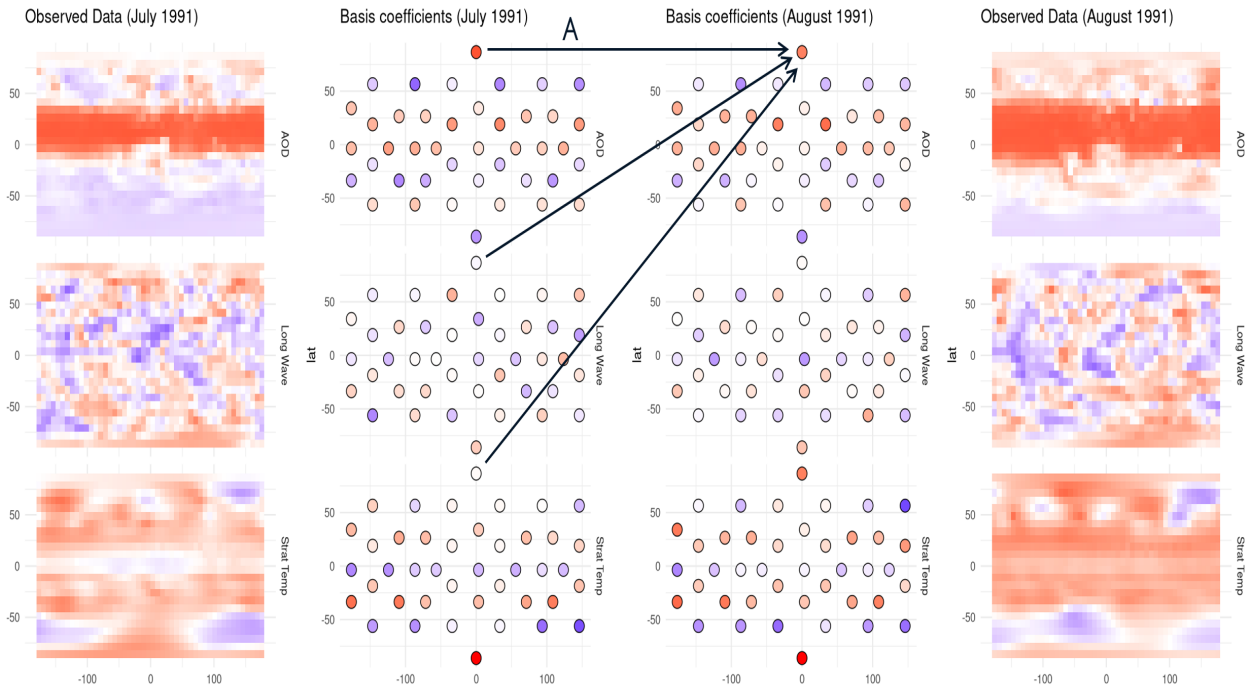


Figure 5-5. Graphical representation of how the evolution matrix of the MV-STDM model operates over a spatial grid domain. The observed data corresponds to spatial fields from MERRA-2 for July 1991 and August 1991 at a 24×48 spatial grid resolution. In this case, the resolution for the basis function coefficients of the RBFs is 42.

The multivariate space-time dynamic model (MV-STDM) involves a high-dimensional parametric representation which is handled thru a sparse representation of the transition matrix $\boldsymbol{A}$ which makes the model tractable. This matrix accounts for correlation in multiple output variables while borrowing information from other variables for prediction. The elements of the basis functions are defined over a spatial resolution grid, so a sparse representation of $\boldsymbol{A}$ captures the dependence of a specific variable in terms of all variables at a previous time point, but only for the same location of each basis grid point. A graphical representation of how $\boldsymbol{A}$ is structured is provided in Figure 5-5 illustrating a model transition from July 1991 to August 1991 where the resolution of the basis coefficients was set to 42 basis functions.

For model implementation and fitting, a Bayesian framework was considered via a Markov Chain Monte Carlo (MCMC) algorithm. The MCMC produces ensembles of model parameters (posterior distribution samples) that provide full uncertainty quantification of the MV-STDM and relies on a Kalman Smoother approach. The MCMC algorithm was verified to work properly on synthetic data sets as described in Section 4 of [45]. A representation of how model fit works is presented in 5-6, where the model fit is broken into the observed data, the estimated spatial mean process and the model residuals. The model fit is illustrated for August 1991 and December 1994, two and forty-one months after Mt. Pinatubo's eruption.

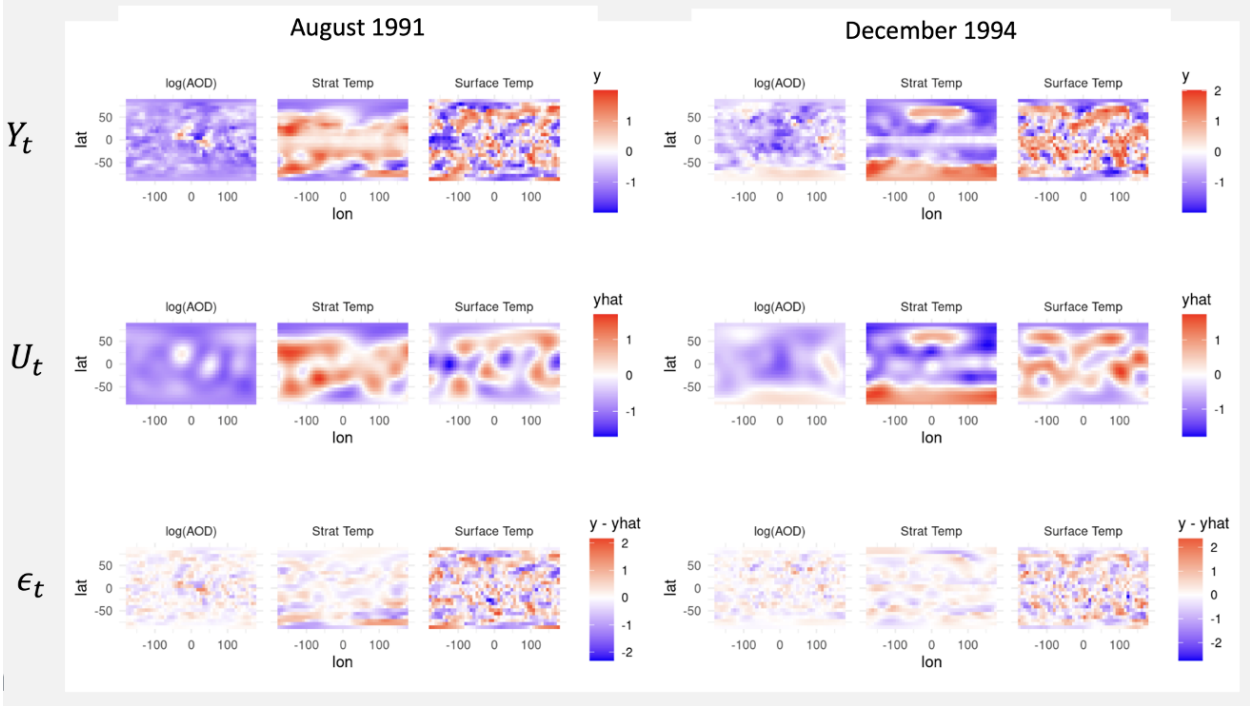


Figure 5-6. Model Fit broken down into observed data Y_t, mean process U_t and residuals ϵ_t , 2 and 41 months after eruption.

5.3.1. Results

For our main application, we studied three different climate variables that had notable observed changes following the Mt. Pinatubo eruption. These variables include aerosol optical depth (AOD), upwelling longwave radiation (LWR), and stratospheric temperature measured at 50mb (T50). This collection of variables represents an important interaction between atmospheric processes following the eruption, where the influx of stratospheric aerosols absorbed a greater than usual amount of outgoing longwave (infrared) radiation, leading to global stratospheric warming which reached up to 3^oC near the equator. We obtain each variable from the Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2). For AOD, we obtain the total aerosol extinction aerosol optical thickness variable at a 550 nm wavelength, or the product labeled TOTEXTTAU. For LWR, we obtain the upwelling longwave flux at top of atmosphere variable, measured in W/m^2, or LWTUP. Finally, for stratospheric temperature, we obtain the air temperature at a pressure level of 50 mb, labeled as T50. A few preprocessing steps were applied to prepare the data for the model. First, the spatial fields for each variable and time point are regridded to a 24×48 spatial grid through spatial averaging. Because our approach is focused on understanding multivariate relationships rather than providing fine-scale spatial predictions, this lower resolution helps to keep the dimension of our data reasonable while still capturing long range spatial trends. Second, we subset a twelve year time period of 1984-1995, centered around the 1991 eruption, to have a sufficient number of time points with which to estimate our model parameters. Lastly, to isolate the effects of the eruption from typical spatial or seasonal trends, we convert the observations for each variable into standardized anomalies. The far left and far right panels of Figure 5-5 show the

MERRA-2 spatial fields for AOD, LWTUP and T50 for July 1991 and August 1991 respectively at a 24×48 spatial grid resolution. Additionally, Figure 1 in [45] shows these anomalies over four 3-month periods before and after the eruption.

For the application to MERRA-2 , [45] provided interpretation of the estimated transition **A** matrix and studied the model's predictive performance for stratosphere temperature while answering to the question, if the MV-STDM provides better predictions than simpler univariate approaches. Figure 5-5 present the estimates of the model's autocorrelation coefficient matrix $\boldsymbol{A}$ projected onto the basis function space. Notable aspects of this figure are that strong positive autocorrelation in almost all regions is present, especially between 30^oS and 30^oN (diagonal blocks of the matrix). Also AOD has a positive impact on T50 in many regions, especially in the mid-upper latitudes (row 3, col 1 of Figure 5-5). We tested our model's performance in predicting stratospheric temperatures over North America for a three-year period following the eruption (August 1991-July 1994). This region is shown in Figure 5-8, and covers the area bounded by 155°W to 35°W longitude and 5°S to 80°N latitude. This holdout set represents a large missing block of observations in both space and time for a single variable and thus highlighted the benefit of our multivariate model for predicted.

We focus on three submodels for predictive comparisons. First, we consider the full MV-STDM with all three variables (AOD, LWR, T50). Then, to evaluate our model's utility as a (univariate) spatiotemporal dynamic model, we fitted a second model for only the T50 anomalies. Lastly, we consider a simplified univariate model where we constrain $\boldsymbol{A} = \boldsymbol{I}_K$, again for T50 only. We denote these three models as "Multivariate", "Univariate", and "Univariate-Random Walk", respectively. The top panel of Figure 5-9 shows the values of the *continuous rank probability score* (CRPS) for each month from August 1991 to July 994, while the bottom panel shows the root mean squared prediction errors (RMPSE) for the the same time scale. Overall, the MV-STDM outperforms simpler models when there is incomplete spatial/ temporal information, as is the case for some observational data. Other variables, especially AOD are important for prediction when spatial information is lacking.

5.3.2. Relative Strengths and Limitation Assessment

The MV-STDM is capable of discovering pathways from observed or simulated data and it allows processing of multivariate output but can also be applicable to single variable output. The model provides estimation of an interpretable multivariate autocorrelation matrix that can support the understanding of spatial interactions of the variables of interest. The model has very good predictive performance and provides uncertainty quantification thru a Bayesian framework. In terms of limitations, the MV-STDM is relatively slow computationally for model fitting in comparison to the Echo State Network (ESN) model. The model's autocorrelation matrix is not time-varying.

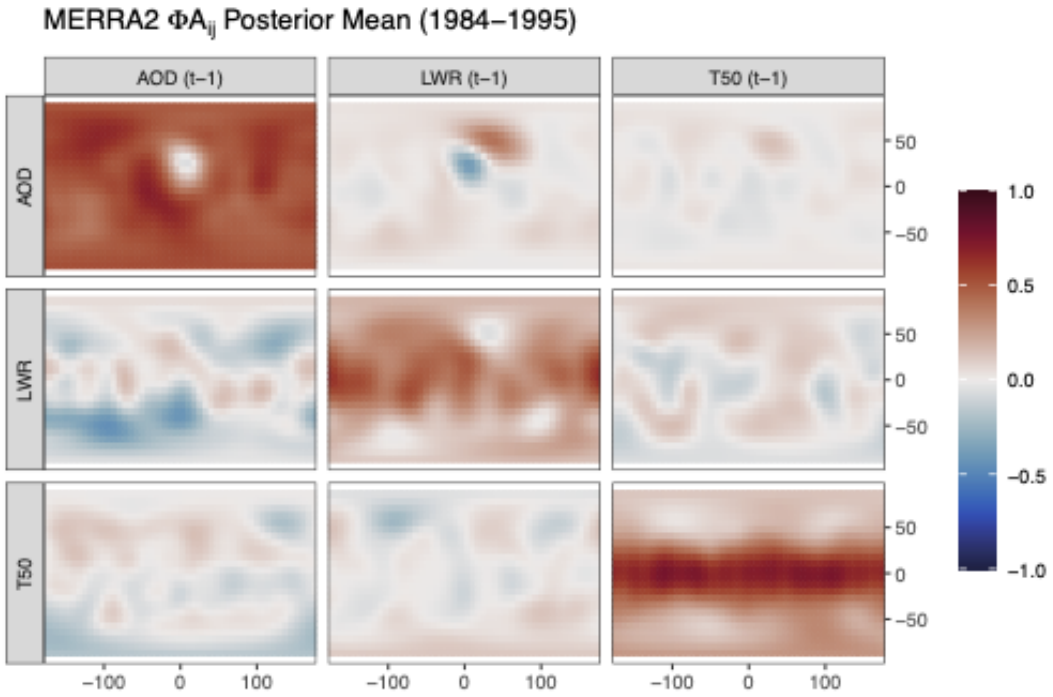


Figure 5-7. Estimates of the autocorrelation coefficient matrix projected onto the basis function space.

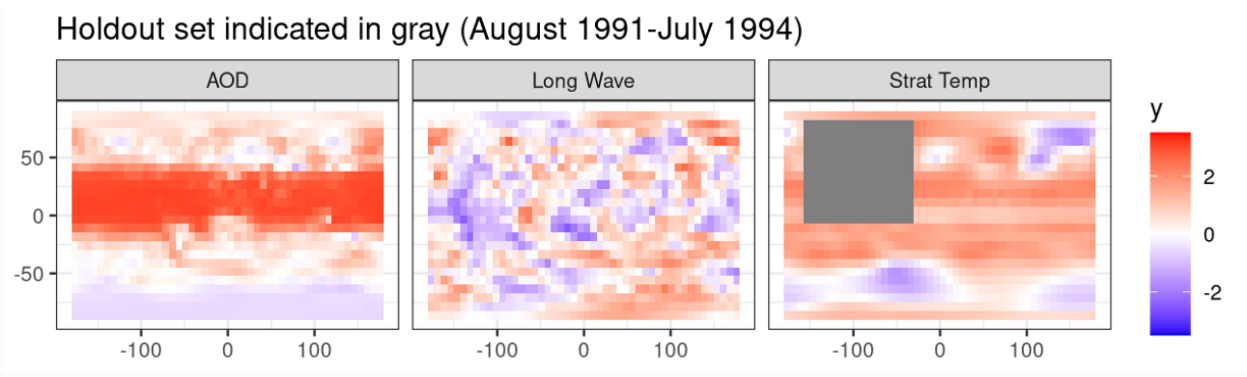


Figure 5-8. Map of the holdout set of observations for stratospheric temperature (T50). The gray region represents the region over which observations are excluded for the August 1991-July 1994 time period when fitting the model. These observations are later used to assess the model's predictive performance. The anomalies shown in the map are for August 1991

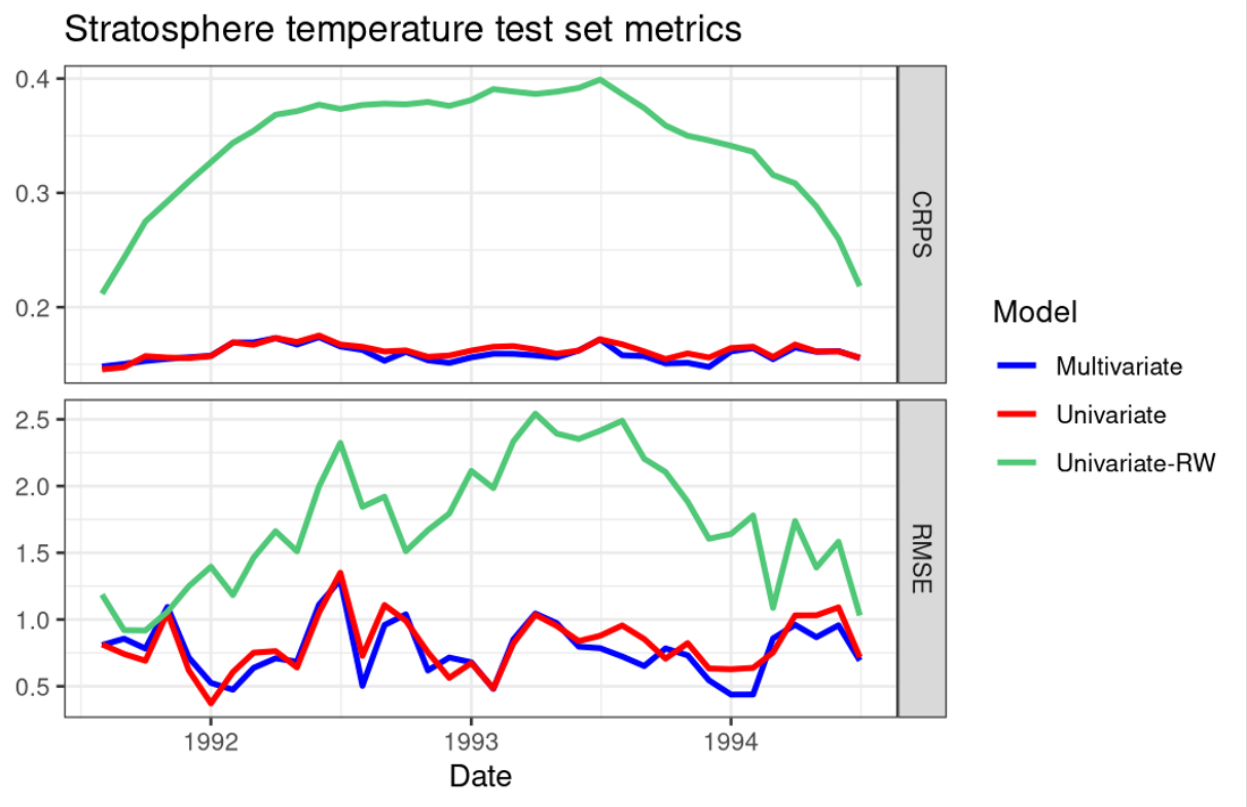


Figure 5-9. Predictive performance for the three submodels on the stratospheric temperature holdout set shown in Figure 5-8. Results are presented for each month from August 1991 through July 1994. The top panel contains the continuous ranked probability score, or CRPS, while the bottom panel contains the root mean squared prediction error, or RMSPE, for each month. Different colors are used to represent each submodel.

5.3.3. Follow-on work

We identified three areas for future work: (1) Apply the MV-STDM to output of climate model simulations to assess the impact of internal variability in the interactions between variables of interest, (2) Assess the similarity of multivariate interactions in climate models versus the same interactions found in observational data and (3) Consider alternative *Kalman smoother algorithms* to improve on the computational efficiency for model fitting of the MV-STDM.

5.4. Space-Time Change Point Detection

Contributing Authors: Lyndsay Shand (5573), Bo Li and Samantha Shi-Jun (University of Illinois, Urbana-Champaign)

Significant events such as volcanic eruptions can have global and long-lasting impacts on climate. These global impacts, however, are not uniform across space and time. Understanding how the Mt. Pinatubo eruption affects global and regional climate is of great interest for predicting the impact on climate due to similar events as well as understanding the possible effect of the Stratospheric Aerosol Injections proposed to combat climate change. While many studies illustrated the impact of the Pinatubo eruption on a global scale, studies at a fine regional scale are scarce. We propose a novel Bayesian spatially-varying changepoint detection and estimation method to trace the impact of Mt Pinatubo eruption on regional climate. Our approach takes into account the diffusing nature and spatial correlation of the climate changes attributed to the volcanic eruption.

Our approach departs from conventional methods by modeling changepoints as a spatial process and further allowing time-after-event of the changepoints to increase with the spatial distance from the event origin. By taking into account the spatial correlation of the data and the anticipated diffusion of the observed impact from the event location, our method is demonstrated to be more effective in capturing the spatial patterns of climate impacts associated with the SAI events. Since our focus is solely on tracing the impact of the Mt. Pinatubo eruption, we consider at most one change after the eruption. Compared to [106], our method extends their model by allowing for spatially-varying changepoints. Furthermore, our approach performs both detection and estimation of changepoint at each location rather than only the estimation. Lastly, while most traditional methods focus on the mean shift, our method can detect changes in either the mean or variance, thus offering greater flexibility. For a detailed description of our model, see [164]:

- Samantha Shi-Jun, Lyndsay Shand, Bo Li. "Tracing the impacts of Mount Pinatubo eruption on global climate using spatially-varying changepoint detection." Accepted *Annals of Applied Statistics*. September 2024. This manuscript is available on arXiv: `https://arxiv.org/abs/2409.08908`

5.4.1. Results

Here, we highlight the results on the MERRA-2 reanalysis data. We apply our method to monthly stratospheric aerosol optical depth (AOD) and latitudinal surface temperature data from 1985 to

1995 to detect and estimate changepoints following the 1991 Mt. Pinatubo eruption. The AOD data is over a 48×16 grid on the sphere excluding latitudes below $60^{c}ircS$ and above $60°N$. The temperature data is aggregated over 24 equally spaced latitude bands from $90°S$ to $90°N$.

Figure 5-10(a) shows the heatmap of estimated changepoints. Our model detects a changepoint at all locations, with estimated values ranging from Jun 1991 to Sep 1991. The estimated changepoints show a pattern driven more by latitude than longitude, with the earliest changepoints predominantly occurring along latitudes 3.5°S through 34°N. Figure 5-11 shows the time series of aerosol after preprocessing and the mean estimates from our model. The estimated changepoints for locations near the latitude of Mt. Pinatubo (15°N) coincide with the month of the eruption, while the changepoints for locations further north or south occur several months later.

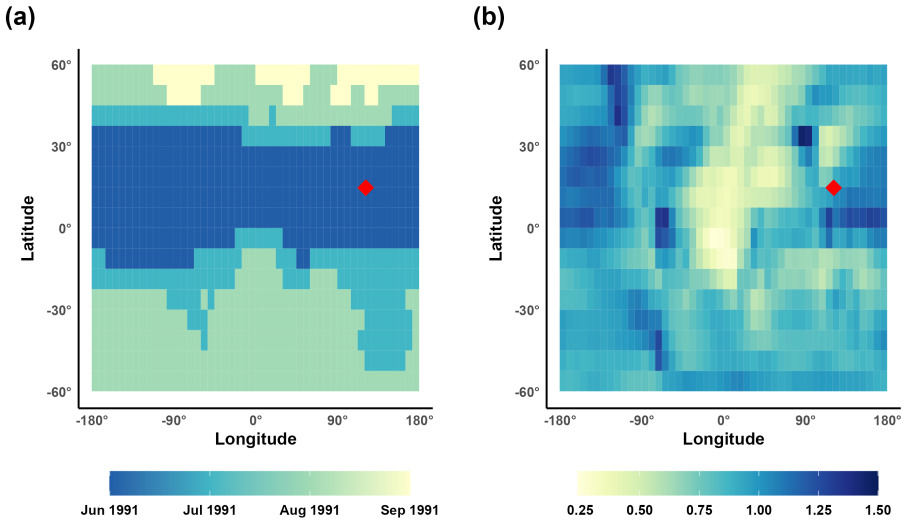


Figure 5-10. (a) Heatmap of detected changepoints for AOD. Darker color indicates earlier change. (b) Heatmap of estimated change amount. Darker color indicates larger magnitude of change. The red diamond marks the location of Mt. Pinatubo.

Figure 5-13 shows the estimated process means before and after a detected changepoint by latitude. Our method detects changepoints for all latitudes in the range $(56°S - 49°N)$ except for latitudes $18.5°S, 26.5°N$, and $34.0°N$. No changepoints were detected in the southern $(86°S - 63.5°S)$ and northern $(56.5°N - 86.5°N)$ ends of the globe. The posterior probability of having no changepoint ranges from 0 to 0.0112 for detected latitudes and 0.7955 to 0.9986 for the undetected latitudes. The detected changepoints range from Sep 1991 - May 1992.

5.4.2. Ongoing work

The methods discussed in the report uses a Metropolis-Hastings sampling for the changepoints and an $N \times N$ covariance matrix generated from exponential covariance function, which is not scalable when N is large. We are currently developing a scalable changepoint detection method for large spatial data on a sphere. The new method uses a latent model for the changepoint process and a covariance function based on Whittle-Matérn SPDE for more efficient sampling. Figure 5-12 shows

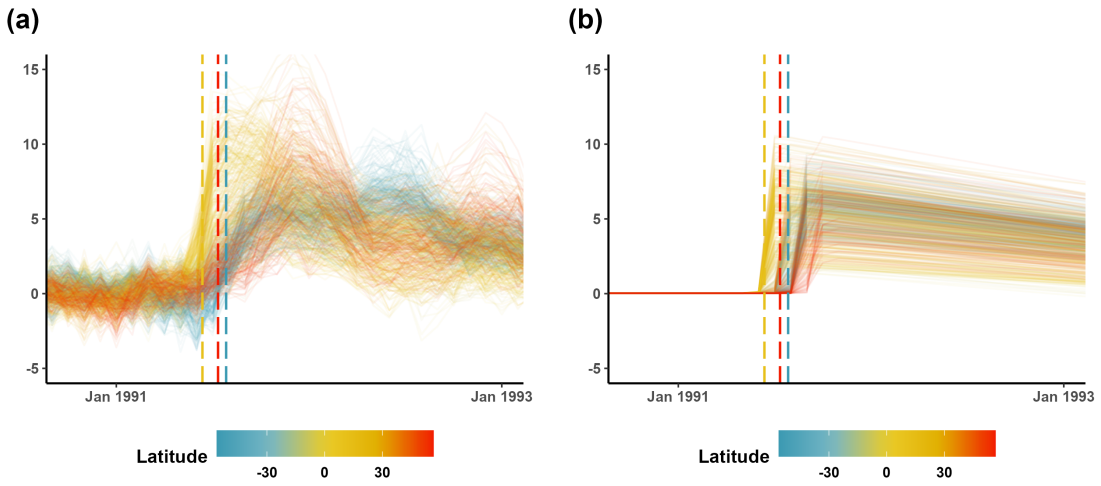


Figure 5-11. Time series of (a) aerosol data after preprocessing and (b) posterior mean of μ_1 and $\mu_2(\mathbf{s}, t)$, zoomed in to years $1991 - 1993$. The color represents the latitude of their location. The blue, yellow and red vertical lines mark the average value of the changepoints in latitude bands $(60°S - 20°S)$, $(20°S - 20°N)$, and $(20°N - 60°N)$.

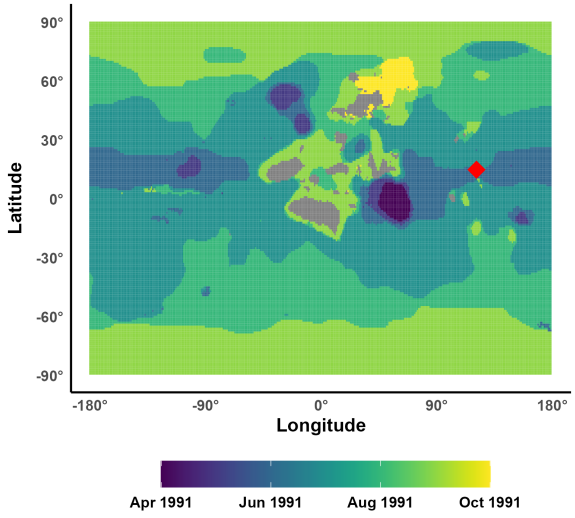


Figure 5-12. Heatmap of detected changepoints

Changepoint	Count
Jul 1991	20916
Sep 1991	19904
Aug 1991	15971
Jun 1991	4845
No CP	1344
May 1991	783
Oct 1991	667
Apr 1991	370

Table 5-1. Summary of detected dates

a preliminary result on the stratospheric AOD data on 360×180 grid. The grey areas indicate that no changepoints were detected. Table 5-1 shows the summary of detected changepoints. Most of the locations have changepoints in the months July to September 1991.

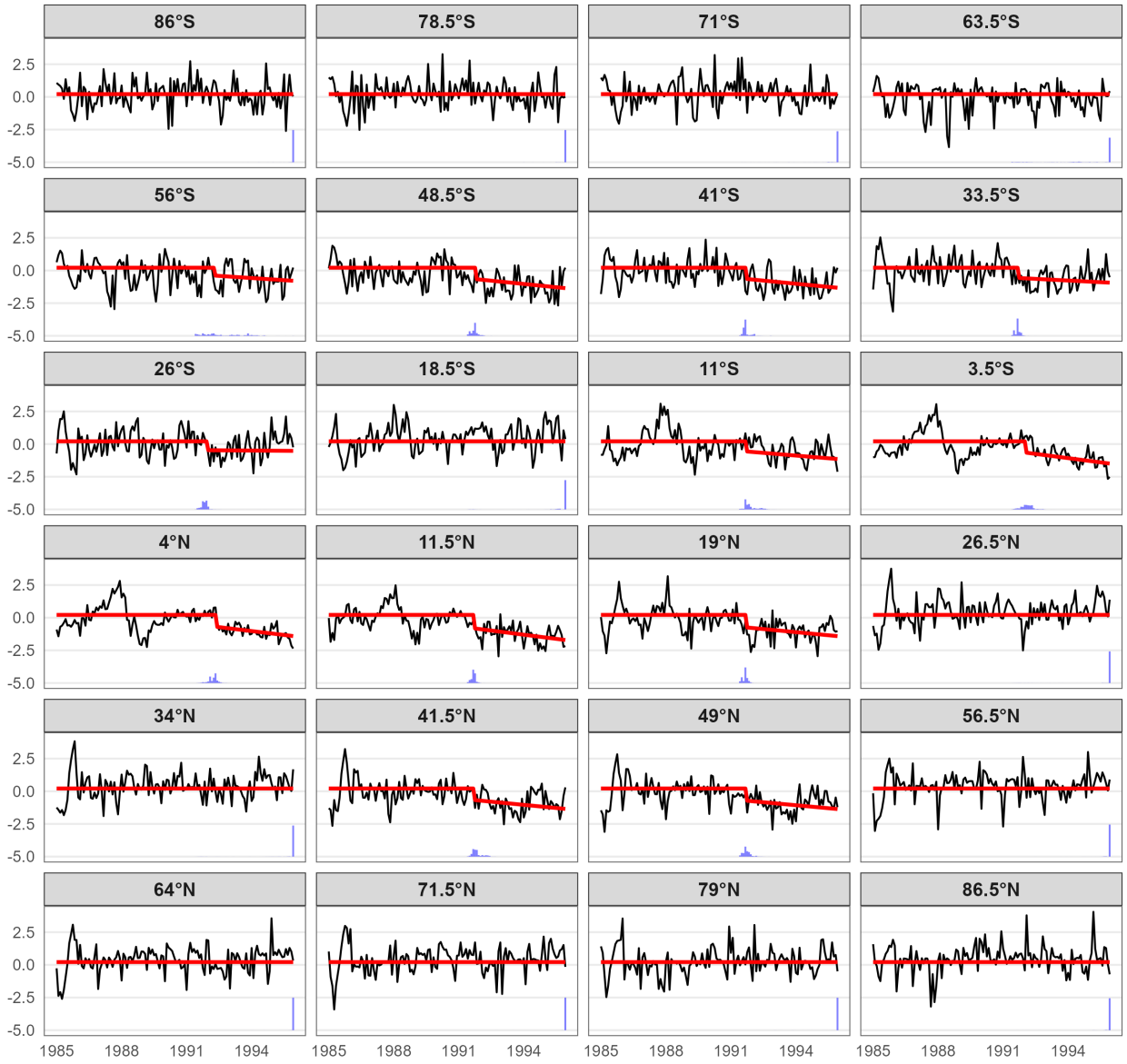


Figure 5-13. Estimated process mean (red llines) before and after a detected changepoint for surface temperature separated by latitude. The black lines are latitudinal mean temperature series after preprocessing. The posterior distribution of the changepoints are overlayed at the bottom of each plot.

5.5. Functional Change Point Detection

Contributing Authors: J. Derek Tucker (5573), Drew Yarger (5573), Carole Hall (5573)

Changepoint detection is a vital tool in the application of climate data analysis. Numerous types of climate observation data are most properly represented by functional timeseries, implying a need for accurate changepoint detection methods applicable to functional timeseries data. Such data taken at a global scale often contain both spatial heterogeneity and dependence as well as phase (time) misalignment. In this report, we present methods which can detect spatially-dependent changepoints while allowing different estimates of change time and change strength depending on location. Additionally, we provide extensions to this spatially-predicted model which controls for phase variability among observations. Our methods provide the ability to detect a single change, or control for epidemic changes (where a "return-to-normal" change is more likely to be detected than the initial change). We showcase results analyzing the June 1991 eruption of Mt. Pinatubo, where our methods demonstrate the ability to accurately detect both single and epidemic changepoints even in the presence of strong seasonal variability. We find that our spatially-predicted model improves the detection of relevant changepoints versus methods which do not take spatial information into account, and we find that controlling for phase variability helps to control the false discovery rate during the detection process.

We present a collection of methods for the purpose of detecting changepoints in global-level climate data. These methods have the ability to detect changepoints in circumstances where the impacts from a large, local perturbation (such as a volcanic eruption) can lead to heterogeneous effects across the globe in terms of impact time and intensity. In this report, we detail two novel methods—one which detects changepoints taking into account spatial dependencies and heterogeneity, and another which extends this approach to also handle phase (time) variability in data (a.k.a. elastic extension model)—and present resulting detected changes focusing on the June, 1991 eruption of Mt. Pinatubo in the Philippines. our methods can detect spatially-heterogeneous changepoints—that is, changes that may occur at different times across locations when caused by the same event, and may have different extents of change. We have demonstrated that the combination of these features in a changepoint detection procedure can help us detect changes due to localized events on global scale, which can be a challenging problem due to the complexity of climate processes. Methods were applied to datasets of various complexity following the tiered verification process: real climate observations for a time period around the eruption (from the MERRA-2 climate reanalysis), simplified climate model output (using the Held- Suarez-Williamson volcano model or HSW-V), as well as simulated, spatially-dependent data.

This method separates itself from the changepoint method discussed in 5.4 in that the data is treated as functional data, i.e. each year as a function of daily values. Thus, the detected changepoint is at a coarser timescale - detected annually - but the phase variability, which is ignored by prior change point methods, in addition to the amplitude variability is accounted for. As an example, we show stratospheric temperature data from the MERRA-2 climate reanalysis at a specific location, represented as a functional time series in Figure 5-14.

5.5.1. Results

When applying our methods to the simulation results, we found that spatially predicting using score-based test statistics increases our changepoint detection ability, and multiple testing corrections helped to control the FDR. The use of four or five principal components during the changepoint

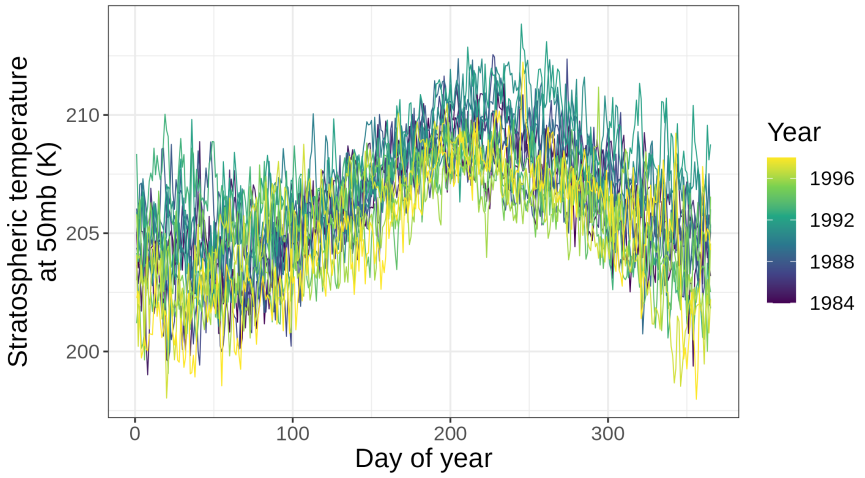


Figure 5-14. MERRA-2 climate reanalysis data presented as a functional time series at 115°E and 10°N.

estimation process in the simulation study yielded better results than using three principal components in the case where spatial dependence is introduced in the simulation model.

The HSW-V study showed that the use of spatially-predicted test statistics increased power of detection versus when we computed test statistics at locations independently. We also found that the epidemic version of our spatially-predicted model was highly useful in ensuring that our method was not inadvertently detecting "return-to-normal" changes while neglecting the source event's changes.

In the MERRA-2 reanalysis study, we have to handle variability stemming from natural seasonality, as well as more complex, long-term trends. Figure 5-15 shows the p-values and detected start and end points for each changepoint detected applied to MERRA-2 stratospheric temperatures. Darker colors indicate a significant epidemic changepoint is detected. We can see an average increase in stratospheric temperature around $0.25 - 1.6$ K after 1989, and then after 1991, we see an average decrease in stratospheric temperature, implying that a "return to normal" change is detected in this case. To see more results and discussion related to the epidemic changepoint model, see [204]. The elastic extension to the spatially-predicted changepoint model provided results consistent with those of the original spatially-predicted model when applied to the MERRA-2 reanalysis data. One difference between these two models we emphasize, however, is the tendency for p-values calculated using our elastic pipeline to be consistently higher than those calculated using the original model. The original model then detected more changepoints than did the elastic extension of the model. Since the use of elastic metrics in functional data analysis helps prevent over-estimation of functional distances due to nuisance phase variation [182], we suggest that the use of our elastic extension can help control the FDR in the changepoint detection process on global climate data which may contain phase variation.

Detailed method descriptions, results of all versions of this model, as well as an evaluation of different testing metric considerations are summarized in [61].

Implementation of the elastic functional changepoint detection has been added to the `fdasrvf`

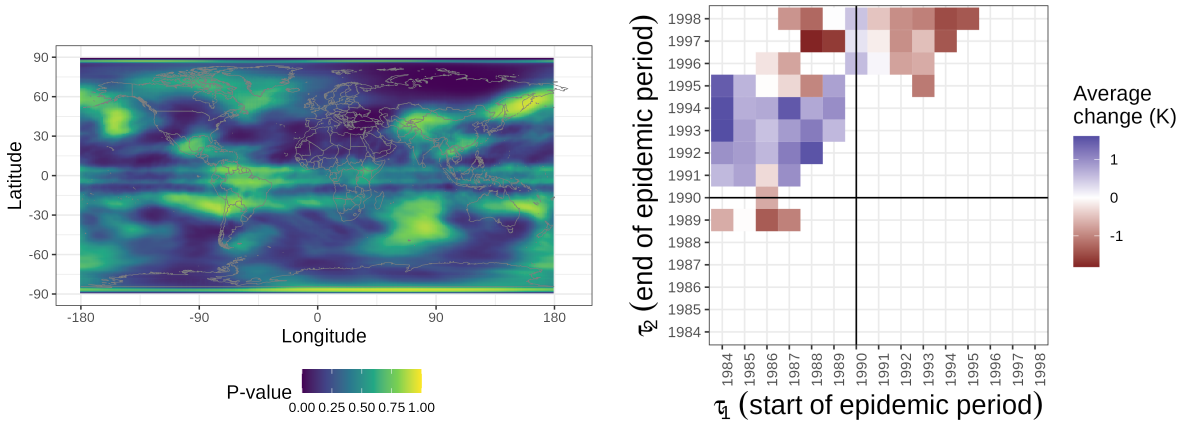


Figure 5-15. (Left) Unadjusted epidemic p-values from the original model. (Right) Mean average change estimate for the original epidemic changepoint model for all locations by year of detected change.

package available in both R [181] and Python [180].

Further details on the research can be found in [182, 204, 61]:

- Tucker, J. D., & Yarger, D. (2023). Elastic functional changepoint detection of climate impacts from localized sources. Environmetrics, e2826. `https://doi.org/10.1002/env.2826`
- Drew Yarger and J. Derek Tucker. "Detecting changepoints in globally-indexed functional time series." Submitted to *Evironmetrics*, August 2024. This manuscript is available on arXiv: `https://arxiv.org/abs/2308.05915`.
- Carole Hall, J. Derek Tucker, and Drew Yarger. "Elastic Changepoint Detection for Globally-indexed Functional Time Series Data with Climate Applications." SAND2024-12470. September 2024, Sandia National Laboratories, Albuquerque NM.

5.6. Data Fusion

Contributing Authors: Audrey McCombs (5574), Justin Li (8732), J. Gabriel Huerta (5573), Lyndsay Shand (5573), Jimmy Hickey (5573), Mauricio Campos (5574)

5.6.1. A Tensor Approach

One increasingly common approach to dealing with the increasing availability and volume of remote-sensing datasets is data fusion, which combines disparate observations together to produce a single higher quality and more complete data product. The many available techniques apply across many different fields, varying based on the modality of the data and the underlying model used to combine the data. For Earth Science, techniques often process images taken using different imaging instrumentation or handle observations for a single quantity of interest. While the emphasis on these multispectral and hyperspectral approaches does contribute to improved quality and coverage for

imaging observations, there is a far larger set of observational data with the different spatio-temporal coverage to be explored.

Here, we instead present a tensor-based method that exploits the correlation across different quantities of interest, with the assumption that the variables related to one another via an underlying common spatial pattern. While tensors have been used for data completion and fusion in other domains, we have not seen significant applications for Earth Science. Notably, we show that the tensor structure suits the variables representing quite disparate physical phenomenon by first using MERRA-2 reanalysis data to have a ground truth for evaluating the fitness of this approach. Then, we extend the implementation to study its utility with changing regimes of data availability which occur frequently in the Earth Science domain which leads to illustrating its use with real observational data from two different satellite instruments.

We also thank Danny Dunlavy (1462) for his input and guidance on the tensor implementation. He is credited as a co-author on our paper which is in preparation [95]:

- Justin D. Li, Audrey McCombs, Daniel Dunlavy, Gabriel Huerta, Lyndsay Shand, "Exploring Multivariate Tensor Data Fusion and Completion." To be submitted to *Special issue of Remote Sensing: Spatiotemporal Fusion of Multi-Source Remote Sensing Data and Its Applications.* October 2024.

5.6.1.1. Results

For a tensor-based method to succeed, the data must share common underlying spatio-temporal structure. We can construct a 3D tensor, with the different variables, time, and the collapsed latitudes and longitudes as space as the dimensions. However, do the different quantities of interest, aerosol optical depth (AOD, equivalent to AOT for thickness, depending on the dataset), radiative flux, and temperature, which measure entirely different phenomenon, exhibit sufficiently coherent patterns of behavior? Using MERRA-2 data, we first confirm that our assumptions are satisfied for this tensor-based method with successful fusion across the different data variables to reconstruct with reasonable fidelity the original reanalysis data.

Moving to real observational data, although these are more processed data products (level 3 and level 4) and we apply downscaling for computational tractability, we use:

- AOD from the Advanced Very High Resolution Radiometer (AVHRR). As a NOAA Climate Data Record (CDR), AVHRR AOD benefits from further post-processing and combines data across multiple instruments.
- Radiative flux, in the form of outgoing longwave radiation (OLR), collected by the TIROS Operational Vertical Sounder (TOVS) aboard NOAA-11. The data includes both morning and evening passes, which we handle as two different variables rather than try to collapse into a single daily value.
- Surface temperature (TREF) collected in the same way as OLR by NOAA-11 TOVS.
- Stratospheric temperature at 50 hPa (T50) collected in the same way as OLR by NOAA-11 TOVS.

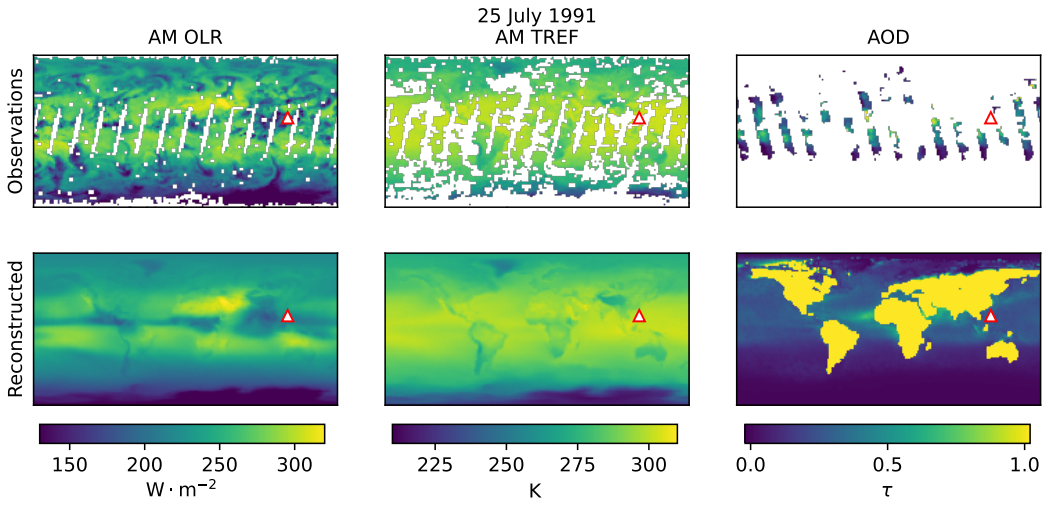


Figure 5-16. Results from the tensor data fusion approach using real observed data, for a single day (July 25, 1991) with the observations shown on the top row and the fused product on the bottom row. Each column shows one of the variables, highlighting the morning pass TOVS OLR (left), the morning pass TOVS TREF (middle), and the daily observation for AVHRR AOD (right). The red triangle marks the location of Mt. Pinatubo.

Applying our tensor-based method, we produce spatially and temporally complete products covering for all variables. Figure 5-16 show the observed data in the top row for three of the variables (the morning / AM pass for OLR and TREF, and AOD), where the occurrence of incomplete observations is especially prominent for AOD (right). The bottom row shows the output from our approach, where the fused product has been reconstructed with complete coverage across the spatial and temporal dimensions to impute values to the missing regions. The modeled values show good overall agreement with the locations where observations occurred, although with a higher degree of spatial smoothness. While AOD shows the same behavior, it presents a more challenging component as AVHRR does not include any measurements over land, due to limitations in the sensor and the algorithm. As such, the values in the fused product here show non-meaningful, saturated values.

Because the tensor method models the underlying behavior through the decomposed tensor components, each component can be individually examined as well. Figure 5-17a shows three components, with different total contributions to the data fusion product, along the three different dimensions. The different components exhibit different degrees of correlation between variables, with differing temporal patterns including varying amounts of seasonal effects, and markedly different spatial patterns. The components themselves could be further examined, to assess how well the phenomenon of interest might be specifically represented, or matching them to modes and oscillations.

Another way to evaluate the results, though less absolutely representative than the same assessment conducted over the complete MERRA-2, is to examine the errors against the given observations to see if the data fusion is overall faithful to the input data. Figure 5-17b shows the median absolute error (blue) across space for each time point, for the seven total variables. The values show largely similar and consistent behavior over time, with some potential seasonal effects. For example, AVHRR AOD uncertainties are given as $+/-0.2$ which captures the range in errors here. Because

the two instruments have different coverage, we also show the number of points (green) used for the error calculation.

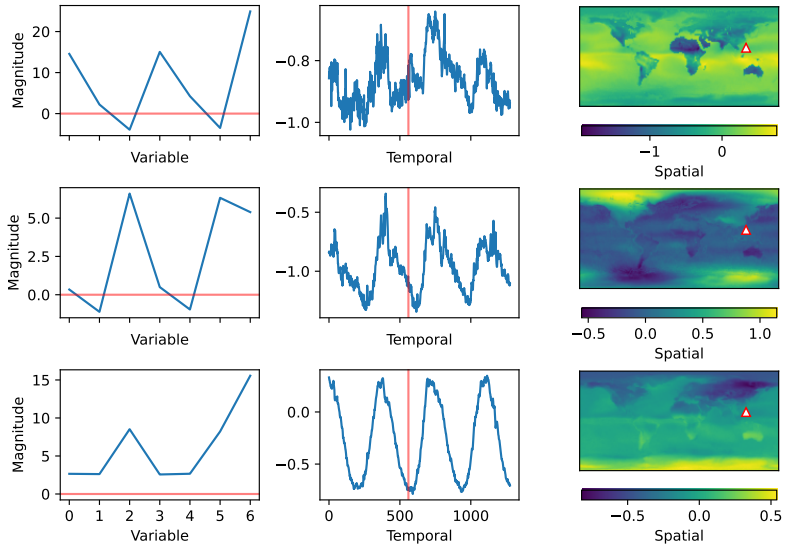
Because the AOD results show such poor results in filling in values over land, with nearly complete saturation, we conduct one further test using different sampling configurations with MERRA-2 data to highlight the impact of data availability in formulating the task of data fusion using the tensor-based method. Because AVHRR never has observations over land, due to the sensor and the retrieval algorithm, there are no points available for correlating between variables when learning the tensor components. If we hypothesize additional datasets generated by new algorithmic data processing capabilities capable of retrieving measurements over land (such as Deep Blue, though with more completeness), the tensor method could then infer the correlations between variables and the trends over time from more data-dense time periods to better predict values from earlier time periods without any data.

Figure 5-18 shows that this is feasible with a simulated scenario with MERRA-2 data (1st row), where no values are collected over land before May 1993 and land values are retrieved after May 1993 (2nd row). If no land data was ever collected, the data fusion reconstruction (3rd row) shows obvious errors over land. (The behavior differs from with real observed data, likely due to the better overall coverage with the synthetic MERRA-2 sampling.) In contrast, this partial land sampling reconstruction using the values observed after May 1993 (4th row) show marked improvement. Under normal conditions, the reconstructed AOD matches the original well (right column) even in the earlier time period without data (left column). Only when there is no data and a larger disturbance is occurring does the reconstruction break down to some extent, such as around the Pinatubo eruption (middle column).

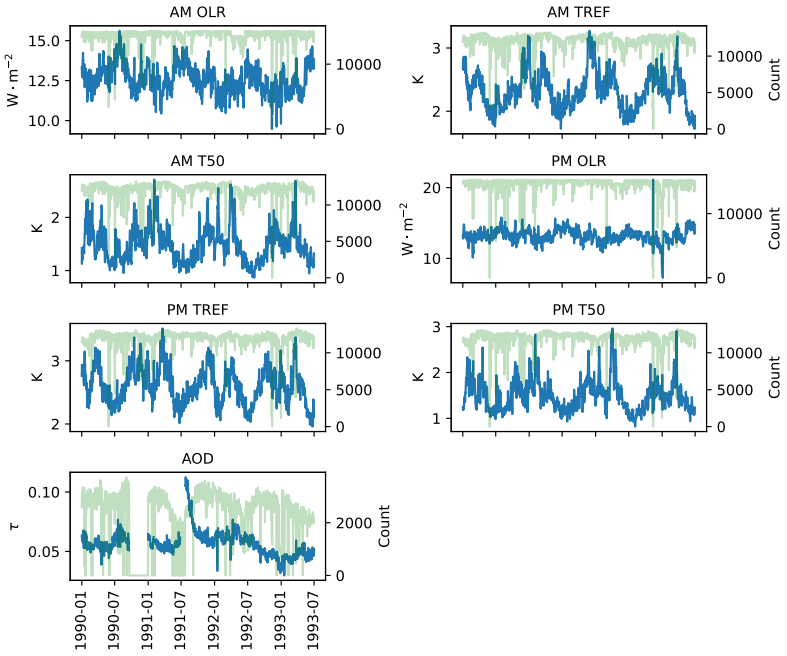
5.6.1.2. Relative Strengths and Limitation Assessment

The tensor-based data fusion method is capable of combining different observations from different instruments for different, but related, physical phenomenon. This leverages correlations between variables to improve the data in-filling due to the different instruments' coverage and capabilities. Relying purely on the input data, without a physics-based model, the tensor fusion produces complete estimates across the defined times and spaces, with a demonstration of its application on real observed data. The proof-of-concept also shows that this method will be useful across regimes of different data and for intelligently in-filling times when little to no data is available.

At this time, the work has several limitations, which could be addressed with future work. The time for computing the tensor decomposition will scale with larger dimensions and more coverage. To avoid excessive computation times, the current work downscaled the data. Higher spatial and temporal resolutions and longer durations will require longer time to complete. More fundamentally, this approach does not currently handle different resolutions, nor does it address conflicting values at the same location from two different instruments. It is a data fusion technique across variables rather than between instruments for a single variable.



(a) Three example tensor components (rows) showing the correlation across variables (left), the trend in time (middle), and the spatial pattern (right). In the left column, the horizontal red line indicates the zero line, where the variables are unrelated for that component. The vertical line in the middle column denotes July 15, 1991, for the month of the Mt. Pinatubo eruption, with the index marking the number of days since January 1, 1990. The red diamond in the right column marks the location of Mt. Pinatubo.



(b) Time-series plots showing the median absolute error (blue) over space at each moment in time for each variable. The number of spatial points available and used for that calculation is plotted with a twined x-axis (faint green) for reference.

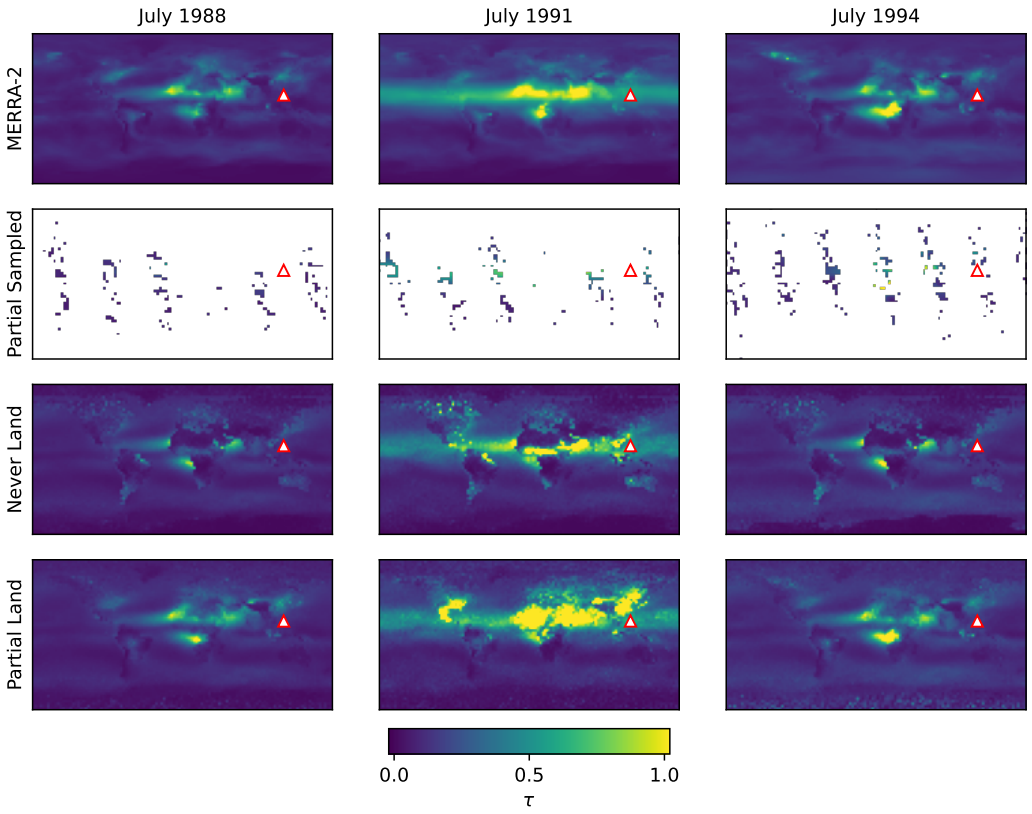


Figure 5-18. Demonstrating a constructed scenario using MERRA-2 data which combines time periods with different data availability. The (1st row) original data is (2nd row) sampled with observations being retrieved over land after May 1993. Data fusion reconstruction results when (3rd row) no land is ever retrieved is shown for comparison with the results using this (4th row) partial sampling. All plots are set to the same color scale, with the red triangle marking the location of Mt. Pinatubo. The three columns show the data and results for three different times, three years (left) before and (right) after and (middle) at the time of eruption.

5.6.1.3. Follow-on work

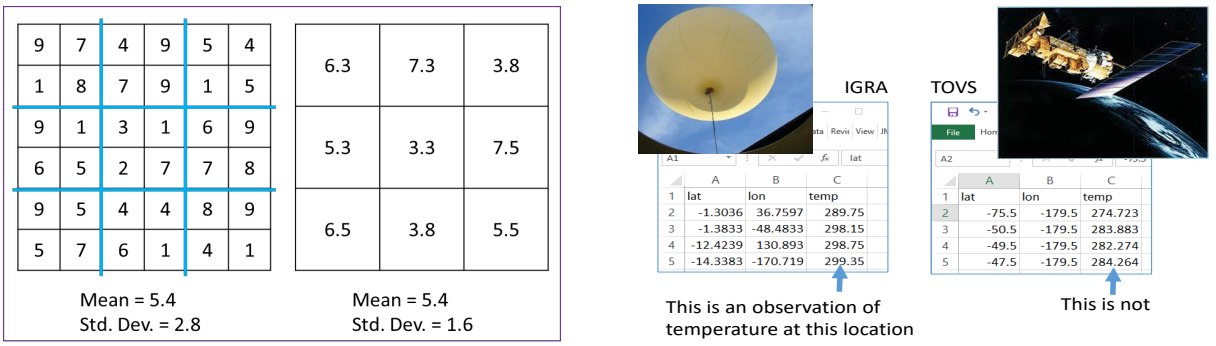
There are several directions for immediate and more extended follow-on work: (1) Contribute the developed tensor code to Sandia's Python tensor toolbox (pyttb), (2) Assess the scalability of the approach with higher resolution data and longer time durations, and (3) Develop extensions to handle multiple sampling resolutions and multiple inputs for the same variable. We note that this method should also generalize to other Earth Science and general applications and would consider demonstration on other exemplars to be valuable for supporting additional data processing tasks for other mission needs.

5.6.2. A Statistical Approach

The fusing of spatial datasets is becoming more common as the availability of remote-sensing datasets grows. Fusing two or more spatial datasets has the potential to produce more accurate estimates of natural processes than modeling each dataset separately. Data fusion is currently practiced in a variety of fields, including climate science, ecology, oceanography, and soils science. For the CLDERA project, fusing temperature datasets produced a spatio-temporally complete observational dataset with uncertainty quantification to support near-real-time data analysis and pathway attribution. While fusing different datasets has advantages, care must be taken when fusing data collected over different spatial scales, since datasets with different spatial supports have different statistical properties. Using areal and point data as an example, areal data (such as remote sensing satellite data) has an added element of uncertainty compared to point data (such as field observations) due to the fact that the values reported are aggregates of information collected over an area, rather than actual measurements at a point in space. Ignoring this added uncertainty results in too-liberal estimates of prediction errors.

5.6.2.1. The change of support problem

Appropriately accounting for different spatial supports in spatial statistical modeling is known as the change of support problem (COSP). The classic reference is [59] who survey the statistical issues associated with inferring a spatial process at one resolution from data at a different resolution. Since that seminal paper, a rich literature has developed introducing different approaches and modeling techniques for addressing the COSP. The extensive literature and highly technical nature of the different methodologies, however, can lead to confusion and frustration in practitioners whose goal is to fuse different datasets in a way that is sensitive to the statistical issues, but who cannot afford to invest in a thorough understanding of the entire field. In this paper we focus primarily on two different methodologies that are implemented in computational tools written for the open-source statistical programming language R [142]. Specifically, we describe and implement COSP models using the R packages `FRK` [205] and `R-INLA` [97]. This study provides practitioners with an understanding of the statistical issues associated with the COSP, and a detailed primer on two tools available for appropriately quantifying uncertainty in spatial models from fused datasets. Our goal is to provide an accessible entrée into the topic that will improve uncertainty quantification from fused datasets across a wide range of scientific fields.



(a) The variability in a population of individuals is higher than the variability of means calculated from subgroups of the population.

(b) Computational tools will happily treat the specified longitude and latitude of an areal datum as if it was a point observation, even though the reported location is an arbitrary point inside an area.

Figure 5-19. Illustration of change of support problem (COSP).

The term "change of support problem" actually refers to several different problems that crop up when we infer a spatial process at one resolution from data at a different resolution. Examples of problems associated with change of support include the modifiable areal unit problem and the ecological fallacy [59]. In this study, we specifically focus on one form of the COSP: the added uncertainty that must be accounted for when combining an areal dataset with a point dataset. The problem is illustrated in Figure 5-19a. On a spatial field made up of 36 areas with the values specified in each area, the mean over the spatial field is 5.4 and the standard deviation of values is 2.8. If we aggregate the field into nine groups of four areas each by taking the mean of the four areas in each group, the mean over the spatial field does not change. However, the standard deviation of the values across the field is reduced, from 2.8 to 1.6. This is a well-known phenomenon in statistics: the variability in a population of individuals is higher than the variability of means calculated from subgroups of the population.

The error associated with a point dataset, in which values represent a measurement at a specific place, is known as measurement error. For our purposes, measurement error captures the error associated with recording a measurement, and can be caused by operator variability, imprecision and/or bias in the measurement instrument, environmental effects, etc. An areal dataset also contains measurement error, but the values reported in an areal dataset include an additional source of uncertainty not present in the point dataset; specifically, the error associated with the aggregation of information over the area. For example, if a satellite dataset reports a temperature value of 310 K over the area between 46 and 45 degrees west and 22 and 23 degrees north, any specific point in that 1 degree-by-1 degree area may or may be 310 K. The difference between the actual (unknown) temperature at a specific location (e.g., exactly 45.2 degrees longitude and 22.4 degrees latitude) and the temperature reported in the dataset is due to the fact that the reported value is an aggregate of all the information gathered over the 1 degree-by-1 degree area. It is this added error due to aggregation that we refer to as the COSP in this study.

It is tempting (and easy) to ignore the extra uncertainty due to aggregation when combining point with areal datasets. In the TOVS dataset we use in this study, temperature values are collected over

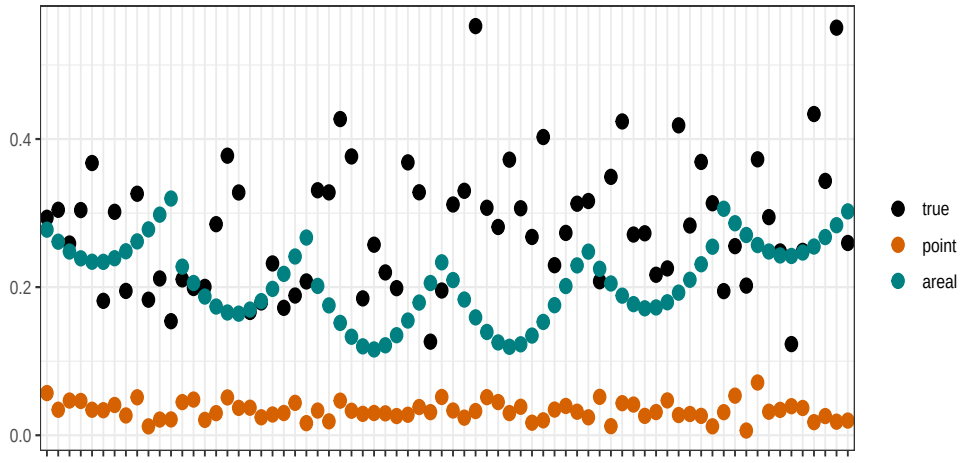


Figure 5-20. True standard deviations (black) and prediction standard errors from a model run as point data (orange) and a model run as areal data (teal).

1 degree-by-1 degree areas and reported at a specific longitudes and latitudes, namely, the centroids of the areas. The choice of centroid, however, is arbitrary—any point in the area would serve to identify the area on the globe as long as the choice is documented for practitioners. However, computational tools will happily accept the specified longitude and latitude and treat the value as if it was a point observation (Figure 5-19b). This is a mistake, for the reasons discussed above.

Figure 5-20 demonstrates the modeling result when areal data is modeled incorrectly as point data, and when it is modeled correctly as areal data. This figure was produced using the LatticeKrig package in R [129], in which a simulated areal dataset was modeled using the standard package functions, (i.e., the areal data was modeled as point data), then modeled accounting for the areal nature of the data using custom functions that numerically integrate over an area. The simulated dataset was created by coarsening a high-resolution spatial field of 20,000 points into 72 areas, by averaging the point values in each of the 72 areas. The standard deviation of the true simulated values (the variability associated with each area) was compared to the prediction standard error for each model. The prediction uncertainty reported from the point model is significantly lower than the prediction uncertainty from the areal model, and while the uncertainty in the areal model is still smaller than the actual variability in the data, it is closer to the true variability relative to the uncertainties from the point data model.

5.6.2.2. Modeling tools

Figure 5-20 was produced using custom functions that will require development before they can be released for general use. However, two published R packages account for the COSP natively: `FRK` and `R-INLA`. The `FRK` package implements a fixed rank kriging modeling framework as a spatial random effects model, in which a spatially-correlated mean-zero random process is decomposed using a linear combination of spatial basis functions with random weights (linear coefficients) plus a mean-zero error term that captures small-scale variation. The random process is modeled using a spatial covariance function that is always nonnegative definite and can be constructed so as

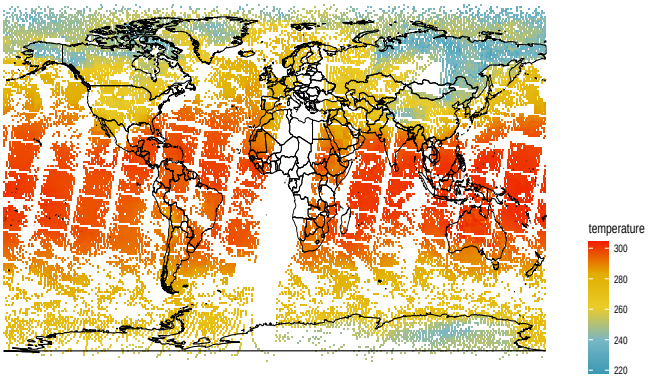
to approximate standard families of spatial covariance functions. The conceptual machinery that allows the model to account for change of support is the basic areal unit (BAU), a discretization, or "tiling," of the spatial domain into a large number of small, non-overlapping areas. The model is implemented on the tiled surface and predictions are made at the level of the BAU. Each areal datum includes at least one (and usually more) BAU, while each point datum falls into a single BAU. Model predictions, with reported uncertainty, are made at the BAU level over the entire spatial domain. A full mathematical description of the model and details of model estimation are included in [206]

The second package that natively accounts for change of support is `R-INLA`, the R implementation of the integrated nested Laplace approximation (INLA) approach. INLA is a method for approximate Bayesian inference that can be applicable for latent Gaussian Markov random fields (GMRFs). The model involves a three-stage Bayesian hierarchical structure in which the observables $\boldsymbol{y}$ are described as conditionally independent given some latent parameters η which define a latent GMRF with a Matérn covariance matrix and hyperparameters θ. Prior distributions on θ specify the final stage of the hierarchical model. A mathematical description of the INLA model and computational details of estimation procedures are in [149]. Computationally, the INLA model constructs a triangulation over the spatial domain, and a projection matrix that maps the observation locations to the triangulation nodes. It is the structure of the projection matrix that accounts for different spatial supports from different datasets. Each row in the projection matrix represents an observation, while each column represents a triangulation node. For a point observation, at most 3 columns in that row contain non-zero entries, associated with the nodes surrounding the observation point and weighted by distance from the node to the point. For areal data, all triangulation nodes that fall within the area have non-zero entries and values are a function of the number of nodes in the area. The model is implemented at the triangulation nodes, but predictions (with uncertainty) can be made at any arbitrary location in the spatial domain.

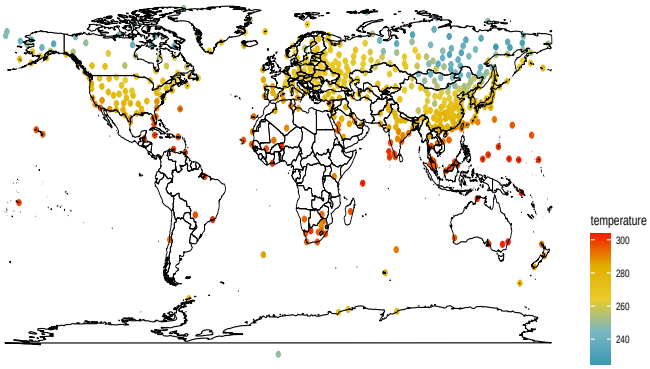
In this paper, we compare the performance of these two packages using three global temperature datasets: 1) Observational data (Figures 5-21a and 5-21b): an observational dataset from January 1, 1991 that combines point-level radiosonde data (instrument carried on a weather balloon) from the Integrated Global Radiosonde Archive (IGRA), with areal data at a 1 degree latitude-by-1 degree longitude scale from the TIROS Operational Vertical Sounder (TOVS) on NOAA weather satellites; 2) Simulated observational data (Figure 5-21c): A simulated dataset that mimics the sample size and spatial structure of the observational data; and 3) Simulated areal data (Figures 5-22b and 5-22c): A simulated areal-only dataset that was generated by first simulating mock temperature values on a high-resolution grid (200 x 100, Figure 5-22a), then coarsening the simulated values into 72 areas (12 x 6) by averaging over all values in each area. It is this third dataset that was used in the LatticeKrig model described above in Section 5.6.2.1 to illustrate the change of support problem.

5.6.2.3. Results

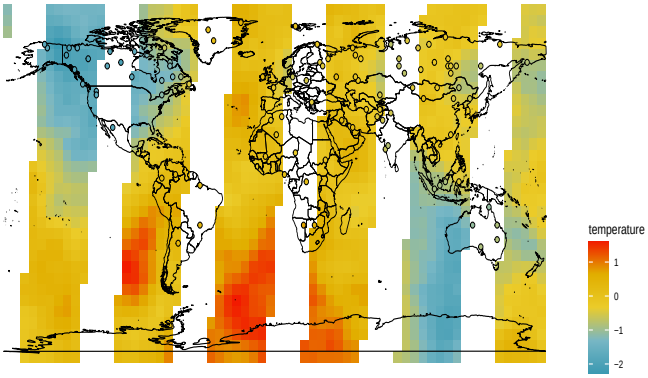
As expected, both models did a good job predicting temperatures from the observational and simulated observational data (see Figure 5-23a for FRK results on observational data and Figure 5-24a for INLA results on observational data). For the simulated observational data, we ran the FRK



(a) TOVS areal observations of surface temperature for January 1, 1991.

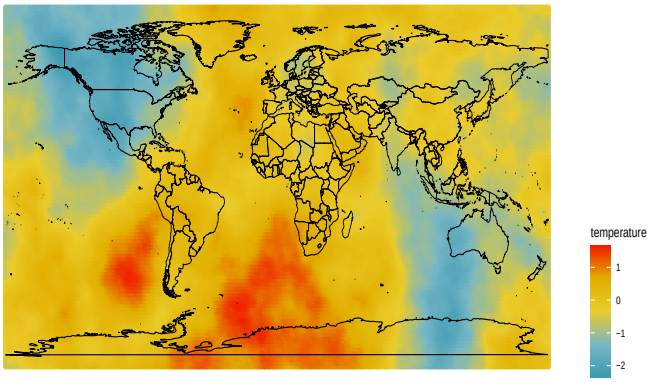


(b) IGRA point observations of surface temperature for January 1, 1991.

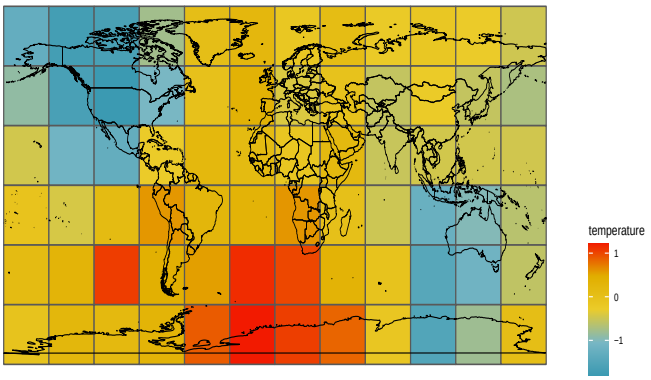


(c) Simulated observational data that mimics the sample size and spatial structure of the observational data.

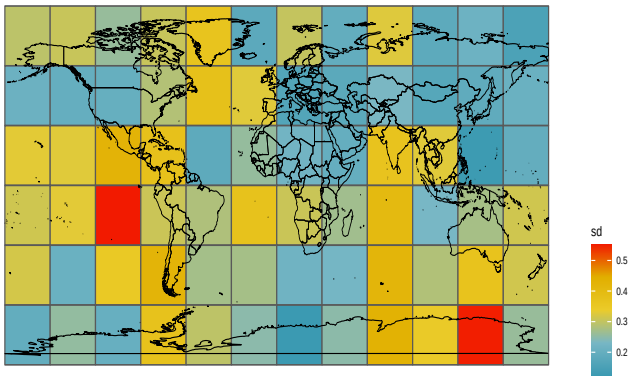
Figure 5-21. Global temperature datasets used in performance comparisons (observational A & B and simulated C).



(a) High-resolution simulated data on a 200 x 100 grid.

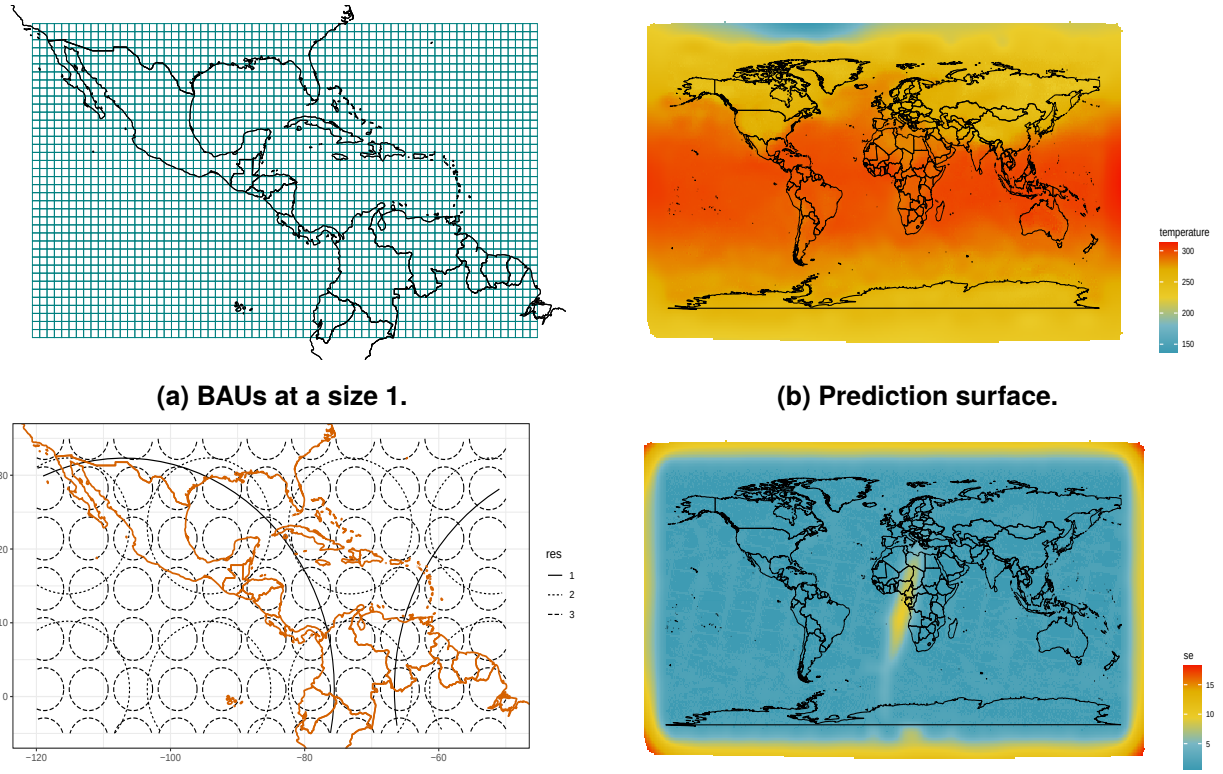


(b) The mean temperature value for 72 areas (12 x 6) when the data has been coarsened by averaging the high-resolution simulated values in each area.



(c) The standard deviation (i.e., true variability) of the simulated values in each of the 72 areas.

Figure 5-22. Simulated areal data example.



(a) BAUs at a size 1.

(b) Prediction surface.

(c) Three resolutions of basis functions.

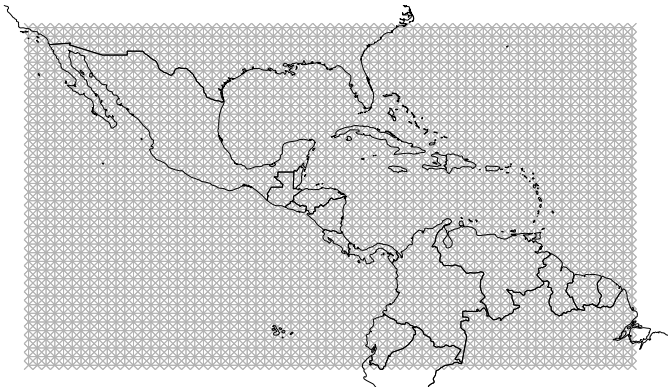
(d) Prediction standard errors.

Figure 5-23. FRK model on observational data.

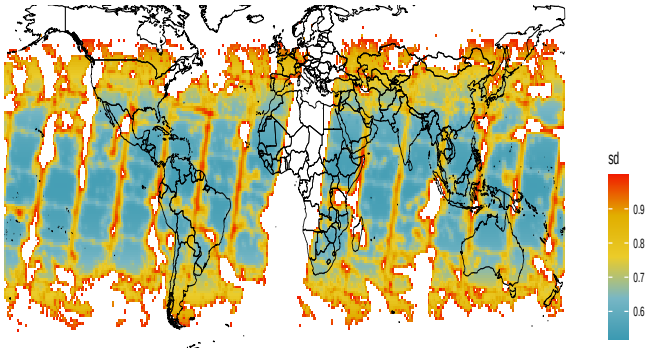
model with BAUs of size 1 and a resolution of 3 for the basis functions. Otherwise we used the default settings for the model with no covariates except latitude and longitude. The point dataset ($n = 102$) was fed into the model as a SpatialPointsDataFrame, while the areal data ($n = 1286$) was a SpatialPolygonsDataFrame, both created with the R package `sp` [9], for a total of 1388 data values. We compared the predictions at the BAU level with the true (simulated) areal and point data, where the true values ranged from -2.3 to 1.6. The mean prediction standard error was 0.0563, nominal 95% coverage was 98.2%, and the RMSE was 0.0433 (see Table 5-2).

We ran the INLA model with a projection matrix that captured the difference between the areal and point simulated observations, using a mesh size of 20. Predictions were made on a 200 x 100 grid, and each observed value was compared to the prediction at the point closest to the observed location (the centroid of an area for the areal observations). The mean posterior standard deviation was 0.0252, about half that of the FRK model. Nominal 95% coverage was 84.9%, much less than both nominal and the FRK model, and the RMSE was 0.0347, very slightly smaller than the FRK model. Table 5-2 compares model results for the simulated observational data.

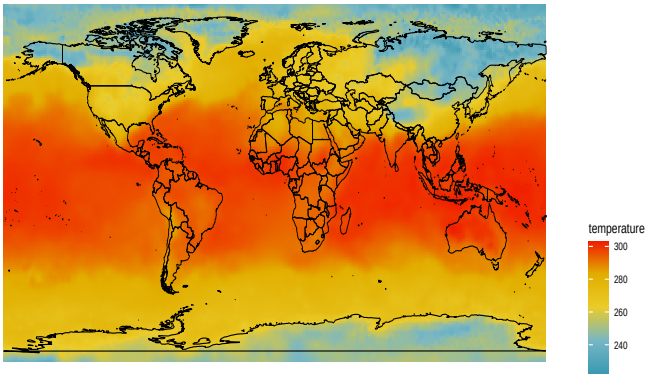
We used the simulated areal dataset to investigate how well each model handles the change of support problem. We ran the simulated areal data through each model, first as point data and then as areal data, and asked for one prediction for each of the 72 areas. We compared the prediction standard errors (for FRK) and the posterior standard deviations (for INLA) from the areal model to those from the point model, and expected the prediction uncertainty to be smaller for the point



(a) INLA mesh of size 1 (grey), TOVS areal data (blue), and IGRA point data (brown).



(b) Posterior standard deviations for predictions, for those less than 1. Areas in white had standard deviations greater than one.



(c) Prediction surface.

Figure 5-24. INLA model for observational data.

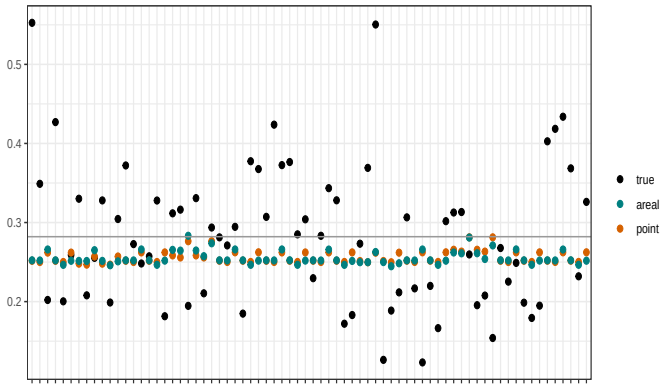
Table 5-2. Comparison of model fit statistics for simulated observation data. Error column is mean prediction error (FRK) or posterior standard deviation (INLA). Nominal coverage is 95%

Model	BAU (FRK) or mesh size (INLA)	Error	Coverage	RMSE
FRK	1	0.0563	98.2%	0.0433
INLA	20	0.0252	84.9%	0.0347

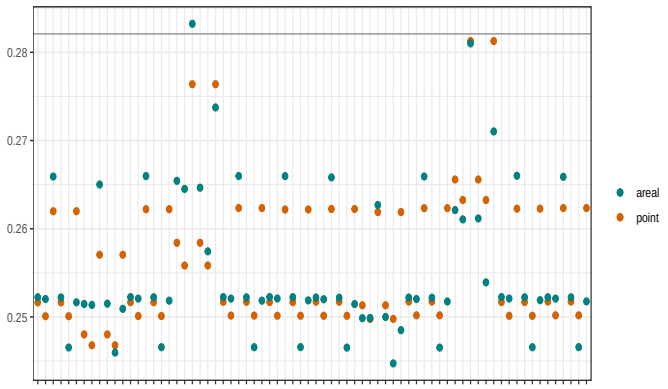
model than the areal. Results are displayed in Figure 5-25a. The INLA model behaved as expected: the posterior standard deviations for the areal model were substantially larger than those for the point model, and while the posterior standard deviations were generally smaller than the true standard deviations, there were within 2σ of the mean of the true standard deviations. These results suggest that the INLA model is appropriately accounting for the added uncertainty in areal data.

The FRK results were less promising. In general, the prediction standard errors from the point model were smaller than those from the areal model, however the differences were small, especially compared to the variability in the true standard deviations. These results seem to indicate that the FRK model is not appropriately accounting for the difference between areal and point data. We suspect the reason might be a mis-match between the model conceptualization and the computational procedures. The BAUs are areas as conceptualized in the model, however they are treated as points in computations. In the `FRK` vignette for R, [205] state that the spatial field $\mathbf{S}$ can be approximated with the set of points $\mathbf{s}$, namely the centroids of the BAUs. "Since small BAUs are always assumed, this approximation is used throughout `FRK`" (p. 4). The INLA framework explicitly incorporates a mechanism to account for the difference between point and areal data, while the FRK model does not. Rather, the FRK model relies on the fact that a point datum will fall into a single BAU, while an areal datum will spatially cover more than one BAU. This may be the reason why there is very little difference between the prediction standard errors in the point model versus the areal model.

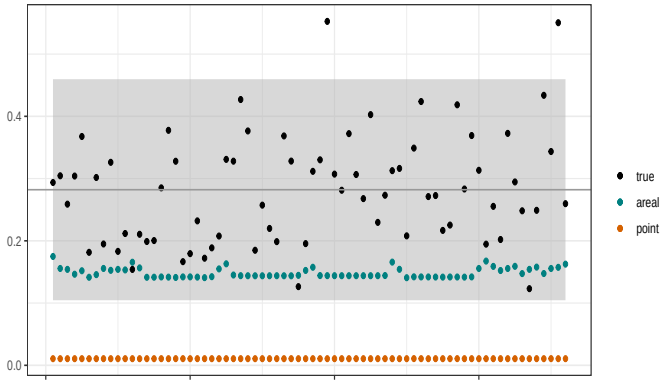
We used the simulated areal data for a second investigation into how the models handle change of support. Because both models allow predictions at arbitrary locations, we predicted on a very fine grid then aggregated the predictions within each of the original 72 areas, and compared the mean and standard deviation of the predictions within each area with the original data values. We did not account for the uncertainty in the predictions. We ran the models as areal data, and results are displayed in Figure 5-26a. The standard deviations of the aggregated predictions more closely matched the true standard deviations than the reported prediction uncertainties (discussed above). This is perhaps not surprising given that this procedure of aggregation is similar to the procedure we used to generate the simulated data. But as the Figures 5-26b and 5-26d show, the aggregated standard deviations were generally within 2σ of the mean of the true standard deviations. As above, the variability in the INLA predictions more closely matched the variability in the true data compared to the variability in the predictions from the FRK model. Qualitatively, the spatial pattern in the model variability generally matched the spatial pattern in the true data (Figures 5-26a and 5-26c), where areas in the upper right had less variability, the area at $(3, 3)$ had high variability, and the two areas that cover Australia had moderate variability. Again, results from the INLA model matched the true values better than results from the FRK model.



(a) True standard deviations (black, grey horizontal line: mean), standard error of predictions from FRK areal model (teal), standard error of predictions from FRK point model (brown).



(b) Same as above except zoomed-in on model results only.



(c) True standard deviations (black; black horizontal line: mean; grey shaded area: ± 2 standard deviations from the mean), standard error of predictions from INLA areal model (teal), standard error of predictions from INLA point model (brown).

Figure 5-25. True standard deviations versus prediction standard errors ($n = 72$) from models on simulated areal data.

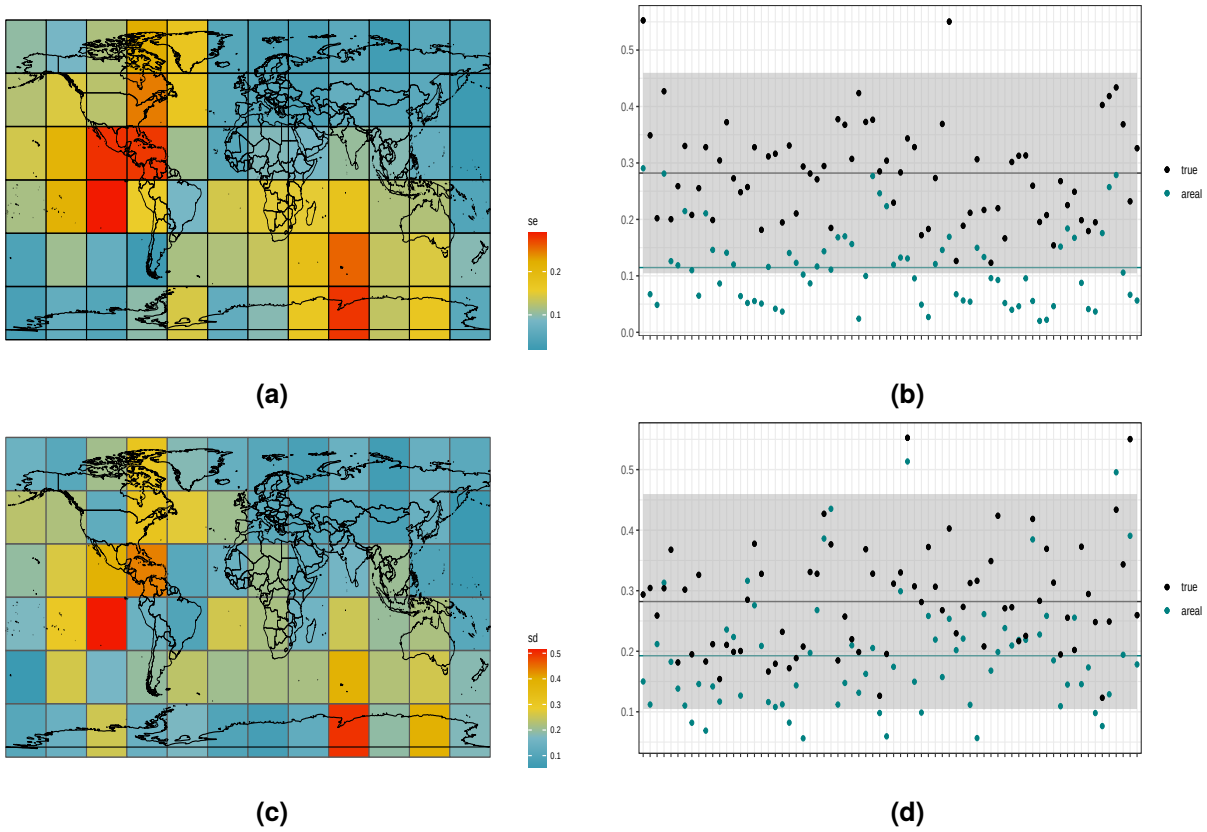


Figure 5-26. True variability in simulated data versus variability in predictions from areal models. Top row: FRK model, Bottom row: INLA model.Left: Standard deviations of predicted values when aggregated into 72 areas. Right: True standard deviations (black, grey horizontal line: mean, grey shaded area: ± 2 standard deviations from the mean) and standard deviations of predicted values when aggregated into 72 areas (blue, blue horizontal line: mean).

5.6.2.4. Conclusions

We conclude that the INLA model does a better job than the FRK model of accounting for the additional uncertainty associated with areal data, and we suspect the reason might be a mis-match between model conceptualization and computation in the FRK framework. From a modeling perspective, therefore, the INLA framework seems to be superior. From a practitioner's point of view, however, we note that in our experience the package `R-INLA` was more difficult to use than the `FRK` package. `R-INLA` was more memory intensive, and the reported error messages were sometimes unhelpful. The package has been recently updated to a new package `inlabru` [6], available on CRAN. We are not certain whether the functions necessary to produce a projection matrix that appropriately accounts for areal versus point data have been ported over to the new package.

As we worked through this study, several lines of further investigation suggested themselves. First, we believe that the custom functions we applied to `LatticeKrig` which allow that package to account for areal data are worth of development and publication. The two existing packages that account for change of support —`FRK` and `R-INLA` —each have strengths and weaknesses, and making another option available to practitioners would be valuable. Second, we ran all datasets on a 2-D lattice, but both `R-INLA` and `FRK` are nominally capable of handling data on a sphere. For global datasets, a spherical model would be an improvement both because great circle distances will be more accurate that the Euclidean distances used here, and because a spherical model will account for connectedness at the far west and east edges of the 2-D lattice. Finally, these models should be tested in the new `inlabru` package using the spatial package `sf` rather than `sp`, as both `R-INLA` and `sp` will likely be decommissioned in the future.

Perhaps the most important finding of our study is the following: when considering the change of support problem as defined in this paper, we obtained the best results by predicting on a fine grid and then aggregating the predictions into the areas defined by the original data. Results from this procedure were more successful at capturing the true variability in the data than relying on the reported prediction uncertainty alone. This procedure is more computationally expensive than predicting at only the locations of the observation areas, and practitioners will need to balance the computational costs with the improvement in results.

6. ATTRIBUTION THRUST

The ability to robustly understand and quantitatively assess the causes of climate impacts is central to scientific, legal, and policy communities. In climate science, this is performed through detection and attribution (D&A). Detection of a change in a climate impact (such as global mean surface temperature, GMST) is defined as the process of demonstrating that a change has occurred in some statistically significant sense without providing a reason for the change. For example, the Intergovernmental Panel on Climate Change (IPCC AR6, working group 1) has published charts showing the detection of an GMST increase of 1.1°C over the period 1850-2020; see the discussion in Chapter 3 of Working Group 1 report, FAQ3.1 [37]. Attribution is defined as the process of evaluating the relative contributions of "multiple causal factors to a change or event with an assignment of statistical confidence." [70] For example, the IPCC attributes this GMST increase to anthropogenic activity: "humans are the dominant cause of observed global warming over recent decades" because the observational temperature increase can only be matched using ensembles of climate models when anthropogenic greenhouse gas forcings as well as anthropogenic aerosols are included. The attribution process also accounts for underlying noise: the temperature increase signal has risen far above the internal variability of the climate system [37]. The statement of the IPCC such as "It is unequivocal that human influence has warmed the atmosphere, ocean and land." [24] are attribution statements.

Due to the chaotic nature of the climate system, detection and attribution of a climate impact to a climate forcing is plagued by variability in any climate variable. The further downstream, shorter duration, or more spatially-localized the change, the larger the role that variability plays. Hence, multiple climate attribution approaches [70, 68, 174, 157, 11] have been developed to decipher a signal from the noise, i.e., a forced response from underlying variability. The existing attribution approaches include the concept of a "fingerprint" which is a spatial and/or temporal signal that helps indicate a change. The following references provide a history and evolution of attribution approaches including fingerprinting and its variation, regional attribution, extreme event attribution, and storyline approaches: [156, 68, 69, 71, 127, 174, 157, 143, 162, 163, 11, 102, 119].

In CLDERA, the Attribution Thrust focused on attributing changes in stratospheric temperature or surface temperature to Mt. Pinatubo (Pathway 1) as well as agriculture productivity changes due to Mt. Pinatubo (Pathway 2). We note a fundamental difference between the Mt. Pinatubo exemplar and long-term climate change. In CLDERA, we are trying to attribute a short-term and spatially localized forcing (Mt. Pinatubo erupted over a few days and its effects lasted a few years) as we were targeting high variability environments. This is quite different from the IPCC focus on long term climate change, where decades of observational records of temperature exist and the greenhouse gas forcing has been increasing for over 200 years. Thus, to perform attribution studies of Mt. Pinatubo impacts, we developed three novel approaches that address different aspects of attribution of localized forcings on impacts that have high variability. The first uses an inverse optimization

framework in which the objective is to infer the magnitude (or location or height, etc.) of the source forcing given downstream observations of impact variables. The second is a multi-step attribution framework that uses features of the downstream variables and relationships between these features in a conditional Bayesian formulation. This multi-step method employs scaling experiments (e.g. simulated Pinatubo runs of increasing SO2 magnitude representing increasing forcing) to achieve more certainty in the attribution. The final approach is a new causal modeling algorithm which accounts for space-time dependencies between variables, in contrast with existing causal modeling approaches, which find causal relationships in random variable distributions or time-series data (e.g. temporal dependencies). Collectively, these attribution approaches have increased our understanding and ability to attribute low signal-to-noise impacts from localized forcings.

6.1. Inverse Optimization

Contributing Authors: Joseph Hart (1441), Indu Manikam (5524), Mamikon Gulian (8734), Laura Swiler (1400)

In complex and large-scale climate systems, important effects are caused by a combination of confounding processes that are not fully observable. The identification of sources from observations of system state is vital for attribution and prediction, which inform critical policy decisions. In the context of CLDERA, our objective is to infer the correct source forcing (e.g. the magnitude of SO_2 that Mt. Pinatubo injected into the atmosphere) given downstream observations of aerosol optical depth (AOD). The primary uncertainties in this process are the injection magnitude and the atmospheric wind variability. The difficulty of these types of inverse problems lies in the inability to isolate sources, the cost of simulating computational models, and the inherent uncertainties and variability in the climate system.

6.1.1. CLDERA contributions in Inverse Optimization methods

To address these challenges, we developed a framework which employs the following:

1. Dimension reduction. To perform inverse optimization, it is necessary to reduce the high-dimensional 3-D global gridded data to a smaller dimensional space. We investigated Principal Component Analysis (PCA, [85]) and Radial Basis Functions (RBFs) as two alternatives for dimension reduction. PCA was ineffective due to the internal climate variability and advective characteristics of the plume transport. This led us to our use of RBFs. Traditionally, RBFs are defined on a fixed spatial grid and their coefficients are fit via linear least squares. In this way, the RBF representation is smoothing but does not necessarily give a significant dimension reduction. We adopted a different approach wherein we fit the RBF location, shape parameter, and coefficient simultaneously via nonlinear least squares. This approach adds to the complexity of fitting the RBFs, but enables a low dimensional representation of the plume advection which is not possible using the traditional approach where the basis functions are on a fixed grid.

2. Surrogate models. Optimization involves querying a "forward model" many times, to understand which inputs lead to the desired objective. In our problem, E3SM is the forward simulation model which maps an injection magnitude and atmospheric wind state to AOD. Due to the high computational cost of querying E3SM, we trained an autoregressive neural operator to serve as a surrogate model for E3SM. The model evolved the state of AOD over the first 2 weeks of the simulation. We enforced physical constraints in the model architecture to ensure that physically realizable solutions were produced. These constraints included enforcing that the mass of SO_2 decreased post-eruption (as SO_2 is converted into sulfate), and enforcing that the the total mass of sulfur from both SO_2 and sulfate is conserved at each timestep.

3. Bayesian inversion. The inverse problem inverts the forward problem: given observations of AOD and a probability distribution over the winds, we estimate the injection magnitude(s) which may have led to the AOD observations. We perform this inversion using a probabilistic Bayesian framework, where we determine a posterior probability on the injection magnitude given AOD observations and wind data. This probabilistic estimate uses a deterministic optimizer to obtain the Maximum a Posteriori (MAP) point of the posterior distribution of injection magnitude and then uses additional stochastic approaches to obtain the full distribution. A Bayesian Approximation Error (BAE) approach was used to embed the wind variability within the Bayesian formulation, thus enhancing our quantification of uncertainty.

4. Ensembles. Our dataset consisted of volcanic simulations of varying injection magnitudes. For each injection magnitude, multiple simulations were run that had different atmospheric states and wind patterns (climate ensembles). We relied on limited variability ensembles to reflect prior knowledge of the climate state around the time of the Mt. Pinatubo eruption and thus reduce the uncertainty in our source magnitude estimate. In our final study, 7 ensemble members, each having 5 injection magnitudes, were used to train the surrogate model. We used 5 additional simulations for hyperparameter tuning the model architecture, and 2 simulations from previously unseen injection magnitudes and atmospheric states as the final test set.

The inverse optimization team followed a careful tiered verification approach to develop the necessary components for the inverse framework. The team started with a Gaussian plume model which captured the time evolution and also demonstrated the wind variability which needed to be accounted for in the inversion process. Then, the approach was demonstrated and tested on the Held-Suarez-Williamson model, where spatial averaging and dimension reduction needed to be addressed more fully. Finally, the inverse method was demonstrated on the full E3SM model with prognostic aerosol treatment. In this model, the source tagging capability was used to separate Pinatubo aerosols from other aerosol forcings (e.g. anthropogenic).

Two journal papers have been prepared summarizing the inverse optimization framework developed under CLDERA. They are listed below [65, 66] :

- Joseph Hart, Mamikon Gulian, Indu Manickam, and Laura P. Swiler. Solving high-dimensional inverse problems with auxiliary uncertainty via operator learning with limited data. *Journal of Machine Learning for Modeling and Computing*, Vol.4, No. 2, pp. 105-133. 2023.

- Joseph Hart, Indu Manickam, Mamikon Gulian, Laura Swiler, Diana Bull, Thomas Ehrmann, Hunter Brown, Benj Wagman, and Jerry Watkins. Stratospheric aerosol source inversion: Noise, variability, and uncertainty quantification, submitted to *Journal of Machine Learning for Modeling and Computing*. This manuscript is available on arXiv, `https://doi.org/10.48550/arXiv.2409.06846`.

6.2. Extensions of Attribution

Contributing Authors: Laura Swiler (1400), Diana Bull (8931), Christopher Wentland (8734), R. Michael Weylandt (5573)

In the CLDERA project, we first examined fingerprinting methods for attribution. Then, we developed a multi-step conditional attribution approach which leverages features from pathways. The paragraphs below place this multi-step attribution approach developed under CLDERA in the context of the larger field of attribution science.

6.2.1. Attribution Background

Attribution of long-term climate impacts to anthropogenic forcings such as greenhouse gases (GHG) and aerosols in the climate system is a well-established practice [116, 37]. These studies rely almost entirely on the "fingerprinting" methodology first introduced by Hasselmann [67] and refined over the following decade [156, 68, 69, 71, 127, 4]. This method determines spatial and/or temporal patterns of impacts under externally forced conditions. Typically, observational results are regressed on simulation results driven by external forcings like GHG or aerosols. The regression coefficients and corresponding significance are used to identify combinations of forcings which "best fit" the observational data; these are the important forcings to which the change is attributed. The forced simulation data might be as simple as a global average time series "signal," though in many studies a signal is computed from the projection of data onto a set of empirical orthogonal functions (derived from a principal component analysis). A large body of work aims to identify optimal fingerprints which best improve the signal-to-noise ratio [143], characterize and accommodate for the internal variability of the climate [157, 195], and perform regional attribution studies [12, 174]. Multivariate fingerprinting approaches utilize multiple climate variables (such as temperature and precipitation together) to identify a fingerprint which more robustly achieves attribution [11, 111]. Bayesian formulations of the fingerprinting problem allow analysts to more deliberately quantify uncertainty in the fitted model [64] and encode prior knowledge of climate forcings [8].

Pattern scaling techniques [5, 130, 72] can be used to evaluate how the magnitude of the climate response changes with respect to the degree of an external forcing (e.g., the amount of GHG); these patterns are normally expressed as a change per degree of global-mean temperature change [130]. This is similar to epidemiological studies which invoke dose-response relationships as evidence of causal relationships [74]. As attribution is inherently retrospective, associative pattern attribution [70] studies evaluating the sensitivity of impacts to changes in the degree of external forcings are not common, as the degree of GHG forcing up to present is generally not contested.

"Multi-step" fingerprinting methods attempt to link downstream impacts to a climate forcing through a series of single-step fingerprinting assessments [70]. By incorporating progressive steps of the process, this method explicitly evaluates the physical mechanism by which a driver influences a downstream impact. The assumed statistically-independent nature of the component steps unfortunately limits the confidence in such analyses to that of the single weakest attribution step. The literature appears devoid of any such multi-step attribution findings following the procedure outlined by Hegerl [70], successful or otherwise, likely due to this limitation. Other multi-step attribution frameworks exist beyond the optimal fingerprinting canon [56], but remain relatively rare.

6.2.2. Attribution challenges for localized forcings

The above attribution methods often struggle to achieve successful attribution in regimes where the forcing response is dominated by the climate's internal variability. In particular, there are many climatic events that are short-lived and characterized by large variability in the forcing response. These can arise from spatially and/or temporally localized external forcings like forest fires [101], industrial fires (e.g., the Kuwait oil fires [17]), volcanic eruptions, and geoengineering projects such as marine cloud brightening or stratospheric aerosol injection [81]. The characteristic time scales of these events and their purported impacts (months or years) are much shorter than those of anthropogenic climate change (decades), and their attribution requires methods capable of overcoming the associated low signal-to-noise ratios for the impacts of interest.

6.2.3. CLDERA contributions in fingerprinting

To address the above challenges, we first focused on dimension reduction approaches, which are an essential part of the climate analyst's toolkit. Due to the enormous scale of climate data, dimension reduction methods are used to identify major patterns of variability within climate dynamics, to create compelling and informative visualizations, and to quantify major named modes such as the El-Niño Southern Oscillation. Principal Components Analysis (PCA), also known as the method of empirical orthogonal functions (EOFs), is the most commonly used form of dimension reduction, characterized by a remarkable confluence of attractive mathematical, statistical, and computational properties. Despite its ubiquity, PCA suffers from several difficulties relevant to climate science: high computational burden with large data sets, decreased statistical accuracy in high-dimensions, and difficulties comparing *across* multiple data sets.

We introduced several variants of PCA that are likely to be of use in climate sciences and address these problems. Specifically, we introduced *non-negative*, *sparse*, and *tensor* PCA and demonstrate how each approach provides superior pattern recognition in climate data. We demonstrated these approaches through an analysis of several runs of the E3SM climate model from 1991 to 1995, focusing on the simulated response to the Mt. Pinatubo eruption; our findings are consistent with a recently-identified stratospheric warming fingerprint associated with this type of stratospheric aerosol injection. These variants of PCA and the demonstration of these dimension reduction approaches have been published in the following paper [193]:

- Michael Weylandt and Laura P. Swiler. Beyond PCA: Additional Dimension Reduction Techniques to Consider in the Development of Climate Fingerprints. *Journal of Climate*, Vol. 37, Issue 5, pp. 1723-1735. 2023.

6.2.4. CLDERA contributions in multi-step attribution

To address the attribution challenges of localized forcings, we developed a multi-step attribution approach in CLDERA for which the inclusion of additional steps can *increase* the confidence of attribution of downstream impacts from a localized source forcing. Rather than the assumed statistical independence of steps as in the method noted by [70], this new method relies on a *conditional* Bayesian approach which leverages significant relationships in each step of a pathway to increase attribution certainty. Under this formulation, when intermediate steps in a proposed causal pathways exhibit a stronger relationship than the equivalent single step from forcing to final impact, the ability to distinguish between different forcing levels is improved and confidence in attributing an observed response to a specific forcing level is increased. While we use the phrase "multi-step" to describe this method, this represents a significant departure from the approach outlined by Hegerl [70].

To demonstrate the proposed methodology for a short-term climate forcing marred by high variability in the impact response, we examine the June 15th 1991 eruption of Mt. Pinatubo. We propose a multi-step pathway derived from the eruption's injection of ~10 teragrams of SO_2 into the atmosphere and the resulting formation of sulfates in the stratosphere. The sulfates modify the radiative fluxes at both the top-of-atmosphere and the Earth's surface, and ultimately lead to changes in the stratospheric and surface temperatures.

A major challenge with a typical fingerprinting approach is the use of time-series data. While aligning time-series signals work for a single variable, the time series of different steps along the P1 pathway exhibit different characteristic time scales based on Mt. Pinatubo's impact. For example, a decreased surface temperature is sustained for much longer than the shortwave radiative flux decrease from the Mt. Pinatubo eruption. This precludes simple analysis of relationships between such variables over entire time series (as is standard in optimal fingerprinting). As such, carefully-designed scalar features of the time series must be specified to more easily link variables in the proposed pathway. The need to align pathway variables with different time-scale characteristics led us to a feature-based approach, where each time-series was reduced to a feature.

In the multi-step approach, the feature used was a peak value: we extracted the peak impact from each time series using a procedure which determines the maximum deviation of the impact time series from pre-forcing (i.e., pre-eruption) conditions. We found that peak response (e.g. peak stratospheric or surface temperature, peak radiative flux) got larger as a function of the magnitude of the Mt. Pinatubo forcing. This is as expected, but nicely demonstrates a "dose-response" behavior which is part of the causal argument for downstream impacts. Finally, another key relationship that we identified and leveraged was the fact that the peak values of downstream variables tended to be linearly related to peak values of upstream variables. Linear regressions of the peaks (as a function of the Mt. Pinatubo forcing magnitude) were used in the multi-step analysis.

Using this novel approach involving peaks, linear relationships, and Bayesian inference, we were able to correctly attribute the eruption size (forcing level) based on downstream observations and increase our confidence in the attribution of forcing level by including conditionally-dependent intermediary steps.

This work is documented in the following paper [191]:

- Christopher R. Wentland, Michael Weylandt, Laura P. Swiler, Thomas S. Ehrmann, Diana Bull. Probabilistic multi-step attribution for short-term climate forcings. Submitted to *Journal of Climate*, Aug. 2024. This manuscript is available on arXiv: `https://doi.org/10.48550/arXiv.2409.01396`.

6.3. Causal Discovery

Contributing Authors: J. Jake Nichol (1441), R. Michael Weylandt (5573), Melanie Moses and Matthew Fricke (University of New Mexico), Mark Smith (5493), Laura Swiler (1400)

Causal discovery (CD) tools [171, 168, 54] enable scientists to infer meaningful relationships from observational data, spurring advances in fields as diverse as biology, economics, and climate science. Despite these successes, the application of causal discovery to space-time systems remains immensely challenging due to the high-dimensional nature of the data. For example, in climate sciences, modern observational temperature records over the past few decades regularly measure thousands of locations around the globe.

To address these challenges, CLDERA researchers developed **Ca**usal **S**pace-Time S**t**encil **Le**arning (CaStLe), a novel algorithm for discovering causal structures in complex space-time systems. CaStLe leverages regularities in local dependence to learn governing global dynamics. This local perspective eliminates spurious confounding and drastically reduces sample complexity, making space-time causal discovery practical and effective. These advances enable causal discovery of geophysical phenomena that were previously unapproachable, including non-periodic, transient phenomena such as volcanic eruption plumes. When applied to ever-larger spatial grids, CaStLe's performance actually improves because it transforms large grids into *informative spatial replicates*.

6.3.1. Causal discovery background

Today's climate observation modeling capabilities provide a wealth of data for studying our planet's complex dynamics. Unfortunately, the immense complexity of those dynamics means that simple analyses only tell a small part of the data's story. Causal discovery tools offer the ability to understand finer mechanistic details via causal graphs' simplicity, interpretability, and flexibility.

Causal discovery is a field that utilizes algorithmic causal inference to identify causal models as dependencies between fields of interest, which are often represented as a directed acyclic graph (DAG). CD seeks to recover causal dynamics from observational data. CD generally has two output classes: a causal graph/network [133] or a structural causal model [134]. We focus on

causal graphs which let us analyze the space-time evolution of fields of interest without specifying hypothesized physical models. The causal graphs are networks of variables (nodes) which are connected by edges that denote a causal dependence. Causal graphs can provide insights that are valuable building blocks to better mechanistic understandings of the climate system. Insights gleaned from causal discovery can further inform physical models, validate simulations against observational data, and identify future research questions. Additionally, causal models visualized as causal graphs provide a useful visual shorthand model of complex physical dynamics.

While CD has the power to impact climate science significantly, the enormous size and scope of climate data have limited its applications. More specifically, climate data often contains tens to hundreds of thousands of grid cells, each with several orders of magnitude fewer observations in time. That imbalance is frequently called *the curse of dimensionality* in machine learning because statistical methodologies fail to estimate relationships when there are more fields than observations to describe them. Despite these challenges, CD has been successfully applied in climate science [32, 155, 128, 22, 23, 87, 43, 177], primarily via dimensionality reduction techniques to reduce the number of relationships to estimate. These studies have identified teleconnection pathways to recover large, periodic climate modes and their effects. While that approach can be practical, analysis of local effects has been considered challenging and generally avoided due to the curse of dimensionality [34, 155, 128]. In contrast to this prior work, our contribution leverages the known locality in space-time systems to harness spatial replicates and identify local causal graphs.

6.3.2. CLDERA contribution in causal discovery: CaStLe

Under CLDERA, we introduced the CaStLe approach to causal discovery for space-time data. CaStLe allows the discovery of causal structures in high-dimensional spatial data, avoiding the need for dimension reduction techniques that dominate causal discovery of space-time data, e.g. the work by [128]. By working in the raw data space, CaStLe's causal graphs are *inherently interpretable* and do not require mapping structures from the dimension-reduced space back onto the original data.

In the work developing CaStLe, we performed the following studies and analysis:

1. We performed a theoretical analysis of CaStLe, showing that it has attractive computational and statistical properties and, rather remarkably, that CaStLe's accuracy actually increases on larger spatial domains.

2. We performed extensive validation experiments to demonstrate the effectiveness of CaStLe over existing causal-discovery frameworks on a range of synthetic benchmark data that are useful proxies for climate data. These include a set of vector autoregression (VAR) problems and the use of CaStLe to identify the governing dynamics of an important non-linear partial differential equation (PDE).

3. We successfully applied CaStLe to discover the atmospheric dynamics governing the climate response to the 1991 Mount Pinatubo volcanic eruption. We demonstrated how CaStLe can be used to better understand how stratospheric winds mediate climate response to volcanic activity.

The CaStLe causal discovery methodology, along with the studies and analysis mentioned above, are documented in the following paper [124]:

- Jake Nichol, Michael Weylandt, G. Matthew Fricke, Melanie E. Moses, Diana Bull, Laura P. Swiler. Causal Space-Time Stencil Learning: Local Causal Dynamics in Complex Systems. Submitted to *Journal of Geophysical Research, Machine Learning and Computation*

The code repository can be found at: `https://github.com/jjakenichol/CaStLe`.

As part of the causal model development, the team developed a large suite of benchmark test cases for tiered verification of causal discovery methods, specifically focusing on datasets that exhibited space-time dependencies. The initial test cases are based on VARS (Vector Autoregressive Models) that modeled 1-D and 2-D spatial dependencies as well as temporal dependencies. The test cases provide an important set of benchmarks for the climate community, as they have tests of increasing difficulty on multiple dimensions: the number of time points, the grid size (e.g. number of spatial grid points in the test data), and neighborhood dependence density (the amount of connectedness between a center grid cell and its eight neighbors), as well as the amount of noise and the amount of autocorrelation in the data. These test cases are documented in the following SAND report [125]:

- J. Jake Nichol, Michael Weylandt, Mark Smith, Laura Swiler. "Benchmarking the PCMCI Causal Discovery Algorithm for Spatiotemporal Systems." SAND2023-05141. June 2023, Sandia National Laboratories, Albuquerque NM.

Finally, we note that the main author and developer of the CaStLe approach, J. Jake Nichol, is now working on a multi-variate version of CaStLe as part of his Ph.D. dissertation.

This page intentionally left blank.

7. SCIENCE ANALYSIS

In addition to the method development performed under the three technical thrust for Tiers 1-3 described in Chapters 4, 5, and 6, research was performed under CLDERA to advance the scientific understanding of stratospheric processes within E3SM and to evaluate the effects of Mt. Pinatubo on debated and further downstream impacts. This chapter provides an overview of further research aimed at advancing scientific understanding, starting with the first detailed analysis of stratospheric circulation in E3SMv2 (Section 7.1) as well as an analysis of Brewer-Dobson circulation using tracers within E3SMv2 (Section 7.2). Using simulations with both prescribed and prognostic volcanic aerosols we evaluated how the variability changed in terms of stratospheric circulation (Section 7.5). The CLDERA E3SMv2-SPA simulations were also used to perform an analysis of changes in cirrus clouds due to the Mt. Pinatubo eruption (Section 7.6) and evaluate a debated impact from the Mt. Pinatubo eruption, which is winter warming in the Northern hemisphere in the year after the eruption (Section 7.4). Under Tier 4, the Echo State Network was used to analyze the impacts of Mt. Pinatubo on the south Asian monsoon (Section 7.7) and multi-step attribution was applied to the impacts of Mt. Pinatubo on agricultural productivity in the temperate North 7.8).

7.1. Characterization of E3SMv2's general circulation

Contributing Authors: Christiane Jablonowski and Joe Hollowed (University of Michigan).

The CLDERA project was an early adopter of the E3SMv2 model configuration which was publicly released in September 2021, just before the start of the CLDERA project. This raised open questions about the physical realism of the low-resolution (LR) E3SMv2 simulations with 1° grid spacing and 72 vertical levels that reach up to 0.1 hPa (about 63 km). This resolution captures the stratospheric circulation with vertical grid spacings between 500 m and 1 km in the tropopause region and lower stratosphere below 50 hPa (about 20 km). The level spacing is then stretched to about 3 km at 1 hPa (or about 50 km). This design resolves important elements of the stratospheric circulation, such as the polar jets, but misses some forcing mechanisms by upward traveling waves with short wavelengths of just a few kilometers. Such upward traveling waves are, for example, important for the realistic representation of the QBO in the tropical stratosphere. Our early results revealed that E3SMv2's QBO is highly biased, and these findings motivated further studies to provide a general understanding of E3SMv2's stratospheric biases.

While E3SMv2's model performance of the coupled system was investigated for the troposphere via the 6th Coupled Model Intercomparison Project (CMIP6) simulation campaign as documented in [55], no published analyses were available to judge the realism of E3SMv2's stratospheric circulation. We therefore conducted a series of studies to inform the CLDERA team about stratospheric biases that can impact the life cycle of the volcanic aerosols and their physical pathways for the

source-to-impact attribution studies. The majority of our climatology studies were built upon the publicly released five-member ensemble of the CMIP6 simulation campaign with the coupled E3SMv2 configuration. In addition, some studies were conducted with the CMIP6 data of the coupled predecessor version E3SMv1 as well as the three-member ensemble of the uncoupled E3SMv2 configuration in so-called AMIP mode. The latter was driven by prescribed sea surface temperatures and sea ice. Both our coupled and uncoupled analyses focused on the historical CMIP6 scenario which incorporates the observed greenhouse gas and aerosol concentrations until 2014. In particular, the enhanced aerosol concentrations due to volcanic eruptions, including the Mt. Pinatubo eruption in June 1991, were prescribed. We focused our studies on the satellite era from 1980 onwards. This enabled us to compare the E3SM climatologies to satellite data, and other observational and reanalyses products.

Our assessments of E3SM's general circulation, as seen in the CMIP6 data sets, focused on the climatology of the zonal wind, temperature, water vapor, and precipitation rates as well as on the stratospheric Brewer-Dobson Circulation (BDC). The latter was assessed via the "Transformed Eulerian Mean" (TEM) framework, which sheds light on the diabatic (residual) circulation and the driving of the zonal-mean zonal flow by resolved waves via the divergence of the Eliassen-Palm flux. The diabatic circulation largely determines the tracer transport velocities in the stratosphere which were investigated further via passive tracer studies (see also section 7.2). In addition, we investigated E3SM's modes of variability, such as the so-called "water vapor tape recorder" and the QBO in the tropical stratosphere, the semi-annual oscillation (SAO) of the zonal wind in the tropical mesosphere, and the tropical wave activity as measured by the wavenumber-frequency spectra of the tropical precipitation. The water vapor tape recorder also provides information about the troposphere-stratosphere exchange processes. This is due to the fact that the main source of water vapor for the stratosphere is the upward advection of the water vapor from the tropical tropopause region.

Our studies found that E3SMv2 exhibits important biases, such as a severe lack of water vapor in the stratosphere, a large-magnitude warm bias near the model top above 1 hPa, a reduced tropical wave activity and degraded representation of the QBO in comparison to E3SMv1, and biased polar jets that extend too far up with overestimated zonal wind speeds. Further analyses suggest that some of these biases are likely due to undesirable properties of E3SM's physical parameterization package. For example, it was found that the stratospheric water vapor gets removed by moist physical processes, which likely come from the cloud and boundary layer schemes. However, cloud formation is not a relevant process for the stratosphere. Therefore, this aspect needs further investigations and corrections in future E3SM versions. In addition, our analyses suggest that the temperature biases are likely caused by the radiation scheme. This was analyzed via the temperature tendencies for longwave and shortwave radiation which showed significant biases with opposite signs.

These biases were revealed when comparing the E3SM CMIP6 data to various observational and model products. They included the SWOOSH satellite data set for stratospheric water vapor [30], ERA5 reanalyses data [73], as well as simulation data from NCAR's whole atmosphere model WACCM. As mentioned above, the primary purpose of the assessments was to understand the realism of E3SMv2's stratospheric circulation and its ensemble spread. This informed CLDERA's pathway verification tasks and the aerosol-related model developments as, for example, discussed

in [13]. Despite the physical biases, E3SMv2 was deemed to capture the Mt. Pinatubo impacts on the circulation at a satisfactory level which was important for the development and tests of the CLDERA tool sets.

This research is currently being summarized in the following articles [83, 76]:

- Jablonowski, C., Hollowed, J., Nguyen, L., Hillman, B., Ehrmann, T., Wagman, B. "Stratospheric Characteristics in E3SMv2." To be submitted by December 2024.
- Hollowed, J., Jablonowski, C., Ehrmann, T., Hillman, B., ... "Volcanic Aerosol Modification of the Stratospheric Circulation in E3SMv2 Part I: Wave-Mean Flow Interaction." To be submitted to *Climatic Dynamics* by October 2024.

7.2. Tracer Analyses of Brewer-Dobson Circulation

Contributing Authors: Joe Hollowed and Christiane Jablonowski (University of Michigan).

The Brewer–Dobson circulation (BDC) describes the global atmospheric circulation pattern that includes the rise of the air from the tropical troposphere into the stratosphere, and subsequent descending poleward motion in the stratosphere. It is this global circulation pattern that determines the mean age of stratospheric air, the residence times of stratospheric gases, the distributions of stratospheric water vapor and ozone, and the tropical tropopause temperatures (important for exchange processes). Because this circulation pattern is so important for determining the chemical and thermodynamic composition of the stratosphere, and thus the global climate as a whole, we are particularly interested in identifying modifications of the BDC by the radiative forcing of stratospheric sulfate aerosols, and the dynamical mechanisms which drive them.

In order to investigate these questions, we instrumented the CLDERA E3SMv2 configuration with a set of idealized tracers, designed to assess the model's BDC and the tracer transport characteristics in the stratosphere. This tracer set includes age-of-air (AOA) tracers, as well as a pair of related tracers known as E90 and ST80. All of these tracers are "passive", meaning that they do not interact in any way with other species, and they do not provide any feedbacks to the atmosphere (i.e. a model run will proceed identically with or without them enabled). Each of these tracers has an associated source and sink, which describe their creation and removal as a function of space. The integrated sum of the source, sink, and tracer advection yields the tracer concentration distribution at a given time, expressed as dimensionless mass mixing-ratios (kg of tracer per kg of air). The source and sink functions for AOA, E90, and ST80 have been designed such that the mean distributions of these tracers are able to diagnose certain aspects of the stratospheric circulation.

The AOA tracer, or sometimes referred to as a "clock tracer", is defined to exist everywhere in the atmosphere, from the surface to the model top [188]. Its source is simply set to 1 s^{-1}, which drives a uniform production of the AOA over the entire atmosphere. Its sink is nonzero only near the surface (below 700 hPa), where it quickly forces the tracer concentration to zero. In the right units, the mixing ratio of the AOA thus gives the average age of air, or the time since the air's last contact with the surface layer. In accordance with the BDC, the AOA is generally smallest in the tropical

troposphere, and largest in the polar stratosphere. Thus, modifications to the global AOA indicates changes to the BDC.

The E90 tracer has a constant emission at the surface, and has a uniform sink throughout the atmosphere which describes an exponential decay with a 90-day e-folding timescale, e.g. see also the description in [2]. Using this simple process the E90 sink quickly establishes a global tracer distribution throughout the troposphere, with the highest concentrations being found near the surface and throughout the tropical vertical column. Because troposphere-stratosphere mass exchange is hindered by the temperature inversion at the tropopause, the E90 tracer accumulates here with a sharp vertical gradient. Thus, modifications of the stratospheric E90 concentration encodes changes to upward cross-tropopause transport.

While E90 is designed to diagnose troposphere-stratosphere exchange, the complimentary ST80 tracer does the same for stratosphere-troposphere exchange. This tracer is held to a fixed concentration in the stratosphere, above 80 hPa, and has a uniform sink below the tropopause which describes an exponential decay with a 25-day e-folding timescale. Thus, modifications of the tropospheric ST80 concentration encode changes to the down cross-tropopause transport.

Currently, these tracers are used to identify modifications to the BDC by volcanic forcing of the Mt. Pinatubo eruption in E3SMv2-SPA [13]. To achieve this, we have enabled these tracers in ensembles of simulations with Pinatubo included, and also ensembles without Pinatubo (known as the “counterfactual" simulations). Each simulation in the Pinatubo ensemble have a “paired" simulation in the counterfactual ensemble, which shares an identical initial condition. The difference in the tracer fields between each pair thus provides the changes to the tracer distributions that are due to the volcanic forcing specifically.

Figure 7-1 shows an example of the impacts that we are able to detect via this method. Here “impact” specifically refers to the difference between each pair of ensemble members. Positive impacts are excesses in the Pinatubo runs, while negative impacts are deficits in the Pinatubo runs. Also shown is the statistical significance of the measured impact, determined by a t-test between the two ensembles. Figures 7-1 (a)-(b) show the unaltered AOA distribution in the counterfactual ensemble for October 1991 (4 months post-eruption), and the impact by the volcanic forcing over this same time period. The dominant feature in the impact is a significant decrease in the mean AOA in the tropical stratosphere, which is indicative of enhanced tropical upwelling. At the same time, there is increased AOA at higher latitudes in the mid- to lower-stratosphere, which indicates enhanced subsidence of older air from the upper stratosphere. Overall, these effects indicate an accelerated BDC. Figures 7-1 (c)-(d) show the analogous information for E90, which demonstrates that E90 concentrations have enhanced throughout the global stratosphere. This further suggests an accelerated BDC, and in particular an enhanced mass exchange from the troposphere to the stratosphere.

Further work is underway to investigate the full time series of these tracer impacts for several years following the Pinatubo eruption, and magnitude scaled versions of the eruption, in E3SMv2-SPA. Theoretical methods of identifying the specific dynamical mechanisms responsible for these circulation changes are also being developed, which will shine light on the fundamental ways in which volcanic forcing enables stratospheric composition changes. These mechanisms are

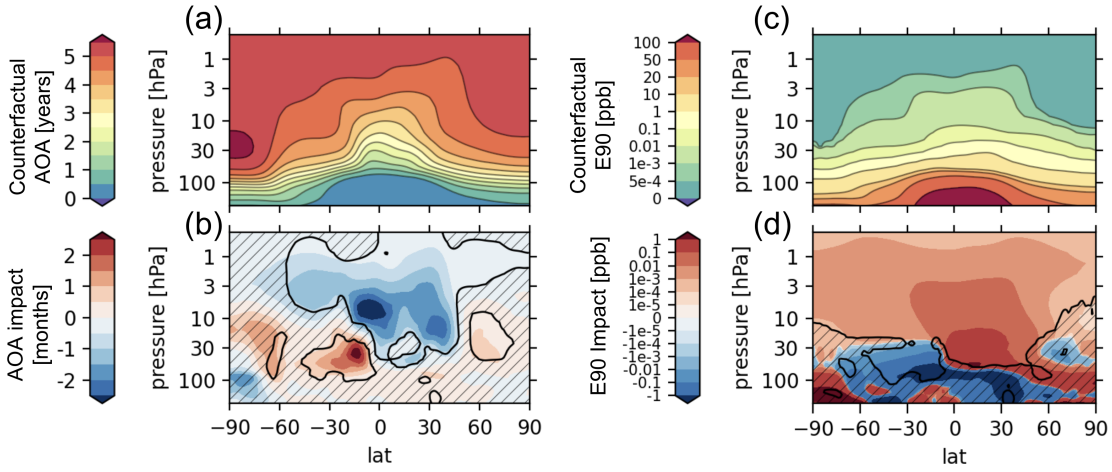


Figure 7-1. October 1991 monthly mean AOA and E90 in the counterfactual ensemble, and the impact in the tracer distributions between the Pinatubo and counterfactual simulations. (a) ensemble-mean, zonal-mean AOA for the counterfactual ensemble in years, (b) ensemble-mean, zonal-mean AOA impact in months, (c) ensemble-mean, zonal-mean E90 for the counterfactual ensemble in ppb, (d) ensemble-mean, zonal mean E90 impact in ppb. In both panels (b) and (d), a black contour shows 95% statistical significance, and hashed regions show statistical significance less than 95%.

important to identify in order to properly understand the dynamics relevant to the pathways between the volcanic source, and far-removed, indirect impacts on the climate.

In addition, the team conducted work to use the dynamic tracers "Potential Vorticity" (PV) and "Potential Temperature" (PT) as markers for the stratospheric circulation. These quantities are conserved for adiabatic conditions. Therefore, the deviation of the PV and PT tracers from their dynamic flow counterparts reveals the impact of the diabatic forcing from the volcanic eruption on the circulation. Both, the PV and PT tracers have been added to CLDERA's E3SMv2 code base and can be used for future studies, including studies of tropopause folds and troposphere-stratosphere exchange processes.

This research is currently being summarized in the following article [75]:

- Hollowed, J., Jablonowski, C., Ehrmann, T., Hillman, B., ... "Volcanic Aerosol Modification of the Stratospheric Circulation in E3SMv2 Part II: Tracer Sensitivity." To be submitted by December 2024.

We note that two papers relating to stratospheric circulation will become two chapters in the main researcher's, Joe Hollowed, Ph.D. dissertation.

7.3. Size-resolved process understanding of stratospheric sulfate aerosol following the Pinatubo eruption

Contributing Authors: Allen Hu and Xiaohong Liu (Texas A& M University).

Stratospheric sulfate aerosol produced by volcanic eruptions plays important roles in atmospheric chemistry and the global radiative balance of the atmosphere. The simulation of stratospheric sulfate concentrations and optical properties is highly dependent on the chemistry scheme and microphysical treatment. In this work, we implemented a sophisticated gas-phase chemistry scheme (full chemistry, FC) and a 5-mode version of the Modal Aerosol Module (MAM5) for the treatment of stratospheric sulfate aerosol in the Department of Energy's Energy Exascale Earth System Model version 2 (E3SMv2) model to better simulate the chemistry-aerosol feedback following the Pinatubo eruption, and to compare it against a simulation using simplified chemistry (SC) and the default 4-mode version of the Modal Aerosol Module (MAM4). MAM5 experiments were found to better capture the stratospheric sulfate burden from the eruption of the volcano to the end of 1992 as compared to the High-resolution Infra Red Sounder (HIRS) observations, and the formation of sulfate in MAM5FC was significantly faster than in MAM4FC due to the addition of a OH replenishment reaction. Analyses of microphysical processes indicate that more sulfate aerosol mass was generated in total in FC experiments than in SC experiments. MAM5 performs better than MAM4 in simulation of aerosol optical depth (AOD); AOD anomalies from the MAM5 experiment have better agreement with AVHRR. The simulated largest changes in global mean net radiative flux at the top of the atmosphere following the eruption were about -3 W/m^2 in MAM5 experiments and roughly -1.5 W/m^2 in MAM4 experiments.

This work is detailed more fully in the following article [79]:

- Hu, A., Liu, X, Ke, Z., Wagman, B., Brown, H., Lu, Z., Bull, D., Peterson, K. Size-resolved process understanding of stratospheric sulfate aerosol following the Pinatubo eruption, Submitted July 2024. *Atmospheric Chemistry and Physics*.

7.4. Northern Hemisphere Winter Warming

Contributing Authors: Thomas Ehrmann (8931), Benjamin Wagman (8931), Diana Bull (8931), Benjamin Hillman (8931), Joseph Hollowed (University of Michigan), Hunter Brown (8931), Kara Peterson (1442), Laura Swiler (1400), Jerry Watkins (8734), and Joseph Hart (1441).

The Mt. Pinatubo eruption on 15 June 1991 is often associated with surface warming in the northern hemisphere during the subsequent winter. The prevailing theory is that increased low latitude stratospheric temperatures caused by the volcanic aerosol injection amplified the meridional temperature gradient and strengthened the northern hemisphere stratospheric polar vortex (SPV). Through the existing teleconnection between the high latitude stratosphere and troposphere the strengthened SPV induces a positive North Atlantic Oscillation (NAO) phase which leads to surface warming over the continental landmasses.

To isolate this localized impact of the eruption we created an ensemble of simulations using the Energy Exascale Earth System Model version 2 (E3SMv2) with prognostic aerosol modifications initialized on 1 June 1991 to limit the intra-ensemble variability at the time of the eruption. The initialization of these simulations was selected to have major climate modes, specifically the quasi-biennial oscillation (QBO) and the El Nino southern oscillation (ENSO), in similar states to those preceding the real-world eruption. We also generated a more traditional ensemble representing the

full range of intra-ensemble variability to compare to our limited variability ensemble. For each ensemble member in both the limited variability and full variability experiment we generated a paired counterfactual simulation with the Pinatubo forcing removed to allow for isolation of the Pinatubo impact. In general, the limited variability ensemble has greater coherence in the Pinatubo impact across ensemble members which leads to more statistically robust signal compared to the full variability ensemble.

Investigating the northern hemisphere winter warming post-Pinatubo, we find significant warm anomalies in our limited variability ensemble consistent with observation. However, similarly significant warm anomalies are also present in our limited variability counterfactual simulation, and the difference between the limited variability ensemble members with Pinatubo versus the counterfactuals do not show any significant mid-winter surface warming. Additionally, the full variability ensemble members do not show significant mid-winter surface warming anomalies. All of this suggests that the northern hemisphere mid-winter surface warming present in observations is more likely a result of the state of the major modes in the atmosphere at the time of the eruption, specifically ENSO and the QBO, than a response to the volcanic aerosol injection.

Although we did not find evidence of the traditionally proposed northern hemisphere warming, we did find a statistically robust high latitude surface warming response in fall months immediately following Pinatubo. There was a significant mid-latitude westerly signature in the stratospheric zonal wind that starts less than a month after the eruption and continues through October 1991. During October 1991, as the SPV begins to spin-up, these increased mid-latitude westerlies are associated with a significant weakening of the high latitude westerlies. This weakening of the SPV core is associated with a positive geopotential height impact which penetrates to the surface driving a high-latitude warm temperature impact. This alternative warming pathway is especially coherent in the limited variability ensemble where it is present in more than 90% of individual ensemble members.

This work is documented in following paper [36]

- Thomas Ehrmann, Benjamin Wagman, Diana Bull, Benjamin Hillman, Joseph Hollowed, Hunter Brown, Kara Peterson, Laura Swiler, Jerry Watkins, and Joseph Hart. "Identifying Northern Hemisphere Temperature Responses to the Mt. Pinatubo Eruption through Limited Variability Ensembles." Submitted to *Climate Dynamics*, September 2024.

7.5. Prescribed Versus Prognostic Aerosol Variability

Contributing Authors: Benjamin Wagman (8931), Thomas Ehrmann (8931), Hunter Brown (8931), Diana Bull (8931).

Fully coupled earth system model (ESM) ensembles are essential tools for diagnosing climate variability. However, many ESMs, including E3SMv2, prescribe aspects of stratospheric volcanic eruptions instead of prognostically simulating the formation and evolution of volcanic stratospheric sulfate aerosol. E3SMv2 prescribes stratospheric sulfate extinction, such that the volcanic forcing for stratospheric eruptions is identical across ensemble members, and the feedback from simulated climate fields to volcanic forcing is missing. The identical volcanic forcing across ensemble

members is an artificial constraint on forcing; in the natural system, and in ESMs with prognostic stratospheric volcanic aerosol, volcanic forcing evolves independently across ensemble members as a function of internally simulated stratospheric circulation, moisture, and other simulated variables.

Here we compare the Mt. Pinatubo eruption in two E3SMv2 ensembles. One ensemble has identical prescribed stratospheric volcanic forcing, and the other uses a modified version of E3SMv2 in which the volcanic stratospheric aerosol and its forcing are simulated prognostically, and therefore varies across ensemble members. We hypothesize that the prognostic ensemble should exhibit greater variability in volcanic forcing and therefore greater variability in other simulated fields, e.g. surface radiative fluxes and temperature. However, we find only brief and localized enhancements to the variability in volcanic forcing, and no statistically significant enhancement in variability in the surface radiative fluxes and temperature. We then consider possible explanations for the relative lack of enhancement in variability, particularly, whether aspects of the stratospheric circulation may be constraining the variability in forcing in the prognostic ensemble.

This work is documented in following draft paper [183] with anticipated submission to Atmospheric Chemistry and Physics:

- Benjamin M. Wagman, Hunter Brown, Diana Bull, Tom Ehrmann, Ben Hillman, Laura Swiler, Kara Peterson, Joe Hollowed. "Does prognostic volcanic aerosol enhance climate variability in Earth System Model simulations of volcanic eruptions?" SAND2024-12725o. September 2024, Sandia National Laboratories, Albuquerque NM.

7.6. Impacts on Cirrus and implications for Cirrus Cloud Thinning

Contributing Authors: Benjamin Hillman (8931), Xiaohong Liu (Texas A& M University), Allen Hu (Texas A& M University).

7.6.1. Background

It has been hypothesized that large volcanic eruptions might impact cirrus clouds by increasing homogeneous freezing due to increased aerosol number concentration, however, there is a bit of conflicting evidence for this in the literature. An initial study by [198] using HIRS data suggested a strong increase in cirrus following the Mount Pinatubo eruption, but a subsequent reexamination by [196] showed a diminished impact. [105] showed no significant impact on ISCCP cirrus amount or brightness temperature. However, more idealized studies have suggested an impact. [100] showed an increase in homogeneous freezing rate, and [84] showed potential for increase in slow updrafts, but this result was later diminished by including new freezing rates computed by [92].

It generally seems to be accepted that Pinatubo probably increased homogeneous freezing, but this does not necessarily lead to substantial changes in cirrus. [103] suggested that homogeneous freezing is not usually limited by aerosol number concentration, so increasing aerosol number would not be expected to increase cirrus. They saw small increase in ice crystal number concentration, and no effect on cirrus cloud radiative effects, consistent with the [105] ISCCP results.

An interesting quote from [103]: "In order to investigate the combined effects of Pinatubo and ENSO on cirrus and potential feedbacks on climate, coupled simulations need to be carried out." Interestingly, we have exactly these kinds of simulations here, but given the small impact on cirrus CRE that the literature seems to suggest, any signal may be difficult to detect.

7.6.2. Changes in cirrus observed in CLDERA-E3SM simulations

In order to evaluate cirrus changes in a way consistent with observations, we derive high-topped cloud metrics from output from the ISCCP simulator [86, 189], which is embedded into E3SM via the Cloud Feedback Model Intercomparison Project Observational Simulator Package (COSP; [10]). The function of the ISCCP simulator is to mimic what ISCCP retrievals would produce, given the model-simulated atmosphere state at a given timestep. Statistics of cloud properties are aggregated in a similar way as the retrievals are to produce time-aggregated statistical summaries that are directly comparable between model and observation. These statistical summaries are output as joint histograms of cloud optical depth and cloud top pressure.

Figure 7-2 shows time and area-weighted global means of ISCCP-simulated joint histograms of cloud top pressure and optical depth from our early limvar simulations for the first five ensemble pairs of 10 Tg Pinatubo eruption simulations compared with their counterfactuals. Only the period from June 1991-December 1993 is shown. Each comparison shows the time-area-averaged joint histogram for each ensemble member pair, along with the difference between the Pinatubo and counterfactual simulations (the "impact analysis"). From the difference histograms, it is evident that high-topped cloud (bins with cloud top pressure ranges greater than 560 hPa in the joint histograms) increases in each ensemble member relative to its counterfactual, while mid and low-topped cloud (bins with cloud top pressure ranges below 560 hPa) decreases in each ensember member relative to the associated counterfactual. Based on this, we define a "high-topped cloud fraction" quantity that is the sum of all bins with cloud top pressure greater than 560 hPa so that we can more closely investigate the time evolution and spatial distribution of high-topped cloud and its changes in response to the eruption. Because the ISCCP/ISCCP-simulated retrievals are 1D, summing the bins of the joint histogram produce an estimate of cloud cover. Summing just the bins with cloud top pressure less than 560 hPa, for example, yields the fractional area covered by clouds with tops above 560 hPa.

The time series of this high-topped cloud estimate is shown in Figure 7-3 for the limvar ensembles. This shows a very clear increase in ISCCP-simulated high-topped cloud which sets up a few months post-eruption and remains for at least two years. Note that not only do each Pinatubo/counter-factual pair show an increase in Figure 7-3, but in fact there is almost no overlap in the extreme spread between the Pinatubo and counter-factual ensembles. This shows a robust increase in high-level cloud, consistent with the hypothesis that a Pinatubo-like eruption could increased cirrus.

7.6.3. Investigating the mechanisms for cirrus cloud changes

The hypothesized pathway for cirrus increase due to Pinatubo in the literature is through an increase in homogeneous freezing associated with an increase in sulfate. Figure 7-4 shows activated

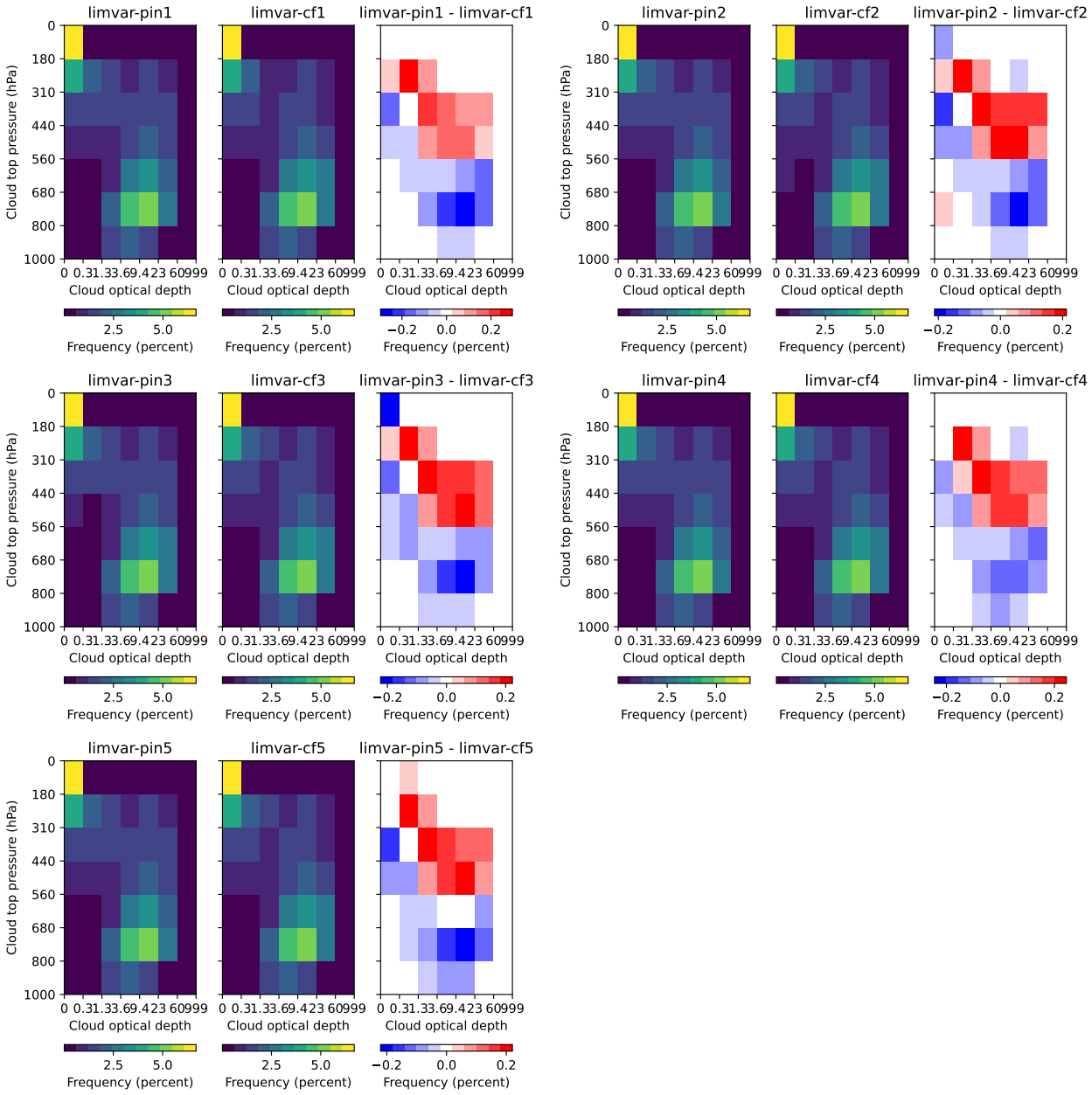


Figure 7-2. ISCCP-simulated area-weighted global mean joint histograms of cloud top pressure and cloud optical depth for June 1991 - December 1993 in limvar experiments.

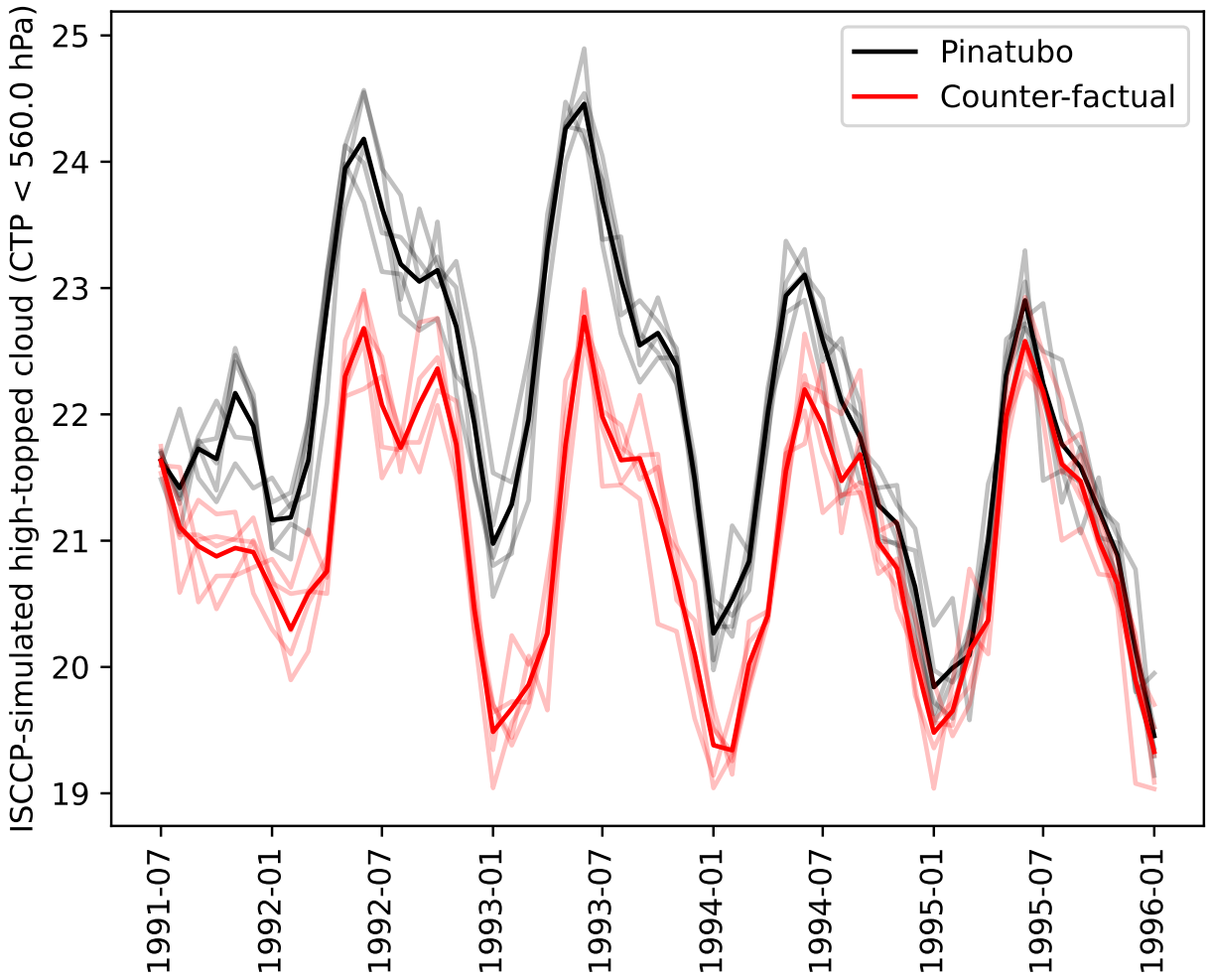


Figure 7-3. ISCCP-simulated cloud area for high-topped clouds (cloud top pressure less than 560 hPa) in limvar Pinatubo and counter-factual ensembles. Ensemble means are indicated by darker lines, while results for each ensemble member are shown in lighter colored lines.

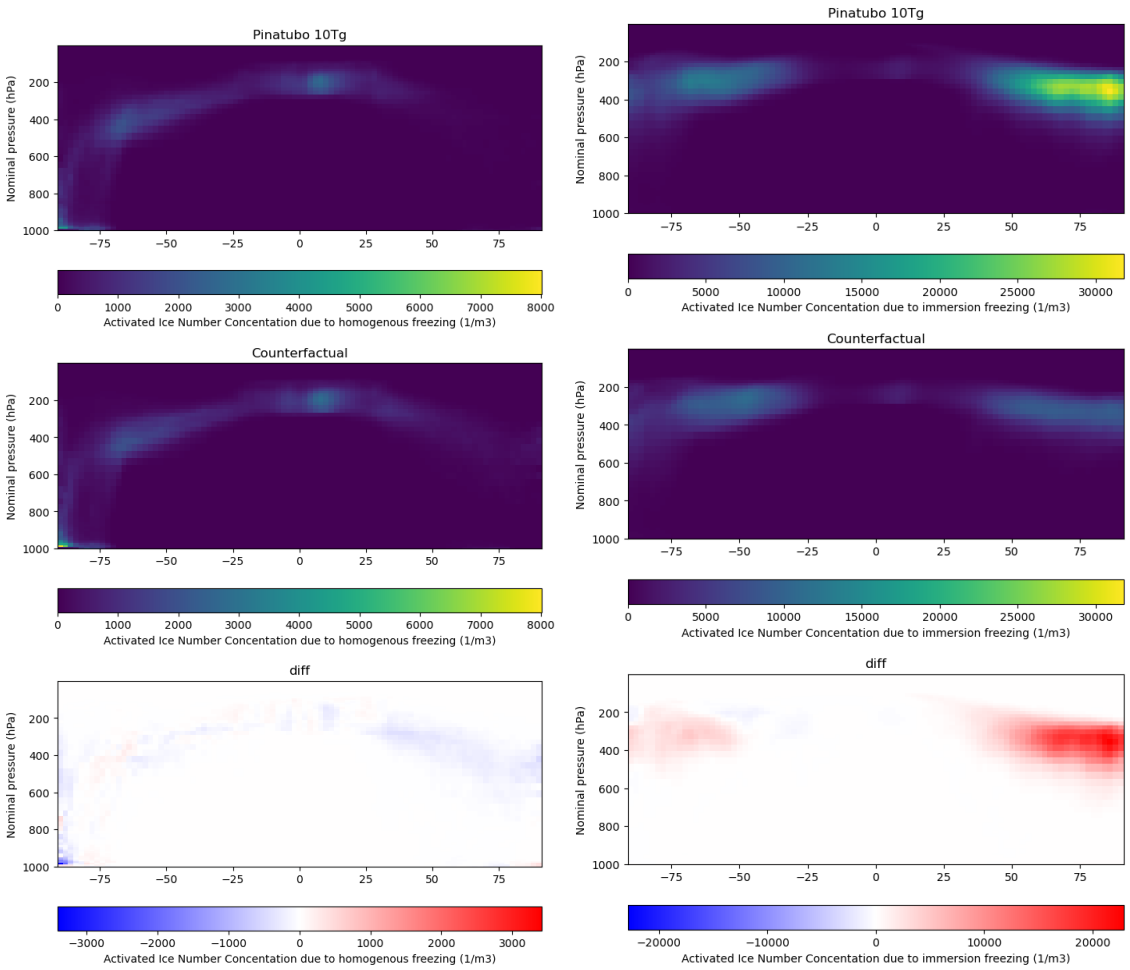


Figure 7-4. Changes in activated ice number concentration originating from homogeneous (left) and heterogeneous (right) freezing in E3SM limvar simulations.

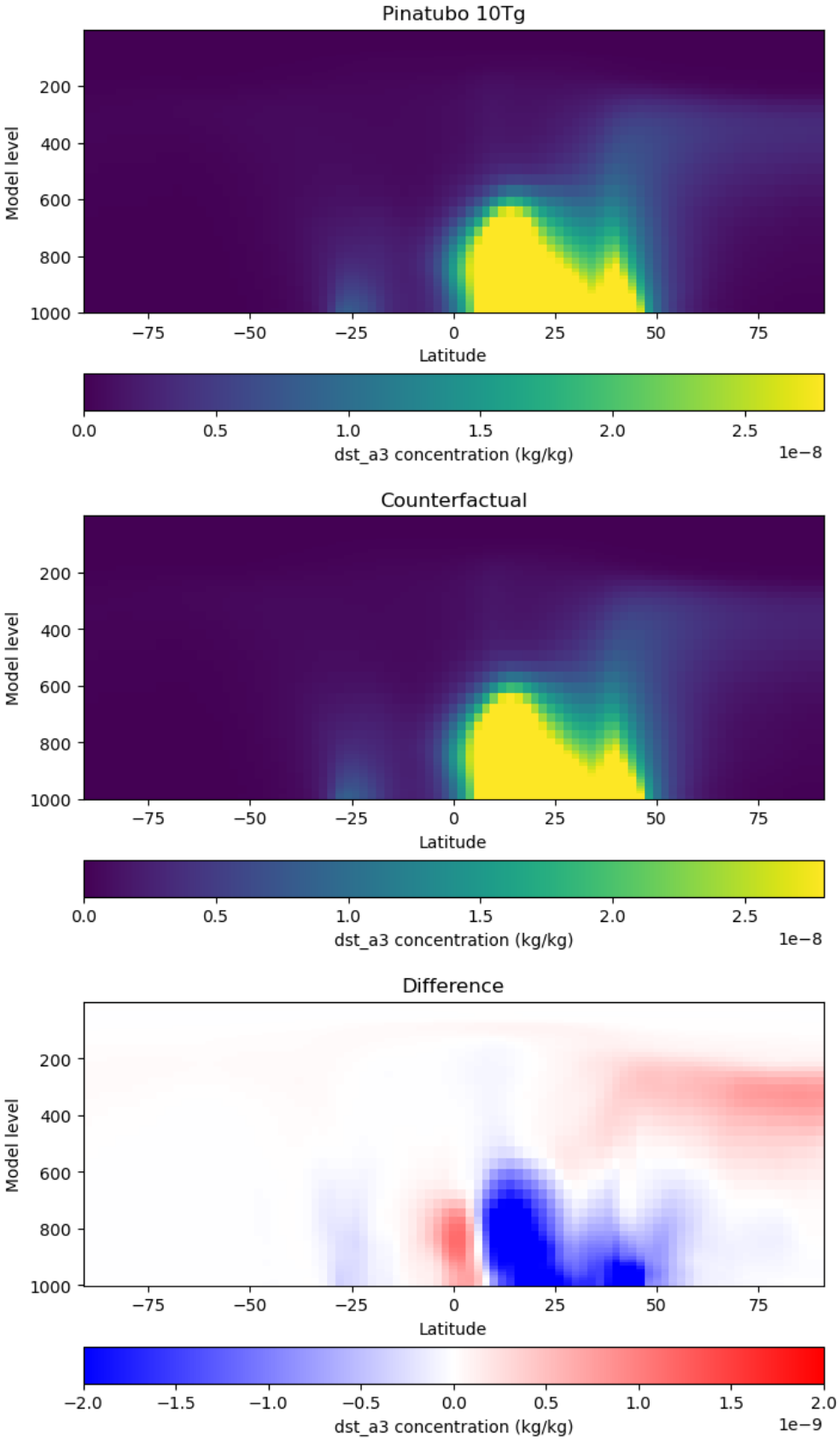


Figure 7-5. Changes in coarse-mode dust between counterfactual and Pinatubo 10 Tg injection simulations.

ice number concentration due to both homogeneous and heterogeneous (immersion) freezing. Interestingly, the increase in activated ice number concentration is seen not in the homogeneous freezing as proposed in the literature on Pinatubo, but rather in the heterogeneous (immersion) freezing. In E3SM, immersion freezing occurs only on the *coarse-mode* dust concentration. But the coarse-mode is treated collectively in MAM4; that is, there is no separate treatment for coarse-mode dust distinct from sulfate. It turns out then that in making the changes to allow for increased lifetime of sulfate aerosol in the stratosphere, coarse-mode dust in the troposphere is also affected. Thus, an increase in sulfate also inadvertently increases coarse-mode dust. This is demonstrated in Figure 7-5, which shows the changes in coarse-mode dust mass concentrations between the counterfactual and Pinatubo 10 Tg injection simulations. It is clear that there is a large increase in coarse-mode dust in the upper troposphere that is correlated with the increase in heterogeneous freezing seen in Figure 7-4.

7.6.4. Separating the impacts of dynamics and aerosol-cloud interaction on changes in cirrus amount

In this section, we explore the different drivers for cirrus formation and evaluate whether changes we see in cirrus due to Pinatubo are driven by aerosol-cloud interaction changes (ice nucleation changes due to presence of aerosol) or due to dynamical drivers in response to heating and circulation changes. In order to do this, a set of *nudged* Pinatubo and counterfactual simulations were performed:

1. Pinatubo free-running limvar ensemble member, writing out winds and temperature at 6-hourly frequency to be used for nudging
2. Counterfactual free-running limvar ensemble member
3. Counterfactual *nudged to Pinatubo U, V, and T*

With these simulations, we should be able to separately evaluate the direct impact of the aerosol changes on cirrus and the dynamical changes on cirrus (influenced due to changes in circulation and heating in response to changes in aerosol). These impacts can be diagnosed by taking the following differences (impact analyses):

$$
\begin{aligned}
X_{\mathrm{Pin}} - X_{\mathrm{CF}} &: \text{ Total Pinatubo impact} \\
X_{\mathrm{Pin}} - X_{\mathrm{CF-nudged}} &: \text{ Pinatubo aerosol impact} \\
X_{\mathrm{CF-nudged}} - X_{\mathrm{CF}} &: \text{ Pinatubo dynamical impact}
\end{aligned}
$$

where X_i is a given quantity of interest for case i. By nudging U, V, and T we hope to more tightly constrain the temperature and wind profiles, but recognize this might make for unrealistic simulation results due to over-constraining the model.

Figure 7-6 shows the total, aerosol, and dynamical impact on zonal means of cloud ice mass, ice number concentration, and ice number concentration due to heterogeneous and homogeneous freezing. There is a clear increase in cloud ice mass between 600 and 200 hPa in the northern hemisphere mid to high latitudes (top panel). The dynamical and aerosol-cloud interaction signals

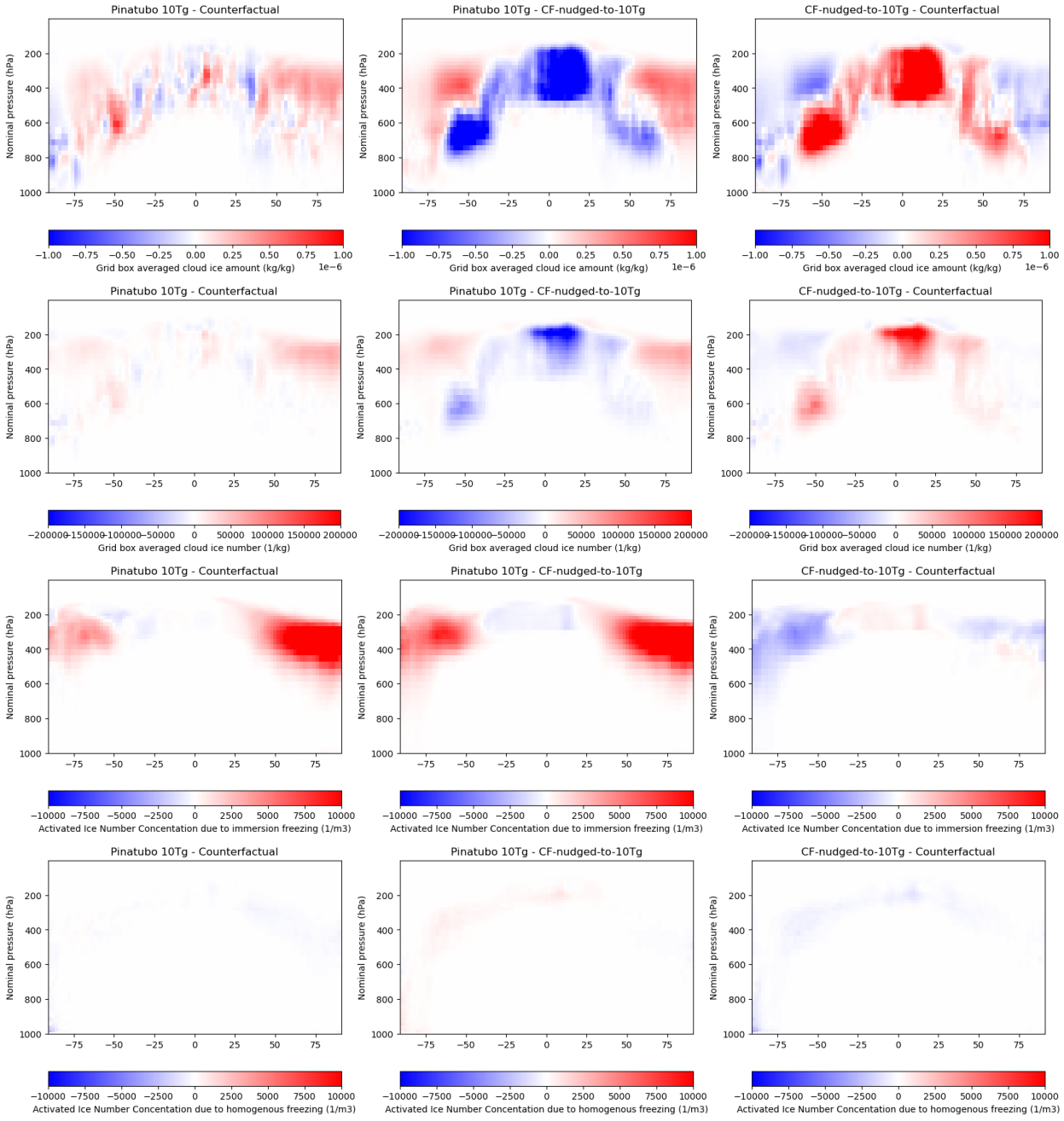


Figure 7-6. From top to bottom, differences in ice mass concentration, ice number concentration, activated ice number concentration due to immersion freezing, and activated ice number concentration due to homogeneous freezing in nudged simulations showing the total Pinatubo impact (left), the aerosol-cloud interaction impact (middle), and the dynamical impact (right).

oppose each other, but the increase in cloud ice mass appears to be dominated by the increase in the aerosol-cloud interaction response (top-center panel). This is also true for the total ice number concentration in high latitudes (north of 50 N), where the increase in upper levels is dominated by an increase in the aerosol-cloud response. The responses in number concentration due to immersion freezing makes it abundantly clear that this is driven by the changes to the aerosol: the aerosol-cloud interaction and dynamical responses oppose one another, but clearly the increase in immersion freezing due to the aerosol-cloud interaction response dominates, with the result being an overall increase in immersion freezing in the total Pinatubo response. These results suggest that while dynamical changes could be playing a role, the increase in high-level clouds in the Pinatubo limvar ensemble appear to be primarily driven by changes in the aerosol-cloud interactions due to the changes in dust. The changes in homogeneous freezing are small in magnitude compared to the changes in heterogeneous freezing (bottom row). This suggests that rather than converting homogeneous freezing to heterogeneous, the increased ice nucleation is perhaps occurring where heterogeneous freezing is already dominant.

7.6.5. Implications for cirrus cloud thinning

While the inadvertent increase to dust concentrations makes it difficult to evaluate realistic impacts on cirrus due to the Pinatubo eruption, it does provide an interesting opportunity to look for evidence of cirrus cloud thinning in E3SM. Cirrus cloud thinning (CCT) is a proposed method for climate intervention. The approach would be to inject a source of heterogeneous nucleation (such as mineral dust) into regions in the atmosphere that are dominated by homogeneous nucleation to shift the dominate source of ice nucleation from homogeneous to heterogeneous freezing. In doing so, larger ice crystals are formed that are a) optically thinner, thereby trapping less outgoing longwave radiation, and b) tend to sediment or precipitate out of suspension faster, thus shortening the lifetime of cirrus clouds. Both of these impacts act to cool the climate system by reducing the longwave heating provided by cirrus.

Figure 7-4 shows a clear increase in heterogeneous freezing and a decrease in homogeneous freezing in the Pinatubo simulations relative to the counterfactual. This is consistent with the increase in coarse-mode dust in the Pinatubo simulation discussed in the previous section. However, it is also evident from Figure 7-3 that cirrus amount tends to *increase* rather than decrease in the Pinatubo simulations. Thus, the first part of the cirrus cloud thinning pathway appears to be consistent with theory (increasing dust leads to increased heterogeneous freezing), but this is not leading to a decrease in cirrus.

It is possible that the change in coarse-mode dust in the 10 Tg eruption simulations is just already too large, and that cirrus are effectively being "over-seeded", which would lead to an increase rather than a decrease in cirrus [173]. In an attempt to evaluate the sensitivity of cirrus to dust seeding, we look at the varying eruption magnitude ensembles. Because we are ultimately interested in changes in the outgoing longwave that result from changes in cirrus, we look at both high-topped cloud fraction (derived from the ISCCP simulator) and the longwave cloud radiative effect (the difference between all-sky and clear-sky outgoing longwave radiative fluxes). These are shown in Figure 7-7 and Figure 7-8, respectively. While there is considerable spread in individual ensemble members, there does not appear to be any evidence for either reduction in cirrus amount,

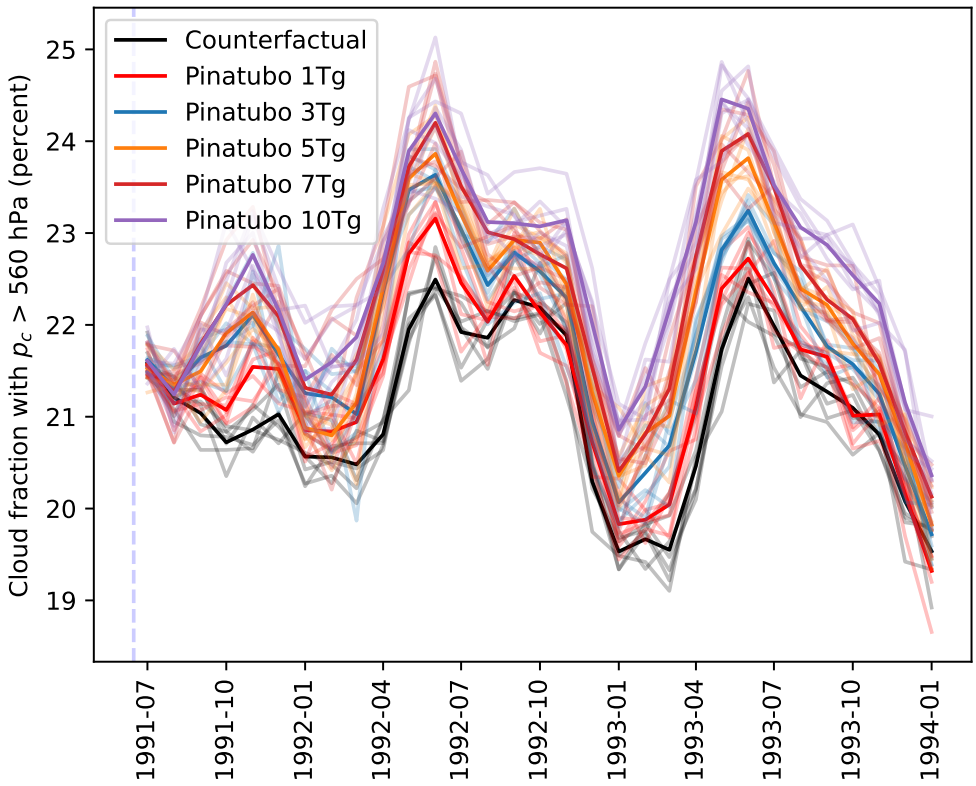


Figure 7-7. High-topped cloud fraction derived from ISCCP simulator in limvar varying eruption magnitude ensembles.

or reduction in longwave cloud radiative effect in these figures even down to the 1 Tg eruption magnitude ensembles. This suggests that the changes in coarse-mode dust at the 1 Tg injection level are already too large to see a cirrus thinning response, and instead the impact is primarily an increase in cirrus amount and increase in longwave heating.

The fact that at 1 Tg we are already seeing a cirrus increase and consequent longwave warming signal makes it difficult to use these simulations as-is for an analysis of cirrus cloud thinning. It does however highlight the delicate balance needed to obtain the desired impact from cloud seeding. This may also be another artifact of the changes made to MAM4 to support longer sulfate lifetimes in the stratosphere. The development of MAM5, with a fifth mode added to better account for this, should enable better representation of the relevant processes without unrealistically shifting the nature of the freezing processes important for cirrus formation and maintenance.

7.7. Pinatubo Impacts on South Asian Monsoon

Contributing Authors: Thomas Ehrmann (8931), Daniel Ries (5574), Kelly McClernon (5573), Katherine Goode (5573), Diana Bull (8931).

The South Asian Summer Monsoon (SASM) is generally considered to be between June and September and accounts for as much as 90% of the annual rainfall in western and central India

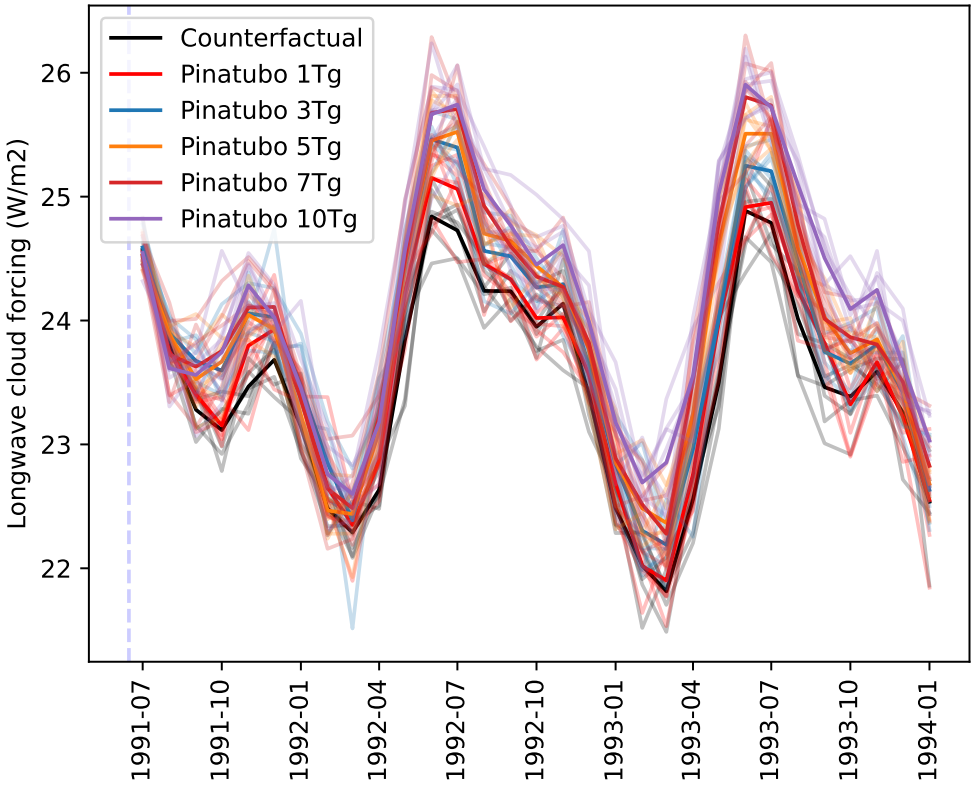


Figure 7-8. Longwave cloud radiative effect in limvar varying eruption magnitude ensembles.

[63]. For most of the last 300 years SASM has been explained as a gigantic land-sea breeze [62, 190, 117]. This theory postulates that, because the specific heat over the ocean is much greater than over the land, as the radiative forcing of the early summer increases, the temperatures of the land heat up faster than over the ocean driving a large-scale thermal wind system. While there is an intuitiveness to this theory, in recent years a consensus has been developing around an alternative theory. The emergent theory is that the SASM, along with all other tropical and subtropical monsoons, is actually driven by the seasonal migration of the intertropical convergence zone (ITCZ) [40, 41, 46]. The ITCZ is generally considered to lie in the ascendent branch of the Hadley Cell and is characterized by enhanced precipitation and a local reduction in out-going longwave radiation due to increased cloud cover [184, 158].

While the exact mechanism is still under debate, most publications believe that large volcanic forcing in the northern hemisphere tropics is likely to lead to a decrease in precipitation in the SASM [110, 80]. Our data also exhibited a noticeable area of significant precipitation reduction as a result of the Pinatubo eruption, especially around the central western coast of India in July 1992. Pinatubo could affect either of the proposed drivers of SASM above and multiple publications have proposed both [26, 207, 148, 108].

We believe that this problem can be at least partially addressed using spatial-temporal zeroed feature importance (stZFI) with an echo state network (ESN) [145, 57]. Our hypothesis is that the ESN will be able to weigh the competing mechanism and stZFI will identify which mechanisms are the most important for driving the Pinatubo impact on SASM precipitation over the landmass of India (IPCC WR6 region SAS) [82]. This is the first application of stZFI to regionally defined variables, particularly where the spatial domain of the output is distinct from that of most of the inputs.

This work explored the drivers of the Indian Monsoon with and without forcings due to the 1991 Mount Pinatubo eruption by investigating the relative role of two of these mechanisms (ITCZ and land-ocean temperature contrast) using regionally refined feature importance with an echo state network (ESN), stZFI [145, 57, 113]. We found that the ESN possesses improved predictive power over baseline models and that temperature as well as radiative indices were important to this predictive power. However, we were unable to uncover conclusive evidence of the underlying mechanism.

This work is documented in following paper [146]:

- Daniel Ries, Kellie McClernon, Thomas Ehrmann, and Diana Bull. "Using Spatio-Temporal Feature Importance to Identify Drivers of Indian Monsoon after the Mount Pinatubo Eruption." SAND2024-12581O. September 2024, Sandia National Laboratories, Albuquerque NM.

7.8. Pinatubo Impacts on Agricultural Productivity

Contributing Authors: Dan Krofcheck (5522), Justin Li (8732), Christopher Wentland (8734), Diana Bull (8931), Joe Crockett (8931) Laura Swiler (1400), Kara Peterson (1442), Michael Weylandt (5573), Ram Singh, Kostas Tsigaridis, and Kate Marvel (Columbia University).

The impacts of Mt. Pinatubo on agricultural productivity are still debated. They involve several interacting or confounding effects, rely on changes to both light-based drivers and water-based drivers of plant productivity, and exhibit spatially varying impacts. For this reason, Pathway 2 (P2) in CLDERA (identification of pathways from Mt. Pinatubo to changes in plant productivity) has been challenging.

7.8.1. Soil-moisture and evapotranspiration deficit indices

We present two lines of research developed in CLDERA. The first one represents a collaboration with Ram Singh, Kostas Tsigaridis, and Kate Marvel. This research focused on how the Mt. Pinatubo eruption affected the hydroclimatic conditions and water-based drivers of plant productivity. We used NASA's Earth system model (the Goddard Institute for Space Studies GISS model) for modeling the eruption and detection of hydroclimate response.

We investigated changes in agricultural drought indices from Pinatubo by considering the store (soil moisture) and flux (evapotranspiration) of water as potential short-term controls over productivity in particular regions. We used NASA's GISS model with interactive aerosol chemistry to conduct the simulation experiments consistent with the counterfactual inference of causation approach for the Pinatubo eruption. The Pinatubo effect in the model-simulated climate is evaluated through the various pathways of climate impacts, from the primary dependent variables to the higher order responses controlling plant productivity. Considering the complexity of modeling the terrestrial system, vegetation demographics, and physiological characteristics, we used the soil moisture and evapotranspiration-based agricultural drought indices SMDI (soil moisture deficit index) and ETDI (evapotranspiration deficit index) to account for agricultural productivity. We evaluated short-term (weekly) and long-term (seasonal) scale changes in SMDI and ETDI relative to statistics over a longer modern time-period. We also evaluated changes in actual and potential evapotranspiration, precipitation, transpiration, and surface temperature.

Using these metrics, we concluded that approximately 10-15% of land region shows statistically significant dry or wet patterns in the volcanically perturbed climate conditions for 1992 and 1993. In equatorial Africa, decreases in both SMDI and ETDI indicated that there was likely a negative impact on plant productivity while a Middle East region showed increases in SMDI and ETDI indicating a positive impact on plant productivity. Northern Asia in comparison exhibited an increase in SMDI with a decrease in ETDI indicating that plant productivity likely decreased, but not because of water-based drivers. We demonstrated that there is an excess of root-zone soil moisture in high latitudes which is not being utilized by plants to grow; it is likely that the more dominant drivers in the high latitude regions are temperature and radiation-based drivers. The intricate nature of the compounded response, particularly regarding the soil moisture-based impact pathways in tropical regions and higher latitudes across the northern hemisphere, underscores the necessity of broadening the scope of the investigation beyond soil moisture and land-atmosphere interactions.

This work is documented in following paper [165]:

- Ram Singh, Kostas Tsigaridis, Diana Bull, Laura P Swiler, Benjamin M Wagman, and Kate Marvel. Pinatubo's effect on the moisture-based drivers of plant productivity. Submitted

to *Atmospheric Chemistry and Physics*, July 2024. The review preprint is available at: https://egusphere.copernicus.org/preprints/2024/egusphere-2024-2280/

7.8.2. Attributing Mt. Pinatubo in spatial patterns of vegetation response

Volcanic eruptions induce significant perturbations within the climate system cascading into consequences for quantities as far downstream as terrestrial ecosystems. Understanding the complex and coupled interactions driving changes in vegetation function and productivity outside of natural variability in response to a volcanic eruption is extremely difficult; not only are the drivers (energy and water) of vegetative function differential and heterogeneous, but so too are the responses in distinct plant representations across the Earth's surface. This heterogeneity has the potential to mask underlying relationships by which a pathway becomes expressed (especially at low magnitudes of primary forcing).

In this research we seek to amplify and align proposed pathways to determine if they can be reliably detected within spatially distributed data. This requires filtering the data for only the growing season responses. Further it requires co-locating grid cells with strong and weak pathways (i.e. a large reduction in downwelling shortwave radiation resulting in a strong reduction in photosynthesis). With analyses proving that strong relationships, for instance, are the most prevalent, we can then employ these grid cells in a multi-step conditional attribution framework [191]. This analysis would be the first to bring probabilistic attribution statements of the size of the eruption to vegetative impacts.

This research is ongoing with the following expected publication [94]

- Li, J., Wentland, C., Swiler, L, Peterson, K., Weylandt, M., Krofcheck, D., Wagman, B., Bull, D. Attributing Mt. Pinatubo in spatial patterns of vegetation response. To be submitted November 2024

7.8.3. Understanding hydroclimatic shifts through a Budyko framing

The eruption of Mt. Pinatubo and similar volcanic eruptions force significant departures in global incoming/outgoing radiation as well as the provision of precipitation. By placing these impacts into a framework that incorporates both water and energy balances, we can examine the ecohydrological effects of this eruption that may have implications for long-term changes in terrestrial function. The Budyko framework, a conceptual model used to understand the relationship between climate and the partitioning of precipitation into runoff and evapotranspiration, accomplishes this goal. We leverage this framework by examining how the eruption altered an evaporative ratio (actual evapotranspiration/ potential evapotranspiration) and a wetness index (precipitation / potential evapotranspiration) between factual and counterfactual E3SM models. A joint change in the evaporative ratio and the wetness index, known as movement in the Budyko space, can signify changing water resources, shifts in vegetation composition, river basin specific changes due to climate change, agricultural expansion or contraction or other departures from regional ecological stability.

Our research focuses on the impact of Mt. Pinatubo in this ecohydrological context. Specifically, we hypothesize that the Pinatubo eruption altered the energy and water balance of areas of the globe differentially, and in some cases inverted the historical constraints on the hydrologic cycle. This would suggest that the eruption may have a bifurcating consequence on terrestrial function, the directionality of which is a function of the initial state of the landscape and the local impacts of the aerosol injections. This research analyzes movement within the Budyko space in response to Mt. Pinatubo, evaluates time dependencies, and consider co-occurring environmental states with these movements to construct an understanding of the ecohydrological effects from Mt. Pinatubo.

This research is ongoing with the following expected publication [89]

- Krofcheck, D., Crockett, J., Peterson, K., Li, J., Bull, D. Volcanic stratospheric aerosol injection may buffer against global ecohydrological impacts of hotter and drier climates. To be submitted December 2024

8. CONCLUSION

In conclusion, the CLDERA project enabled multi-step attribution in the climate through development of new methods that quantitatively assess pathways between a climate forcing and its downstream impacts. To do this there have been three major technical outcomes.

CLDERA Developed Stratospheric Expertise and Aerosol Modeling Capabilities in E3SM:

Under CLDERA, a stratospheric prognostic aerosol capability in E3SMv2-SPA was developed which allows for evolution of volcanic aerosols as well as tagging and diagnostics of various types of aerosols. We performed a massive E3SMv2-SPA simulation campaign, generating 400+ TB of data encompassing source magnitude varying ensembles representing 1750 simulated years. We uncovered and communicated significant biases in E3SM's stratosphere and studied how volcanic eruptions can change circulation.

CLDERA Created Original Methods to Detect and Model Pathways from Source-to-Impact:

We extended detection methods for spatio-temporally evolving changes to better understand where and when changes occur as well as their start and end. CLDERA has developed state-of-the-art machine learning, reduced order modeling, and statistical tools to elucidate and represent pathways, which are a foundational means for understanding dependent causal-like relationships in many complex nonlinear systems. These eight methods have been described in detail in this report and encompass both exploratory and confirmatory approaches. The breadth of methods (multivariate, multi-variable, in situ, data driven, etc.) explored is part of the strength of CLDERA enabling future method applications that are most suited to the method approach.

CLDERA Advanced Climate Attribution through Novel Methods, Cases, and Approaches:

CLDERA has tackled and been successful in developing methods to determine forcing magnitude given an impact. Inversion approaches of this type have been avoided due to the complexity of the climate system, however under CLDERA two approaches have been successfully demonstrated. We also developed experiments well suited to attribution frameworks in order to expand the cases and approaches by which we pursue attribution in the climate system. We employed pattern-scaling to identify relationships that are robust to variability in the system in the same vein as epidemiological dose-response studies. We also constrained natural variability in the simulations to sharpen impacts. By initializing fully-coupled simulation to the historical ENSO3.4 and QBO characteristics, we drew upon the heritage of storylines in the extreme weather event community and constrained our variability in the early months making our results conditional upon those natural modes.

Further, CLDERA developed a tiered verification process that can be utilized in other domains. Simplified and controlled data sets were developed with key characteristics of the multi-step pathways. These datasets were used to prototype, verify, and refine the CLDERA methods.

CLDERA increased Sandia's footprint in the climate analytics community and developed new climate collaborations. CLDERA created a cadre of climate analysts, representing a significant workforce development effort.

New methods and tools developed through CLDERA offer a framework that can be translated to other localized or episodic sources such as large wildfires, changes to ocean currents (e.g., Atlantic Meridional Overturning Circulation), or climate interventions (e.g., stratospheric aerosol injection). The CLDERA methods could also be extended to understand and attribute impacts from tipping points in the climate system. In the future, CLDERA methods could be used to advance climate science and analyze climate impacts. CLDERA methods can provide decision support using ESMs, potentially helping inform policies and regulations as well as requirements on monitoring systems.

The products from CLDERA have been extensive with a total of 9 journal articles published, 12 articles submitted and under review, and an additional 8 articles in preparation[1]. We have produced 1750 simulated years[2] and developed 9 code-bases. Additionally, CLDERA staff have given 79 conference presentations and posters and 17 invited talks. The following sections highlight a concise list of the CLDERA products.

8.1. CLDERA Products

8.1.1. Published Articles

1. I. Tezaur, K. Peterson, A. Powell, J. Jakeman, E. Roesler. "Global sensitivity analysis using the ultra-low resolution Energy Exascale Earth System Model," *Journal of Advances in Modeling Earth Systems*, 14, e2021MS002831, 2022 [175]

2. Joseph Hart, Mamikon Gulian, Indu Manickam, and Laura P. Swiler. Solving high-dimensional inverse problems with auxiliary uncertainty via operator learning with limited data. *Journal of Machine Learning for Modeling and Computing*, Vol.4, No. 2, pp. 105-133. 2023. [65]

3. Tucker, J. D., & Yarger, D. (2023). Elastic functional changepoint detection of climate impacts from localized sources. Environmetrics, e2826. `https://doi.org/10.1002/env.2826` [182]

4. Michael Weylandt and Laura P. Swiler. Beyond PCA: Additional Dimension Reduction Techniques to Consider in the Development of Climate Fingerprints. *Journal of Climate*, Vol. 37, Issue 5, pp. 1723-1735. 2024. [193]

[1] Our website `https://www.sandia.gov/cldera/` will continue to be updated with publication information

[2] Data from the full E3SMv2-SPA simulation campaign including pre-industrial control, historical, and Mt. Pinatubo ensembles will be hosted at Sandia National Laboratories with location and download instructions announced on `https://www.sandia.gov/cldera/e3sm-simulations-data/` when available.

5. McClernon, Kellie. Goode, Katherine. Ries, Daniel. "A comparison of model validation approaches for echo state networks using climate model replicates". *Spatial Statistics*. 2024. https://doi.org/10.1016/j.spasta.2024.100813 [112]

6. Brown, H. Y., Wagman, B., Bull, D., Peterson, K., Hillman, B., Liu, X., Ke, Z., and Lin, L. "Validating a microphysical prognostic stratospheric aerosol implementation in E3SMv2 using observations after the Mount Pinatubo eruption." *Geoscientific Model Development*, 17, 5087–5121, https://doi.org/10.5194/gmd-17-5087-2024, 2024. [13]

7. Joseph P. Hollowed, Christiane Jablonowski, Hunter Y. Brown, Benjamin R. Hillman, Diana L. Bull, and Joseph L. Hart. "HSW-V v1.0: localized injections of interactive volcanic aerosols and their climate impacts in a simple general circulation model." *Geoscientific Model Development*, 17, 5913–5938, https://doi.org/10.5194/gmd-17-5913-2024, 2024. [77]

8. K. Goode, D. Ries, and K. McClernon, "Characterizing climate pathways using feature importance on echo state networks", *Stat. Anal. Data Min.: ASA Data Sci. J.* 17 (2024), e11706. https://doi.org/10.1002/sam.11706. [57]

9. Samantha Shi-Jun, Lyndsay Shand, Bo Li. "Tracing the impacts of Mount Pinatubo eruption on global climate using spatially-varying changepoint detection." Accepted *Annals of Applied Statistics*. September 2024. This manuscript is available on arXiv: https://arxiv.org/abs/2409.08908 [164]

8.1.2. Submitted Articles

1. Jake Nichol, Michael Weylandt, G. Matthew Fricke, Melanie E. Moses, Diana Bull, Laura P. Swiler. Causal Space-Time Stencil Learning: Local Causal Dynamics in Complex Systems. Submitted to *Journal of Geophysical Research, Machine Learning and Computation* July 2024. [124]

2. Ram Singh, Kostas Tsigaridis, Diana Bull, Laura P Swiler, Benjamin M Wagman, and Kate Marvel. Pinatubo's effect on the moisture-based drivers of plant productivity. Submitted to *Atmospheric Chemistry and Physics*, July 2024. The review preprint is available at: https://egusphere.copernicus.org/preprints/2024/egusphere-2024-2280/ [165]

3. Daniel Ries, Katherine Goode, Kellie McClernon, and Ben Hillman. "Using feature importance as exploratory data analysis tool on earth system models." Submitted to *GMD Special Issue: Theoretical and computational aspects of ensemble design, implementation, and interpretation in climate science*, July 2024. The review preprint is available at: https://gmd.copernicus.org/preprints/gmd-2024-133/ [145]

4. Hu, A., Liu, X, Ke, Z., Wagman, B., Brown, H., Lu, Z., Bull, D., Peterson, K. Size-resolved process understanding of stratospheric sulfate aerosol following the Pinatubo eruption, Submitted July 2024. *Atmospheric Chemistry and Physics*. [79]

5. Drew Yarger and J. Derek Tucker. "Detecting changepoints in globally-indexed functional time series." Submitted to *Evironmetrics*, August 2024. This manuscript is available on arXiv: https://arxiv.org/abs/2308.05915. [204]

6. Andrew Steyer, Luca Bertagna, Graham Harper, Jerry Watkins, Irina Tezaur, and Diana Bull. "In-situ data extraction for pathway analysis in an idealized atmosphere configuration of E3SM." Submitted to *Computing in Science & Engineering*, August 2024. [172]

7. Christopher R. Wentland, Michael Weylandt, Laura P. Swiler, Thomas S. Ehrmann, Diana Bull. "Conditional multi-step attribution for climate forcings." Submitted to *Journal of Climate*, Aug. 2024. This manuscript is available on arXiv: https://doi.org/10.48550/arXiv.2409.01396. [191]

8. Garrett, R., Shand, L. and Huerta, J. G.. A Multivariate Space-Time Dynamic Model for Characterizing the Atmospheric Impacts Following the Mt. Pinatubo Eruption. Submitted to *Environmetrics*. Aug 2024. [45]

9. Joseph Hart, Indu Manickam, Mamikon Gulian, Laura Swiler, Diana Bull, Thomas Ehrmann, Hunter Brown, Benj Wagman, and Jerry Watkins. Stratospheric aerosol source inversion: Noise, variability, and uncertainty quantification, submitted to *Journal of Machine Learning for Modeling and Computing*. This manuscript is available on arXiv, https://doi.org/10.48550/arXiv.2409.06846. [66]

10. Thomas Ehrmann, Benjamin Wagman, Diana Bull, Benjamin Hillman, Joseph Hollowed, Hunter Brown, Kara Peterson, Laura Swiler, Jerry Watkins, and Joseph Hart. "Identifying Northern Hemisphere Temperature Responses to the Mt. Pinatubo Eruption through Limited Variability Ensembles." Submitted to *Climate Dynamics*, September 2024. [36]

11. Jerry Watkins, Luca Bertagna, Graham Harper,Andrew Steyer, Irina Tezaur, and Diana Bull. "Entropy-based feature selection for capturing impacts in Earth system models with extreme forcing." Submitted to *Journal of Computational and Applied Mathematics*, September 2024. [187]

12. Meredith G. L. Brown, Matt Peterson, Irina Tezaur, Kara Peterson, and Diana Bull. "Random Forest Regression Feature Importance for Climate Impact Pathway Detection." Submitted to *Journal of Computational and Applied Mathematics*, September 2024. [16]

8.1.3. Articles in preparation to be submitted

1. Warren L. Davis, Max Carlson, Irina Tezaur, and Diana Bull. "Spatio-temporal Multivariate Cluster Evolution Analysis for Detecting and Tracking Climate Impacts." Intended submission to *Journal of Computational and Applied Mathematics*, September 2024. [31]

2. Justin D. Li, Audrey McCombs, Daniel Dunlavy, Gabriel Huerta, Lyndsay Shand, "Exploring Multivariate Tensor Data Fusion and Completion." To be submitted to *Special issue of Remote Sensing: Spatiotemporal Fusion of Multi-Source Remote Sensing Data and Its Applications.* October 2024. [95]

3. Hollowed, J., Jablonowski, C., Ehrmann, T., Hillman, B., ... "Volcanic Aerosol Modification of the Stratospheric Circulation in E3SMv2 Part I: Wave-Mean Flow Interaction." To be submitted to *Climatic Dynamics* by October 2024. [76]

4. Li, J., Wentland, C., Swiler, L, Peterson, K., Weylandt, M., Krofcheck, D., Wagman, B., Bull, D. "Attributing Mt. Pinatubo in spatial patterns of vegetation response." To be submitted November 2024. [94]

5. Krofcheck, D., Crockett, J., Peterson, K., Li, J., Bull, D. "Volcanic stratospheric aerosol injection may buffer against global ecohydrological impacts of hotter and drier climates." To be submitted December 2024. [89]

6. Jablonowski, C., Hollowed, J., Nguyen, L., Hillman, B., Ehrmann, T., Wagman, B. "Stratospheric Characteristics in E3SMv2." To be submitted by December 2024. [83]

7. Hollowed, J., Jablonowski, C., Ehrmann, T., Hillman, B., ... "Volcanic Aerosol Modification of the Stratospheric Circulation in E3SMv2 Part II: Tracer Sensitivity." To be submitted by December 2024. [75]

8. Hu, A., Liu, X, Hillman, B., Wagman, B., Bull, D., Peterson, K. "Aerosol Indirect Effect: a study of cirrus implications with MAM5." To be submitted by December 2024. [78]

8.1.4. SAND Reports

1. J. Jake Nichol, Michael Weylandt, Mark Smith, Laura Swiler. "Benchmarking the PCMCI Causal Discovery Algorithm for Spatiotemporal Systems." Technical Report. SAND2023-05141. June 2023, Sandia National Laboratories, Albuquerque NM. [125]

2. Thomas Ehrmann, Benjamin Wagman, Diana Bull, Benjamin Hillman, Joseph Hollowed, Hunter Brown, Kara Peterson, Laura Swiler, Jerry Watkins, and Joseph Hart. "Identifying Northern Hemisphere Stratospheric and Surface Temperature Responses to the Mt. Pinatubo Eruption within E3SMv2-SPA." Technical Report. SAND2024-12730. September 2024, Sandia National Laboratories, Albuquerque NM. [35]

3. Carole Hall, J. Derek Tucker, and Drew Yarger. "Elastic Changepoint Detection for Globally-indexed Functional Time Series Data with Climate Applications." Technical Report. SAND2024-12470. September 2024, Sandia National Laboratories, Albuquerque NM. [61]

4. Daniel Ries, Kellie McClernon, Thomas Ehrmann, and Diana Bull. "Using Spatio-Temporal Feature Importance to Identify Drivers of Indian Monsoon after the Mount Pinatubo Eruption." SAND2024-12581o. September 2024, Sandia National Laboratories, Albuquerque NM. [146]

5. Benjamin M. Wagman, Hunter Brown, Diana Bull, Tom Ehrmann, Ben Hillman, Laura Swiler, Kara Peterson, Joe Hollowed. "Does prognostic volcanic aerosol enhance climate variability in Earth System Model simulations of volcanic eruptions?" SAND2024-12725o. September 2024, Sandia National Laboratories, Albuquerque NM. [183]

6. McClernon, Kellie, Ries, Daniel, and Goode, Katherine. "Verifying regional feature importance on echo state networks for climate pathways." SAND2024-12404o. September 2024, Sandia National Laboratories, Albuquerque NM. [113]

8.1.5. Codes & Data

- **CLDERA Simulation campaign:** Data from the full E3SMv2-SPA simulation campaign including pre-industrial control, historical, and Mount Pinatubo ensembles will be hosted at Sandia National Laboratories with location and download instructions announced on `https://www.sandia.gov/cldera/e3sm-simulations-data/` when available
- **Prognostic aerosol code base:** E3SMv2-SPA at `https://github.com/sandialabs/CLDERA-E3SM`
- **Profiling in-situ co-analysis:** CLDERA-Tools at `https://github.com/sandialabs/cldera-tools`
- **Inverse optimization with operator neural network:** SAILo is currently undergoing copyright assertion and is only available on SNL internal resources at gitlab-ex repo CLDERA/sailo
- **Conditional Multivariate Attribution:** at `https://github.com/sandialabs/conditional-multistep-attribution`
- **Pair-wise RFR:** currently undergoing copyright assertion and is only available on SNL internal resources at gitlab repo mgpeter/RFR-CLDERA
- **Explainability for Echo State Networks:** listner at `https://github.com/sandialabs/listenr`
- **Multivariate space-time dynamic model:** MV-STDM at `github.com/garretrc/MV-STDM`
- **Elastic Functional Data Analysis:** fdasrvf_python/r at `https://pypi.org/project/fdasrsf` or `https://cran.r-project.org/web/packages/fdasrvf`
- **Causal Space-Time Stencil Learning:** CaStLe at `https://github.com/jjakenichol/CaStLe`

REFERENCES

[1] Climate change: The ipcc scientific assessment. Technical report, Intergovernmental Panel on Climate Change (IPCC), Assessment Report 1, 1990.

[2] Marta Abalos, William J. Randel, Douglas E. Kinnison, and Rolando R. Garcia. Using the Artificial Tracer e90 to Examine Present and Future UTLS Tracer Transport in WACCM. *Journal of the Atmospheric Sciences*, 74(10):3383–3403, 2017.

[3] James F. Allen. *Natural language processing*, page 1218–1222. John Wiley and Sons Ltd., GBR, 2003.

[4] Myles R. Allen and Peter A. Stott. Estimating signal amplitudes in optimal fingerprinting, part I: Theory. *Climate Dynamics*, 21(5):477–491, 2003.

[5] N. W. Arnell, J. A. Lowe, A. J. Challinor, and T. J. Osborn. Global and regional impacts of climate change at different levels of global temperature increase. *Climatic Change*, 155:377–391, 2019.

[6] Fabian E Bachl, Finn Lindgren, David L Borchers, and Janine B Illian. inlabru: an r package for bayesian spatial modelling from ecological survey data. *Methods in Ecology and Evolution*, 10(6):760–766, 2019.

[7] Elizabeth A Barnes, Susan Solomon, and Lorenzo M Polvani. Robust wind and precipitation responses to the mount pinatubo eruption, as simulated in the cmip5 models. *Journal of Climate*, 29(13):4763–4778, 2016.

[8] L. Mark Berliner, Richard A. Levine, and Dennis J. Shea. Bayesian climate change assessment. *Journal of Climate*, 13(21):3805–3820, 2000.

[9] Roger S Bivand, Edzer J Pebesma, Virgilio Gómez-Rubio, and Edzer Jan Pebesma. *Applied spatial data analysis with R*, volume 747248717. Springer, 2008.

[10] A. Bodas-Salcedo, M. J. Webb, S. Bony, H. Chepfer, J.-L. Dufresne, S. A. Klein, Y. Zhang, R. Marchand, J. M. Haynes, R. Pincus, and V. O. John. Cosp: Satellite simulation software for model assessment. *Bulletin of the American Meteorological Society*, 92(8):1023 – 1043, 2011.

[11] Céline Bonfils, Benjamin D. Santer, John C. Fyfe, Kate Marvel, Thomas J. Phillips, and Susan R. H. Zimmerman. Human influence on joint changes in temperature, rainfall and continental aridity. *Nature Climate Change*, 10:726–731, 2020.

[12] Céline Bonfils, Benjamin D. Santer, David W. Pierce, Hugo G. Hidalgo, Govindasamy Bala, Tapash Das, Tim P. Barnett, Daniel R. Cayan, Charles Doutriaux, Andrew W. Wood, Art Mirin, and Toru Nozawa. Detection and attribution of temperature changes in the mountainous western united states. *Journal of Climate*, 21(23):6404–6424, 2008.

[13] H. Y. Brown, B. Wagman, D. Bull, K. Peterson, B. Hillman, X. Liu, Z. Ke, and L. Lin. Validating a microphysical prognostic stratospheric aerosol implementation in e3smv2 using observations after the mount pinatubo eruption. *Geoscientific Model Development*, 17(13):5087–5121, 2024.

[14] Hunter Brown, Hailong Wang, Mark Flanner, Xiaohong Liu, Balwinder Singh, Rudong Zhang, Yang Yang, and Mingxuan Wu. Brown carbon fuel and emission source attributions to global snow darkening effect. *Journal of Advances in Modeling Earth Systems*, 14(4):e2021MS002768, 2022. e2021MS002768 2021MS002768.

[15] M. Brown, M. Peterson, I. Tezaur, K. Peterson, and D. Bull. Random Forest Regression Feature Importance for Climate Impact Pathway Detection. ArXiv pre-print, 2024.

[16] Meredith G. L. Brown, Matt Peterson, Irina Tezaur, Kara Peterson, and Diana Bull. Random forest regression feature importance for climate impact pathway detection. J. Comput. Appl. Math. (submitted), September 2024.

[17] K. A. Browning, R. J. Allam, S. P. Ballard, R. T.H. Barnes, D. A. Bennetts, R. H . Maryon, P. J. Mason, D. S. McKenna, J. F.B. Mitchell, C. A. Senior, A. Slingo, and F. B. Smith. Environmental Effects From Burning Oil Wells in the Gulf. *Weather*, 47(6):201–212, 1992.

[18] Michael Burger, Jessica Wentz, and Radley Horton. The law and science of climate change attribution. *Colum. J. Envtl. L.*, 45:57.

[19] J.M. Burgers. A Mathematical Model Illustrating the Theory of Turbulence. *Advances in Applied Mechanics*, 1:171–199, 1948.

[20] Christopher W. Callahan and Justin S. Mankin. Globally unequal effect of extreme heat on economic growth. *Science Advances*, 8(43):eadd3726, 2022.

[21] Mauricio Campos, Bo Li, Guillaume de Lafontaine, Joseph Napier, and Feng Hu. Integrating different data sources using a bayesian hierarchical model to unveil glacial refugia. *Journal of Agricultural, Biological and Environmental Statistics*, 29, 11 2023.

[22] Giorgia Di Capua, Marlene Kretschmer, Reik V. Donner, Bart van den Hurk, Ramesh Vellore, Raghavan Krishnan, and Dim Coumou. Tropical and mid-latitude teleconnections interacting with the Indian summer monsoon rainfall: a theory-guided causal effect network approach. *Earth System Dynamics*, 11(1):17–34, 2019.

[23] Giorgia Di Capua, Jakob Runge, Reik V. Donner, Bart van den Hurk, Andrew G. Turner, Ramesh Vellore, Raghavan Krishnan, and Dim Coumou. Dominant patterns of interaction between the tropics and mid-latitudes in boreal summer: causal relationships and the role of timescales. *Weather and Climate Dynamics*, 1(2):519–539, 2020.

[24] D Chen, M Rojas, BH Samset, K Cobb, A Diongue Niang, P Edwards, S Emori, SH Faria, E Hawkins, P Hope, P Huybrechts, M Meinshausen, SK Mustafa, GK Plattner, and AM Tréguier. Framing, context, and methods. climate change 2021, the physical science basis. contribution of working group I to the sixth assessment report of the Intergovernmental Panel on Climate Change, Chapter 1. *IPCC Sixth Assessment Report*, 2021.

[25] John A Church, Neil J White, and Julie M Arblaster. Significant decadal-scale impact of volcanic eruptions on sea level and ocean heat content. *Nature*, 438(7064):74–77, 2005.

[26] C. M. Colose, A. N. LeGrande, and M. Vuille. Hemispherically asymmetric volcanic forcing of tropical hydroclimate during the last millennium. *Earth System Dynamics*, 7(3):681–696, 2016.

[27] Noel Cressie, Matthew Sainsbury-Dale, and Andrew Zammit-Mangion. Basis-Function Models in Spatial Statistics. *Annual Review of Statistics and Its Application*, 9(1):373–400, 2022.

[28] G. Danabasoglu, J.-F. Lamarque, J. Bacmeister, D. A. Bailey, A. K. DuVivier, J. Edwards, L. K. Emmons, J. Fasullo, R. Garcia, A. Gettelman, C. Hannay, M. M. Holland, W. G. Large, P. H. Lauritzen, D. M. Lawrence, J. T. M. Lenaerts, K. Lindsay, W. H. Lipscomb, M. J. Mills, R. Neale, K. W. Oleson, B. Otto-Bliesner, A. S. Phillips, W. Sacks, S. Tilmes, L. van Kampenhout, M. Vertenstein, A. Bertini, J. Dennis, C. Deser, C. Fischer, B. Fox-Kemper, J. E. Kay, D. Kinnison, P. J. Kushner, V. E. Larson, M. C. Long, S. Mickelson, J. K. Moore, E. Nienhouse, L. Polvani, P. J. Rasch, and W. G. Strand. The community earth system model version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, 12(2):e2019MS001916, 2020. e2019MS001916 2019MS001916.

[29] MK Davey, A Brookshaw, and S Ineson. The probability of the impact of enso on precipitation and near-surface temperature. *Climate Risk Management*, 1:5–24, 2014.

[30] S. M. Davis, K. H. Rosenlof, B. Hassler, D. F. Hurst, W. G. Read, H. Vömel, H. Selkirk, M. Fujiwara, and R. Damadeo. "the stratospheric water and ozone satellite homogenized (swoosh) database: a long-term database for climate studies". *Earth System Science Data*, 8(2):461–490, 2016.

[31] W. Davis, M. Carlson, I. Tezaur, and D. Bull. Spatio-temporal Multivariate Cluster Evolution Analysis for Detecting and Tracking Climate Impacts. in preparation for submission to J. Comput. Appl. Math., 2024.

[32] Yi Deng and Imme Ebert-Uphoff. Weakening of atmospheric information flow in a warming climate in the Community Climate System Model. *Geophysical Research Letters*, 41(1):193–200, 2014.

[33] Jack Dunn, Luca Mingardi, and Ying Daisy Zhuo. Comparing interpretability and explainability for feature selection. *arXiv preprint arXiv:2105.05328*, 2021.

[34] Imme Ebert-Uphoff and Yi Deng. A new type of climate network based on probabilistic graphical models: Results of boreal winter versus summer. *Geophysical Research Letters*, 39(19), 2012.

[35] Thomas Ehrmann, Benjamin Wagman, Diana Bull, Benjamin Hillman, Joseph Hollowed, Hunter Brown, Kara Peterson, Laura Swiler, Jerry Watkins, and Joseph Hart. Identifying northern hemisphere stratospheric and surface temperature responses to the mt. pinatubo eruption within e3smv2-spa. Technical Report SAND2024-12730, Sandia National Laboratories, 2024.

[36] Thomas Ehrmann, Benjamin Wagman, Diana Bull, Benjamin Hillman, Joseph Hollowed, Hunter Brown, Kara Peterson, Laura Swiler, Jerry Watkins, and Joseph Hart. Identifying northern hemisphere temperature responses to the mt. pinatubo eruption through limited variability ensembles. *Climate Dynamics*, 2024:1–40, 2024.

[37] Veronika Eyring, NP Gillett, Krishna Achutarao, Rondrotiana Barimalala, Marcelo Barreiro Parrillo, Nicolas Bellouin, Christophe Cassou, Paul Durack, Yu Kosaka, Shayne McGregor, et al. Human influence on the climate system: Contribution of working group I to the sixth assessment report of the Intergovernmental Panel on Climate Change, chapter 3. *IPCC Sixth Assessment Report*, 2021.

[38] Suvarna Fadnavis, Rolf Müller, Tanusri Chakraborty, TP Sabin, Anton Laakso, Alexandru Rap, Sabine Griessbach, Jean-Paul Vernier, and Simone Tilmes. The role of tropical volcanic eruptions in exacerbating indian droughts. *Scientific reports*, 11(1):2714, 2021.

[39] D James Fulton and Gabriele C Hegerl. Testing methods of pattern extraction for climate data using synthetic modes. *Journal of Climate*, 34(18):7645–7660, 2021.

[40] Sulochana Gadgil. The indian monsoon and its variability. *Annual Review of Earth and Planetary Sciences*, 31(Volume 31, 2003):429–467, 2003.

[41] Sulochana Gadgil. The monsoon system: Land–sea breeze or the ITCZ? *Journal of Earth System Science*, 127:1–29, 2018.

[42] Sulochana Gadgil and Siddhartha Gadgil. The indian monsoon, gdp and agriculture. *Economic and Political Weekly*, 41(47):4887–4895, 2006.

[43] Evgenia Galytska, Katja Weigel, Dörthe Handorf, Ralf Jaiser, Raphael Harry Köhler, Jakob Runge, and Veronika Eyring. Causal model evaluation of Arctic-midlatitude teleconnections in CMIP6. *Journal of Geophysical Research: Atmospheres*, 128(17), 2022.

[44] C. C. Gao and Y. J. Gao. Revisited asian monsoon hydroclimate response to volcanic eruptions. *Journal of Geophysical Research: Atmospheres*, 123(15):7883–7896, 2018.

[45] Robert Garrett, Lyndsay Shand, and Gabriel Huerta. A Multivariate Space-Time Dynamic Model for Characterizing the Atmospheric Impacts Following the Mt. Pinatubo Eruption. *journal=arXiv preprint arXiv:2408.13392*, 2024.

[46] Ruth Geen, Simona Bordoni, David S. Battisti, and Katrina Hui. Monsoons, itczs, and the concept of the global monsoon. *Reviews of Geophysics*, 58(4):e2020RG000700, 2020. e2020RG000700 10.1029/2020RG000700.

[47] A. Gettelman, M. J. Mills, D. E. Kinnison, R. R. Garcia, A. K. Smith, D. R. Marsh, S. Tilmes, F. Vitt, C. G. Bardeen, J. McInerny, H.-L. Liu, S. C. Solomon, L. M. Polvani, L. K. Emmons, J.-F. Lamarque, J. H. Richter, A. S. Glanville, J. T. Bacmeister, A. S. Phillips, R. B. Neale, I. R. Simpson, A. K. DuVivier, A. Hodzic, and W. J. Randel. The whole atmosphere community climate model version 6 (waccm6). *Journal of Geophysical Research: Atmospheres*, 124(23):12380–12403, 2019.

[48] S. J. Ghan. Technical note: Estimating aerosol effects on cloud radiative forcing. *Atmospheric Chemistry and Physics*, 13(19):9971–9974, 2013.

[49] NP Gillett, AJ Weaver, FW Zwiers, and MF Wehner. Detection of volcanic influence on global precipitation. *Geophysical Research Letters*, 31(12), 2004.

[50] Global Modeling and Assimilation Office (GMAO). MERRA-2 instM_3d_asm_Np: 3d,3-Hourly,Time-Averaged,Model-Level,Assimilation,Assimilated Meteorological Fields V5.12.4, 2015. Accessed: 2023-05-11.

[51] Global Modeling and Assimilation Office (GMAO). MERRA-2 tavg1_2d_aer_Nx: 2d,1-Hourly,Time-averaged,Single-Level,Assimilation,Aerosol Diagnostics V5.12.4, 2015. Accessed: 2023-05-11.

[52] Global Modeling and Assimilation Office (GMAO). MERRA-2 tavg1_2d_rad_Nx: 2d,1-Hourly,Time-Averaged,Single-Level,Assimilation,Radiation Diagnostics V5.12.4, 2015. Accessed: 2023-07-31.

[53] Global Modeling and Assimilation Office (GMAO). MERRA-2 tavg1_2d_slv_Nx: 2d,1-Hourly,Time-Averaged,Single-Level,Assimilation,Single-Level Diagnostics, 2015. Accessed: 2023-07-31.

[54] Clark Glymour, Kun Zhang, and Peter Spirtes. Review of Causal Discovery Methods Based on Graphical Models. *Frontiers in Genetics*, 10:524, 2019.

[55] Jean-Christophe Golaz, Luke P. Van Roekel, Xue Zheng, Andrew F. Roberts, Jonathan D. Wolfe, Wuyin Lin, Andrew M. Bradley, Qi Tang, Mathew E. Maltrud, Ryan M. Forsyth, Chengzhu Zhang, Tian Zhou, Kai Zhang, Charles S. Zender, Mingxuan Wu, Hailong Wang, Adrian K. Turner, Balwinder Singh, Jadwiga H. Richter, Yi Qin, Mark R. Petersen, Azamat Mametjanov, Po-Lun Ma, Vincent E. Larson, Jayesh Krishna, Noel D. Keen, Nicole Jeffery, Elizabeth C. Hunke, Walter M. Hannah, Oksana Guba, Brian M. Griffin, Yan Feng, Darren Engwirda, Alan V. Di Vittorio, Cheng Dang, LeAnn M. Conlon, Chih-Chieh-Jack Chen, Michael A. Brunke, Gautam Bisht, James J. Benedict, Xylar S. Asay-Davis, Yuying Zhang, Meng Zhang, Xubin Zeng, Shaocheng Xie, Phillip J. Wolfram, Tom Vo, Milena Veneziani, Teklu K. Tesfa, Sarat Sreepathi, Andrew G. Salinger, J. E. Jack Reeves Eyre, Michael J. Prather, Salil Mahajan, Qing Li, Philip W. Jones, Robert L. Jacob, Gunther W. Huebler, Xianglei Huang, Benjamin R. Hillman, Bryce E. Harrop, James G. Foucar, Yilin Fang, Darin S. Comeau, Peter M. Caldwell, Tony Bartoletti, Karthik Balaguru, Mark A. Taylor, Renata B. McCoy, L. Ruby Leung, and David C. Bader. The doe e3sm model version 2: Overview of the physical model and initial model evaluation. *Journal of Advances in Modeling Earth Systems*, 14(12):e2022MS003156, 2022. e2022MS003156 2022MS003156.

[56] Andrew Gonzalez, Jonathan M. Chase, and Mary I. O'Connor. A framework for the detection and attribution of biodiversity change. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 378(1881), 2023.

[57] Katherine Goode, Daniel Ries, and Kellie McClernon. Characterizing climate pathways using feature importance on echo state networks. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 17(4):e11706, 2024.

[58] Katherine Goode, Daniel Ries, and Kellie McClernon. *listenr: Explainability for Echo State Networks*, 2024. R package version 0.4.0.

[59] Carol A Gotway and Linda J Young. Combining incompatible spatial data. *Journal of the American Statistical Association*, 97(458):632–648, 2002.

[60] Guojun Gu, Robert F Adler, George J Huffman, and Scott Curtis. Tropical rainfall variability on interannual-to-interdecadal and longer time scales derived from the gpcp monthly product. *Journal of Climate*, 20(15):4033–4046, 2007.

[61] Carole Hall, J. Derek Tucker, and Drew Yarger. Elastic changepoint detection for globally-indexed functional time series data with climate applications. Technical Report SAND2024-12470, Sandia National Laboratories, 2024.

[62] Edmond Halley. An historical account of the trade winds, and monsoons, observable in the seas between and near the tropicks, with an attempt to assign the physical cause of the said winds. *Philosophical Transactions of the Royal Society of London*, 16(183):153–168, 1686.

[63] Michael S. Halpert and Gerald D. Bell. Climate assessment for 1996. *Bulletin of the American Meteorological Society*, 78(5s):S1 – S50, 1997.

[64] Alexis Hannart. Integrated optimal fingerprinting: Method description and illustration. *Journal of Climate*, 29(6):1977–1998, 2016.

[65] Joseph Hart, Mamikon Gulian, I. Manickam, and Laura P. Swiler. Solving high-dimensional inverse problems with auxiliary uncertainty via operator learning with limited data. *Journal of Machine Learning for Modeling and Computing*, 4(2):105–133, 2023.

[66] Joseph Hart, Indu Manickam, Mamikon Gulian, Laura Swiler, Diana Bull, Thomas Ehrmann, Benjamin Wagman, and Jerry Watkins. Stratospheric aerosol source inversion: Noise, variability, and uncertainty quantification. *arXiv preprint arXiv:2409.06846*, 2024.

[67] Klaus Hasselmann. Optimal Fingerprints for the Detection of Time-dependent Climate Change. *Journal of Climate*, 6(10):1957–1971, 1993.

[68] Klaus Hasselmann. Multi-pattern fingerprint method for detection and attribution of climate change. *Climate dynamics*, 13(9):601–611, 1997.

[69] Gabriele C Hegerl, Klaus Hasselmann, Ulrich Cubasch, John FB Mitchell, Erich Roeckner, Reinhard Voss, and Jürgen Waszkewitz. Multi-fingerprint detection and attribution analysis of greenhouse gas, greenhouse gas-plus-aerosol and solar forced climate change. *Climate Dynamics*, 13(9):613–634, 1997.

[70] Gabriele C Hegerl, Ove Hoegh-Guldberg, Gino Casassa, Martin Hoerling, Sari Kovats, Camille Parmesan, David Pierce, and Peter Stott. Good practice guidance paper on detection and attribution related to anthropogenic climate change. 2010.

[71] Gabriele C Hegerl and Gerald R North. Comparison of statistically optimal approaches to detecting anthropogenic climate change. *Journal of Climate*, 10(5):1125–1133, 1997.

[72] Nadja Herger, Benjamin M. Sanderson, and Reto Knutti. Improved pattern scaling approaches for the use in climate impact studies. *Geophysical Research Letters*, 42(9):3486–3494, 2015.

[73] Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, Adrian Simmons, Cornel Soci, Saleh Abdalla, Xavier Abellan, Gianpaolo Balsamo, Peter Bechtold, Gionata Biavati, Jean Bidlot, Massimo Bonavita, Giovanna De Chiara, Per Dahlgren, Dick Dee, Michail Diamantakis, Rossana Dragani, Johannes Flemming, Richard Forbes, Manuel Fuentes, Alan Geer, Leo Haimberger, Sean Healy, Robin J. Hogan, Elías Hólm, Marta Janisková, Sarah Keeley, Patrick Laloyaux, Philippe Lopez, Cristina Lupu, Gabor Radnoti, Patricia de Rosnay, Iryna Rozum, Freja Vamborg, Sebastien Villaume, and Jean-Noël Thépaut. The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, 146(730):1999–2049, 2020.

[74] Sir Austin Bradford Hill. The Environment and Disease: Association or Causation? *Proceedings of the Royal Society of Medicine*, 58(5):295–300, 1965.

[75] J. Hollowed, C. Jablonowski, T. Ehrmann, B. Hillman, and ... Volcanic aerosol modification of the stratospheric circulation in e3smv2 part ii: Tracer sensitivity, To be submitted December 2024.

[76] J. Hollowed, C. Jablonowski, T. Ehrmann, B. Hillman, and ... Volcanic aerosol modification of the stratospheric circulation in e3smv2 part i: Wave-mean flow interaction. in preparation for submission to Climate Dynamics, To be submitted October 2024.

[77] Joseph P. Hollowed, Christiane Jablonowski, Hunter Y. Brown, Benjamin R. Hillman, Diana L. Bull, and Joseph L. Hart. Localized injections of interactive volcanic aerosols and their climate impacts in a simple general circulation model. *EGUsphere [preprint]*, 2024.

[78] A. Hu, X. Liu, B. Hillman, B. Wagman, D. Bull, and K. Peterson. Aerosol indirect effect: a study of cirrus implications with mam5, To be submitted December 2024.

[79] A. Hu, X. Liu, Z. Ke, B. Wagman, H. Brown, Z. Lu, D. Bull, and K. Peterson. Size-resolved process understanding of stratospheric sulfate aerosol following the pinatubo eruption. *EGUsphere*, 2024:1–28, 2024.

[80] Carley E. Iles, Gabriele C. Hegerl, Andrew P. Schurer, and Xuebin Zhang. The effect of volcanic eruptions on global precipitation. *Journal of Geophysical Research: Atmospheres*, 118(16):8770–8786, 2013.

[81] Peter J. Irvine, Ben Kravitz, Mark G. Lawrence, and Helene Muri. An overview of the Earth system science of solar geoengineering. *WIREs Climate Change*, 7(6):815–833, 2016.

[82] M. Iturbide, J. M. Gutiérrez, L. M. Alves, J. Bedia, R. Cerezo-Mota, E. Cimadevilla, A. S. Cofiño, A. Di Luca, S. H. Faria, I. V. Gorodetskaya, M. Hauser, S. Herrera, K. Hennessy, H. T. Hewitt, R. G. Jones, S. Krakovska, R. Manzanas, D. Martínez-Castro, G. T. Narisma, I. S. Nurhati, I. Pinto, S. I. Seneviratne, B. van den Hurk, and C. S. Vera. An update of IPCC climate reference regions for subcontinental analysis of climate model data: definition and aggregated datasets. *Earth System Science Data*, 12(4):2959–2970, 2020.

[83] C. Jablonowski, L. Hollowed, J.and Nguyen, B. Hillman, T. Ehrmann, and B. Wagman. Stratospheric characteristics in e3smv2., To be submitted December 2024.

[84] Eric J. Jensen and Owen B. Toon. The potential effects of volcanic aerosols on cirrus cloud microphysics. *Geophysical Research Letters*, 19(17):1759–1762, 1992.

[85] Ian T Jolliffe and Jorge Cadima. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065):20150202, 2016.

[86] Stephen A. Klein and Christian Jakob. Validation and sensitivities of frontal clouds simulated by the ecmwf model. *Monthly Weather Review*, 127(10):2514 – 2531, 1999.

[87] Christopher Krich, Jakob Runge, Diego G. Miralles, Mirco Migliavacca, Oscar Perez-Priego, Tarek El-Madany, Arnaud Carrara, and Miguel D. Mahecha. Estimating causal networks in biosphere–atmosphere interaction with the PCMCI approach. *Biogeosciences*, 17(4):1033–1061, 2020.

[88] RH Kripalani, Ashwini Kulkarni, SS Sabade, and ML Khandekar. Indian monsoon variability in a global warming scenario. *Natural hazards*, 29:189–206, 2003.

[89] D. Krofcheck, J. Crockett, K. Peterson, J. Li, and D. Bull. Volcanic stratospheric aerosol injection may buffer against global ecohydrological impacts of hotter and drier climates, To be submitted December 2024.

[90] I Elizabeth Kumar, Suresh Venkatasubramanian, Carlos Scheidegger, and Sorelle Friedler. Problems with Shapley-value-based explanations as feature importance measures. In *International conference on machine learning*, pages 5491–5500. PMLR, 2020.

[91] K. Krishna Kumar, Balaji Rajagopalan, Martin Hoerling, Gary Bates, and Mark Cane. Unraveling the mystery of indian monsoon failure during el niño. *Science*, 314(5796):115–119, 2006.

[92] B. Kärcher and U. Lohmann. A parameterization of cirrus cloud formation: Homogeneous freezing of supercooled aerosols. *Journal of Geophysical Research: Atmospheres*, 107(D2):AAC 4–1–AAC 4–10, 2002.

[93] K. Labitzke and M. P. McCormick. Stratospheric temperature increases due to pinatubo aerosols. *Geophysical Research Letters*, 19(2):207–210, 1992.

[94] J. Li, C. Wentland, L Swiler, K. Peterson, M. Weylandt, D. Krofcheck, B. Wagman, and D. Bull. Attributing mt. pinatubo in spatial patterns of vegetation response, To be submitted November 2024.

[95] Justin D. Li, Audrey McCombs, Daniel Dunlavy, Gabriel Huerta, and Lyndsay Shand. Exploring multivariate tensor data fusion and completion. in preparation for submission to Special Issue of Remote Sensing, To be submitted October 2024.

[96] Tao Lian and Dake Chen. An evaluation of rotated eof analysis and its application to tropical pacific sst variability. *Journal of Climate*, 25(15):5361–5373, 2012.

[97] Finn Lindgren and Håvard Rue. Bayesian spatial modelling with R-INLA. *Journal of statistical software*, 63(19), 2015.

[98] Fei Liu, Jing Chai, Bin Wang, Jian Liu, Xiao Zhang, and Zhiyuan Wang. Global monsoon precipitation responses to large volcanic eruptions. *Scientific reports*, 6(1):24331, 2016.

[99] X. Liu, P.-L. Ma, H. Wang, S. Tilmes, B. Singh, R. C. Easter, S. J. Ghan, and P. J. Rasch. Description and evaluation of a new four-mode version of the modal aerosol module (mam4) within version 5.3 of the community atmosphere model. *Geoscientific Model Development*, 9(2):505–522, 2016.

[100] Xiaohong Liu and Joyce E. Penner. Effect of mount pinatubo h2so4/h2o aerosol on ice nucleation in the upper troposphere using a global chemistry and transport model. *Journal of Geophysical Research: Atmospheres*, 107(D12):AAC 2–1–AAC 2–18, 2002.

[101] Yongqiang Liu, Scott Goodrick, and Warren Heilman. Wildland fire emissions, carbon, and climate: Wildfire-climate interactions. *Forest Ecology and Management*, 317:80–96, 2014.

[102] Elisabeth A Lloyd and Theodore G Shepherd. Climate change attribution and legal contexts: evidence and the role of storylines. *Climatic Change*, 167(3-4):28, 2021.

[103] U. Lohmann, B. Kärcher, and C. Timmreck. Impact of the mount pinatubo eruption on cirrus clouds formed by homogeneous freezing in the echam4 gcm. *Journal of Geophysical Research: Atmospheres*, 108(D18), 2003.

[104] George Luber and Michael McGeehin. Climate Change and Extreme Heat Events. *American Journal of Preventive Medicine*, 35(5):429–435, 2008.

[105] Zhengzhao Luo, William B. Rossow, Toshiro Inoue, and Claudia J. Stubenrauch. Did the eruption of the mt. pinatubo volcano affect cirrus properties? *Journal of Climate*, 15(19):2806 – 2820, 2002.

[106] Anandamayee Majumdar, Alan E Gelfand, and Sudipto Banerjee. Spatio-temporal change-point modeling. *Journal of Statistical Planning and Inference*, 130(1-2):149–166, 2005.

[107] Antonios Mamalakis, Imme Ebert-Uphoff, and Elizabeth A Barnes. Neural network attribution methods for problems in geoscience: A novel synthetic benchmark dataset. *Environmental Data Science*, 1:e8, 2022.

[108] Wenmin Man, Tianjun Zhou, and Johann H. Jungclaus. Effects of large volcanic eruptions on global summer climate and east asian monsoon changes during the last millennium: Analysis of mpi-esm simulations. *Journal of Climate*, 27(19):7394 – 7409, 2014.

[109] Wilson E Marcílio and Danilo M Eler. From explanations to feature selection: assessing SHAP values as feature selection mechanism. In *2020 33rd SIBGRAPI conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 340–347. Ieee, 2020.

[110] Lauren R Marshall, Elena C Maters, Anja Schmidt, Claudia Timmreck, Alan Robock, and Matthew Toohey. Volcanic effects on climate: recent advances and future avenues. *Bulletin of Volcanology*, 84(5):54, 2022.

[111] Kate Marvel, Michela Biasutti, and Céline Bonfils. Fingerprints of external forcings on Sahel rainfall: aerosols, greenhouse gases, and model-observation discrepancies. *Environmental Research Letters*, 15(8):084023, 2020.

[112] Kellie McClernon, Katherine Goode, and Daniel Ries. A comparison of model validation approaches for echo state networks using climate model replicates. *Spatial Statistics*, 59:100813, 2024.

[113] Kellie McClernon, Daniel Ries, and Katherine Goode. Verifying regional feature importance on echo state networks for climate pathways. Technical Report SAND2024-12404O, Sandia National Laboratories, 2024.

[114] A. Menon, A. Levermann, J. Schewe, J. Lehmann, and K. Frieler. Consistent increase in indian monsoon rainfall and its variability across cmip-5 models. *Earth System Dynamics*, 4(2):287–300, 2013.

[115] Michael J. Mills, Anja Schmidt, Richard Easter, Susan Solomon, Douglas E. Kinnison, Steven J. Ghan, Ryan R. Neely III, Daniel R. Marsh, Andrew Conley, Charles G. Bardeen, and Andrew Gettelman. Global volcanic aerosol properties derived from emissions, 1990–2014, using cesm1(waccm). *Journal of Geophysical Research: Atmospheres*, 121(5):2332–2348, 2016.

[116] JFB Mitchell, DJ Karoly, GC Hegerl, FW Zwiers, MR Allen, and J Marengo. Detection of climate change and attribution of causes. Technical report, Intergovernmental Panel on Climate Change (IPCC), Assessment Report 3, 2001.

[117] Masato Murakami. Satellite cloudiness in the monsoon area. *Monsoon meteorology*, pages 354–402, 1987.

[118] Davison Mwale, Thian Yew Gan, Kevin Devito, Carl Mendoza, Uldis Silins, and Richard Petrone. Precipitation variability and its relationship to hydrologic variability in alberta. *Hydrological Processes: An International Journal*, 23(21):3040–3056, 2009.

[119] National Academies of Sciences, Engineering, and Medicine. *Attribution of extreme weather events in the context of climate change*. National Academies Press, 2016.

[120] National Intelligence Council. Climate Change and International Responses Increasing Challenges to US National Security Through 2040. Technical report, 2021.

[121] Ryan R. Neely III and Anja Schmidt. Volcaneesm: Global volcanic sulphur dioxide (so2) emissions database from 1850 to present, 2016.

[122] C. Newhall. The Cataclysmic 1991 Eruption of Mount Pinatubo, Philippines. Vol. 113, U.S. Geological Survey, 1997.

[123] Hai Nguyen, Noel Cressie, and Amy Braverman. Spatial statistical data fusion for remote sensing applications. *JASA. Journal of the American Statistical Association*, 107, 09 2012.

[124] J. Nichol, M. Weylandt, M. Fricke, M. Moses, D. Bull, and L. Swiler. Space-Time Causal Discovery in Climate Science: A Local Stencil Learning Approachpace-Time Causal Discovery in Climate Science: A Local Stencil Learning Approach. JGR Machine Learning and Computation (submitted), 2024.

[125] Jeffrey Jake Nichol, Michael Weylandt, Mark Smith, and Laura P Swiler. Benchmarking the pcmci causal discovery algorithm for spatiotemporal systems. Technical Report 2023-05141, Sandia National Laboratories, Albuquerque, NM, 2023.

[126] Sharon E. Nicholson. The itcz and the seasonal cycle over equatorial africa. *Bulletin of the American Meteorological Society*, 99(2):337 – 348, 2018.

[127] Gerald R. North and Mark J. Stevens. Detecting climate signals in the surface temperature record. *Journal of Climate*, 11(4):563–577, 1998.

[128] Peer Nowack, Jakob Runge, Veronika Eyring, and Joanna D. Haigh. Causal networks for climate model evaluation and constrained projections. *Nature Communications 2020 11:1*, 11(1):1—11, 2020.

[129] Douglas Nychka, Soutir Bandyopadhyay, Dorit Hammerling, Finn Lindgren, and Stephan Sain. A multiresolution gaussian process model for the analysis of large spatial datasets. *Journal of Computational and Graphical Statistics*, 24(2):579–599, 2015.

[130] Timothy J. Osborn, Craig J. Wallace, Ian C. Harris, and Thomas M. Melvin. Pattern scaling using ClimGen: monthly-resolution future climate scenarios including changes in the variabil ity of precipitation. *Climatic Change*, 134(3):353–369, 2016.

[131] Seungmok Paik, Seung-Ki Min, Carley E. Iles, Erich M. Fischer, and Andrew P. Schurer. Volcanic-induced global monsoon drying modulated by diverse el niño responses. *Science Advances*, 6(21):eaba1212, 2020.

[132] David E Parker, H Wilson, Phil D Jones, JR Christy, and Chris K Folland. The impact of mount pinatubo on world-wide temperatures. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 16(5):487–497, 1996.

[133] Judea Pearl. Causal Diagrams for Empirical Research. *Biometrika*, 82(4):669, 1995.

[134] Judea Pearl. Graphs, Causality, and Structural Equation Models. *Sociological Methods & Research*, 27(2):226–284, 1998.

[135] Judea Pearl. Causal inference in statistics: An overview. *Statistics Surveys*, 3(September):96–146, 2009.

[136] Judea Pearl. *Causality*. Cambridge University Press, University of California, Los Angeles, 2 edition, 2009.

[137] Judea Pearl and Dana Mackenzie. *The Book of Why*. Basic Books, New York, 2018.

[138] L. M. Polvani, A. Banerjee, and A. Schmidt. Northern Hemisphere continental winter warming following the 1991 Mt. Pinatubo eruption: reconciling models and observations. *Atmospheric Chemistry and Physics*, 19(9):6351–6366, 2019.

[139] Evgeniya Predybaylo, Georgiy L. Stenchikov, Andrew T. Wittenberg, and Fanrong Zeng. Impacts of a pinatubo-size volcanic eruption on ENSO. *Journal of Geophysical Research: Atmospheres*, 122(2):925–947, 2017.

[140] Jonathan Proctor, Solomon Hsiang, Jennifer Burney, Marshall Burke, and Wolfram Schlenker. Estimating global agricultural effects of geoengineering using volcanic eruptions. *Nature*, 560(7719):480–483, 2018.

[141] Weihong Qian. *Temporal climatology and anomalous weather analysis*. Springer, 2017.

[142] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2023.

[143] Aurélien Ribes, Serge Planton, and Laurent Terray. Application of regularised optimal fingerprinting to attribution. Part I: method, properties and idealised analysis. *Climate dynamics*, 41(11):2817–2836, 2013.

[144] Jadwiga H Richter, Daniele Visioni, Douglas G MacMartin, David A Bailey, Nan Rosenbloom, Brian Dobbins, Walker R Lee, Mari Tye, and Jean-Francois Lamarque. Assessing responses and impacts of solar climate intervention on the earth system with stratospheric aerosol injection (arise-sai): protocol and initial results from the first simulations. *Geoscientific Model Development*, 15(22):8221–8243, 2022.

[145] Daniel Ries, Katherine Goode, Kellie McClernon, and Benjamin Hillman. Validating climate pathways with feature importance applied to earth system models. *Submitted to Geoscientific Model Development*, 2024.

[146] Daniel Ries, Kellie McClernon, Thomas Ehrmann, and Diana Bull. Using spatio-temporal feature importance to identify drivers of indian monsoon after the mount pinatubo eruption. Technical Report SAND2024-12581O, Sandia National Laboratories, 2024.

[147] Alan Robock. Volcanic eruptions and climate. *Reviews of geophysics*, 38(2):191–219, 2000.

[148] Shinto Roose, Govindasamy Bala, KS Krishnamohan, Long Cao, and Ken Caldeira. Quantification of tropical monsoon precipitation changes in terms of interhemispheric differences in stratospheric sulfate aerosol optical depth. *Climate Dynamics*, 61(9):4243–4258, 2023.

[149] Håvard Rue, Andrea Riebler, Sigrunn H Sørbye, Janine B Illian, Daniel P Simpson, and Finn K Lindgren. Bayesian computing with inla: a review. *Annual Review of Statistics and Its Application*, 4(1):395–421, 2017.

[150] J. Runge. Causal network reconstruction from time series: From theoretical assumptions to practical estimation. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28(7):075310, 2018.

[151] Jakob Runge. Quantifying information transfer and mediation along causal pathways in complex systems. *Physical Review E*, 92(6):062829, 2015.

[152] Jakob Runge, Sebastian Bathiany, Erik Bollt, Gustau Camps-Valls, Dim Coumou, Ethan Deyle, Clark Glymour, Marlene Kretschmer, Miguel D. Mahecha, Jordi Munoz-Mari, Egbert H. van Nes, Jonas Peters, Rick Quax, Markus Reichstein, Marten Scheffer, Bernhard Scholkopf, Peter Spirtes, George Sugihara, Jie Sun, Kun Zhang, and Jakob Zscheischler. Inferring causation from time series in Earth system sciences. *Nature Communications*, 10(2553), 2019.

[153] Jakob Runge, Reik V. Donner, and Jürgen Kurths. Optimal model-free prediction from multivariate time series. *Physical Review E*, 91(5):052909, 2015.

[154] Jakob Runge, Peer Nowack, Marlene Kretschmer, Seth Flaxman, and Dino Sejdinovic. Detecting and quantifying causal associations in large nonlinear time series datasets. *Science Advances*, 5(11):4996—5023, 2019.

[155] Jakob Runge, Vladimir Petoukhov, Jonathan F. Donges, Jaroslav Hlinka, Nikola Jajcay, Martin Vejmelka, David Hartman, Norbert Marwan, Milan Paluš, and Jürgen Kurths. Identifying causal gateways and mediators in complex spatio-temporal systems. *Nature Communications*, 6(1):8502, 2015.

[156] BD Santer, TML Wigley, and PD Jones. Correlation methods in fingerprint detection studies. *Climate Dynamics*, 8(6):265–276, 1993.

[157] Benjamin D Santer, C Mears, C Doutriaux, Peter Caldwell, Peter J Gleckler, TML Wigley, Susan Solomon, NP Gillett, D Ivanova, Thomas R Karl, et al. Separating signal and noise in atmospheric temperature changes: The importance of timescale. *Journal of Geophysical Research: Atmospheres*, 116(D22), 2011.

[158] Tapio Schneider, Tobias Bischoff, and Gerald H Haug. Migrations and dynamics of the intertropical convergence zone. *Nature*, 513(7516):45–53, 2014.

[159] Mark R Segal. Machine learning benchmarks and random forest regression. *UCSF: Center for Bioinformatics and Molecular Biostatistics*, Retrieved from https://escholarship.org/uc/item/35x3v9t4, 2004.

[160] Lloyd S Shapley. A value for n-person games. *Contributions to the Theory of Games*, 2, 1953.

[161] T. Shead, I. Tezaur, W. Davis, M. Carlson, D. Dunlavy, E. Parish, P. Blonigan, J. Tencer, F. Rizzi, and H. Kolla. A novel in situ machine learning framework for intelligent data capture and event detection. in Machine Learning and Its Application to Reacting Flows, N. Swaminathan and A. Parente, eds., pp. 53-87, 2023.

[162] Theodore G Shepherd. A common framework for approaches to extreme event attribution. *Current Climate Change Reports*, 2:28–38, 2016.

[163] Theodore G Shepherd, Emily Boyd, Raphael A Calel, Sandra C Chapman, Suraje Dessai, Ioana M Dima-West, Hayley J Fowler, Rachel James, Douglas Maraun, Olivia Martius, et al. Storylines: an alternative approach to representing uncertainty in physical aspects of climate change. *Climatic change*, 151:555–571, 2018.

[164] Samantha Shi-Jun, Lyndsay Shand, and Bo Li. Tracing the impacts of mount pinatubo eruption on global climate using spatially-varying changepoint detection. *arXiv preprint arXiv:2409.08908*, 2024.

[165] Ram Singh, Kostas Tsigaridis, Diana Bull, Laura P Swiler, Benjamin M Wagman, and Kate Marvel. Mount pinatubo's effect on the moisture-based drivers of plant productivity. *EGUsphere*, 2024:1–54, 2024.

[166] Adam B. Smith and Richard W. Katz. US billion-dollar weather and climate disasters: Data sources, trends, accuracy and biases. *Natural Hazards*, 67:387–410, 2013.

[167] Brian J Soden, Richard T Wetherald, Georgiy L Stenchikov, and Alan Robock. Global cooling after the eruption of mount pinatubo: A test of climate feedback by water vapor. *science*, 296(5568):727–730, 2002.

[168] Peter Spirtes. Introduction to Causal Inference. *Journal of Machine Learning Research*, 11(54):1643—1662, 2010.

[169] Peter Spirtes and Clark Glymour. An Algorithm for Fast Recovery of Sparse Causal Graphs. *Social Science Computer Review*, 9(1):62–72, 1991.

[170] Peter Spirtes, Clark Glymour, and Richard Scheines. Causation, Prediction, and Search. *Lecture Notes in Statistics*, 1993.

[171] Peter Spirtes and Kun Zhang. Causal discovery and inference: concepts and recent methodological advances. *Applied Informatics*, 3(1):3, 2016.

[172] Andrew Steyer, Luca Bertagna, Graham Harper, Jerry Watkins, Irina Tezaur, and Diana Bull. In-situ data extraction for pathway analysis in an idealized atmosphere configuration of E3SM, 2024.

[173] T Storelvmo and N Herger. Cirrus cloud susceptibility to the injection of ice nuclei in the upper troposphere. *Journal of Geophysical Research: Atmospheres*, 119(5):2375–2389, 2014.

[174] Peter A Stott, Nathan P Gillett, Gabriele C Hegerl, David J Karoly, Dáithí A Stone, Xuebin Zhang, and Francis Zwiers. Detection and attribution of climate change: a regional perspective. *Wiley Interdisciplinary Reviews: Climate Change*, 1(2):192–211, 2010.

[175] Irina Tezaur, Kara Peterson, Amy Powell, John Jakeman, and Erika Roesler. Global sensitivity analysis using the ultra-low resolution energy exascale earth system model. *Journal of Advances in Modeling Earth Systems*, 14(8):e2021MS002831, 2022.

[176] L. W. Thomason, N. Ernest, L. Millan, L. Rieger, A. Bourassa, J.-P. Vernier, G. Manney, B. Luo, F. Arfeuille, and T. Peter. A global space-based stratospheric aerosol climatology: 1979–2016. *Earth System Science Data*, 10(1):469–492, 2018.

[177] Xavier-Andoni Tibau, Christian Reimers, Andreas Gerhardus, Joachim Denzler, Veronika Eyring, and Jakob Runge. A spatiotemporal stochastic climate model for benchmarking causal discovery methods for teleconnections. *Environmental Data Science*, 1:e12, 2022.

[178] Benjamin A. Toms, Elizabeth A. Barnes, and Imme Ebert-Uphoff. Physically interpretable neural networks for the geosciences: Applications to earth system variability. *Journal of Advances in Modeling Earth Systems*, 12(9):e2019MS002002, 2020.

[179] Kevin E. Trenberth and Aiguo Dai. Effects of mount pinatubo volcanic eruption on the hydrological cycle as an analog of geoengineering. *Geophysical Research Letters*, 34(15), 2007.

[180] J. Derek Tucker. *fdasrvf: Elastic Functional Data Analysis*, 2024. Python package version 2.6.1.

[181] J. Derek Tucker and Aymeric Stamm. *fdasrvf: Elastic Functional Data Analysis*, 2024. R package version 2.3.2.

[182] J. Derek Tucker and Drew Yarger. Elastic functional changepoint detection of climate impacts from localized sources. *Envirometrics*, 35:1–28, 2023.

[183] Benjamin M. Wagman, Hunter Brown, Diana Bull, Tom Ehrmann, Ben Hillman, Laura Swiler, Kara Peterson, and Joe Hollowed. Does prognostic volcanic aerosol enhance climate variability in earth system model simulations of volcanic eruptions? Technical Report SAND2024-12725o, Sandia National Laboratories, 2024.

[184] Duane E. Waliser and Catherine Gautier. A satellite-derived climatology of the ITCZ. *Journal of Climate*, 6(11):2162–2174, 1993.

[185] Hailong Wang, Philip J. Rasch, Richard C. Easter, Balwinder Singh, Rudong Zhang, Po-Lun Ma, Yun Qian, Steven J. Ghan, and Nathaniel Beagley. Using an explicit emission tagging method in global modeling of source-receptor relationships for black carbon in the arctic: Variations, sources, and transport pathways. *Journal of Geophysical Research: Atmospheres*, 119(22):12,888–12,909, 2014.

[186] Huanjing Wang, Qianxin Liang, John T Hancock, and Taghi M Khoshgoftaar. Feature selection strategies: a comparative analysis of SHAP-value and importance-based methods. *Journal of Big Data*, 11(1):44, 2024.

[187] Jerry Watkins, Luca Bertagna, Graham Harper, Andrew Steyer, Irina Tezaur, and Diana Bull. Entropy-based feature selection for capturing impacts in Earth system models with extreme forcing. in preparation for submission to J. Comput. Appl. Math., 2024.

[188] D. W. Waugh and T. M. Hall. Age of stratospheric air: Theory, observations, and models. *Review of Geophysics*, 40(4):1.1–1.26, 2002.

[189] M. Webb, C. Senior, S. Bony, and J. J. Morcrette. Combining erbe and isccp data to assess clouds in the hadley centre, ecmwf and lmd atmospheric climate models. *Climate Dynamics*, 17(12):905–922, 2001.

[190] Peter J Webster. The elementary monsoon. *Monsoon*, pages 3–32, 1986.

[191] Christopher R Wentland, Michael Weylandt, Laura P Swiler, Thomas S Ehrmann, and Diana Bull. Conditional multi-step attribution for climate forcings. *arXiv preprint arXiv:2409.01396*, 2024.

[192] Mike West and Jeff Harrison. *Bayesian Forecasting and Dynamic Models*. Springer Series in Statistics, 2nd edition, 1997.

[193] Michael Weylandt and Laura P Swiler. Beyond PCA: Additional dimension reduction techniques to consider in the development of climate fingerprints. *Journal of Climate*, 37(5):1723–1735, 2024.

[194] D. L. Williamson, J. G. Olson, and B. A. Boville. A comparison of semi-Lagrangian and Eulerian tropical climate simulations. *Monthly Weather Review*, 126:1001–1012, 1998.

[195] Robert C. J. Wills, David S. Battisti, Kyle C. Armour, Tapio Schneider, and Clara Deser. Pattern recognition methods to separate forced responses from internal variability in climate model ensembles and observations. *Journal of Climate*, 33(20):8693–8719, 2020.

[196] Donald P. Wylie and W. Paul Menzel. Eight years of high cloud statistics using hirs. *Journal of Climate*, 12(1):170 – 184, 1999.

[197] Donald P Wylie, W Paul Menzel, Harold M Woolf, and Kathleen I Strabala. Four years of global cirrus cloud statistics using hirs. *Journal of Climate*, 7(12):1972–1986, 1994.

[198] Donald P. Wylie, W. Paul Menzel, Harold M. Woolf, and Kathleen I. Strabala. Four years of global cirrus cloud statistics using HIRS. *Journal of Climate*, 7(12):1972 – 1986, 1994.

[199] Hong-Bo Xie, Yong-Ping Zheng, Jing-Yi Guo, and Xin Chen. Cross-fuzzy entropy: A new method to test pattern synchrony of bivariate time series. *Information Sciences*, 180(9):1715–1724, 2010.

[200] Lilian Yamamoto and Miguel Esteban. Vanishing Island States and sovereignty. *Ocean and Coastal Management*, 53:1–9, 2010.

[201] Y. Yang, S. Mou, H. Wang, P. Wang, B. Li, and H. Liao. Global source apportionment of aerosols into major emission regions and sectors over 1850–2017. *Atmospheric Chemistry and Physics*, 24(11):6509–6523, 2024.

[202] Y. Yang, H. Wang, S. J. Smith, R. Easter, P.-L. Ma, Y. Qian, H. Yu, C. Li, and P. J. Rasch. Global source attribution of sulfate concentration and direct and indirect radiative forcing. *Atmospheric Chemistry and Physics*, 17(14):8903–8922, 2017.

[203] Y. Yang, H. Wang, S. J. Smith, R. Zhang, S. Lou, Y. Qian, P.-L. Ma, and P. J. Rasch. Recent intensification of winter haze in china linked to foreign emissions and meteorology. *Scientific Reports*, 8:2791, 2018.

[204] Drew Yarger and J. Derek Tucker. Detecting changepoints in globally-indexed functional time series. *arXiv:2308.05915*, 2023.

[205] Andrew Zammit-Mangion and Noel Cressie. Frk: An R package for spatial and spatio-temporal prediction with large datasets. *arXiv preprint arXiv:1705.08105*, 2017.

[206] Andrew Zammit-Mangion and Noel Cressie. Introduction to fixed rank kriging: The R package. 2024.

[207] Z. Zhuo, I. Kirchner, S. Pfahl, and U. Cubasch. Climate impact of volcanic eruptions: the sensitivity to eruption season and latitude in MPI-ESM ensemble experiments. *Atmospheric Chemistry and Physics*, 21(17):13425–13442, 2021.

This page intentionally left blank.

APPENDIX A. CLDERA Simulation Campaign Output Data

Under the CLDERA project, we selected specialized high temporal-resolution output relevant to impacts from the Mt. Pinatubo eruption for both the atmosphere and land models in E3SMv2-SPA. In the following sections, the daily and subdaily output from the atmosphere and land used in the CLDERA simulation campaign are listed. Each entry includes the variable name, description, units, and number of model levels. Variables with level equal to one are provided at a single level or column integrated values. The entries also include the simulation set and ensemble numbers for which the variable is available. Not all fields were included in all simulation ensembles.

Name	Description	Units	Levels	In Simulations
ABSORB	Aerosol absorption	/m	72	all
AEROD_v	Total Aerosol Optical Depth in visible band	1	1	all
AOA	Age-of-air tracer	kg/kg	72	Limvar_src_tag ens 1-15
AODABS	Aerosol absorption optical depth 550 nm	-	1	all
AODSO401	Aerosol optical depth 550 nm SO4 - Mt. Pinatubo	-	1	Limvar_src_tag ens 1-15
AODSO402	Aerosol optical depth 550 nm SO4 - Cerro Hudson	-	1	Limvar_src_tag ens 1-15
AODSO403	Aerosol optical depth 550 nm SO4 - All other sources	-	1	Limvar_src_tag ens 1-15
AODVIS	Aerosol optical depth 550 nm	-	1	all
AREI	Average ice effective radius	micron	72	Fullvar ens 12-15, Limvar_src_tag ens 10-15
AREL	Average droplet effective radius	micron	72	Fullvar ens 12-15, Limvar_src_tag ens 10-15
BURDEN1	Aerosol burden mode 1	kg/m2	1	all
BURDEN2	Aerosol burden mode 2	kg/m2	1	all
BURDEN3	Aerosol burden mode 3	kg/m2	1	all
BURDEN4	Aerosol burden mode 4	kg/m2	1	all
BURDENSO401	Sulfate aerosol burden - Mt. Pinatubo	kg/m2	1	Limvar_src_tag ens 1-15
BURDENSO402	Sulfate aerosol burden - Cerro Hudson	kg/m2	1	Limvar_src_tag ens 1-15
BURDENSO403	Sulfate aerosol burden - All other sources	kg/m2	1	Limvar_src_tag ens 1-15
BURDENSO4	Sulfate aerosol burden	kg/m2	1	Fullvar ens 1-15, Limvar ens 1-6
CDNUMC	Vertically-integrated droplet concentration	1/m2	1	all
CLDHGH	Vertically-integrated high cloud	1	1	all
CLDICE	Grid box averaged cloud ice amount	kg/kg	72	all
CLDLIQ	rid box averaged cloud liquid amount	kg/kg	72	Fullvar ens 12-15
CLDLOW	Vertically-integrated low cloud	1	1	all
CLDMED	Vertically-integrated mid-level cloud	1	1	all
CLDTOT	Vertically-integrated total cloud	1	1	all
CLDTOT_ISCCP	Total Cloud Fraction Calculated by the ISCCP Simulator	percent	1	all
CLOUD	Cloud Fraction	1	72	Limvar_src_tag ens 1-15
EMIS	Cloud longwave emissivity	1`	72	Fullvar ens 12-15
E90j	E90 Tracer	kg/kg	72	Limvar_src_tag ens 1-15
EXTINCT	Aerosol extinction	1/m	72	Limvar_src_tag ens 1-15
FISCCP1_COSP	Grid-box fraction covered by each ISCCP D level cloud type	percent	72	Fullvar ens 12-15, Limvar_src_tag ens 10-15
FLDS	Downwelling longwave flux at surface	W/m2	1	all
FLDSC	Clearsky Downwelling longwave flux at surface	W/m2	1	all

Table A-1. Daily averaged output fields from E3SMv2-SPA atmosphere model for CLDERA simulation sets.

Name	Description	Units	Levels	In Simulations
FLNS	Net longwave flux at surface	W/m2	1	all
FLNSC	Clearsky net longwave flux at surface	W/m2	1	all
FLNT	Net longwave flux at top of model	W/m2	1	all
FLNTC	Clearsky net longwave flux at top of model	W/m2	1	all
FLUT	Upwelling longwave flux at top of model	W/m2	1	all
FLUTC	Clearsky upwelling longwave flux at top of model	W/m2	1	all
FSDS	Downwelling solar flux at surface	W/m2	1	all
FSDS_d2	Downwelling solar flux at surface - No Pinatubo aerosol	W/m2	1	Limvar_src_tag ens 1-15
FSDSC	Clearsky downwelling solar flux at surface	W/m2	1	all
FSNS	Net solar flux at surface	W/m2	1	all
FSNSC	Clearsky net solar flux at surface	W/m2	1	all
FSNT	Net solar flux at top of model	W/m2	1	all
FSNTC	Clearsky net solar flux at top of model	W/m2	1	all
FSNTOA	Net solar flux at top of atmosphere	W/m2	1	all
FSNTOAC	Clearsky net solar flux at top of atmosphere	W/m2	1	all
FSUTOA	Upwelling solar flux at top of atmosphere	W/m2	1	all
FSUTOAC	Clearsky upwelling solar flux at top of atmosphere	W/m2	1	all
H2SO4	H2SO4 concentration	mol/mol	72	Fullvar ens 1-15, Limvar ens 1-6
LHFLX	Surface latent heat flux	W/m2	1	all
LWCF	Longwave cloud forcing	W/m2	1	all
Mass_SO4	Sum of SO4 mass concentration	kg/kg	72	Fullvar ens 7-15, Limvar ens 6
Mass_SO401	Sum of SO4 mass concentration - Mt. Pinatubo	kg/kg	72	Limvar_src_tag ens 1-15
Mass_SO402	Sum of SO4 mass concentration - Cerro Hudson	kg/kg	72	Limvar_src_tag ens 1-15
Mass_SO403	Sum of SO4 mass concentration - All other sources	kg/kg	72	Limvar_src_tag ens 1-15
NIIMM	Activated Ice Number Concentation due to immersion freezing	1/m3	72	Fullvar ens 12-15, Limvar_src_tag ens 10-15
NIHF	Activated Ice Number Concentation due to homogenous freezing	1/m3	72	Fullvar ens 12-15, Limvar_src_tag ens 10-15
NUMICE	Grid box averaged cloud ice number	1/kg	72	all
NUMLIQ	Grid box averaged cloud liquid number	1/kg	72	Fullvar ens 12-15, Limvar_src_tag ens 10-15
OMEGA	Vertical velocity (pressure)	Pa/s	72	all
OMEGA500	Vertical velocity at 500 mbar pressure surface	Pa/s	1	all
PRECC	Convective precipitation rate (liq + ice)	m/s	1	all
PRECT	Total (convective and large-scale) precipitation rate (liq + ice)	m/s	1	all
PRECTMX	Maximum (convective and large-scale) precipitation rate (liq+ice)	m/s	1	all
PS	Surface pressure	Pa	1	all
PSL	Sea level pressure	Pa	1	all
PV	Ertel potential vorticity	m2 K/kg/s	72	Fullvar ens 7-15, Limvar ens 1-6, Limvar_src_tag ens 1-15
Q010	Specific Humidity at 010 mbar pressure surface	kg/kg	1	all
Q050	Specific Humidity at 050 mbar pressure surface	kg/kg	1	all
Q100	Specific Humidity at 100 mbar pressure surface	kg/kg	1	all
Q1000	Specific Humidity at 1000 mbar pressure surface	kg/kg	1	all
Q200	Specific Humidity at 200 mbar pressure surface	kg/kg	1	all
Q500	Specific Humidity at 500 mbar pressure surface	kg/kg	1	all
Q700	Specific Humidity at 700 mbar pressure surface	kg/kg	1	all
Q850	Specific Humidity at 850 mbar pressure surface	kg/kg	1	all
QBOT	Lowest model level water vapor mixing ratio	kg/kg	1	all
QREFHT	Reference height humidity	kg/kg	1	all
QREFHTMN	Minimum reference height humidity	kg/kg	1	all
QREFHTMX	Maximum reference height humidity	kg/kg	1	all
QRL	Longwave heating rate	K/s	72	all
QRS	Solar heating rate	K/s	72	all
REI	MG REI stratiform cloud effective radius ice	micron	72	Fullvar ens 12-15, Limvar_src_tag ens 10-15
REL	MG REL stratiform cloud effective radius liquid	micron	72	Fullvar ens 12-15, Limvar_src_tag ens 10-15
RHREFHT	Reference height relative humidity	1	1	all
SFso4_a1	so4_a1 surface flux	kg/m2/s	1	Fullvar ens 1-15, Limvar ens 1-6
SFso4_a2	so4_a2 surface flux	kg/m2/s	1	Fullvar ens 1-15, Limvar ens 1-6
SFso401_a1	so4_a1 surface flux - Mt. Pinatubo	kg/m2/s	1	Limvar_src_tag ens 1-15
SFso402_a1	so4_a1 surface flux - Cerro Hudson	kg/m2/s	1	Limvar_src_tag ens 1-15
SFso403_a1	so4_a1 surface flux - All other soures	kg/m2/s	1	Limvar_src_tag ens 1-15
SFso401_a2	so4_a2 surface flux - Mt. Pinatubo	kg/m2/s	1	Limvar_src_tag ens 1-15
SFso402_a2	so4_a2 surface flux - Cerro Hudson	kg/m2/s	1	Limvar_src_tag ens 1-15
SFso403_a2	so4_a2 surface flux - All other sources	kg/m2/s	1	Limvar_src_tag ens 1-15
SHFLX	Surface sensible heat flux	W/m2	1	all
SO2	SO2 concentration	mol/mol	72	Fullvar ens 1-15, Limvar ens 1-6
SO201	SO2 concentration - Mt. Pinatubo	mol/mol	72	Limvar_src_tag ens 1-15
SO202	SO2 concentration - Cerro Hudson	mol/mol	72	Limvar_src_tag ens 1-15
SO203	SO2 concentration - All other sources	mol/mol	72	Limvar_src_tag ens 1-15
SO2_XFRC	external forcing for SO2	molec/cm3/s	72	Fullvar ens 1-15, Limvar ens 1-6
SO201_XFRC	external forcing for SO2 - Mt. Pinatubo	molec/cm3/s	72	Limvar_src_tag ens 1-15
SO202_XFRC	external forcing for SO2 - Cerro Hudson	molec/cm3/s	72	Limvar_src_tag ens 1-15
SO203_XFRC	external forcing for SO2 - All other sources	molec/cm3/s	72	Limvar_src_tag ens 1-15

Table A-2. Daily averaged output fields from E3SMv2-SPA atmosphere model for CLDERA simulation sets (Continued).

Name	Description	Units	Levels	In Simulations
SOLIN	Solar insolation	W/m2	1	all
SOLL	Solar downward near infrared direct to surface	W/m2	1	all
SOLL_d2	Solar downward near infrared direct to surface - No Pinatubo	W/m2	1	Limvar_src_tag ens 1-15
SOLLD	Solar downward near infrared diffuse to surface	W/m2	1	all
SOLLD_d2	Solar downward near infrared diffuse to surface - No Pinatubo	W/m2	1	Limvar_src_tag ens 1-15
SOLS	Solar downward visible direct to surface	W/m2	1	all
SOLS_d2	Solar downward visible direct to surface - No Pinatubo	W/m2	1	Limvar_src_tag ens 1-15
SOLSD	Solar downward visible diffuse to surface	W/m2	1	all
SOLSD_d2	Solar downward visible diffuse to surface - No Pinatubo	W/m2	1	Limvar_src_tag ens 1-15
ST80_25j	ST80_25j tracer	kg/kg	72	Corrected initialization in Limvar_src_tag ens 10-15
SWCF	Shortwave cloud forcing	W/m2	1	all
T	Temperature	K	72	all
T001	Temperature at 1 mbar pressure surface	K	1	all
T002	Temperature at 2 mbar pressure surface	K	1	all
T005	Temperature at 5 mbar pressure surface	K	1	all
T010	Temperature at 10 mbar pressure surface	K	1	all
T025	Temperature at 25 mbar pressure surface	K	1	all
T050	Temperature at 50 mbar pressure surface	K	1	all
T100	Temperature at 100 mbar pressure surface	K	1	all
T1000	Temperature at 1000 mbar pressure surface	K	1	all
T150	Temperature at 150 mbar pressure surface	K	1	all
T200	Temperature at 200 mbar pressure surface	K	1	all
T250	Temperature at 250 mbar pressure surface	K	1	all
T500	Temperature at 500 mbar pressure surface	K	1	all
T700	Temperature at 700 mbar pressure surface	K	1	all
T850	Temperature at 850 mbar pressure surface	K	1	all
TGCLDIWP	Total grid-box cloud ice water path	kg/m2	1	all
TGCLDLWP	Total grid-box cloud liquid water path	kg/m2	1	all
TMH2SO4	H2SO4 column burden	kg/m2	1	Fullvar ens 1-15, Limvar ens 1-6
TMH2SO401	H2SO4 column burden - Mt. Pinatubo	kg/m2	1	Limvar_src_tag ens 1-15
TMH2SO402	H2SO4 column burden - Cerro Hudson	kg/m2	1	Limvar_src_tag ens 1-15
TMH2SO403	H2SO4 column burden - All other sources	kg/m2	1	Limvar_src_tag ens 1-15
TMQ	Total (vertically integrated) precipitable water	kg/m2	1	all
TMSO2	SO2 column burden	kg/m2	1	Fullvar ens 1-15, Limvar ens 1-6
TMSO201	SO2 column burden - Mt. Pinatubo	kg/m2	1	Limvar_src_tag ens 1-15
TMSO202	SO2 column burden - Cerro Hudson	kg/m2	1	Limvar_src_tag ens 1-15
TMSO203	SO2 column burden - All other sources	kg/m2	1	Limvar_src_tag ens 1-15
TOT_CLD_VISTAU	Total gbx cloud extinction visible sw optical depth	1	72	Fullvar ens 12-15, Limvar_src_tag ens 10-15
TOT_ICLD_VISTAU	Total in-cloud extinction visible sw optical depth	1	72	Fullvar ens 12-15, Limvar_src_tag ens 10-15
TOZ	Total column ozone	DU	1	all
TREFHT	Reference height temperature	K	1	all
TREFHTMN	Minimum reference height temperature over output period	K	1	all
TREFHTMX	Maximum reference height temperature over output period	K	1	all
TROPF_P	Tropopause Pressure (cold point)	Pa	1	all
TROPF_T	Tropopause Temperature (cold point)	K	1	all
TROPF_Z	Tropopause Height (cold point)	m	1	all
TROP_P	Tropopause Pressure	Pa	1	all
TROP_T	Tropopause Temperature	K	1	all
TROP_Z	Tropopause Height	m	1	all
TS	Surface temperature (radiative)	K	1	all
TUQ	Total (vertically integrated) zonal water flux	kg/m/s	1	all
TVQ	Total (vertically integrated) meridional water flux	kg/m/s	1	all
U	Zonal wind	m/s	72	all
U001	Zonal wind at 1 mbar pressure surface	m/s	1	all
U002	Zonal wind at 2 mbar pressure surface	m/s	1	all
U005	Zonal wind at 5 mbar pressure surface	m/s	1	all
U010	Zonal wind at 10 mbar pressure surface	m/s	1	all
U025	Zonal wind at 25 mbar pressure surface	m/s	1	all
U050	Zonal wind at 50 mbar pressure surface	m/s	1	all
U10	10m wind speed	m/s	1	all
U100	Zonal wind at 100 mbar pressure surface	m/s	1	all
U1000	Zonal wind at 1000 mbar pressure surface	m/s	1	all
U150	Zonal wind at 150 mbar pressure surface	m/s	1	all
U200	Zonal wind at 200 mbar pressure surface	m/s	1	all
U250	Zonal wind at 250 mbar pressure surface	m/s	1	all
U500	Zonal wind at 500 mbar pressure surface	m/s	1	all
U700	Zonal wind at 700 mbar pressure surface	m/s	1	all
U850	Zonal wind at 850 mbar pressure surface	m/s	1	all
V	Meridional wind	m/s	72	all

Table A-3. Daily averaged output fields from E3SMv2-SPA atmosphere model for CLDERA simulation sets (Continued).

Name	Description	Units	Levels	In Simulations
V010	Meridional wind at 10 mbar pressure surface	m/s	1	all
V050	Meridional wind at 50 mbar pressure surface	m/s	1	all
V100	Meridional wind at 100 mbar pressure surface	m/s	1	all
V1000	Meridional wind at 1000 mbar pressure surface	m/s	1	all
V200	Meridional wind at 200 mbar pressure surface	m/s	1	all
V500	Meridional wind at 500 mbar pressure surface	m/s	1	all
V700	Meridional wind at 700 mbar pressure surface	m/s	1	all
V850	Meridional wind at 850 mbar pressure surface	m/s	1	all
Z010	Geopotential Z at 10 mbar pressure surface	m	1	all
Z050	Geopotential Z at 50 mbar pressure surface	m	1	all
Z100	Geopotential Z at 100 mbar pressure surface	m	1	all
Z1000	Geopotential Z at 1000 mbar pressure surface	m	1	all
Z200	Geopotential Z at 200 mbar pressure surface	m	1	all
Z3	Geopotential Height (above sea level)	m	72	all
Z500	Geopotential Z at 500 mbar pressure surface	m	1	all
Z700	Geopotential Z at 700 mbar pressure surface	m	1	all
Z850	Geopotential Z at 850 mbar pressure surface	m	1	all

Table A-4. Daily averaged output fields from E3SMv2-SPA atmosphere model for CLDERA simulation sets (Continued).

Name	Description	Units	Levels	In Simulations
O3	O3 concentration	mol/mol	72	all
PS	Surface pressure	Pa	1	all
TROP_P	Tropopause Pressure	Pa	1	all

Table A-5. Daily instantaneous output fields from E3SMv2-SPA atmosphere model for CLDERA simulation sets.

Name	Description	Units	Levels	In Simulations
AEROD_v	Total Aerosol Optical Depth in visible band	1	1	all
FLDS	Downwelling longwave flux at surface	W/m2	1	all
FLNS	Net longwave flux at surface	W/m2	1	all
FLNT	Net longwave flux at top of model	W/m2	1	all
FLUT	Upwelling longwave flux at top of model	W/m2	1	all
FSDS	Downwelling solar flux at surface	W/m2	1	all
FSNS	Net solar flux at surface	W/m2	1	all
FSNT	Net solar flux at top of model	W/m2	1	all
OMEGA500	Vertical velocity at 500 mbar pressure surface	Pa/s	1	all
PRECC	Convective precipitation rate (liq + ice)	m/s	1	all
PRECT	Total (convective and large-scale) precipitation rate (liq + ice)	m/s	1	all
PSL	Sea level pressure	Pa	1	all
QREFHT	Reference height humidity	kg/kg	1	all
RHREFHT	Reference height relative humidity	1	1	all
SOLL	Solar downward near infrared direct to surface	W/m2	1	all
SOLLD	Solar downward near infrared diffuse to surface	W/m2	1	all
SOLS	Solar downward visible direct to surface	W/m2	1	all
SOLSD	Solar downward visible diffuse to surface	W/m2	1	all
T050	Temperature at 50 mbar pressure surface	K	1	all
T200	Temperature at 200 mbar pressure surface	K	1	all
T500	Temperature at 500 mbar pressure surface	K	1	all
TMQ	Total (vertically integrated) precipitable water	kg/m2	1	all
TREFHT	Reference height temperature	K	1	all
U200	Zonal wind at 200 mbar pressure surface	m/s	1	all
V200	Meridional wind at 200 mbar pressure surface	m/s	1	all
Z500	Geopotential Z at 500 mbar pressure surface	m	1	all

Table A-6. Six-hourly average output fields from E3SMv2-SPA atmosphere model for CLDERA simulation sets.

Name	Description	Units	Levels	In Simulations
AEROD_v	Total Aerosol Optical Depth in visible band	1	1	all
FLDS	Downwelling longwave flux at surface	W/m2	1	all
FLNS	Net longwave flux at surface	W/m2	1	all
FLNT	Net longwave flux at top of model	W/m2	1	all
FLUT	Upwelling longwave flux at top of model	W/m2	1	all
FSDS	Downwelling solar flux at surface	W/m2	1	all
FSNS	Net solar flux at surface	W/m2	1	all
FSNT	Net solar flux at top of model	W/m2	1	all
OMEGA500	Vertical velocity at 500 mbar pressure surface	Pa/s	1	all
PRECC	Convective precipitation rate (liq + ice)	m/s	1	all
PRECT	Total (convective and large-scale) precipitation rate (liq + ice)	m/s	1	all
PSL	Sea level pressure	Pa	1	all
QREFHT	Reference height humidity	kg/kg	1	all
RHREFHT	Reference height relative humidity	1	1	all
SOLL	Solar downward near infrared direct to surface	W/m2	1	all
SOLLD	Solar downward near infrared diffuse to surface	W/m2	1	all
SOLS	Solar downward visible direct to surface	W/m2	1	all
SOLSD	Solar downward visible diffuse to surface	W/m2	1	all
T050	Temperature at 50 mbar pressure surface	K	1	all
T200	Temperature at 200 mbar pressure surface	K	1	all
T500	Temperature at 500 mbar pressure surface	K	1	all
TMQ	Total (vertically integrated) precipitable water	kg/m2	1	all
TREFHT	Reference height temperature	K	1	all
U200	Zonal wind at 200 mbar pressure surface	m/s	1	all
V200	Meridional wind at 200 mbar pressure surface	m/s	1	all
Z500	Geopotential Z at 500 mbar pressure surface	m	1	all

Table A-7. Six-hourly average output fields from E3SMv2-SPA atmosphere model for CLDERA simulation sets.

Name	Description	Units	Levels	In Simulations
PRECC	Convective precipitation rate (liq + ice)	m/s	1	all
PRECT	Total (convective and large-scale) precipitation rate (liq + ice)	m/s	1	all
QFLX	Surface water flux	kg/m2/s	1	all
SHFLX	Surface sensible heat flux	W/m2	1	all
TREFHT	Reference height temperature	K	1	all
TUQ	Total (vertically integrated) zonal water flux	kg/m/s	1	all
TVQ	Total (vertically integrated) meridional water flux	kg/m/s	1	all

Table A-8. Three-hourly average output fields from E3SMv2-SPA atmosphere model for CLDERA simulation sets.

Name	Description	Units	Levels	In Simulations
EFLX_LH_TOT	total latent heat flux [+ to atm]	W/m^2	1	all
FGR	heat flux into soil/snow including snow melt and lake / snow light transmission	W/m^2	1	all
FIRA	net infrared (longwave) radiation	W/m^2	1	all
FIRE	emitted infrared (longwave) radiation	W/m^2	1	all
FPSN	photosynthesis	umol/m2s	1	all
FSA	absorbed solar radiation	W/m^2	1	all
FSD24	direct radiation (last 24hrs)	K	1	Fullvar ens 1-6, Limvar ens 1-6
FSDSND	direct nir incident solar radiation	W/m^2	1	Fullvar ens 7-15, Limvar_src_tag 1-15
FSDSNI	diffuse nir incident solar radiation	W/m^3	1	Fullvar ens 7-15, Limvar_src_tag 1-15
FSDSVD	direct vis incident solar radiation	W/m^2	1	Fullvar ens 7-15, Limvar_src_tag 1-15
FSDSVI	diffuse vis incident solar radiation	W/m^3	1	Fullvar ens 7-15, Limvar_src_tag 1-15
FSH	sensible heat	W/m^2	1	all
FSI24	indirect radiation (last 24hrs)	K	1	Fullvar ens 1-6, Limvar ens 1-6
FSNO	fraction of ground covered by snow	1	1	all
FSR	reflected solar radiation	W/m^2	1	all
H2OSNO	snow depth (liquid water)	mm	1	all
H2OSOI	volumetric soil water (vegetated landunits only)	mm3/mm3	15	all
QDRAI	sub-surface drainage	mm/s	1	all
QDRAI_XS	saturation excess drainage	mm/s	1	all
QH2OSFC	surface water runoff	mm/s	1	all
QIRRIG_REAL	actual water added through irrigation (surface + ground)	mm/s	1	all
QRUNOFF	total liquid runoff (does not include QSNWCPICE)	mm/s	1	all
QSNOMELT	snow melt	mm/s	1	all
QSOIL	Ground evaporation (soil/snow evaporation + soil/snow sublimation - dew)	mm/s	1	all
QTOPSOIL	water input to surface	mm/s	1	all
QVEGE	canopy evaporation	mm/s	1	all
QVEGT	canopy transpiration	mm/s	1	all
RAIN	atmospheric rain	mm/s	1	all
SNOW	atmospheric snow	mm/s	1	all
SNOW_DEPTH	snow height of snow covered area	m	1	all
SOILICE	soil ice (vegetated landunits only)	kg/m2	15	all
SOILLIQ	soil liquid water (vegetated landunits only)	kg/m2	15	all
SOILWATER_10CM	soil liquid water + ice in top 10cm of soil (veg landunits only)	kg/m2	1	all
TG	ground temperature	K	1	all
TLAI	total projected leaf area index	1	1	all
TSAI	total projected stem area index	1	1	all
TSOI	soil temperature (vegetated landunits only)	K	15	all
TV	vegetation temperature	K	1	all
TWS	total water storage	mm	1	all

Table A-9. Daily average output fields from E3SMv2-SPA land model for CLDERA simulation sets.

Name	Description	Units	Levels	In Simulations
EFLX_LH_TOT	total latent heat flux [+ to atm]	W/m^2	1	all
FPSN	photosynthesis	umol/m2s	1	all
FSH	sensible heat	W/m^2	1	all
QH2OSFC	surface water runoff	mm/s	1	all
QSOIL	Ground evaporation (soil/snow evaporation + soil/snow sublimation - dew)	mm/s	1	all
QVEGE	canopy evaporation	mm/s	1	all
QVEGT	canopy transpiration	mm/s	1	all
SOILWATER_10CM	soil liquid water + ice in top 10cm of soil (veg landunits only)	kg/m2	1	all

Table A-10. Six-hourly average output fields from E3SMv2-SPA land model for CLDERA simulation sets.

DISTRIBUTION

Email—Internal

Name	Org.	Sandia Email Address
Diana Bull	8931	dlbull@sandia.gov
Kara Peterson	1442	kjpeter@sandia.gov
Technical Library	1911	sanddocs@sandia.gov

This page intentionally left blank.

Sandia National Laboratories

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.