Sandia National Laboratories

Machine Learning and Reservoir Computing for Watershed-Scale Drought Prediction

Nicole D. Jackson, Ph.D.

Climate Security Center

American Geophysical Union Fall Meeting 2024

09 December 2024

U.S. DEPARTMENT OF ENERGY NNSA

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-NA0003525.

SAND2024-16489C

Acknowledgements

The EPIC PLANN (Enhanced Prediction in Chaotic Systems with Physics-learned Autonomous Neural Networks) team

- William Chapman[†]
- Josh Mott[‡]
- J. Darby Smith
- Corinne Teeter

‡

- Louis Scuderi
- Zach Strasberg[‡]

† postdoc , ‡ Ph.D. student

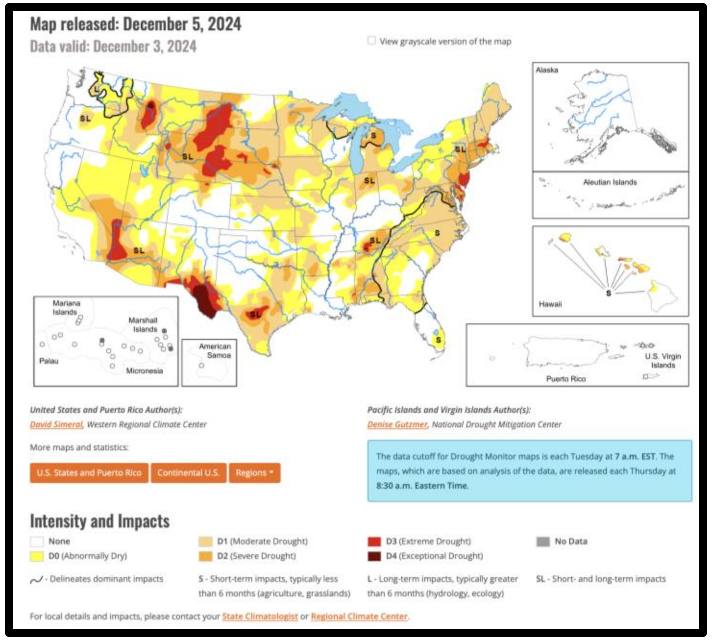
Sandia Earth Science Research Foundation

- Program Manager: Stephanie Kuzio

The US Drought Monitor provides weekly drought estimates

U.S. Drought Monitor

Last week's report

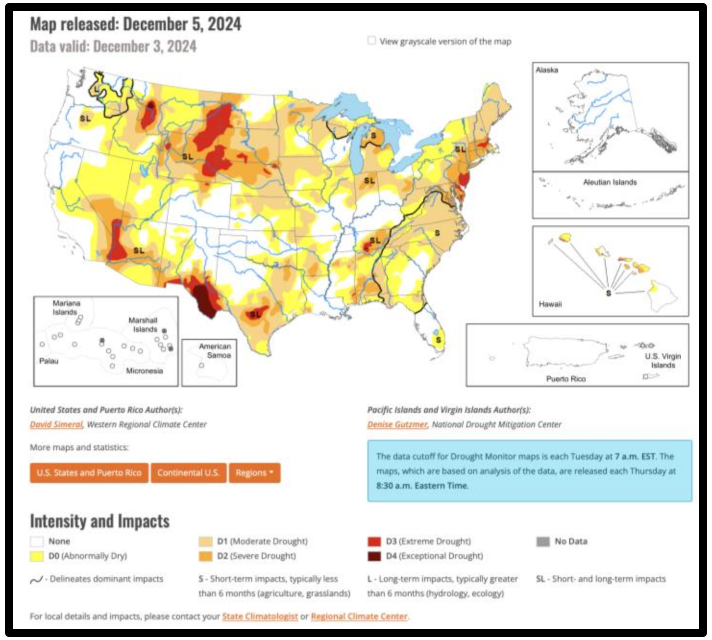


Source: [U.S. Drought Monitor]

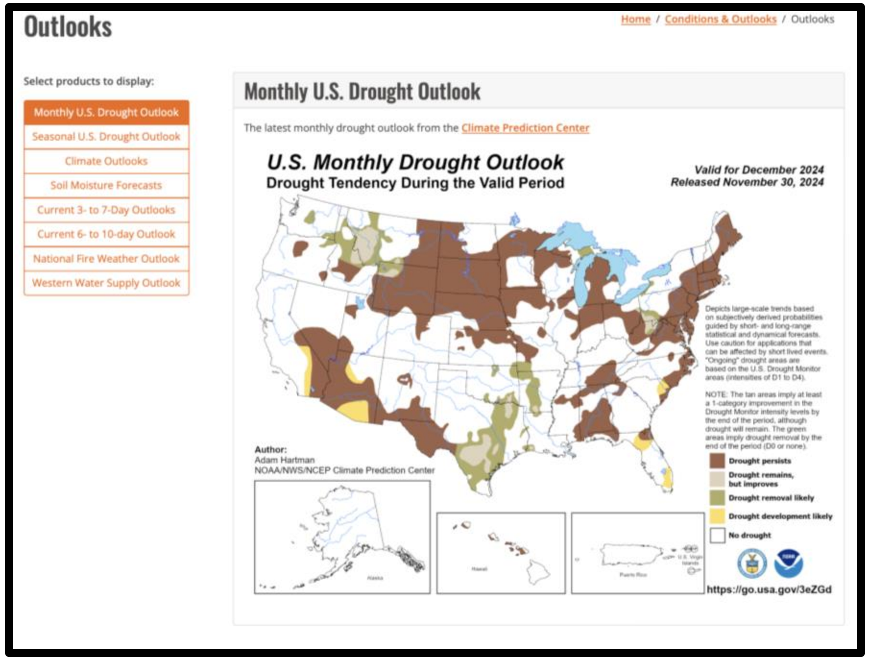
The US Drought Monitor provides weekly drought estimates as well as monthly forecasts

U.S. Drought Monitor

Last week's report



Monthly Outlook

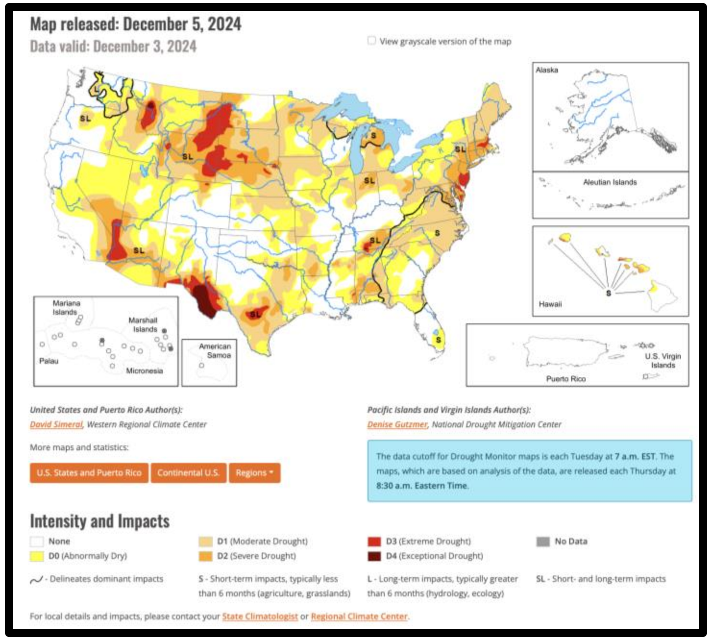


Source: U.S. Drought Monitor

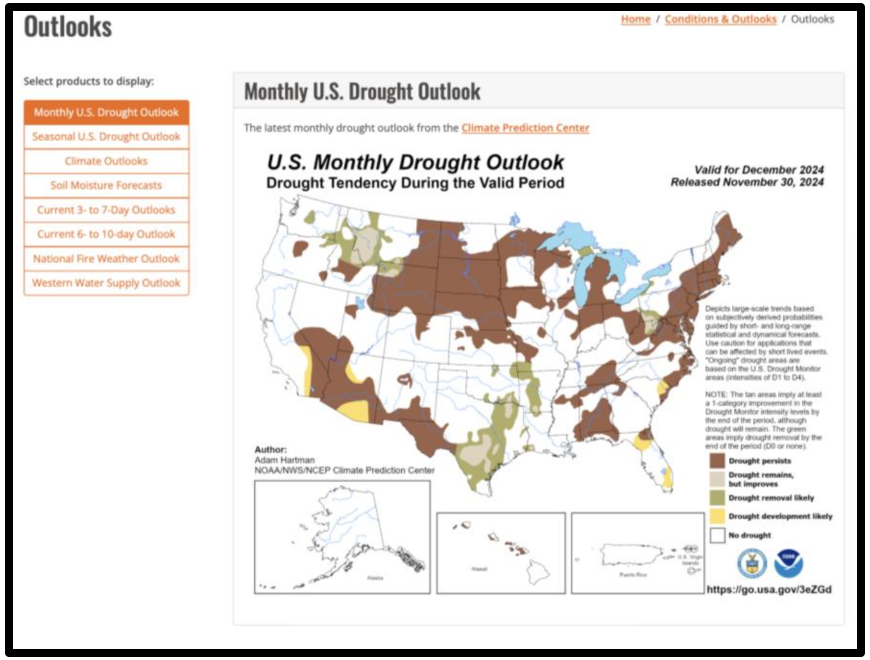
The US Drought Monitor provides weekly drought estimates as well as monthly and seasonal forecasts

U.S. Drought Monitor

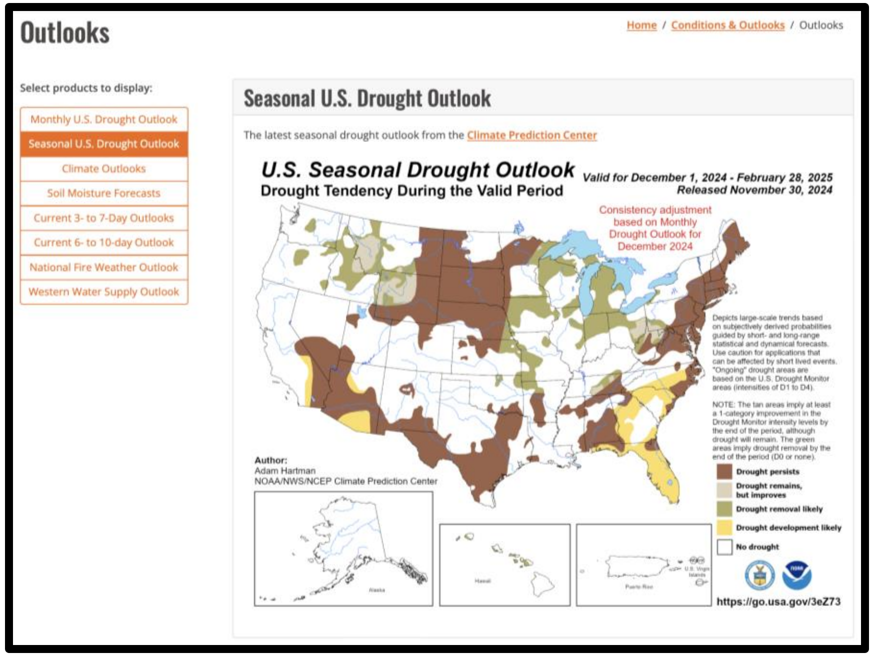
Last week's report



Monthly Outlook



Seasonal Outlook

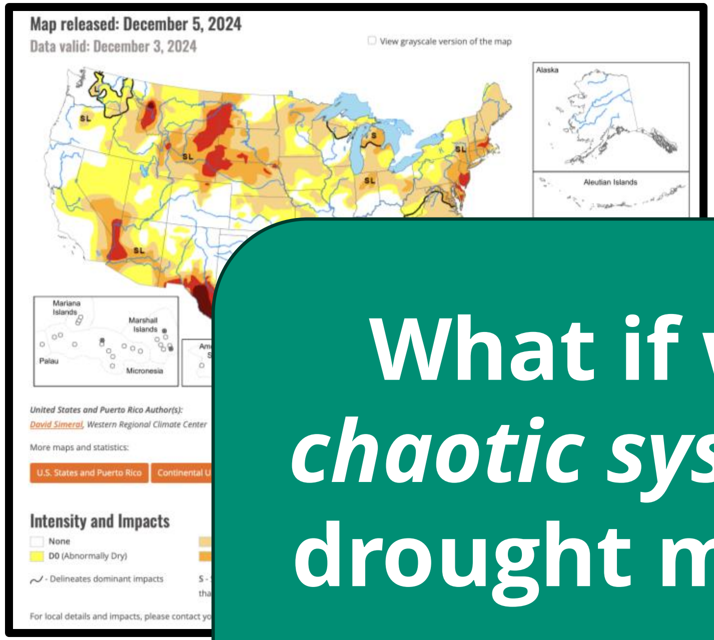


Source: U.S. Drought Monitor

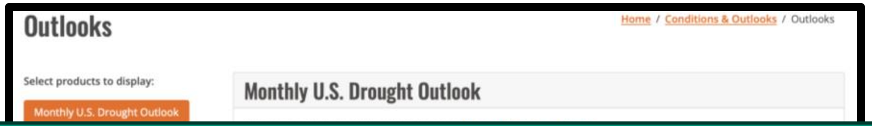
The US Drought Monitor provides weekly drought estimates as well as monthly and seasonal forecasts

U.S. Drought Monitor

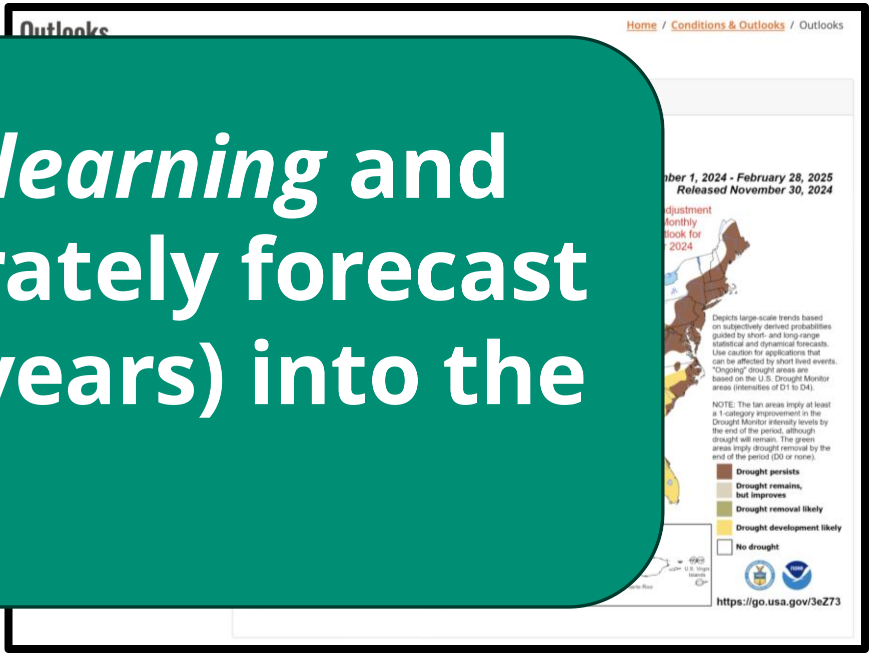
Last week's report



Monthly Outlook



Seasonal Outlook



What if we could use *machine learning* and *chaotic system modeling* to accurately forecast drought much farther (e.g., 1-2 years) into the future?

Source: U.S. Drought Monitor

Multiple machine learning—based algorithms have been used to forecast meteorological drought

Source: Agana and Homaifar (2017)

- Multilayer Perceptron (MLP)
- Support Vector Regression (SVR)

Source: Li et al. (2020)

- SVR
- Random Forest (RF)
- Extreme Learning Machine (ELM)

Source: Mokhtar et al. (2021)

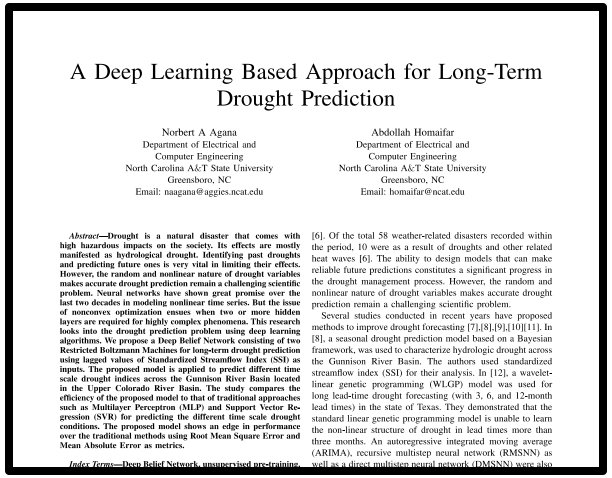
- RF
- Extreme Gradient Boost (XGB)
- Convolutional neural network (CNN)
- Long-term short memory (LSTM)

Source: Prodhan et al. (2022)

Meteorological Drought:

- Artificial neural network (ANN)
- SVR
- K-nearest neighbor
- Adaptive Neuro Fuzzy (ANF)
- SVM-Copula
- Wavelet Boosting-Support Vector Regression (WBS-SVM)

Multiple machine learning—based algorithms have been used to forecast meteorological drought

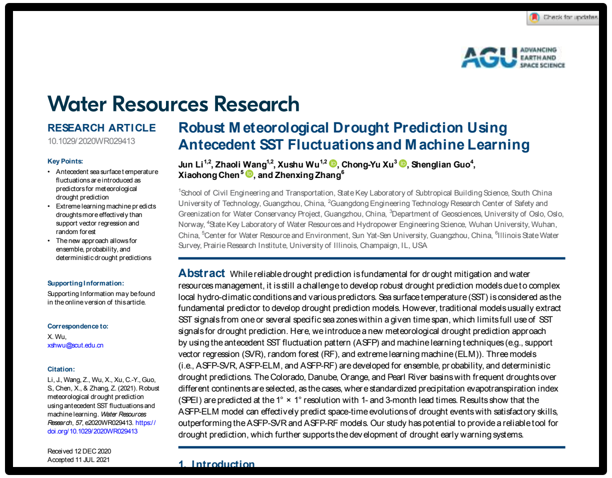


A Deep Learning Based Approach for Long-Term Drought Prediction

Norbert A Agana, Department of Electrical and Computer Engineering, North Carolina A&T State University, Greensboro, NC

Abdollah Homaifar, Department of Electrical and Computer Engineering, North Carolina A&T State University, Greensboro, NC

Source: [Agana and Homaifar (2017)]

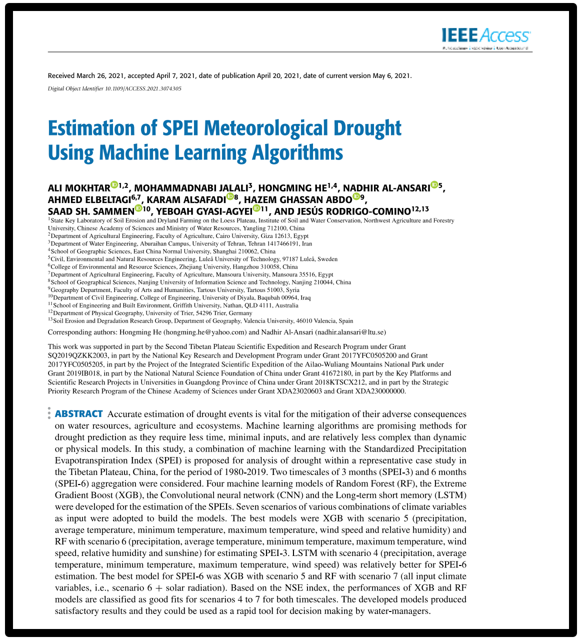


Water Resources Research

RESEARCH ARTICLE

Robust Meteorological Drought Prediction Using Antecedent SST Fluctuations and Machine Learning

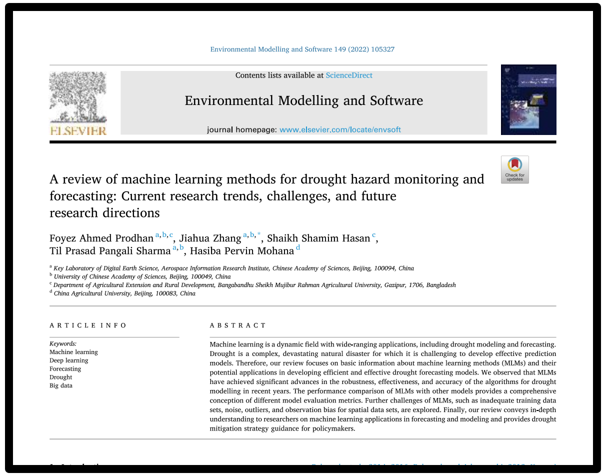
Source: [Li et al. (2020)]



IEEE Access

Estimation of SPEI Meteorological Drought Using Machine Learning Algorithms

Source: [Mokhtar et al. (2021)]



Environmental Modelling and Software

A review of machine learning methods for drought hazard monitoring and forecasting: Current research trends, challenges, and future research directions

Source: [Prodhan et al. (2022)]

- Multilayer Perceptron (MLP)
- Support Vector Regression (SVR)

- SVR
- Random Forest (RF)
- Extreme Learning Machine (ELM)

- RF
- Extreme Gradient Boost (XGB)
- Convolutional neural network (CNN)
- Long-term short memory (LSTM)

Meteorological Drought:

- Artificial neural network (ANN)
- SVR
- K-nearest neighbor
- Adaptive Neuro Fuzzy (ANF)
- SVM-Copula
- Wavelet Boosting-Support Vector Regression (WBS-SVM)

Gap: forecasting for periods beyond 3-6 months

Modeling supported by the high spatial resolution (4 km) NCAR/USGS CONUS404 dataset for period 1979-2021

- Climate Data (CONUS404)
- Drought Metrics

>100 including…

Precipitation

Temperature

Transpiration

Wind speed

Calculated **meteorological drought** metrics:

- Standardized Precipitation Index (SPI)
- Standardized Precipitation Evapotranspiration Index (SPEI)

Source: [Rasmussen et al. (2023)]

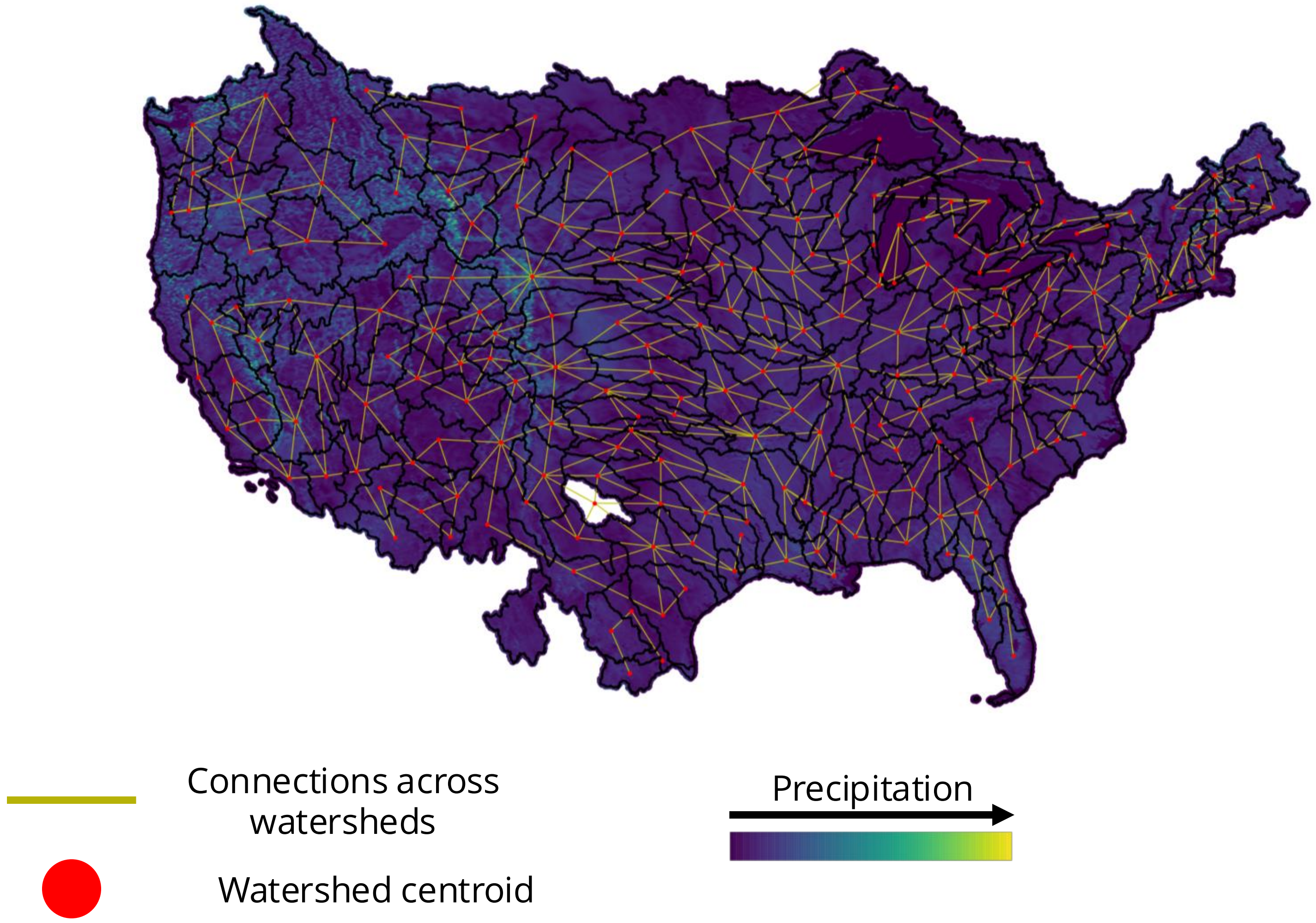
Data is statistically reduced to the weekly timescale at the HUC4 and HUC8 spatial scales

- Climate Data (CONUS404)
- Drought Metrics

↓

Watershed-scale Spatial Aggregation

Sample HUC8 aggregation for precipitation

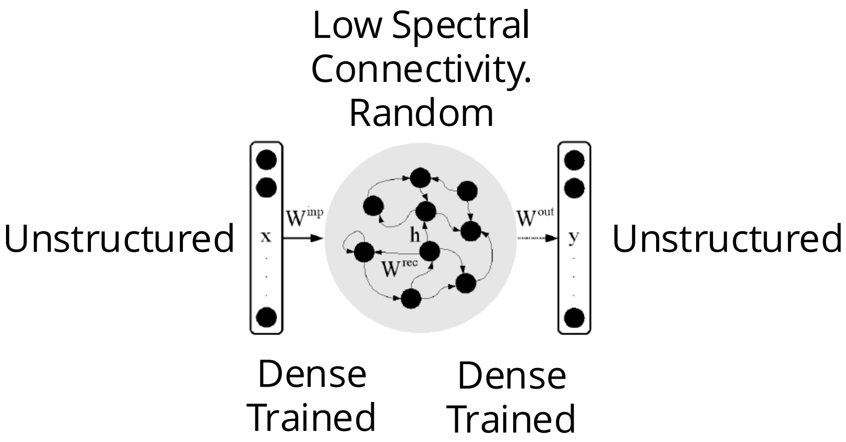


Reservoir networks use spectral normalization and random connectivity to autonomously mimic chaotic system

- Climate Data (CONUS404)
- Drought Metrics

Watershed-scale Spatial Aggregation

Reservoir Computing-based Predictive Drought Model



Graph-based reservoirs reduce normalized root mean square error during prediction period versus a single reservoir

- Climate Data (CONUS404)
- Drought Metrics

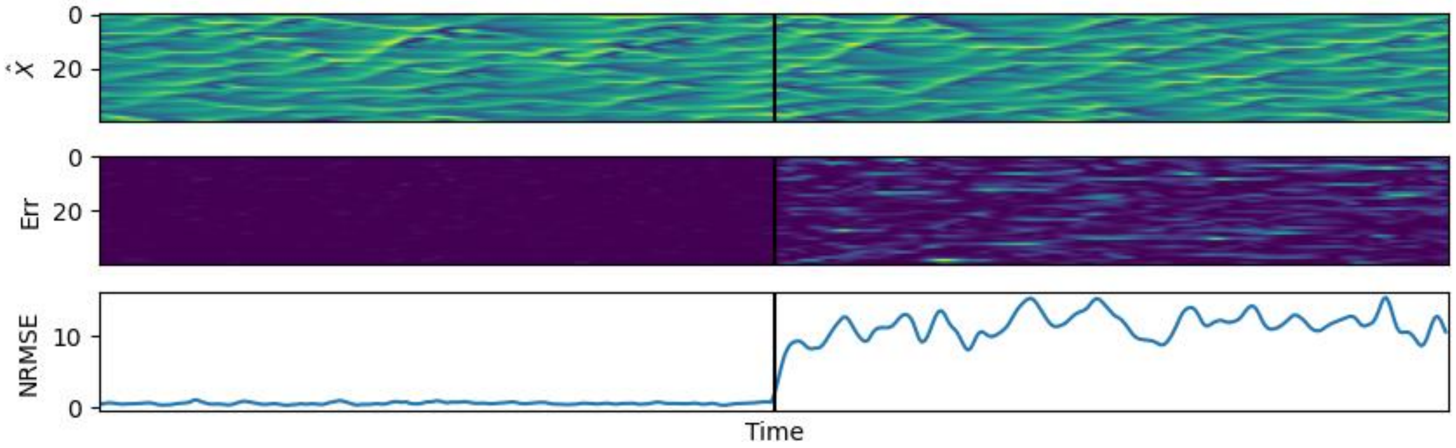
↓

Watershed-scale Spatial Aggregation

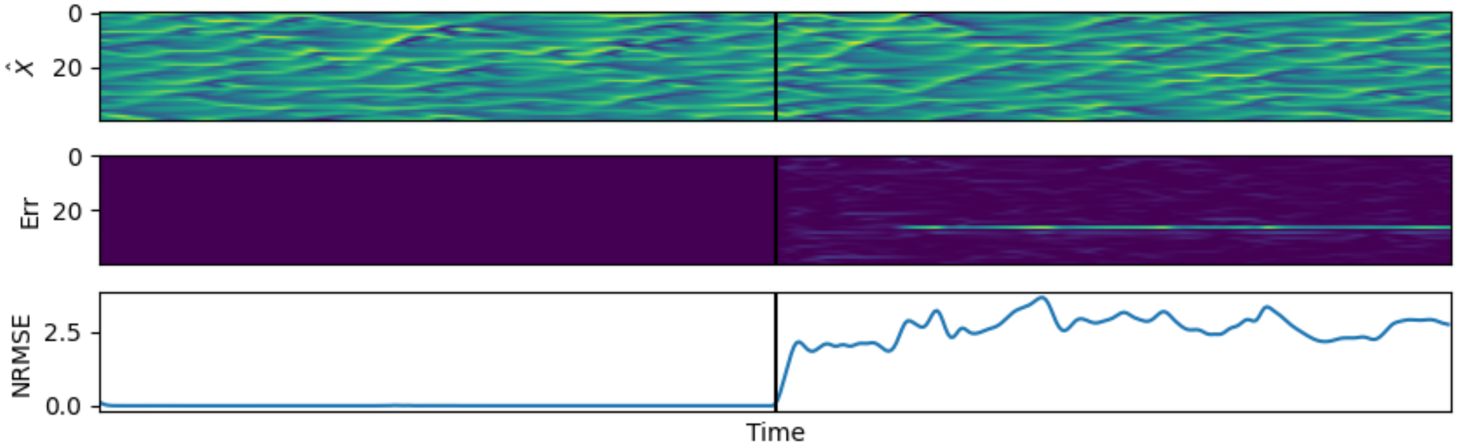
↓

Reservoir Computing-based Predictive Drought Model

Single Reservoir



Graph Reservoir



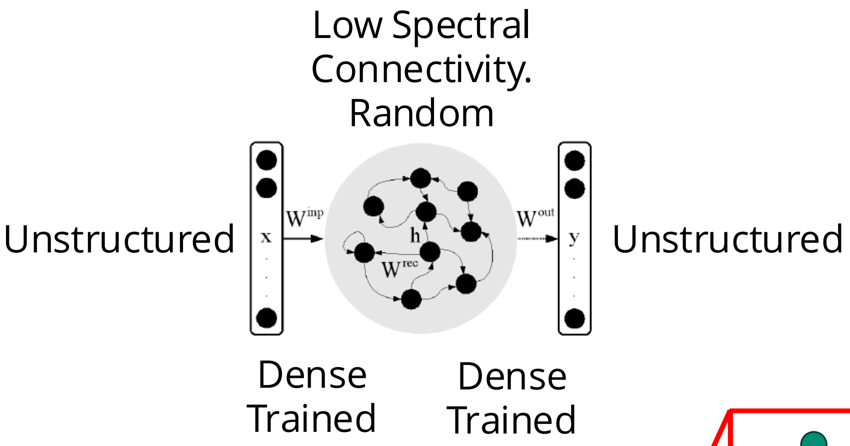
Source: Chapman et al. (*in prep*)

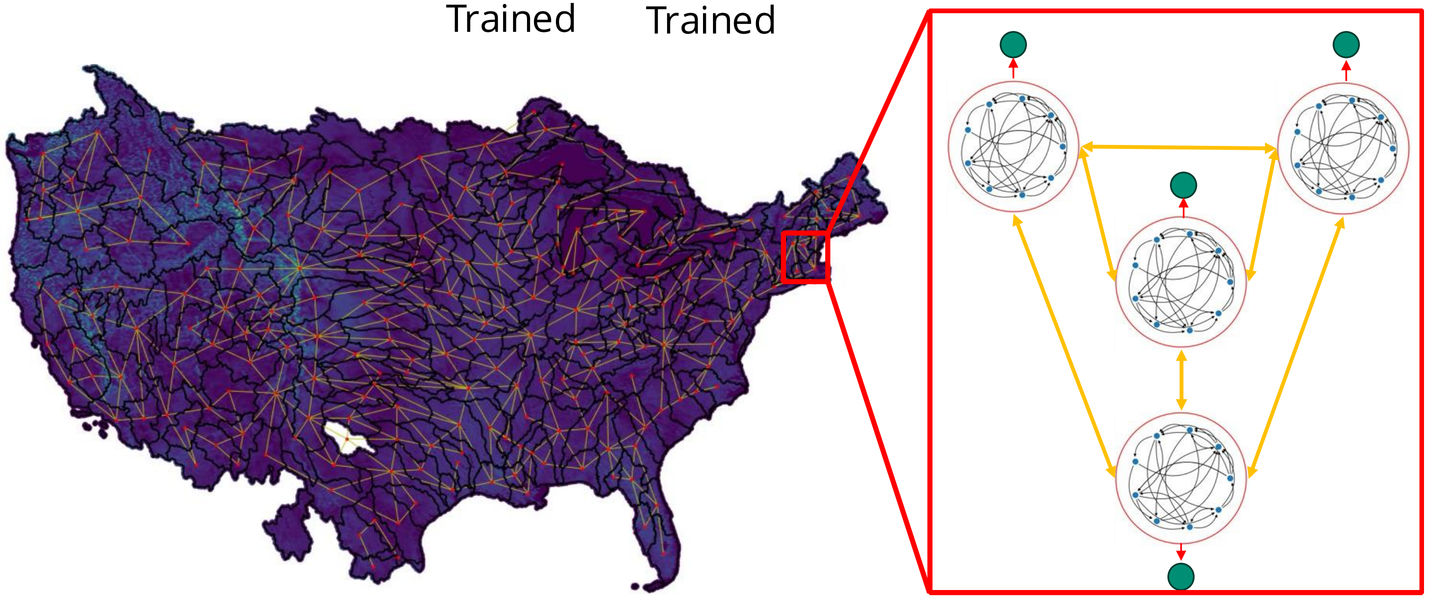
Graph reservoir nodes are generated for each watershed while maintaining the input layer’s spatial structure

- Climate Data (CONUS404)
- Drought Metrics

Watershed-scale Spatial Aggregation

Reservoir Computing-based Predictive Drought Model





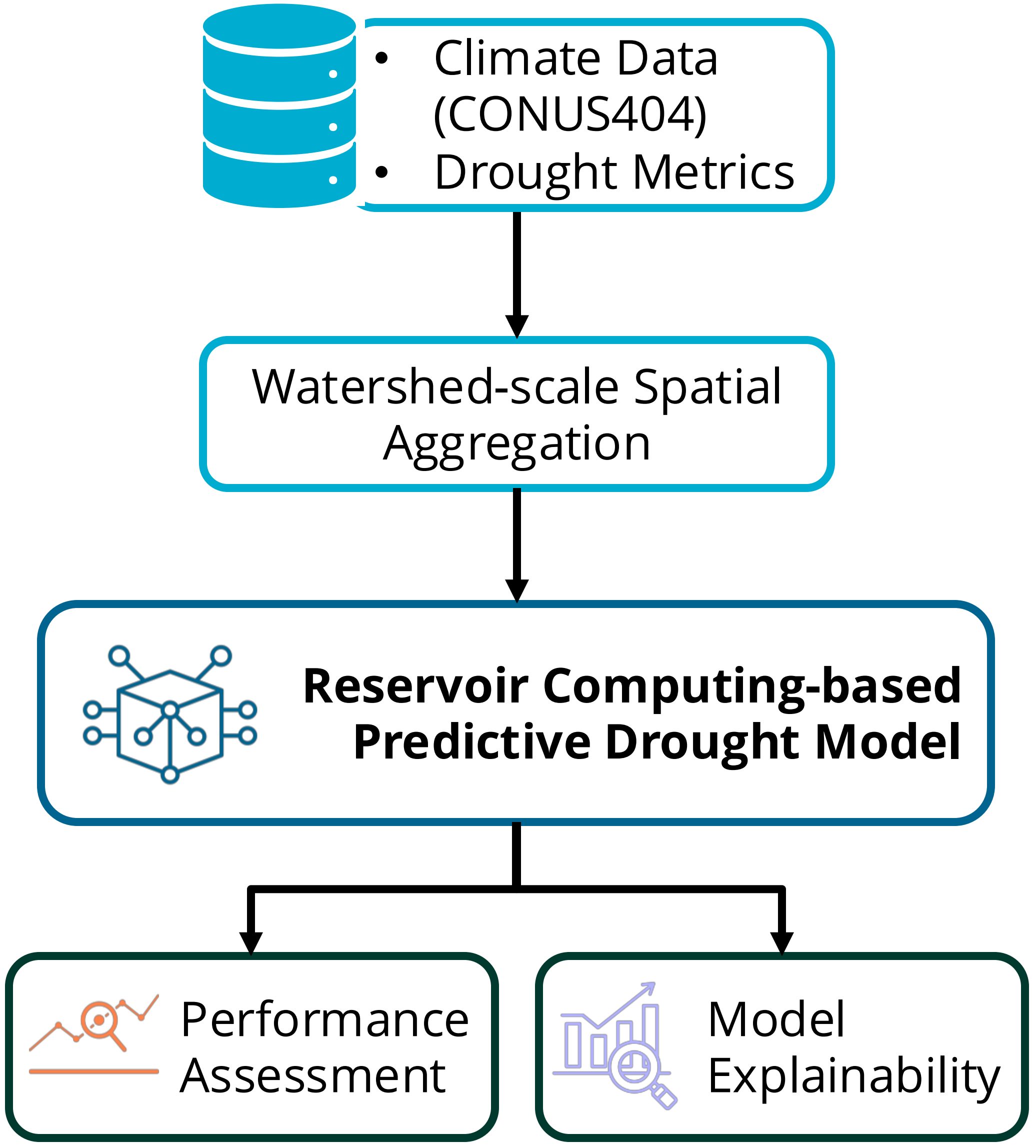
Connections across watersheds

Watershed centroid

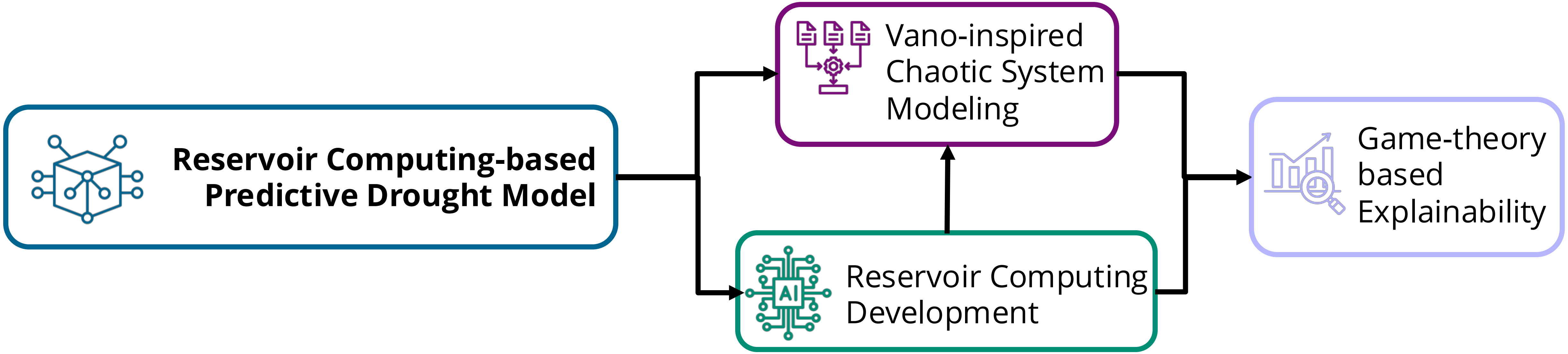
Connection to spatially close watershed

Local observable variable

The EPIC PLANN project is using a standard framework to develop its ML-based predictive model for drought



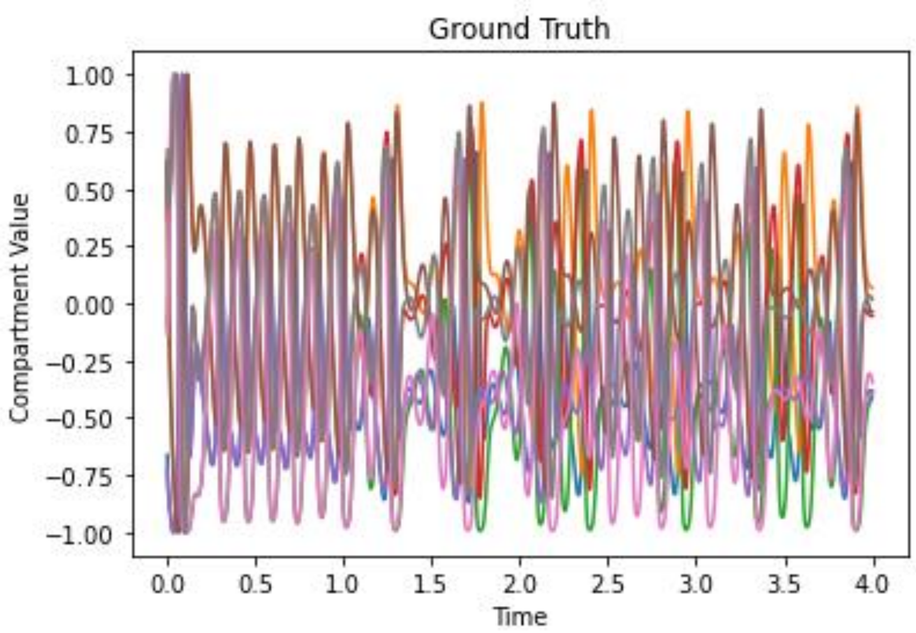
Reservoir model development inspired by chaotic systems modeling and game theory



A graph-based reservoir network is trained to predict two sets of Vano models side by side

Input layer

$X_1(t_0), \ldots, X_1\left(t_{N_k}\right)$
$X_2(t_0), \ldots, X_2\left(t_{N_k}\right)$
$X_3(t_0), \ldots, X_3\left(t_{N_k}\right)$
$X_4(t_0), \ldots X_4\left(t_{N_k}\right)$
$Y_1(t_0), \ldots, Y_1\left(t_{N_k}\right)$
$Y_2(t_0), \ldots, Y_2\left(t_{N_k}\right)$
$Y_3(t_0), \ldots, Y_3\left(t_{N_k}\right)$
$Y_4(t_0), \ldots, Y_4\left(t_{N_k}\right)$



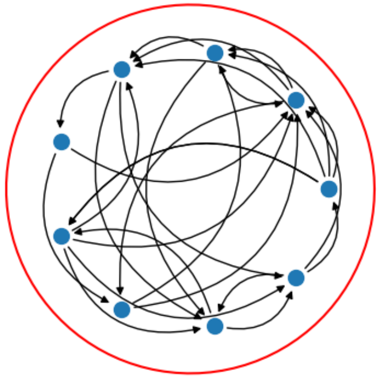
Source: Smith et al. (*in prep*)

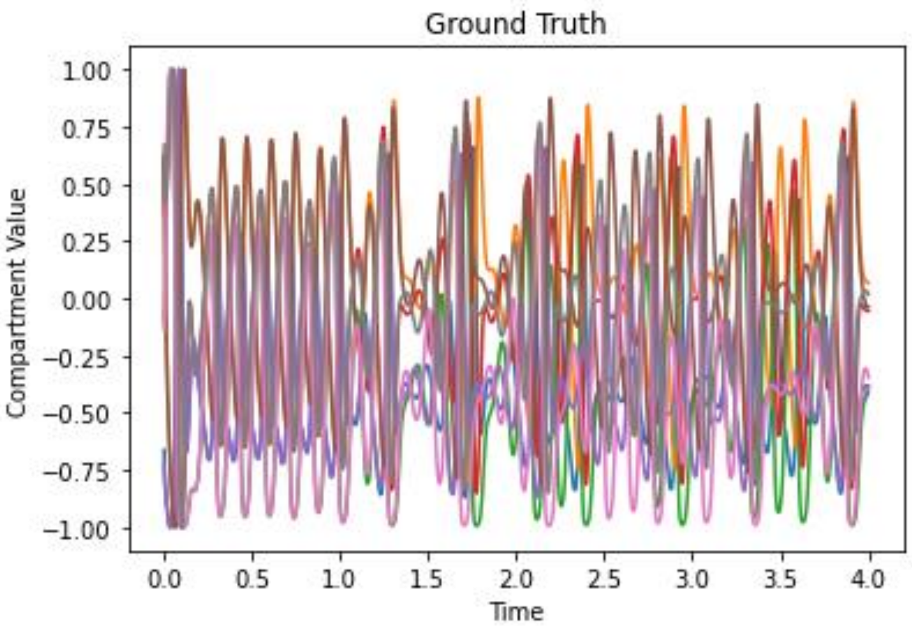
A graph-based reservoir network is trained to predict two sets of Vano models side by side

Input layer

$X_1(t_0), \ldots, X_1\left(t_{N_k}\right)$
$X_2(t_0), \ldots, X_2\left(t_{N_k}\right)$
$X_3(t_0), \ldots, X_3\left(t_{N_k}\right)$
$X_4(t_0), \ldots X_4\left(t_{N_k}\right)$
$Y_1(t_0), \ldots, Y_1\left(t_{N_k}\right)$
$Y_2(t_0), \ldots, Y_2\left(t_{N_k}\right)$
$Y_3(t_0), \ldots, Y_3\left(t_{N_k}\right)$
$Y_4(t_0), \ldots, Y_4\left(t_{N_k}\right)$

Trained Reservoirs





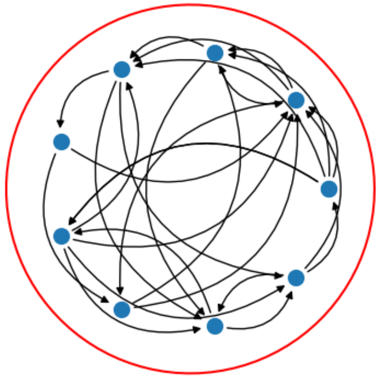
Source: Smith et al. (*in prep*)

A graph-based reservoir network is trained to predict two sets of Vano models side by side

Input layer

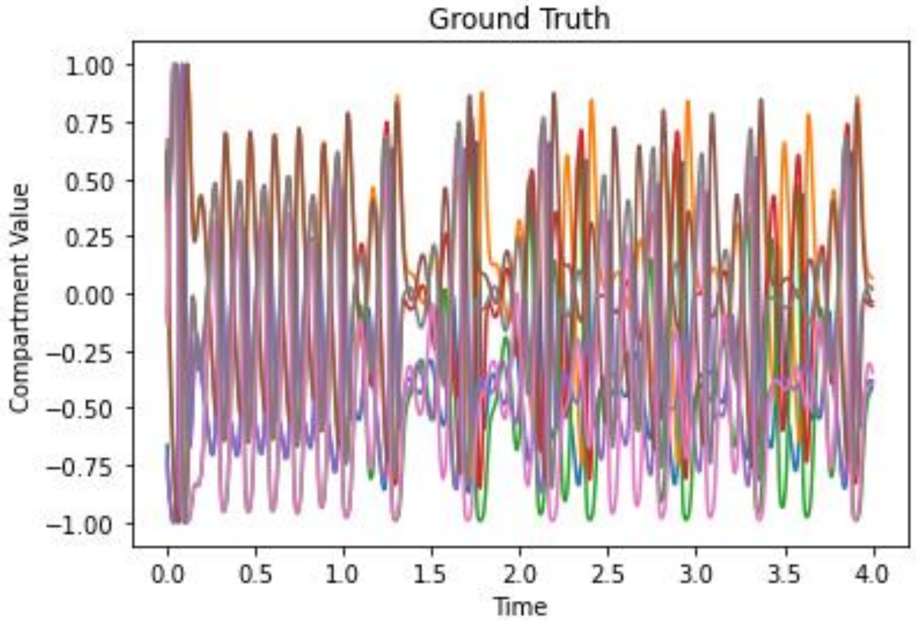
$X_1(t_0), \dots, X_1\left(t_{N_k}\right)$
$X_2(t_0), \dots, X_2\left(t_{N_k}\right)$
$X_3(t_0), \dots, X_3\left(t_{N_k}\right)$
$X_4(t_0), \dots X_4\left(t_{N_k}\right)$
$Y_1(t_0), \dots, Y_1\left(t_{N_k}\right)$
$Y_2(t_0), \dots, Y_2\left(t_{N_k}\right)$
$Y_3(t_0), \dots, Y_3\left(t_{N_k}\right)$
$Y_4(t_0), \dots, Y_4\left(t_{N_k}\right)$

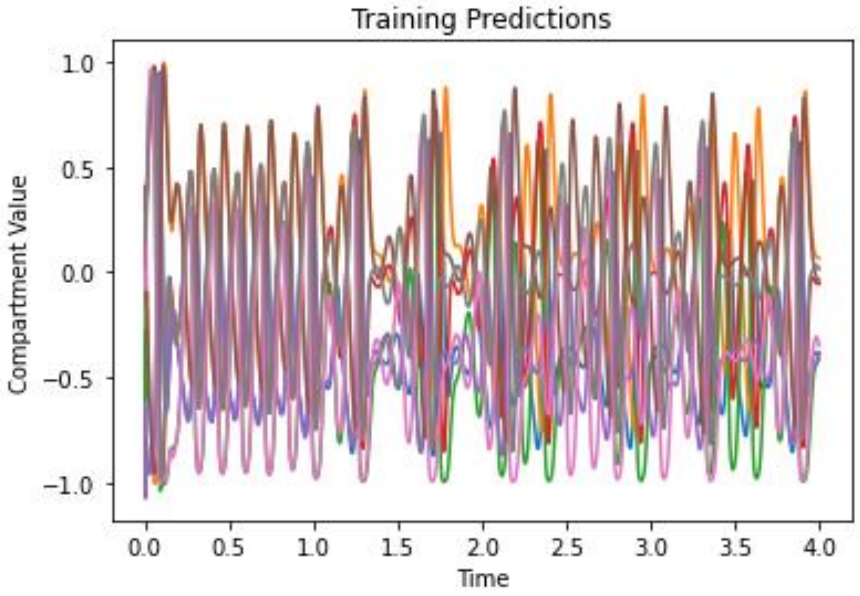
Trained Reservoirs



Output layer

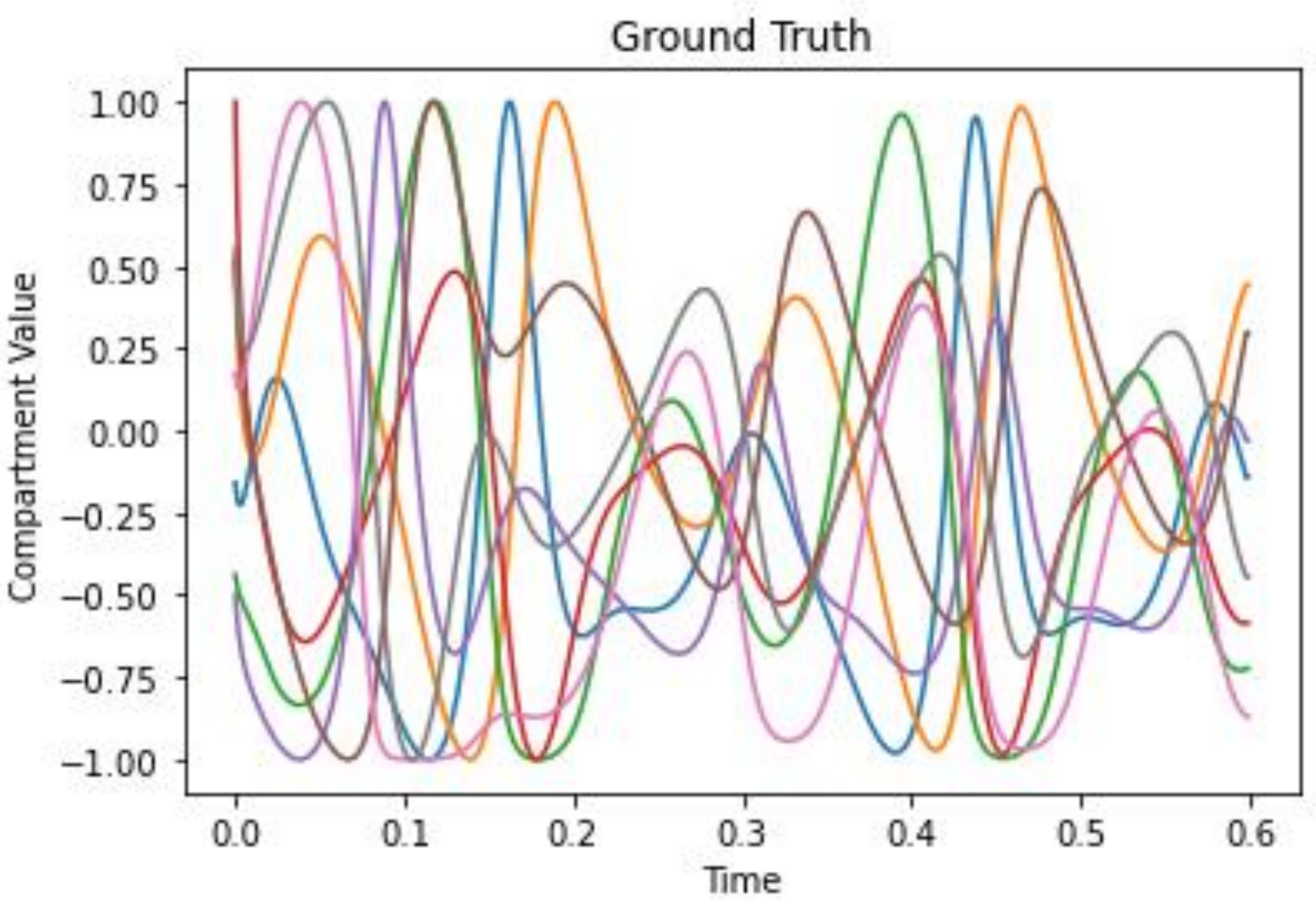
$X_1\left(t_{N_k+1}\right), \dots, X_1\left(t_{N_p}\right)$
$X_2\left(t_{N_k+1}\right), \dots, X_2\left(t_{N_p}\right)$
$X_3\left(t_{N_k+1}\right), \dots, X_3\left(t_{N_p}\right)$
$X_4\left(t_{N_k+1}\right), \dots, X_4\left(t_{N_p}\right)$
$Y_1\left(t_{N_k+1}\right), \dots, Y_1\left(t_{N_p}\right)$
$Y_2\left(t_{N_k+1}\right), \dots, Y_2\left(t_{N_p}\right)$
$Y_3\left(t_{N_k+1}\right), \dots, Y_3\left(t_{N_p}\right)$
$Y_4\left(t_{N_k+1}\right), \dots, Y_4\left(t_{N_p}\right)$
$\boxed{g\left(t_{N_k+1}\right), \dots, g\left(t_{N_p}\right)}$ Mystery function



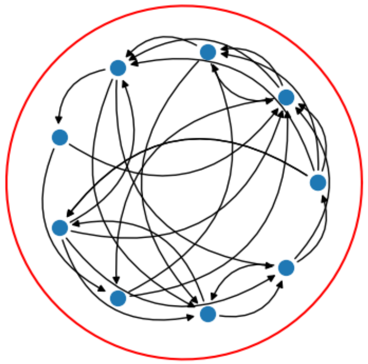


Source: Smith et al. (*in prep*)

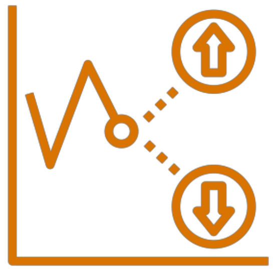
A new time series using perturbed initial conditions is generated to support prediction with trained reservoirs



Previously trained reservoirs

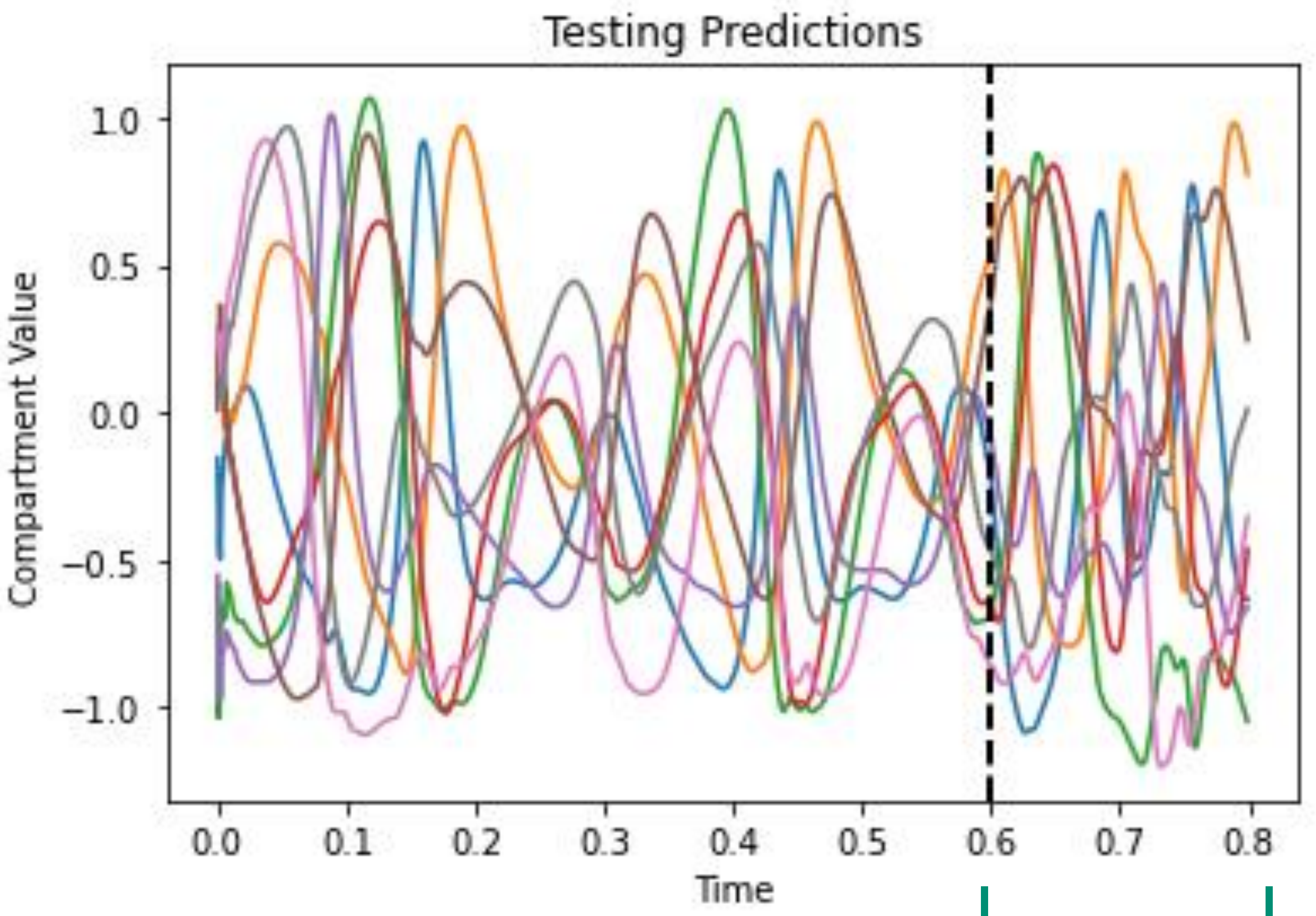


Predicted time series



Source: Smith et al. (*in prep*)

Predicted time series using perturbed data and trained reservoirs with rollout



Rollout using only trained reservoirs

Source: Smith et al. (*in prep*)

Shapley Additive Explanation Values (SHAP) are used to determine influence of inputs on changes to outputs

Input layer

$X_1(t_0), \ldots, X_1(t_{N_k})$
$X_2(t_0), \ldots, X_2(t_{N_k})$
$X_3(t_0), \ldots, X_3(t_{N_k})$
$X_4(t_0), \ldots X_4(t_{N_k})$
$Y_1(t_0), \ldots, Y_1(t_{N_k})$
$Y_2(t_0), \ldots, Y_2(t_{N_k})$
$Y_3(t_0), \ldots, Y_3(t_{N_k})$
$Y_4(t_0), \ldots, Y_4(t_{N_k})$

$\rightarrow$

Output layer

$X_1(t_{N_k+1}), \ldots, X_1(t_{N_p})$
$X_2(t_{N_k+1}), \ldots, X_2(t_{N_p})$
$X_3(t_{N_k+1}), \ldots, X_3(t_{N_p})$
$X_4(t_{N_k+1}), \ldots, X_4(t_{N_p})$
$Y_1(t_{N_k+1}), \ldots, Y_1(t_{N_p})$
$Y_2(t_{N_k+1}), \ldots, Y_2(t_{N_p})$
$Y_3(t_{N_k+1}), \ldots, Y_3(t_{N_p})$
$Y_4(t_{N_k+1}), \ldots, Y_4(t_{N_p})$
$\boxed{g(t_{N_k+1}), \ldots, g(t_{N_p})}$

$\rightarrow$

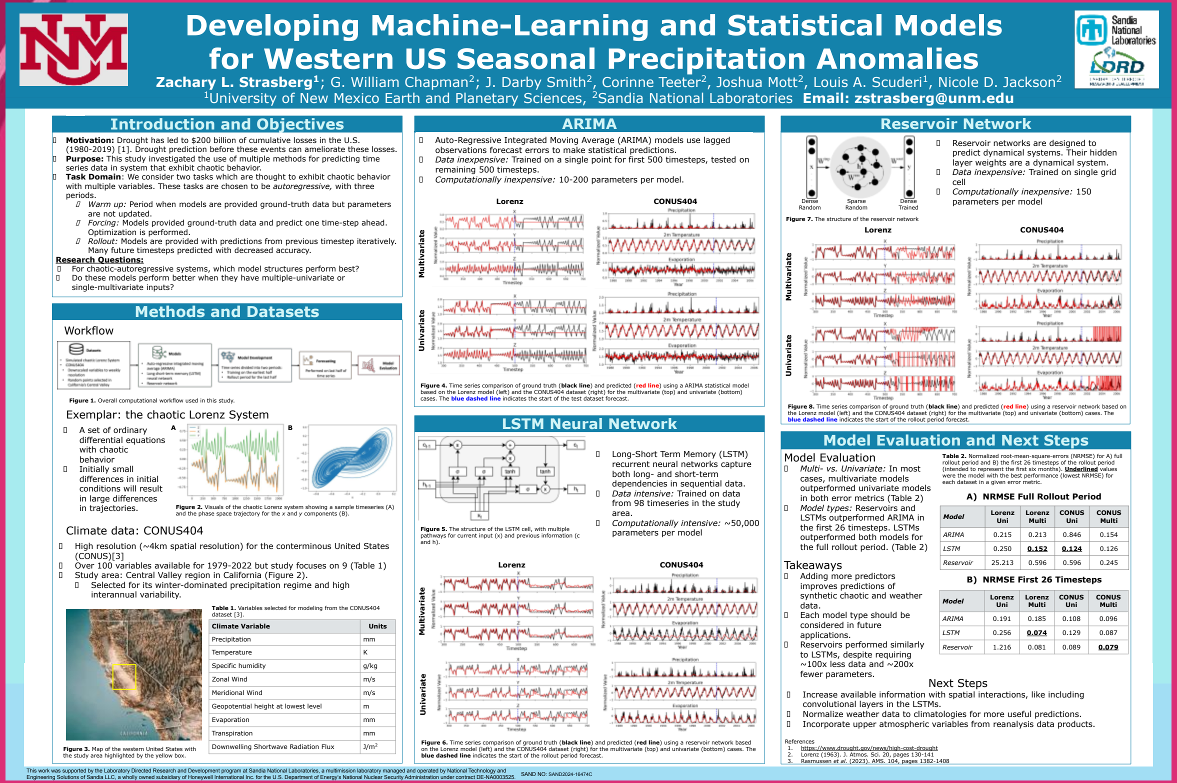
SHAP Values

$$\begin{bmatrix} 0.02147524 \\ 0.02505524 \\ 0.02095524 \\ 0.01256381 \\ 0.00879333 \\ 0.00837357 \\ 0.00999976 \\ 0.00241643 \end{bmatrix}$$

Source: Smith et al. (*in prep*)

A51L-1838: Developing Machine-Learning and Statistical Models for Western US Seasonal Precipitation Anomalies

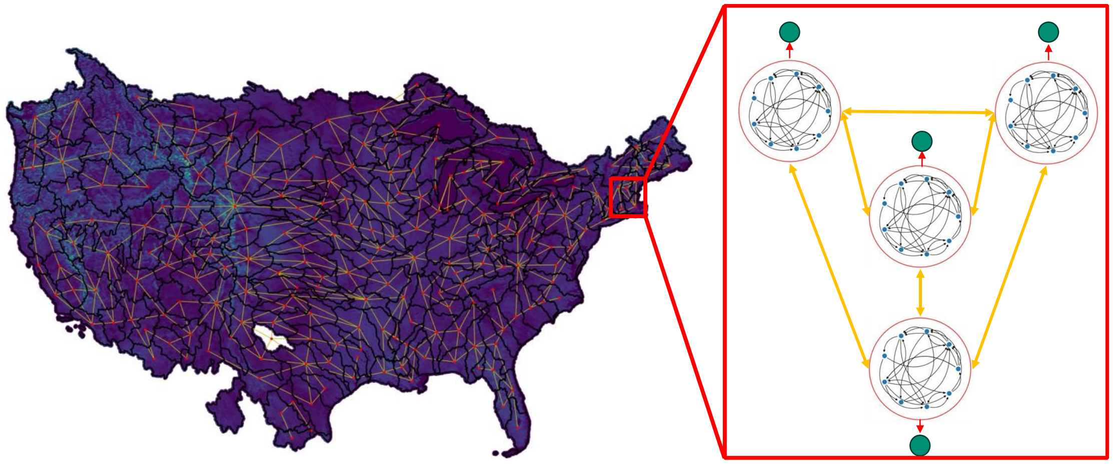
- Presenter: Zach Strasberg
- Friday, 13 December 2024
- 08:30 – 12:20
- Poster Halls B-C

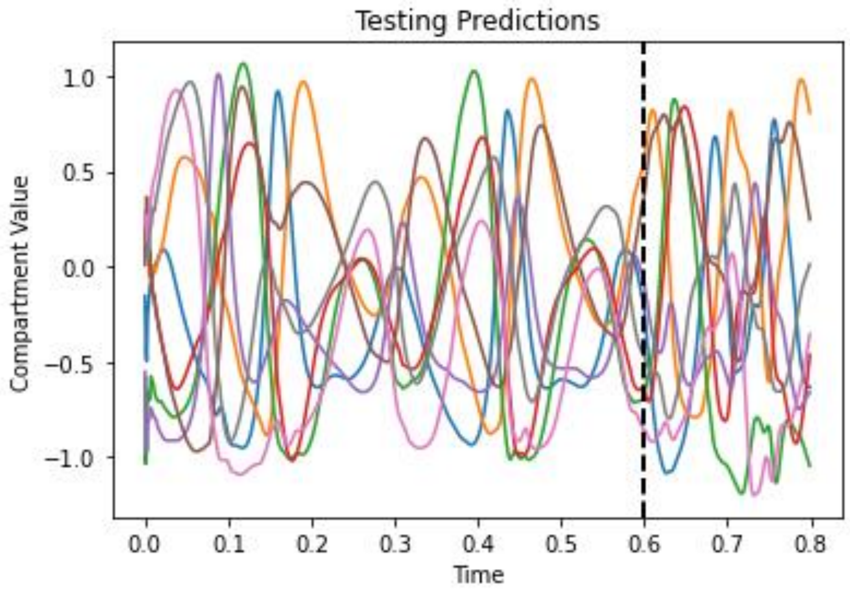


AGU24

In conclusion, we have developed reservoir computing-based predictive time series models for chaotic, dynamic systems

- Graph reservoirs reduce error compared to single reservoirs
- Reservoir-based predictive model developed using the Vano system with chaotic behavior
- SHAP identified 3 key variables driving Vano model performance





Future Work

- Increase variables and explainability techniques for the Vano model
- Consideration of multiple drought metrics (SPEI, SPI, etc.)
- Increase spatial extent of modeled watersheds

Nicole D. Jackson <njacks@sandia.gov>