National Aeronautics and Space Administration

FY23 Innovative Spontaneous Concepts Research and Technology Development (ISC)

Machine Learning for Detecting Hidden Patterns in Data for Air-Land Processes

Principal Investigator: Alphan Altinok (398)**; Co-Investigators:** Nicholas Parazoo (329), Daniel Wenkert (398), Gary Doran (398), Derek Posselt (329), Kerry-Anne Cawse-Nicholson (329), Steffen Mauceri (398), Xubin Zeng (University of Arizona)

Objectives:

The objective of this work was to develop automated machine learning based approaches to find unknown relationships between variables in large datasets. As a case study, we aim to demonstrate the applicability of such methods by using explainable machine learning models and non-linear dependence estimation to investigate Earth land surface and atmospheric variables that explain present and future drought conditions in North America. Identifying these sources of predictability can improve both data- and model-driven approaches for seasonal-to-subseasonal forecasting of drought.

Background:

Earth Science Decadal Survey Key Priority:

Understand the factors affecting the timing, severity, and persistence of droughts

Relevant Variables:

- **Land Surface**: temperature, water/heat exchange, vegetation, etc.
- **Atmospheric**: precipitation, humidity, wind, planetary boundary layer dynamics, etc.
- **Climate Indices**: El Niño Southern Oscillation (ENSO), North Atlantic Oscillation (NAO), etc.

Challenges: non-linear, multivariate, unknown interactions between variables

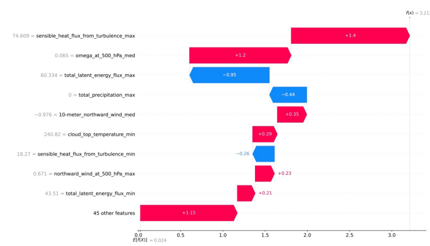

Figure 1: An example Shapley explanation for an individual grid cell and month, January 1980. The total estimated anomaly along the horizontal axis is decomposed into contributions from each feature along the vertical axis.

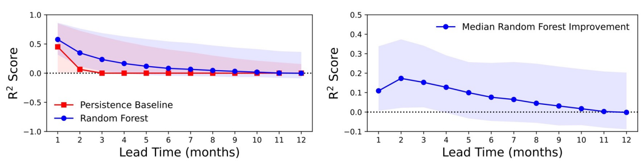

Figure 2: Model improvement over a persistence baseline that uses previous months' anomaly values to predict future anomalies. The random forest model does consistently better than this baseline, with maximum improvement for prediction lead times of 2-3 months.

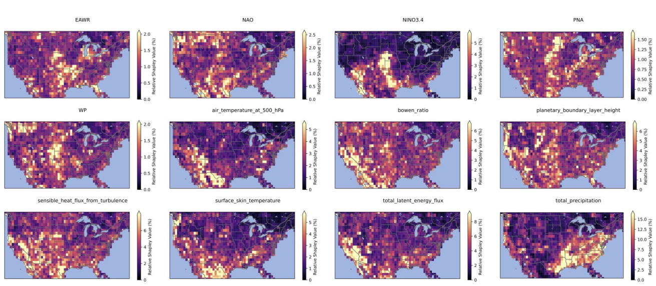

Figure 3: Geographic distribution of relative Shapley values for climate indices and variables in predicting surface soil moisture anomaly.

Approach and Results:

Data: MERRA-2 (1° grid, monthly, Jan. 1980 – Dec. 2022)

Model: Random Forest (ensemble of decision trees)

Shapley Explanations: decompose model's predictions into additive contributions from each variable (Figure 1)

Findings:

- **Improvement over persistence baseline**: nonlinear model compared with using previous month's value as prediction (Figure 2)
- **Feature Selection**: 3-5 dominate in their predictive power over lead times studied (0-6 months)
- **Coherent Spatial Patterns**: (Figure 3)
  - El Niño climate index (NINO3.4): Central US
  - Atmospheric variables (e.g., total precipitation): South East US
  - Land surface drivers (e.g., surface temperature): South West US
- **Poor model performance during the most anomalous periods**

Significance: The outcomes suggest that these techniques exhibit effectiveness in supporting a wide range of **Earth System applications**. The potential for near-real-time processing opens the door to diverse monitoring and predictive applications, for example in **water management**. The evolution and refinement of these capabilities will bolster JPL's efforts to establish comprehensive data production and analysis platforms, which are of particular interest to the state of California in support of the **California Climate Information System (CalCIS)**.

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

www.nasa.gov

Clearance Number: CL#00-0000
Poster Number: RPC#

Publications:

G. Doran, A. Altinok, D. Wenkert, N. Parazoo, K. Cawse-Nicholson, S. Mauceri, D. Posselt, and X. Zeng, *Measuring Relationships Between Atmospheric and Surface Variables to Understand Soil Moisture Processes With a Data-Driven Approach,* AGU 2023 (*submitted*).

PI/Task Mgr. Contact Information:

Alphan Altinok (alphan.altinok@jpl.nasa.gov)