Integrating Empirical Orthogonal Functions (EOFs) into Machine Learning Model

Kachinga Silwimba

Boise State University
Integrating Empirical Orthogonal Functions (EOFs) into Machine Learning Model

Abstract
The representation of soil moisture in Earth System Models, like the Community Earth System Model (CESM), is an essential facet in modeling the response of the Earth System to climate change. Since their inception, land models have grown to represent critical processes like carbon cycling, ecosystem dynamics, terrestrial hydrology, and agriculture. They serve as a lower boundary condition for atmospheric general circulation models. With increasing process representation, they are computationally expensive. Hydrologists and modelers use several parameterization schemes to describe the water and energy balance. However, this is regarded as computationally expensive. Alternative tools called emulators (e.g., machine learning and artificial intelligence) incorporated with the empirical orthogonal function analysis can represent soil moisture.

This student presentation is available at ScholarWorks: https://scholarworks.boisestate.edu/rcd_2023/14
Integrating Empirical Orthogonal Functions (EOFs) into Machine Learning Model

Kachinga Silwimba
Department of Geoscience, Boise State University

Introduction

EOFs are a popular tool for analyzing large datasets in various fields, such as climate science, oceanography, and geology. They identify dominant spatial and temporal patterns in data and reduce dimensionality while preserving essential information. Recently, there has been growing interest in integrating EOF analysis into machine learning models to enhance their performance and interpretability by extracting key features and relationships between variables. This study explores the effectiveness of incorporating EOF analysis into machine learning models using Climate Land Model output to predict and investigate the performance of the EOF-based model.

EOF Analysis

- EOF analysis finds a structure (or pattern) that conveys as much information as the original dataset without redundancy.
- EOF analysis provides means of filtering and compression of the data for better understanding.

**Mathematical Description**

The singular value Decomposition (SVD) technique is used to compute the EOF for climate dataset. This avoids having to compute the covariance matrix directly and is optimal for data sets with large spatial dimensions. The decomposition of A is

\[
X(t, s) = \sum_{k=1}^{M} \xi_k(t) \eta_k(s)
\]

where:

- X - gridded climate dataset,
- \( \xi_k(t) \) - principal component,
- \( \eta_k(s) \) - empirical orthogonal functions,
- M - number of modes in a field

**Integration of EOF into Machine Learning Model**

The extracted EOFs were then used as input features in various machine learning models, including artificial neural networks (ANNs), to predict soil moisture. The performance of the models was evaluated using various statistical metrics, including root mean square error (RMSE), mean absolute error (MAE), and Nash-Sutcliffe efficiency (NSE). The results showed that the EOF-based machine learning models outperformed the traditional machine learning models using raw data as input features.

**Conclusion**

The extracted EOFs were then used as input features in various machine learning models, including artificial neural networks (ANNs), to predict soil moisture. The performance of the models was evaluated using various statistical metrics, including root mean square error (RMSE), mean absolute error (MAE), and Nash-Sutcliffe efficiency (NSE). The results showed that the EOF-based machine learning models outperformed the traditional machine learning models using raw data as input features.

**References**