Characterizing climate pathways using feature importance on echo state networks

Katherine Goode

ASA Albuquerque Chapter Meeting

September 29, 2023

SAND2023-10060C

Sandia National Laboratories

Climate Interventions

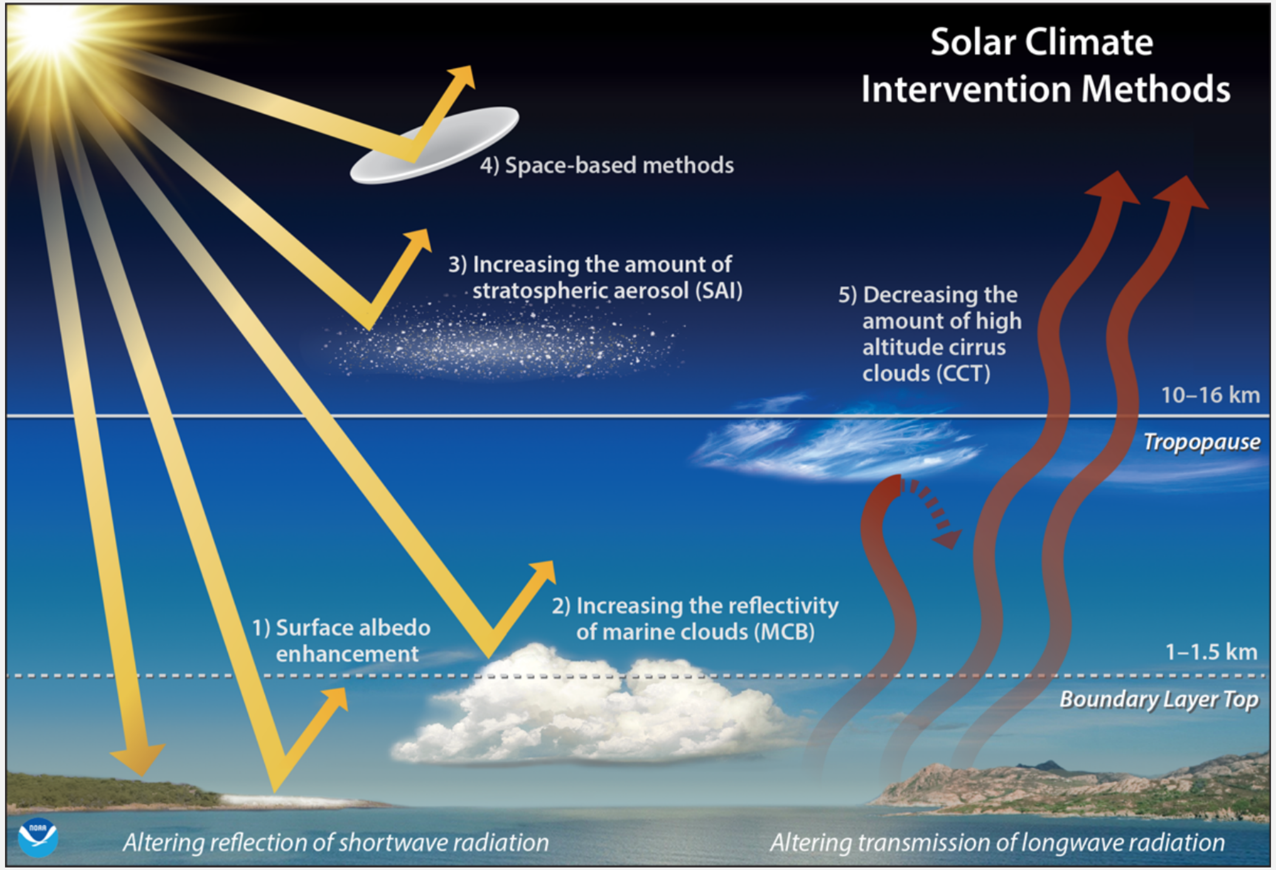


Image source: https://eos.org/science-updates/improving-models-for-solar-climate-intervention-research

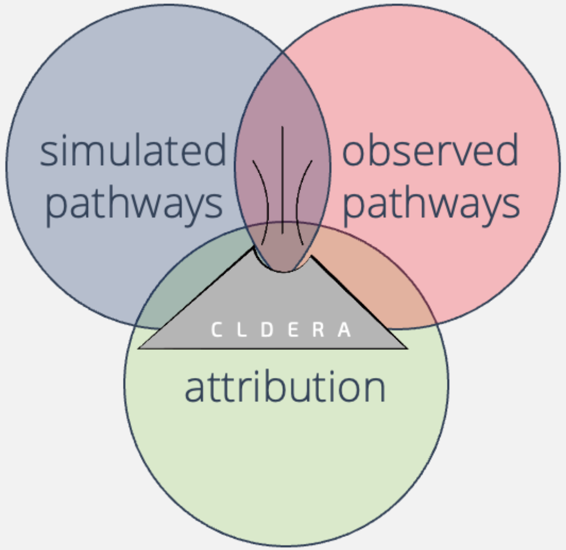
Threat of climate change has led to proposed interventions...

- Stratospheric aerosol injections
- Marine cloud brightening
- Cirrus cloud thinning
- etc.

What are the downstream effects of such mitigation strategies?

CLDERA Grand Challenge

Steve Scharre (ODNI), chair of USG's Climate Security Advisory Council, affirmed the need for attributable knowledge of downstream impacts and early indicators of those impacts



Need

Understand the "cause" (etiology) and extent of climate impacts

Hypothesis

Tracing pathways between sources and impacts will increase certainty of attribution

Pathways represent the simultaneous spatio-temporal dynamics and chain of physical processes acting to evolve the climate trajectory.

Outcome

Develop new climate statistical approaches for attributing positive and negative climate impacts

Novel foundational approach that can facilitate climate decisions

Image credit: CLDERA leadership

Climate Event Exemplar

- Mount Pinatubo eruption in 1991
- Released 18-19 Tg of sulfur dioxide
- Proxy for anthropogenic stratospheric aerosol injection

Image source: https://volcano.si.edu/volcano.cfm?vn=273083

Image source: https://pubs.usgs.gov/fs/1997/fs113-97/

Observational Thrust

Objective: Develop algorithms to characterize (i.e., quantify) relationships between climate variables related to a climate event using observational data

Research Composition:

Data Fusion: strategically source and fuse relevant data of varying resolutions and fidelities to create a "near-global" picture of the relevant processes

Space-Time Statistical Methods: adapt Bayesian hierarchical approaches to represent process dependencies and their dynamic spatio-temporal evolution over the first 3-4months post eruption employing novel changepoint detection algorithms

Hybrid Statistical & ML Methods: develop hierarchical statistical approaches embedded with ML techniques to trace secondary and tertiary temporally-lagged effects

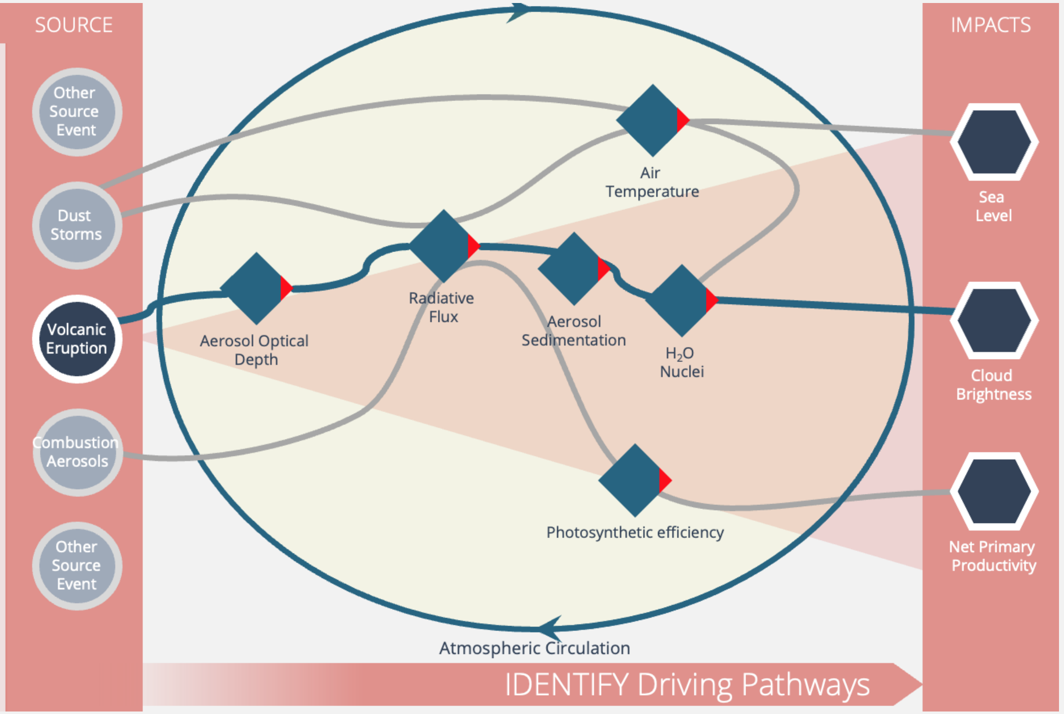


Image credit: CLDERA leadership

Mount Pintabuo Pathway

Sulfur dioxide

- Injection of sulfur dioxide (18-19 Tg) into atmosphere [1]

⬇

Aerosol optical depth (AOD)

- Vertically integrated measure of aerosols in air from surface to stratosphere [2]
- AOD increased as a result of injection of sulfur dioxide [1; 2]

⬇

Stratospheric temperature

- Temperatures at pressure levels of 30-50 mb rose 2.5-3.5 degrees centigrade compared to 20-year mean [3]

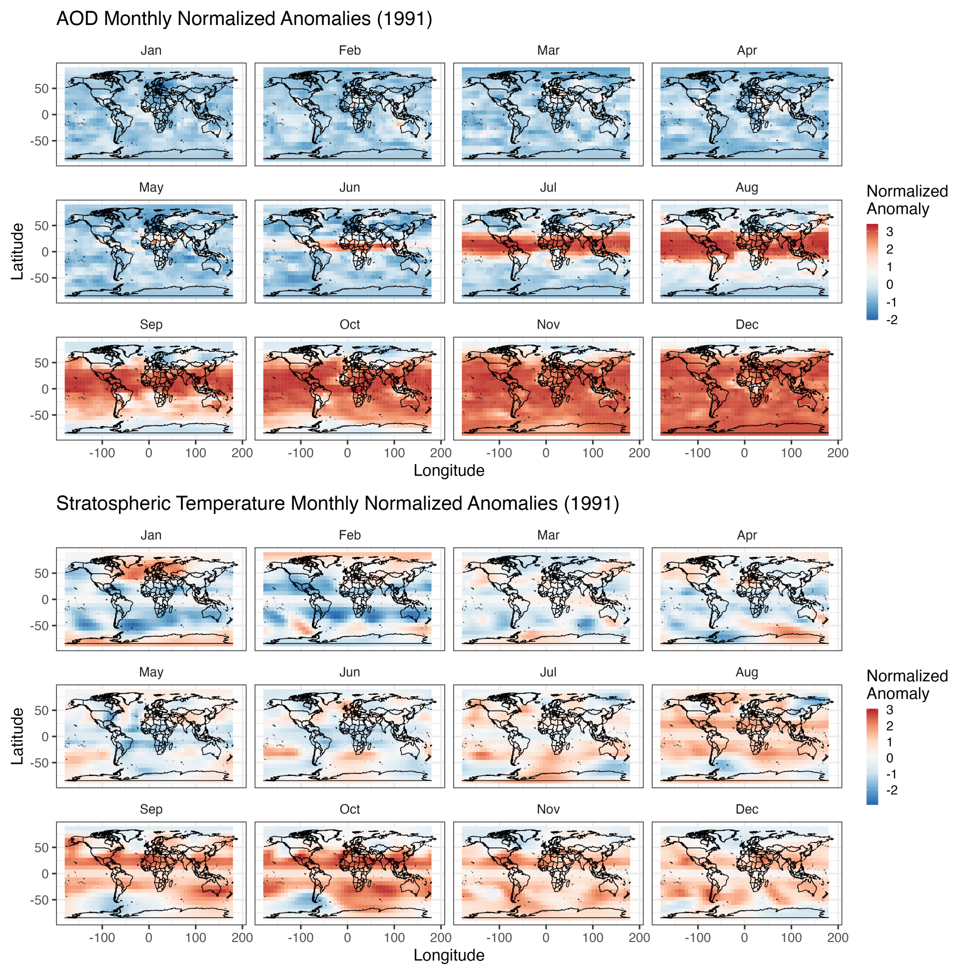


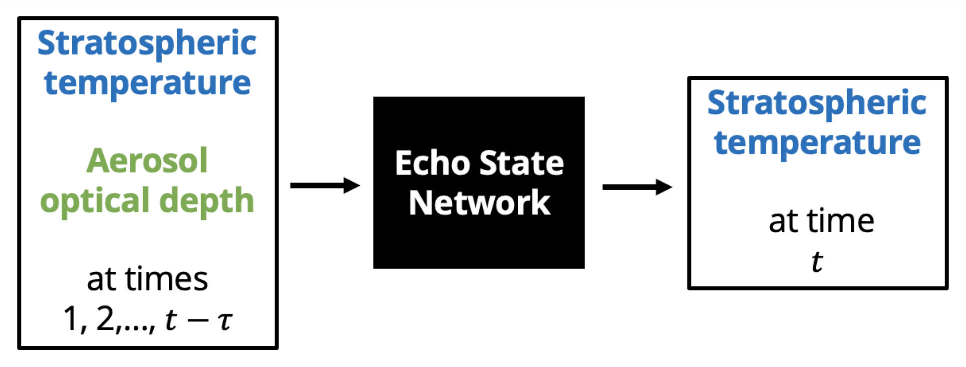
Figure generated using Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA- 2) data [4]

Our Approach

Use machine learning…

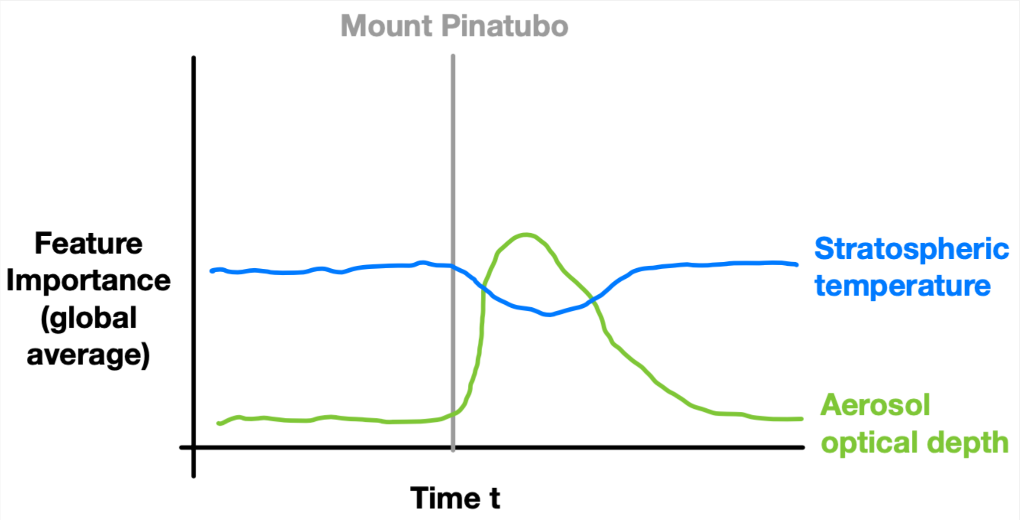
Step 1: Model climate event variables with echo state network

Allow complex machine learning model to capture complex variable relationships



Step 2: Quantify relationships via explainability

Apply feature importance to understand relationships captured by model



Remainder of Talk...

- Approach: Echo State Networks and Feature Importance

- Climate Application: Mount Pinatubo

- Conclusions and Future/Current Work

Acknowledgements

- **Sub-thrust team**: Daniel Ries, Kellie McClernon
- **Thrust lead**: Lyndsay Shand
- **Advise**: Gabriel Huerta, Derek Tucker

Approach

Echo State Networks and Feature Importance

Echo-State Networks

Overview

- Nonlinear machine learning model for temporal data
  - Sibling to recurrent neural network (RNN)
  - Known for forecasting with chaotic systems
- Computationally efficient model
  - Compared to RNNs and spatio-temporal statistical models
  - ESN reservoir parameters randomly sampled instead of estimated

Recent work

ESN for spatio-temporal forecasting:

- Sea surface temperature: McDermott and Wikle [5]
- Soil moisture: McDermott and Wikle [6]
- Wind power: Huang, Castruccio, and Genton [7]
- Air pollution: Bonas and Castruccio [8]

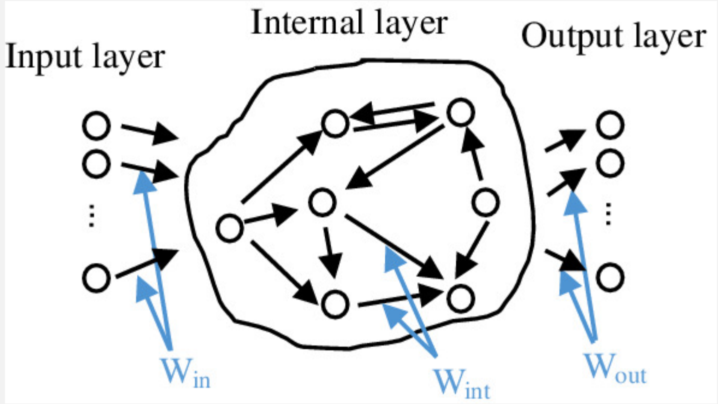


Image source: https://onlinelibrary.wiley.com/doi/10.1002/cpe.5719

Echo-State Networks

Single-Layer Echo State Network

Output stage: ridge regression

$$\mathbf{y}_t = \mathbf{V}\mathbf{h}_t + \boldsymbol{\epsilon}_t \qquad \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I})$$

Hidden stage: nonlinear stochastic transformation

$$\mathbf{h}_t = g_h\left(\frac{\nu}{|\lambda_w|}\mathbf{W}\mathbf{h}_{t-1} + \mathbf{U}\tilde{\mathbf{x}}_{t-\tau}\right)$$

$$\tilde{\mathbf{x}}_{t-\tau} = \left[\mathbf{x}'_{t-\tau}, \mathbf{x}'_{t-\tau-\tau^*}, \ldots, \mathbf{x}'_{t-\tau-m\tau^*}\right]'$$

Only parameters estimated are in $\mathbf{V}$.

Elements of $\mathbf{W}$ and $\mathbf{U}$ randomly sampled…

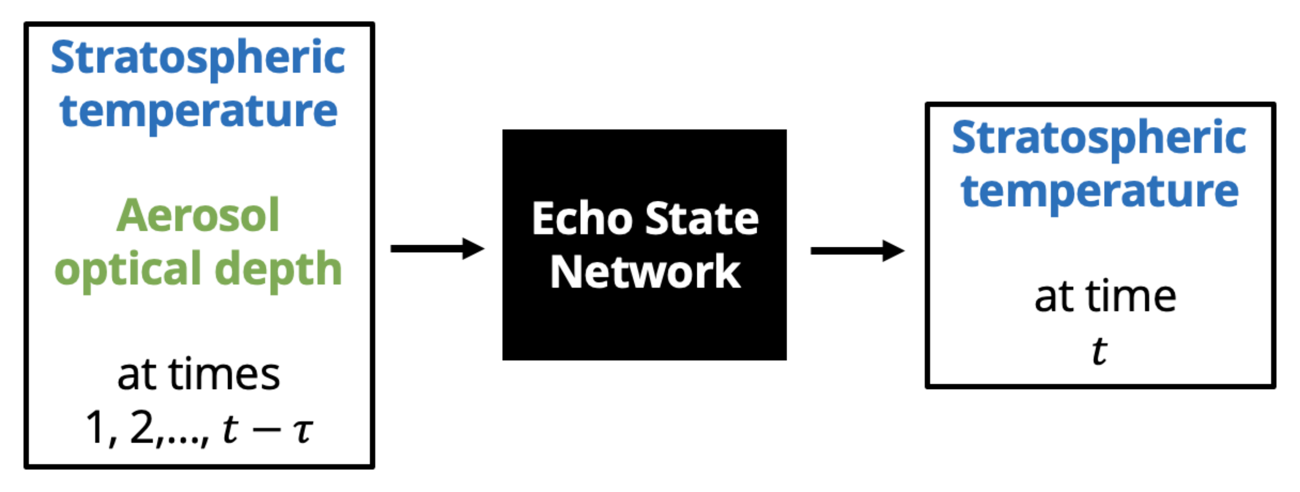
$$\mathbf{W}[h, c_w] = \gamma^w_{h,c_w}\mathrm{Unif}(-a_w, a_w) + (1 - \gamma^w_{h,c_w})\delta_0,$$
$$\mathbf{U}[h, c_u] = \gamma^u_{h,c_u}\mathrm{Unif}(-a_u, a_u) + (1 - \gamma^u_{h,c_u})\delta_0,$$

where

- $\gamma^w_{h,c_w} \sim Bern(\pi_w)$
- $\gamma^u_{h,c_u} \sim Bern(\pi_u)$
- δ_0 is a Dirac function

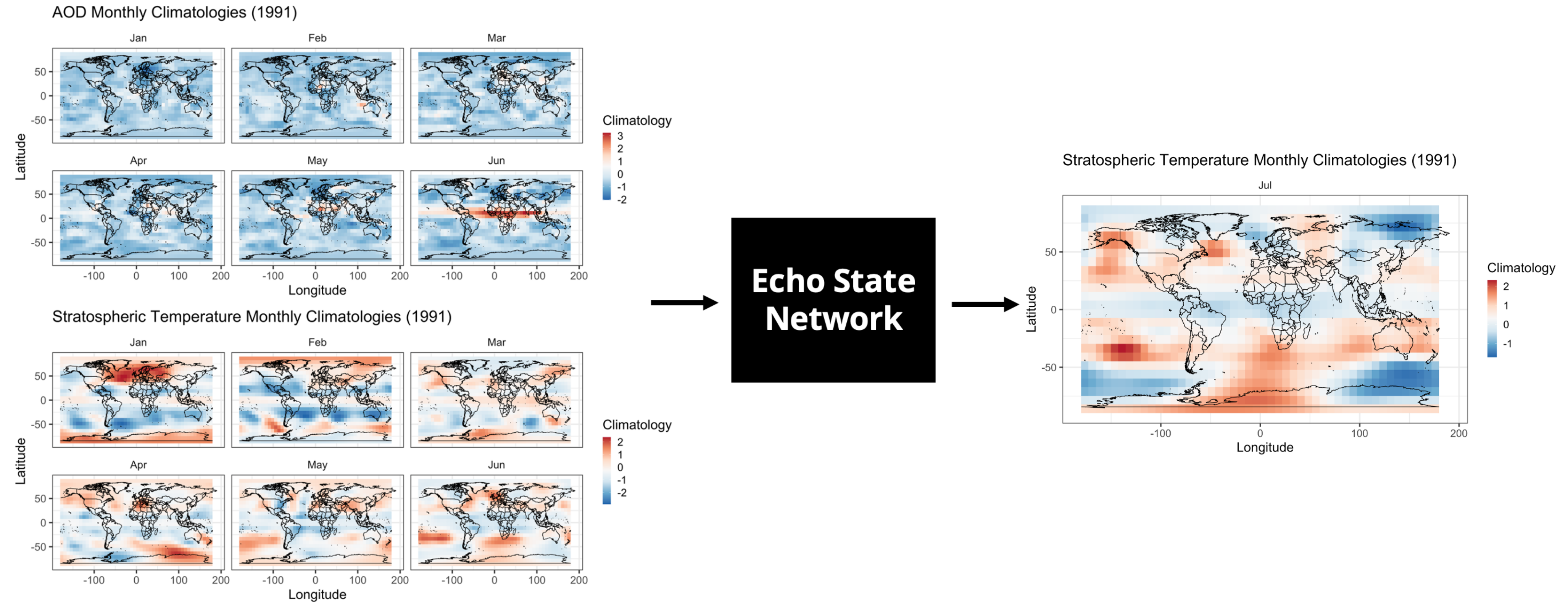
and values of a_w, a_u, π_w, and π_u are pre-specified and set to small values.

Echo-State Networks



Echo-State Networks: Spatio-Temporal Context

Recall that we are working with spatio-temporal data...



Echo-State Networks: Spatio-Temporal Context

Spatio-temporal processes at spatial locations $\{\mathbf{s}_i \in \mathcal{D} \subset \mathbb{R}^2; i = 1, \ldots, N\}$ over times $t = 1, \ldots, T$...

Output variable (e.g., stratospheric temperature):

$$\mathbf{Z}_{Y,t} = (Z_{Y,t}(\mathbf{s}_1), Z_{Y,t}(\mathbf{s}_2), \ldots, Z_{Y,t}(\mathbf{s}_N))'$$

Input variables (e.g., lagged aerosol optical depth and stratospheric temperature): For $k = 1, \ldots, K$

$$\mathbf{Z}_{k,t} = (Z_{k,t}(\mathbf{s}_1), Z_{k,t}(\mathbf{s}_2), \ldots, Z_{k,t}(\mathbf{s}_N))'$$

Stage	Formula	Description
Output data stage	$\mathbf{Z}_{Y,t} \approx \boldsymbol{\Phi}_Y \mathbf{y}_t$	Basis function decomposition (e.g., PCA)
Output stage	$\mathbf{y}_t = \mathbf{V}\mathbf{h}_t + \boldsymbol{\epsilon}_t$	Ridge regression
Hidden stage	$\mathbf{h}_t = g_h\left(\frac{\nu}{\lvert\lambda_w\rvert}\mathbf{W}\mathbf{h}_{t-1} + \mathbf{U}\tilde{\mathbf{x}}_{t-\tau}\right)$ $\tilde{\mathbf{x}}_{t-\tau} = \left[\mathbf{x}'_{t-\tau}, \mathbf{x}'_{t-\tau-\tau^*}, \ldots, \mathbf{x}'_{t-\tau-m\tau^*}\right]'$	Nonlinear stochastic transformation
Input data stage	$\mathbf{Z}_{k,t} \approx \boldsymbol{\Phi}_k \mathbf{x}_{k,t}$ $\quad \mathbf{x}_t = [\mathbf{x}'_{1,t}, \ldots, \mathbf{x}'_{K,t}]'$	Basis function decomposition (e.g., PCA)

Feature Importance for ESNs

Goal

- Feature importance aims to quantify effect of input variable on a model's predictions

Background

- Permutation feature importance [9]
- Pixel absence affect with ESNs [10]
- Temporal permutation feature importance [11]

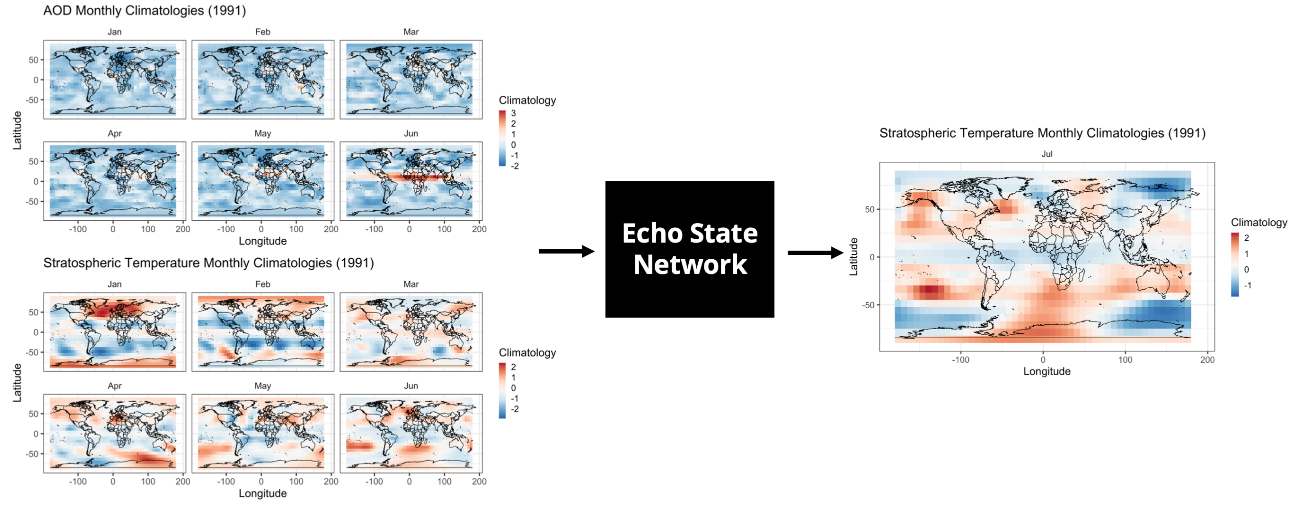
Our Work

- Adapt for ESNs in context of spatio-temporal data

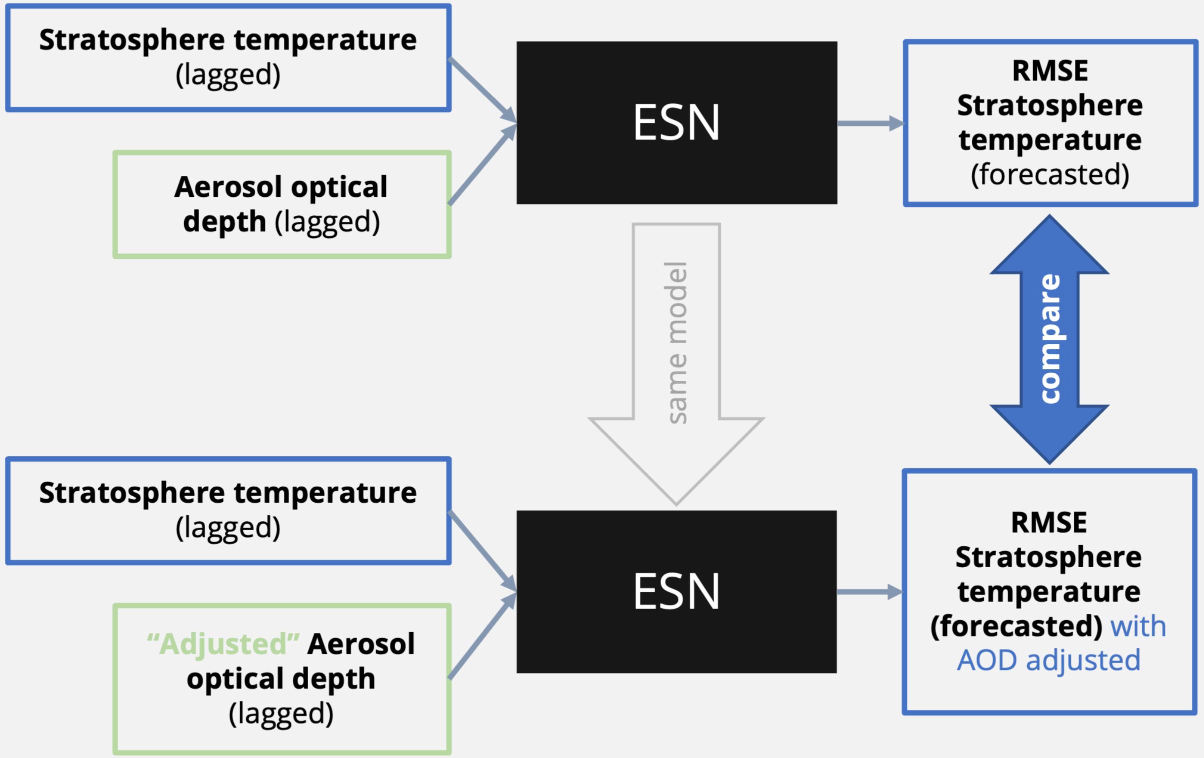
In particular…

Compute feature importance on trained ESN model for:

- input variable over block of times
- on forecasts of response variable at a time



Feature Importance for ESNs



Concept: "Adjust" inputs at times(s) of interest and quantify effect on model performance

- **Permute values**: spatio-temporal permutation feature importance (stPFI)
- **Set values to zero**: spatio-temporal zeroed feature importance (stZFI)

Feature Importance: Difference in RMSEs from "adjusted" and observed spatial predictions:

$$RMSE_{adj,t} - RMSE_{obs,t}$$

Interpretation: Large feature importance indicates "adjusted" inputs lead to a decrease in model performance indicating the model uses those inputs for prediction (i.e., inputs 'important' to model)

Climate Application

Mount Pinatubo

Mount Pinatubo Example: Data

Source

- Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA- 2)

Training Years

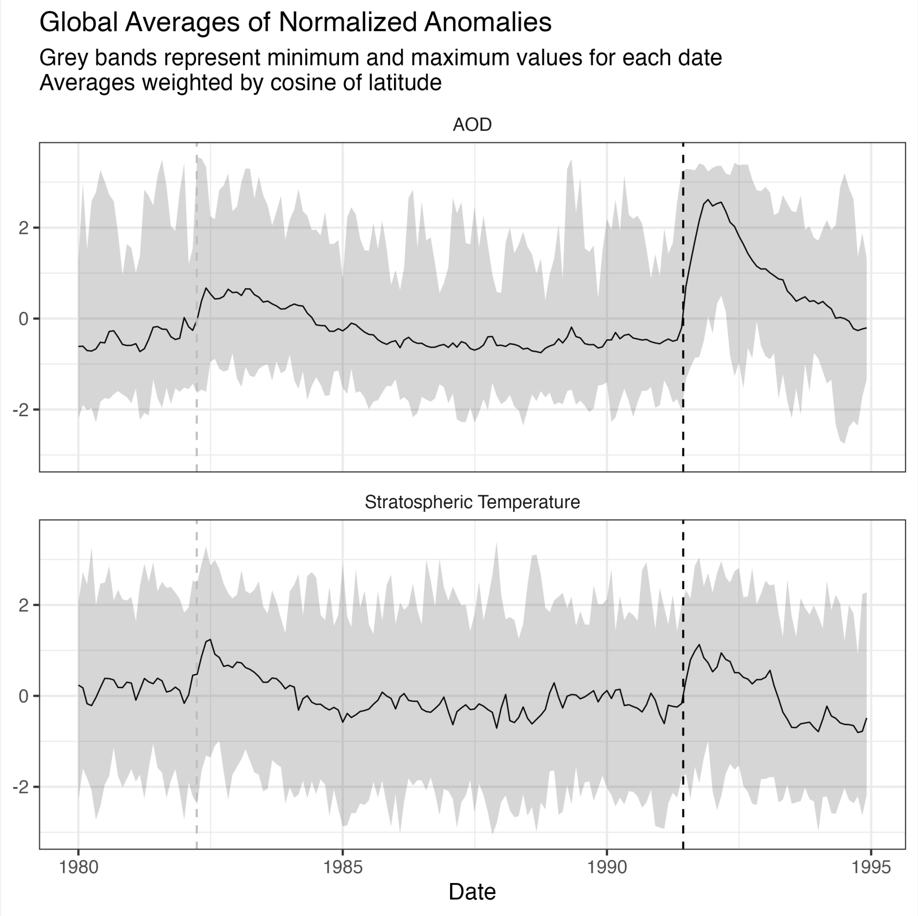
- 1980 to 1995
- Includes eruptions of Mount Pinatubo (1991) and El Chichón (1982)

Time Interval

- Monthly

Latitudes

- -86 to 86 degrees



Mount Pinatubo Example: Model

ESN Output

- Stratospheric Temperature (50mb)

ESN Inputs

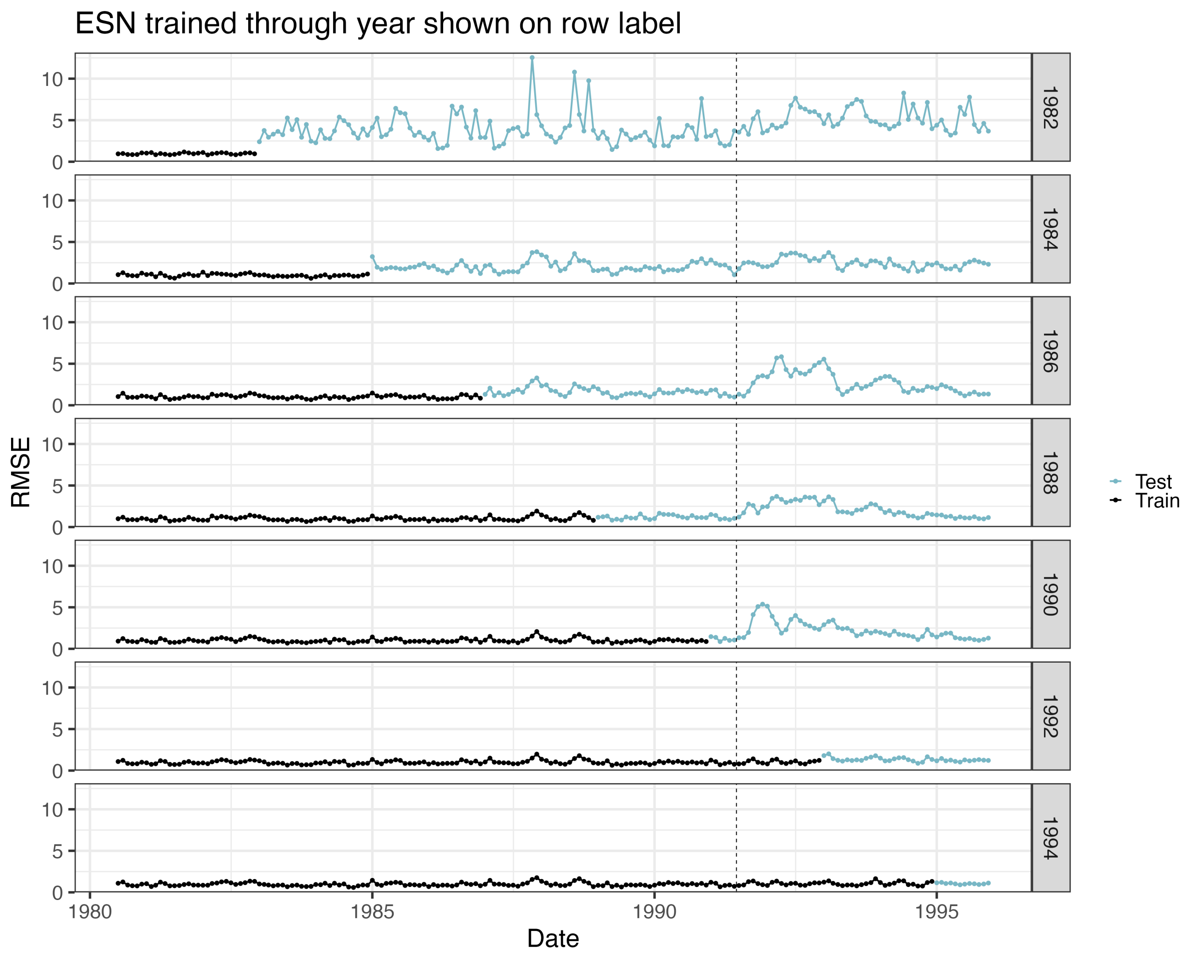
- Lagged Stratospheric Temperature (50mb)
- Lagged AOD

Forecast Lag

- One month

Preprocessing (all variables)

- Normalized Anomalies
- Principal components (first 5)



Mount Pinatubo Example: Feature Importance

- Block size of 3 months
  - e.g., How important are May, June, and July in 1991 for forecasting August in 1991?

- Using weighted RMSE to compute feature importance (weighted by cos latitude):

$$\text{weighted RMSE}_t = \sqrt{\frac{\sum_{loc} w_{loc}(y_{t,loc} - \hat{y}_{t,loc,adj})^2}{\sum_{loc} w_{loc}}} - \sqrt{\frac{\sum_{loc} w_{loc}(y_{t,loc} - \hat{y}_{t,loc})^2}{\sum_{loc} w_{loc}}}$$

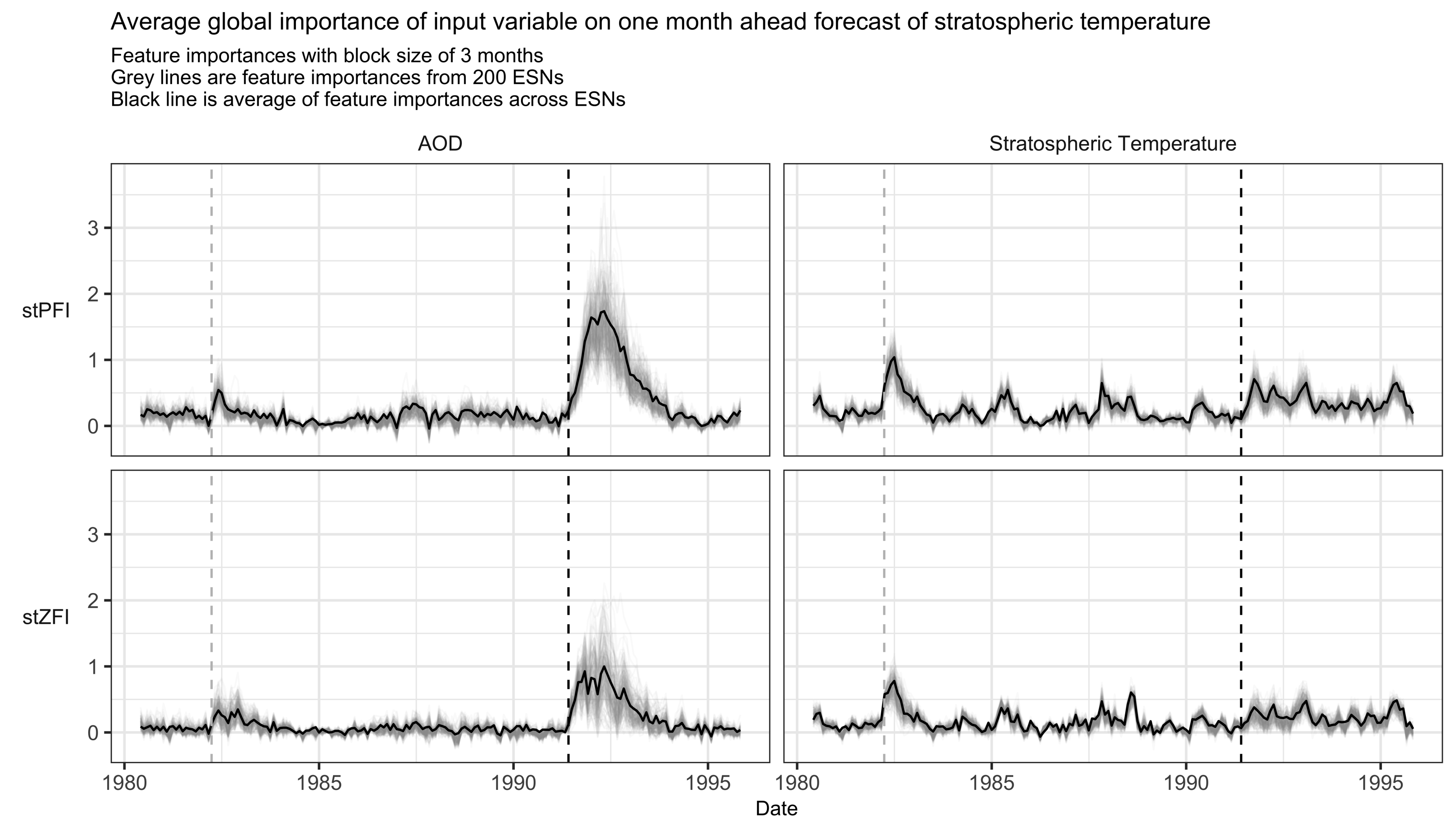
Mount Pinatubo Example: Feature Importance

Key Point

Peak of importance for AOD (and lack of peak of importance for lagged stratospheric temperatures), provides evidence that volcanic eruption impact on temperature can be traced through AOD

FI Metric

Weighted RMSE (weighted by cosine of the latitude)



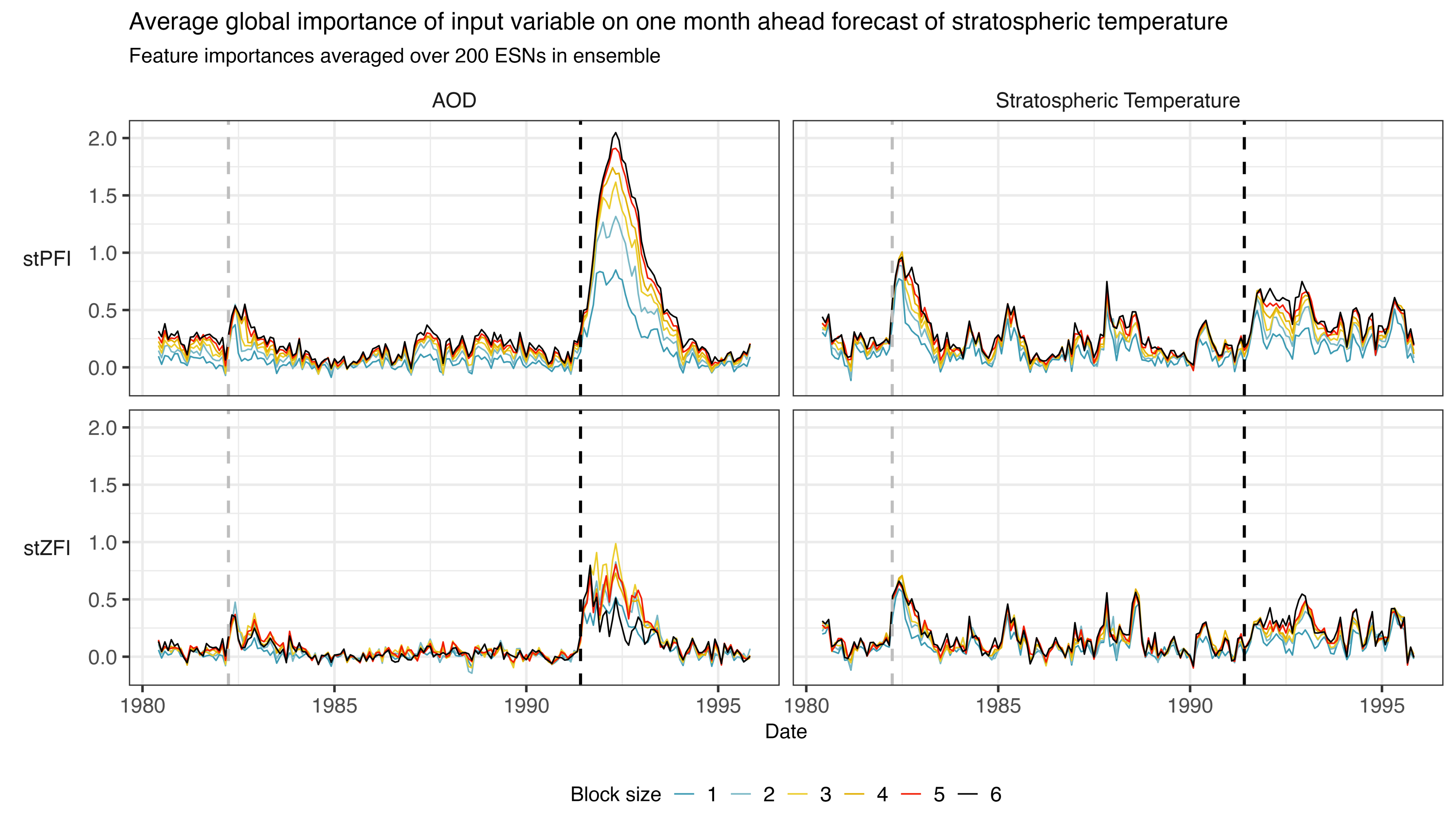
Mount Pinatubo Example: Feature Importance

Key Point

Larger block size (mostly) leads to larger feature importance magnitudes and less noisy trends

FI Metric

Weighted RMSE (weighted by cosine of the latitude)



Conclusions and Future Work

Summary and Conclusions

Key Take Aways

- Interested in quantifying relationships between climate variables associated with pathway of climate event
- Motivated by increasing possibility of climate interventions
- Our machine learning approach:
    - Use ESN to model variable relationships
    - Understand variable relationships using proposed spatio-temporal feature importance
- Approach provided evidence of AOD being an intermediate variable in Mount Pinatubo climate pathway affecting stratospheric temperature

Future (Current) Work

ESN extensions

- ESN ensembles
- Addition of multiple layers

Feature importance

- Implement proposed retraining technique [12] to lessen detection of spurious relationships due to correlation
- Adapt to visualize on spatial scale
- Comparison to other newly proposed explainability techniques for ESNs (layer-wise relevance propagation) [13]

Mount Pinatubo application

- Inclusion of additional pathway variables (e.g., SO2, radiative flux, surface temperature)
- Importance of grouped variables
- Application to climate simulation ensembles
- Use of climate simulation ensembles for method assessment

References

[1] S. Guo, G. J. Bluth, W. I. Rose, et al. "Re-evaluation of SO2 *release of the 15 June 1991 Pinatubo eruption using ultraviolet and infrared satellite sensors". In: _Geochemistry, Geophysics, Geosystems* 5 (4 2004), pp. 1-31. DOI: 10.1029/2003GC000654.

[2] M. Sato, J. E. Hansen, M. P. McCormick, et al. "Stratospheric aerosol optical depths, 1850-1990". In: *Journal of Geophysical Research: Atmospheres* 98.D12 (1993), pp. 22987-22994. DOI: https://doi.org/10.1029/93JD02553. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/93JD02553. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/93JD02553.

[3] K. Labitzke and M. McCormick. "Stratospheric temperature increases due to Pinatubo aerosols". In: *Geophysical Research Letters* 19 (2 1992), pp. 207-210. DOI: 10.1029/91GL02940.

[4] R. Gelaro, W. McCarty, M. J. Suarez, et al. "The ModernEra Retrospective Analysis for Research and Applications, Version 2 (MERRA-2)". In: *Journal of Climate* 30 (14 2017), pp. 5419-5454. DOI: 10.1175/JCLI-D-16-0758.1.

[5] P. L. McDermott and C. K. Wikle. "An ensemble quadratic echo state network for non-linear spatio-temporal forecasting". In: *Stat* 6.1 (2017), pp. 315-330.

[6] P. L. McDermott and C. K. Wikle. "Deep echo state networks with uncertainty quantification for spatio-temporal forecasting". In: *Environmetrics* 30.3 (2019). ISSN: 1180-4009. DOI: 10.1002/env.2553.

[7] H. Huang, S. Castruccio, and M. G. Genton. "Forecasting high-frequency spatio-temporal wind power with dimensionally reduced echo state networks". In: *Journal of the Royal Statistical Society Series C: Applied Statistics* 71.2 (2022), pp. 449-466.

[8] M. Bonas and S. Castruccio. "Calibration of SpatioTemporal forecasts from citizen science urban air pollution data with sparse recurrent neural networks". In: *The Annals of Applied Statistics* 17.3 (2023), pp. 1820-1840.

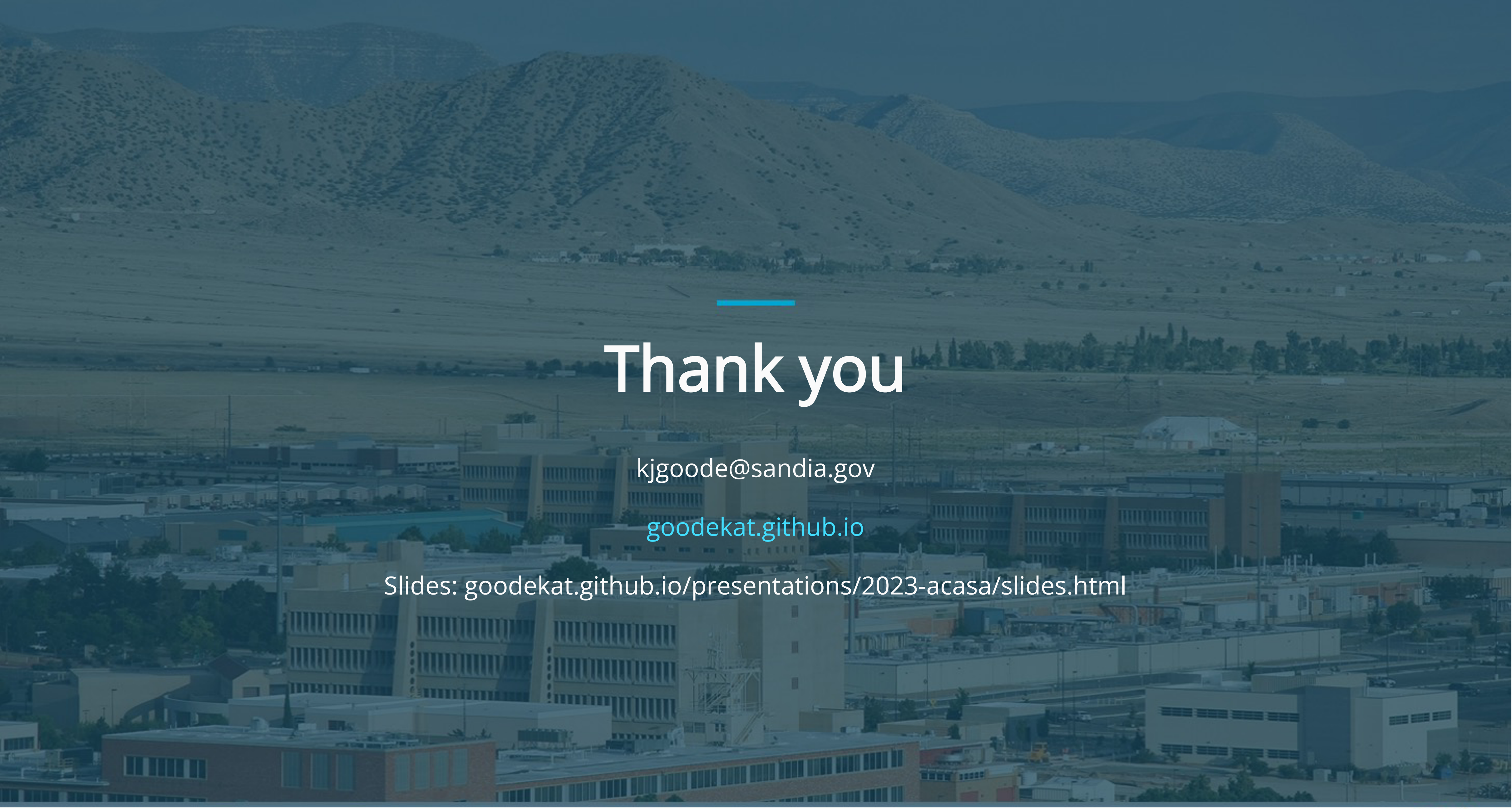
[9] A. Fisher, C. Rudin, and F. Dominici. "All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously". In: *Journal of Machine Learning Research*. 177 20 (2019), pp. 1-81. eprint: 1801.01489. URL: http://jmlr.org/papers/v20/18-760.html.

[10] A. B. Arrieta, S. Gil-Lopez, I. Laña, et al. "On the post-hoc explainability of deep echo state networks for time series forecasting, image and video classification". In: *Neural Computing and Applications* 34.13 (2022), pp. 10257-10277. ISSN: 0941-0643. DOI: 10.1007/s00521-021-06359-y.

[11] A. Sood and M. Craven. "Feature Importance Explanations for Temporal Black-Box Models". In: *arXiv* (2021). DOI: 10.48550/arxiv.2102.11934. eprint: 2102.11934.

[12] G. Hooker, L. Mentch, and S. Zhou. "Unrestricted permutation forces extrapolation: variable importance requires at least one more model, or there is no free variable importance". In: *Statistics and Computing* 31 (2021), pp. 1-16.

[13] M. Landt-Hayen, P. Kröger, M. Claus, et al. "Layer-Wise Relevance Propagation for Echo State Networks Applied to Earth System Variability". In: *Signal, Image Processing and Embedded Systems Trends*. Ed. by D. C. Wyld. Computer Science & Information Technology (CS & IT): Conference Proceedings 20. ARRAY(0x55588c8d8680), 2022, pp. 115-130. ISBN: 978-1-925953-80-0. DOI: doi:10.5121/csit.2022.122008. URL: https://doi.org/10.5121/csit.2022.122008.



Thank you

kjgoode@sandia.gov

goodekat.github.io

Slides: goodekat.github.io/presentations/2023-acasa/slides.html

Back-Up Slides

Feature Importance: Spatio-Temporal Context

Compute FI on the trained ESN model for…

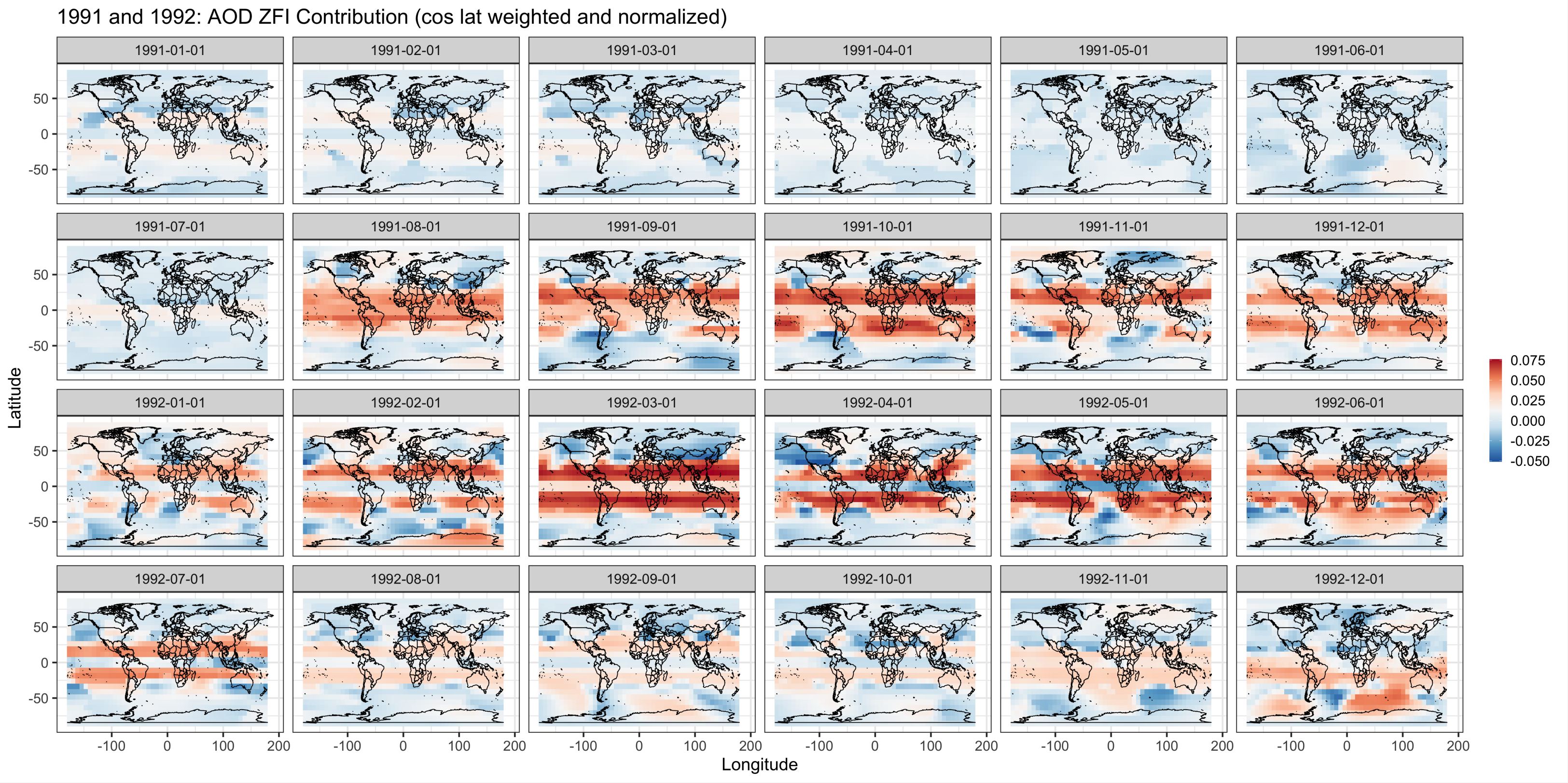
- spatio-temporal input variable k
- over the block of times $\{t, t-1, \ldots, t-b+1\}$
- on the forecasts of the spatio-temporal response variable at time $t+\tau$.

	$x_{1,t,1}$	…	x_{1,t,P_1}	$x_{2,t,1}$	…	x_{2,t,P_2}	…	$x_{K,t,1}$	…	x_{K,t,P_K}
$t=1$										
$t=2$										
$t=3$										
$t=4$										
$t=5$										
…										
$t=T$										

	$y_{1,t}$	…	$y_{Q,t}$
$t=1$			
$t=2$			
$t=3$			
$t=4$			
$t=5$			
…			
$t=T$			

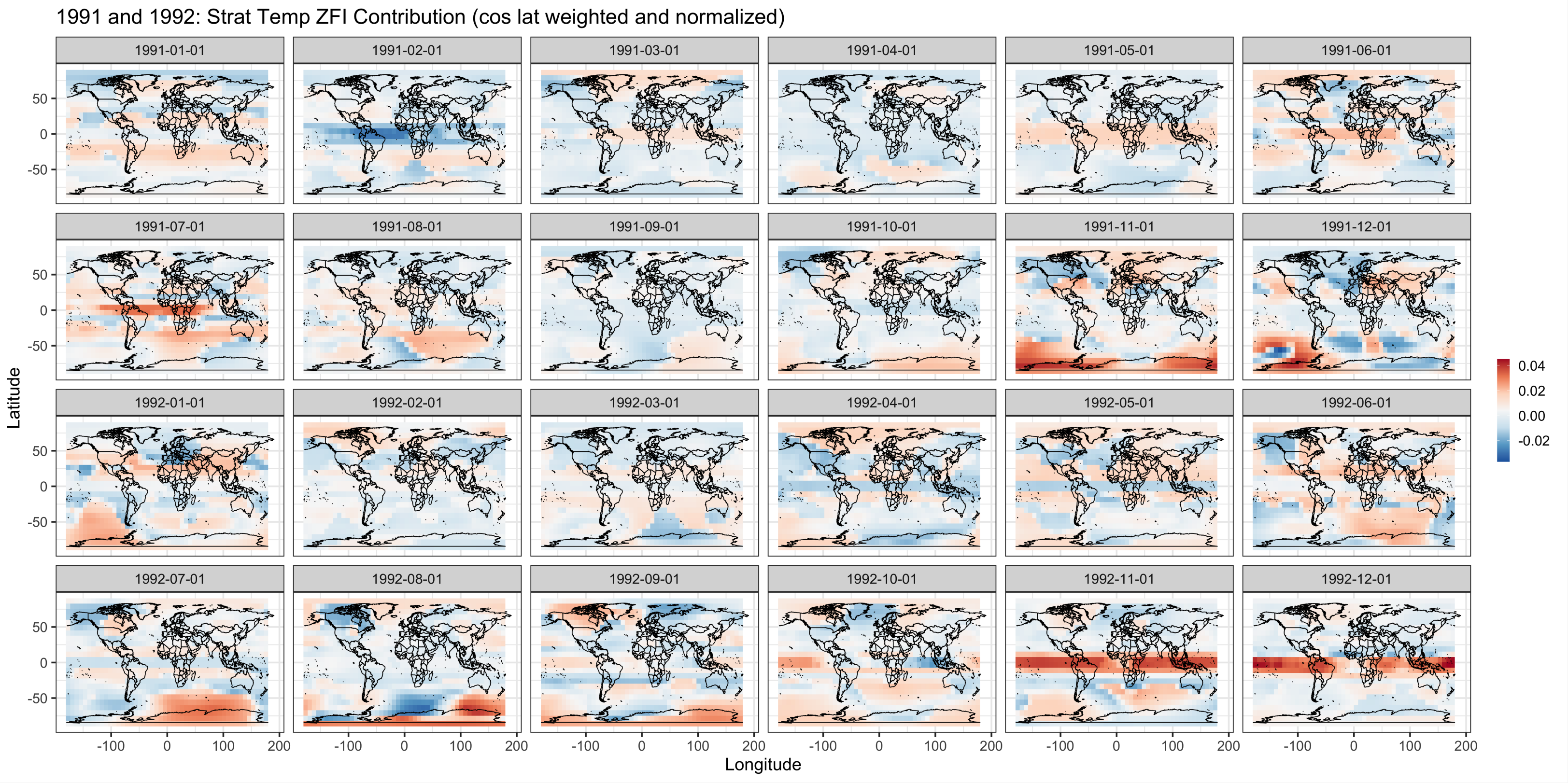
Feature Importance: Spatio-Temporal Context

Contribution of spatial locations to ZFI: $\sqrt{\dfrac{w_{loc}(y_{t,loc}-\hat{y}_{t,loc,zeroed})^2}{\sum_{loc} w_{loc}}} - \sqrt{\dfrac{w_{loc}(y_{t,loc}-\hat{y}_{t,loc})^2}{\sum_{loc} w_{loc}}}$



Feature Importance: Spatio-Temporal Context

Contribution of spatial locations to ZFI: $\sqrt{\frac{w_{loc}(y_{t,loc}-\hat{y}_{t,loc,zeroed})^2}{\sum_{loc} w_{loc}}} - \sqrt{\frac{w_{loc}(y_{t,loc}-\hat{y}_{t,loc})^2}{\sum_{loc} w_{loc}}}$



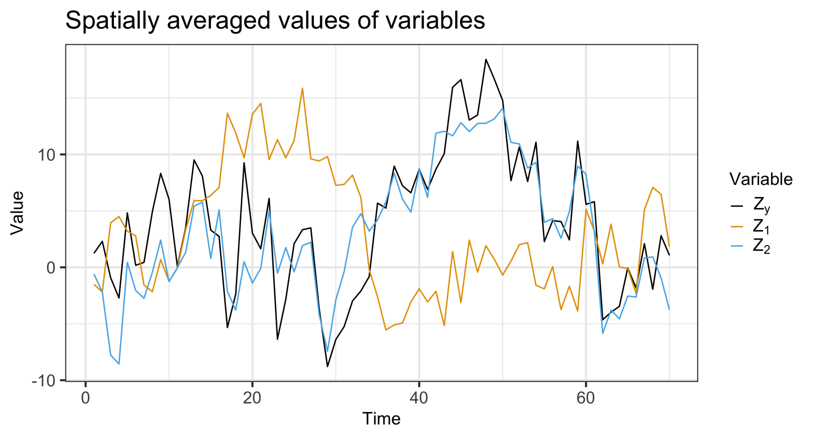
Simulated Data Demonstration

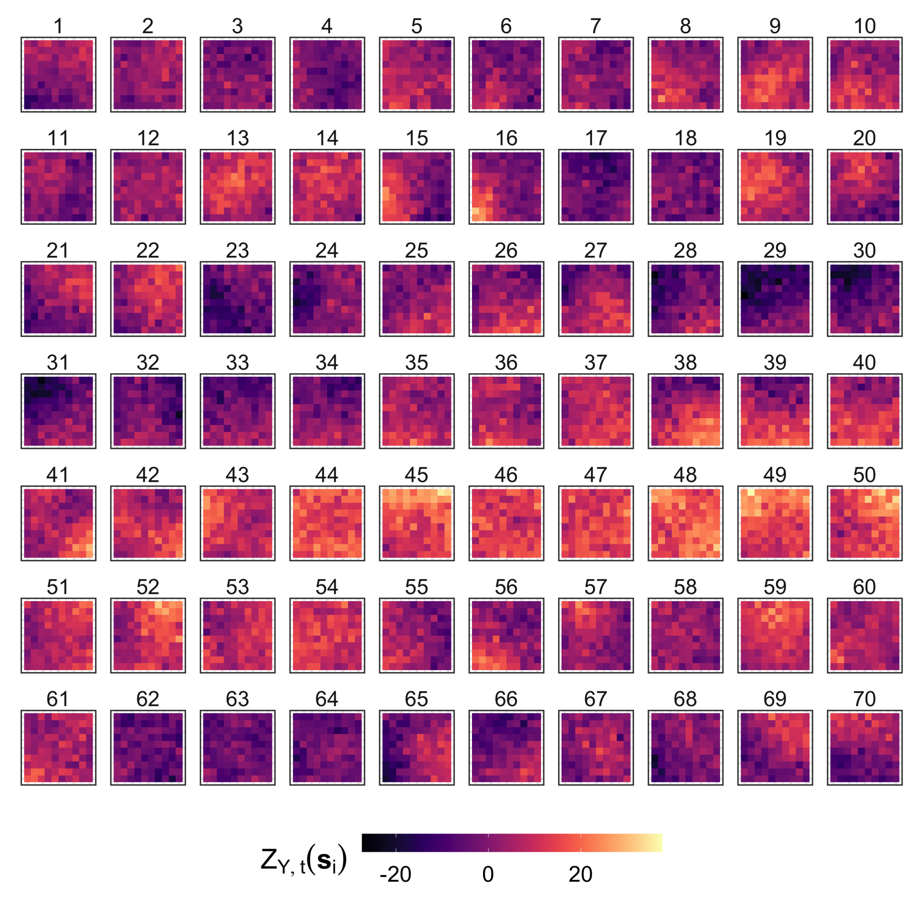
Simulated response

$$Z_{Y,t}(\mathbf{s}_i) = Z_{2,t}(\mathbf{s}_i)\beta + \delta_t(\mathbf{s}_i) + \epsilon_t(\mathbf{s}_i)$$

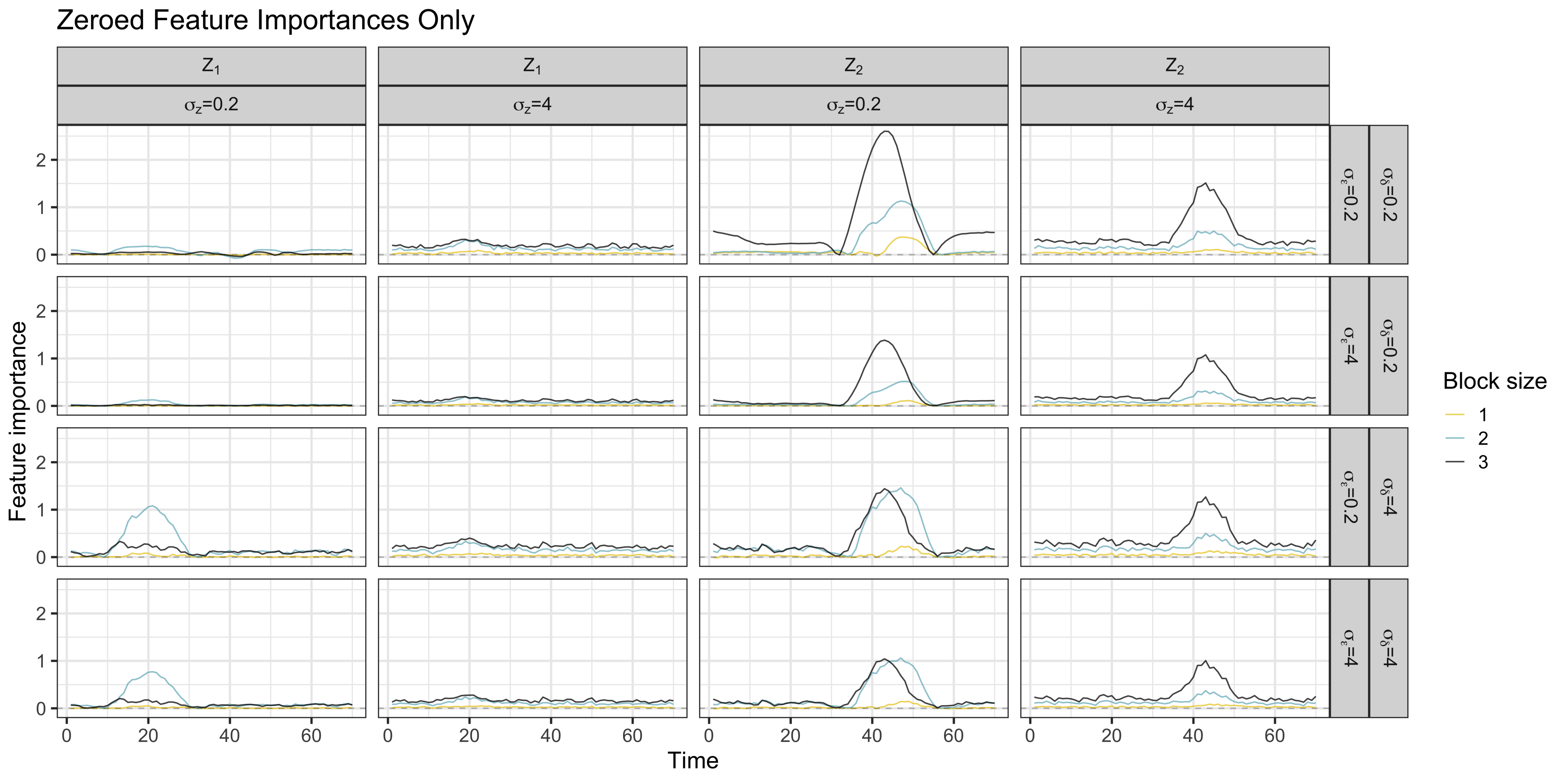
where

- $Z_{2,t}$ spatio-temporal covariate
- $\delta_t(\mathbf{s}_i)$ spatio-temporal random effect
- $\epsilon_t(\mathbf{s}_i) \overset{iid}{\sim} N(0, \sigma_\epsilon^2)$





Simulated Data: Effect of Variability on FI



Effect of Correlation on FI

Effect of Correlation on PFI

Correlation between features can lead to biased PFI values dues to the model being forced to extrapolate

- When a correlated variable is permuted, it can lead to observations not in the training data
- Model is forced to extrapolate for that observation
- **Extrapolation can lead to a major effect on prediction making a variable seem more important than it is**

Example

Data is simulated so that X1 affects Y but X2 does not:

(Left) Within training data (stars) random forest correctly determines relationship between X1, X2, and Y (contour lines) but incorrect outside of training data

(Right) When X2 is permuted, observation could land outside training data and lead to change in prediction (i.e., large PFI)

Source: Hooker, Mentch, and Zhou (2021)