Explainable Echo-State Networks for Climate Data Applications

Presenter: Katherine Goode

Sandia National Laboratories

Overview

Motivation

- Echo-State Networks (ESNs) are being more commonly used to model climate data due to their ability to capture non-linear relationships in spatio-temporal data [1-4].
- However, ESNs lack interpretability due to algorithmic complexity (i.e., "black box" models).
- There has been recent interest in using explainability approaches for climate applications to provide insight into black-box machine learning models [5-6].
- We present work that contributes to the emerging field by exploring the application of feature importance explainability techniques for ESNs in two climate applications.

Echo State Networks ESNs are nonlinear machine learning models for temporal data. A single layer ESN is composed of two stages:

Output stage: ridge regression

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

Hidden stage: nonlinear stochastic transformation

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

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

ESNs are computationally efficient since the only estimated parameters are those in $\mathbf{V}$. The elements of $\mathbf{W}$ and $\mathbf{U}$ are randomly sampled to create sparse matrices. All other parameters are tunable. See [7-8].

Discussion on Evaluating Explainability

- While both applications here demonstrate the use of explainability to gain insight into ESNs, work has identified the need for more technical evaluations of explainability approaches (includes work from the climate space [9]).
- Future work is needed to identify how to appropriately use and evaluate explainability methods with spatio-temporal data.

References

1 Bonas, M. and Castruccio, S. (2023). Calibration of spatiotemporal forecasts from citizen science urban air pollution data with sparse recurrent neural networks. The Annals of Applied Statistics, 17(3):1820–1840.

2 Hassanibesheli, F., Kurths, J., and Boers, N. (2022). Long-term enso prediction with echo-state networks. Environmental Research: Climate, 1(1):011002.

3 Huang, H., Castruccio, S., and Genton, M. G. (2022). Forecasting high-frequency spatio-temporal wind power with dimensionally reduced echo state networks. Journal of the Royal Statistical Society Series C: Applied Statistics, 71(2):449–466.

4 Zhang, M., Zhou, Y., and Liu, Y. (2023). Deep reservoir calculation model and its application in the field of temperature and humidity prediction. Applied Intelligence, 53(4):4393–4414.

5 Ebert-Uphoff, I. and Hilburn, K. (2020). Evaluation, tuning, and interpretation of neural networks for working with images in meteorological applications. Bulletin of the American Meteorological Society, 101(12):E2149 – E2170.

6 Mamalakis, A., Barnes, E. A., and Hurrell, J. W. (2023). Using explainable artificial intelligence to quantify "climate distinguishability" after stratospheric aerosol injection. Geophysical Research Letters, 50(20):e2023GL106137. e2023GL106137 2023GL106137.

7 McDermott, P. L. and Wikle, C. K. (2017). An ensemble quadratic echo state network for non-linear spatio-temporal forecasting. Stat, 6(1):315–330.

8 McDermott, P. L. and Wikle, C. K. (2019). Deep echo state networks with uncertainty quantification for spatio-temporal forecasting. Environmetrics, 30(3):e2553. e2553 env.2553.

9 A. Mamalakis, E. A. Barnes, and I. Ebert-Uphoff, 2022: Investigating the Fidelity of Explainable Artificial Intelligence Methods for Applications of Convolutional Neural Networks in Geoscience. Artif. Intell. Earth Syst., 1, e220012, https://doi.org/10.1175/AIES-D-22-0012.1.

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. 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.

Characterizing Pathways

Collaborators: Daniel Ries and Kellie McClernon

This work aimed to develop algorithms to characterize (i.e., quantify) relationships between climate variables associated with the 1991 volcanic eruption of Mount Pinatubo. The eruption serves as a proxy for a stratospheric aerosol injection. The approach taken developed spatio-temporal feature importance for ESNs to quantify the importance of input variables over time.

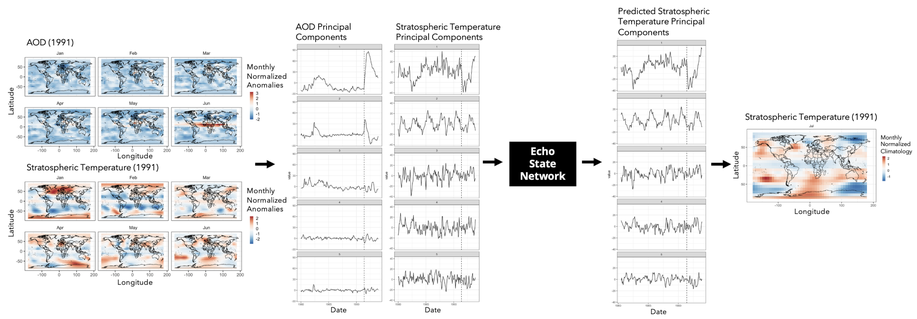

Figure 1: Modeling Process. This example considers the relationship between aerosol optical depth (AOD) and lagged stratospheric temperature (50 mb) on predicting stratospheric temperatures one month ahead. The analysis used Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2) monthly values from 1980 to 1995 within ~86 to 86 degrees latitude. The principal components were computed on the normalized anomalies of the variables. The first five principle components from both variables were used for modeling.

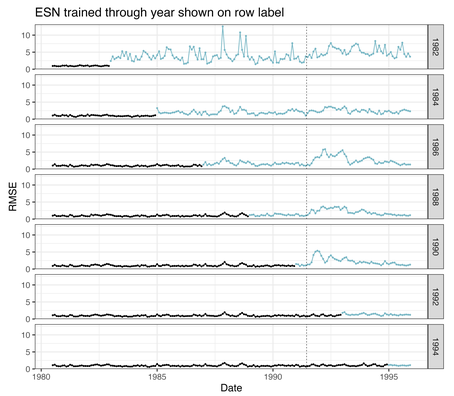

Figure 2: Model Performance. An ensemble of 25 ESNs was fit to account for random variability (after hyper-parameter tuning). This figure shows results from a temporal cross validation analysis.

Computing feature importance of AOD for predicting temperature at time 3 using a block of size 2:

$\text{RMSE}_{\text{zeroed},3}$ - $\text{RMSE}_{\text{obs},3}$

Input Matrix	AOD PC 1	...	AOD PC 5	Temp PC 1	...	Temp PC 5
t = 1	0	0	0			
t = 2	0	0	0			
t=3						
...						
t = T-τ						

Output Matrix	Temp PC 1	...	Temp PC 5
t = 1			
t = 2			
t=3			
...			
t = T-τ			

Figure 3: Spatio-Temporal Zeroed Feature Importance (stZFI). The goal of stZFI is to quantify the effect of input variable k over a block of times $(t-b, \ldots, t-1)$ on forecasts at time t. stZFI is computed as the difference between root mean squared erros (RMSEs) from "zeroed" and observed spatial predictions at time t: $RMSE_{zeroed,t} - RMSE_{obs,t}$. Large feature importance values indicate "zeroed" inputs lead to a decrease in model performance, which suggests those inputs are important for prediction. The method is inspired by work in [10].

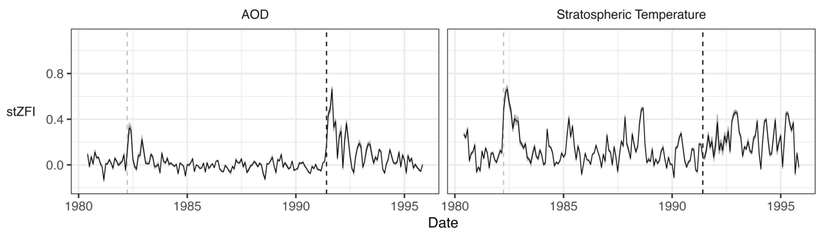

Figure 4: Feature Importance Results. These stZFI values are computed for AOD and stratospheric temperature using a block size of 6 months. The RMSEs used to compute the stZFI values are weighted by cos latitude. The grey lines represent the variability across the 25 ESNs, and the black line is the mean. Spikes in feature importance for AOD are seen after Pinatubo (1991) and the eruption of El Chichón (1982). A large spike in feature importance is seen in temperature after El Chichón.

For more details...

Goode, K., Ries, D., and McClernon, K. (2024). Characterizing climate pathways using feature importance on echo state networks. Statistical Analysis and Data Mining: The ASA Data Science Journal, 17(4):e11706.

Ries, D., Goode, K., McClernon, K., and Hillman, B. (2024). Using feature importance as exploratory data analysis tool on earth system models. Geoscientific Model Development Discussions, 2024:1–35.

Subseasonal Extreme Temperature Forecasting

Collaborators: Thoman Ehrmann, Maike Holthuijzen, Meredith Brown, Jacob Johnson

The goal of this project is to use machine learning models to predict large-scale temperature extremes over the continental US on subseasonal time scales (2-8 weeks). Traditional physics-based weather models are too chaotic to predict extreme events beyond 15 days in advance. Improved forecasts may help with preparation for extreme temperature events. We are currently applied random forests, quantile random forests, and ESNs for prediction. We are employing feature importance to identify important variables for prediction (aggregated over time).

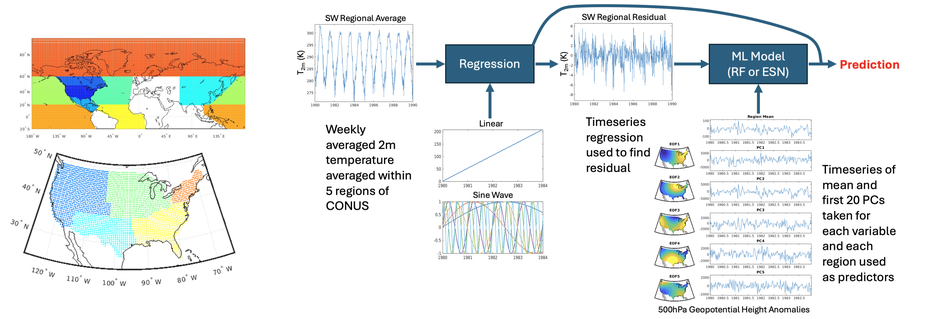

Figure 5: Modeling Process. Data were sourced from the MERRA-2. We are working with weekly averages from 1980 through 2022. Our target variable is 2m temperature averaged within 5 regions of the continental US (CONUS). Input variables were averaged over 9 global regions for 8 different fields: surface temperate, sea-level pressure, geopotential height at 850 hPa, 500 hPa, and 200 hPa, and air temperature at 850 hPa, 500 hPa, and 200 hPa.

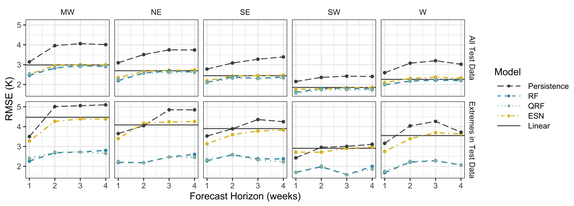

Figure 6: Preliminary Model Performance. A key finding is that the ESN and random forest models perform similarly on all test data, but the random forest outperforms the ESN on the extreme temperatures (i.e., temperatures more than 1 standard deviation away from the mean).

Input Matrix	Temp PC 1	...	Temp PC 20	...	SLP PC 1	...	SLP PC 20
t = 1							
t = 2							
t=3							
...							
t = T-τ							

Output Matrix	Temp Residual
t = 1	
t = 2	
t=3	
...	
t = T-τ	

Figure 7: Permutation Feature Importance (PFI). We apply PFI [11] to identify variables that are important for forecasts (accounting for all times). PFI is computed as the difference between the RMSE when one input matrix column is permuted and the RMSE from observed predictions: $RMSE_{perm} - RMSE_{obs}$. Ongoing work is exploring grouping of variables to account for correlated input variables.

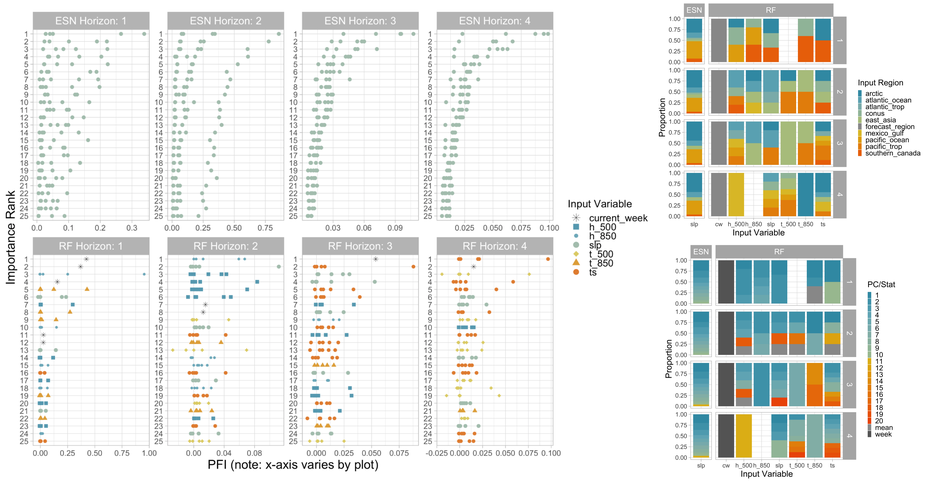

Figure 8: Preliminary Feature Importance Results. (Left) Variable importances ranked by average importance across target regions for the top 25 variables. Noticeably, all top 25 variables for the ESN are sea level pressure variables unlike with the random forest. (Right) Proportion of the top 25 variables that are associated with an input region (top) and principle component/statistic (bottom). The proportions are computed within a model type, input climate variable, and forecast horizon. Note that the ESN has high proportions of arctic and pacific regions in the top 25 variables. Further, the ESN only has PCs 1-11 in the top 25 variables .

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and 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-NA-0003525. SAND NO. 2025-07557C