Research Computing Days 2023 Research Computing Days

3-28-2023

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.

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.

$$\mathbf{X}(t, \mathbf{s}) = \sum_{k=1}^{M} \mathbf{c_k}(\mathbf{t})\mathbf{u_k}(\mathbf{s})$$

$\mathbf{X}$ - gridded climate dataset, $\mathbf{c_k}(\mathbf{t})$ - principal component, $\mathbf{u_k}(\mathbf{s})$ - empirical orthogonal functions, M - number of modes in a field

Mathematical Description

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

$$\mathbf{A} = \mathbf{URV}^T \quad \mathbf{U} \in \mathbb{R}^{m \times m}, \mathbf{V} \in \mathbb{R}^{n \times n}, \mathbf{R} \in \mathbb{R}^{n \times n}$$

- The singular values along the R diagonal are customarily arranged in decreasing order.

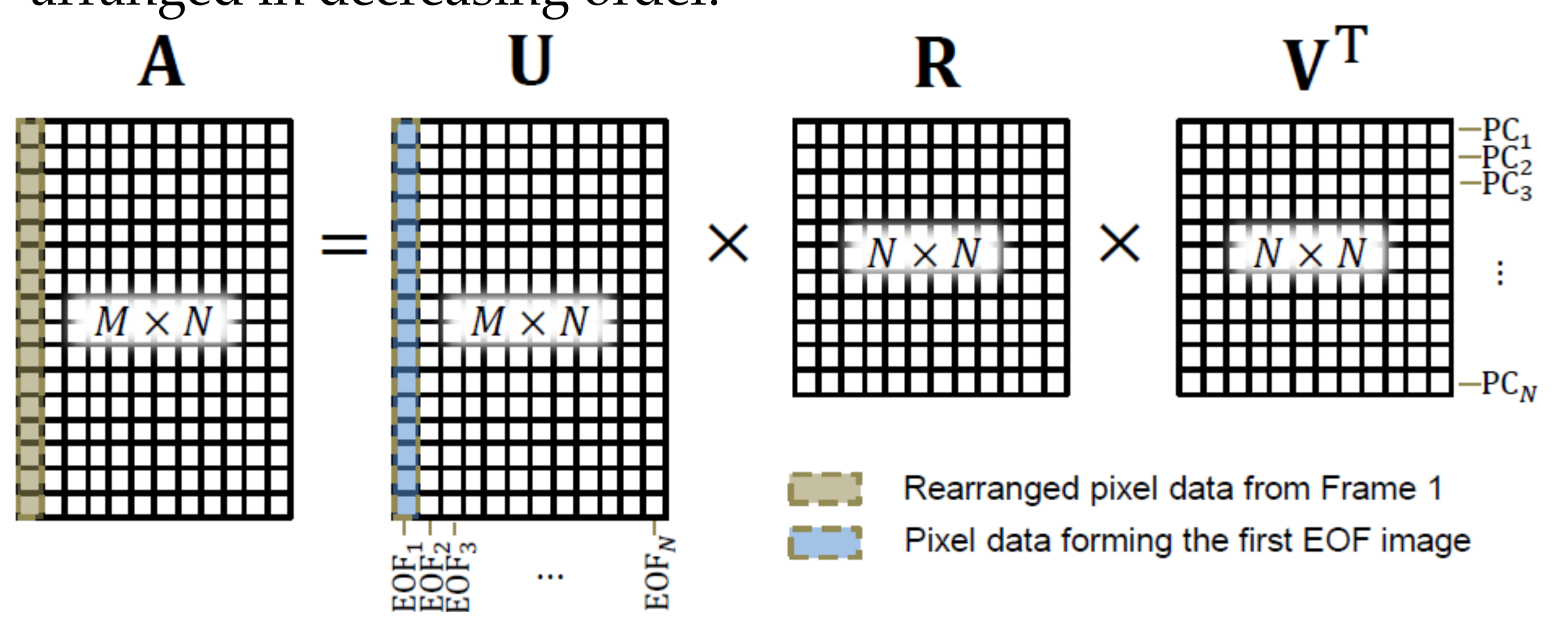

Fig.1. Shows a sample of the schematic diagram of the EOF and PC representation after the decomposition of the climate data using the (SVD).

Integration of EOF into Machine Learning Model

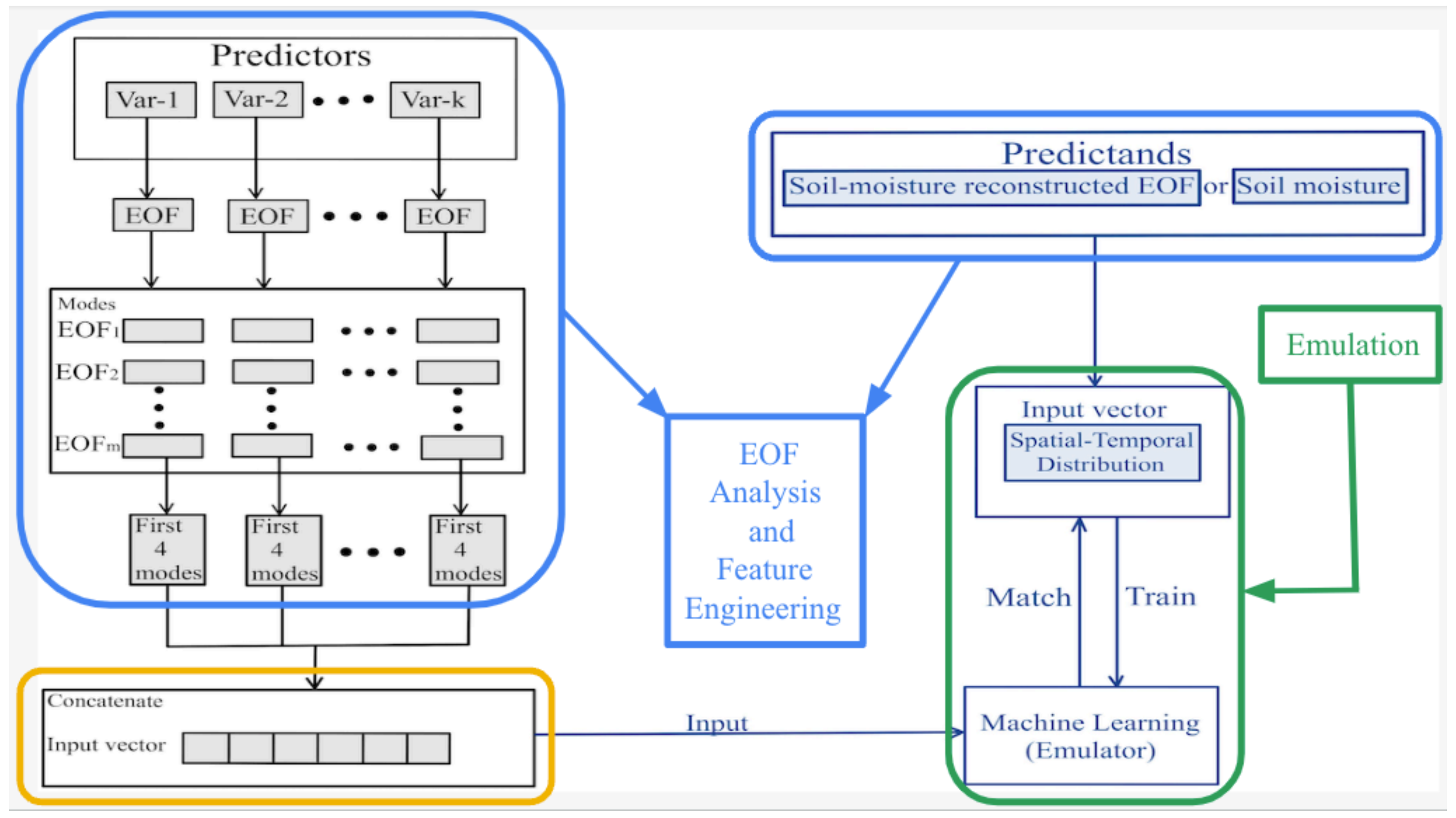

Fig.2. Schematic workflow diagram of the machine learning model and decomposition of the hydrological soil variable using the EOF analysis and the integration of EOF into the machine learning model to predict soil moisture.

Neural Network Model

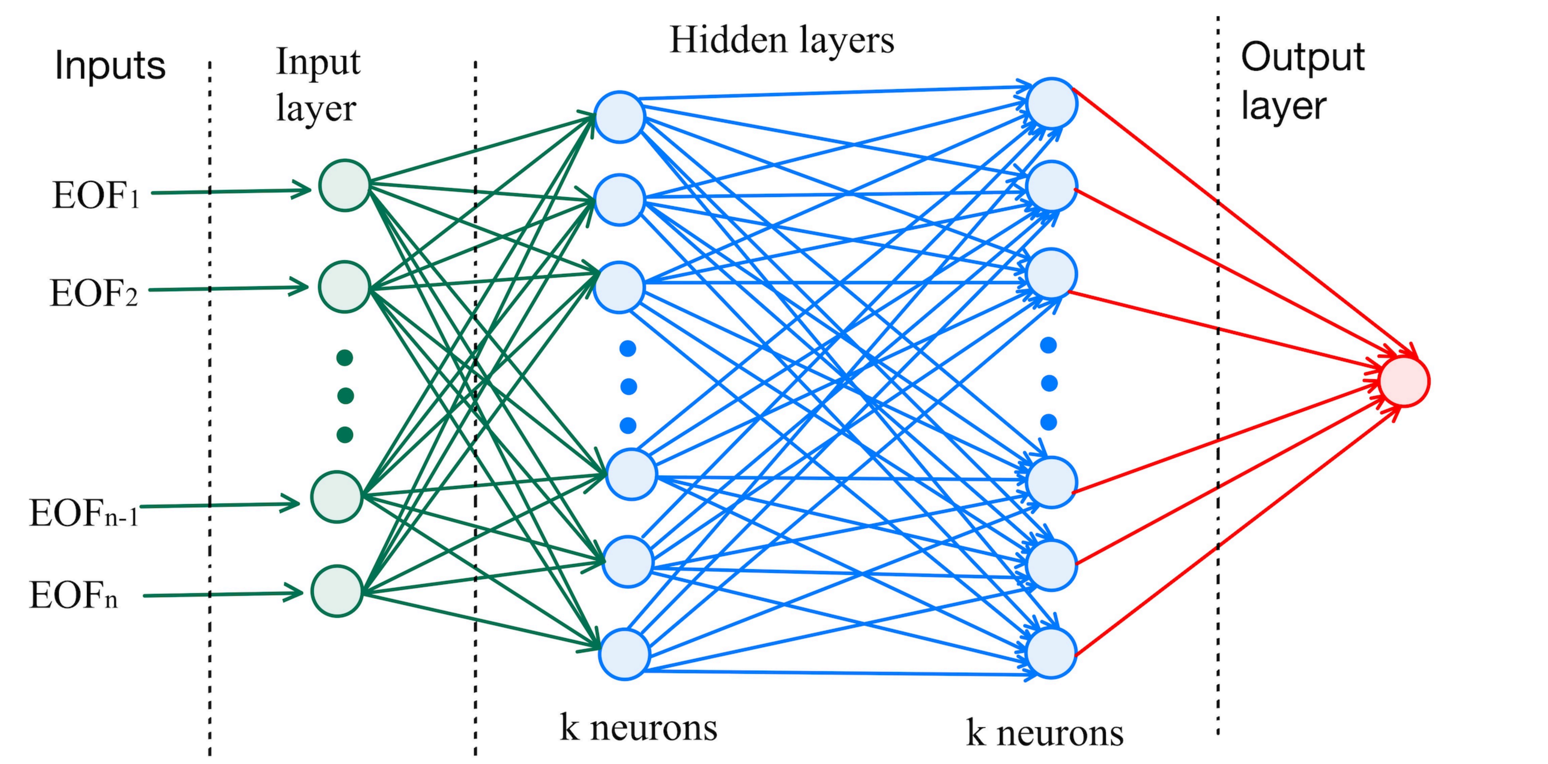

Fig.3. Schematic workflow of the machine learning model to predict soil moisture.

EOF Analysis Modes

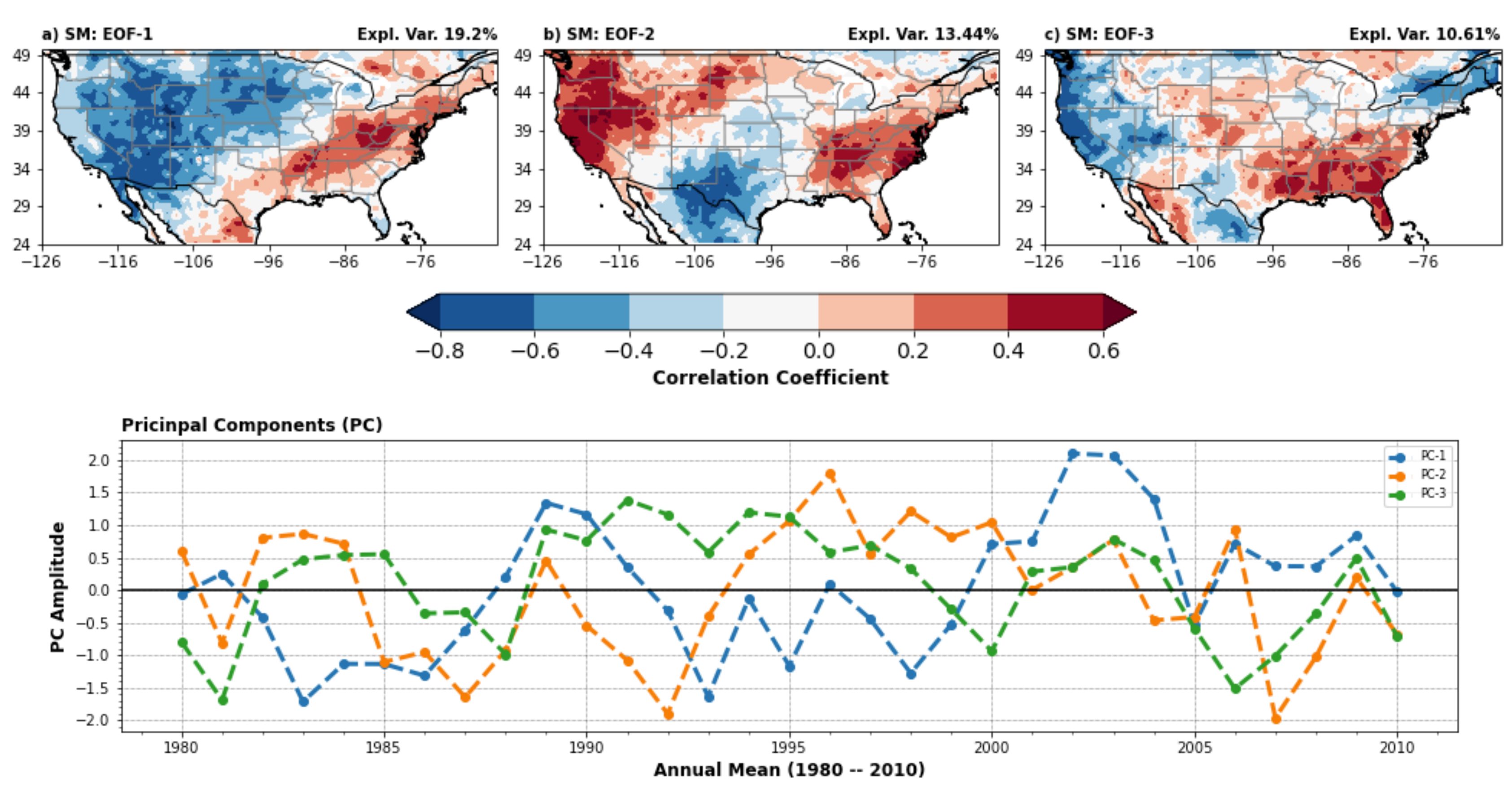

Fig.4. Results for the first three EOF modes for the 30-year soil moisture gridded data. The modes show the variation in soil moisture signals across the US.

Constructed Soil Moisture Data

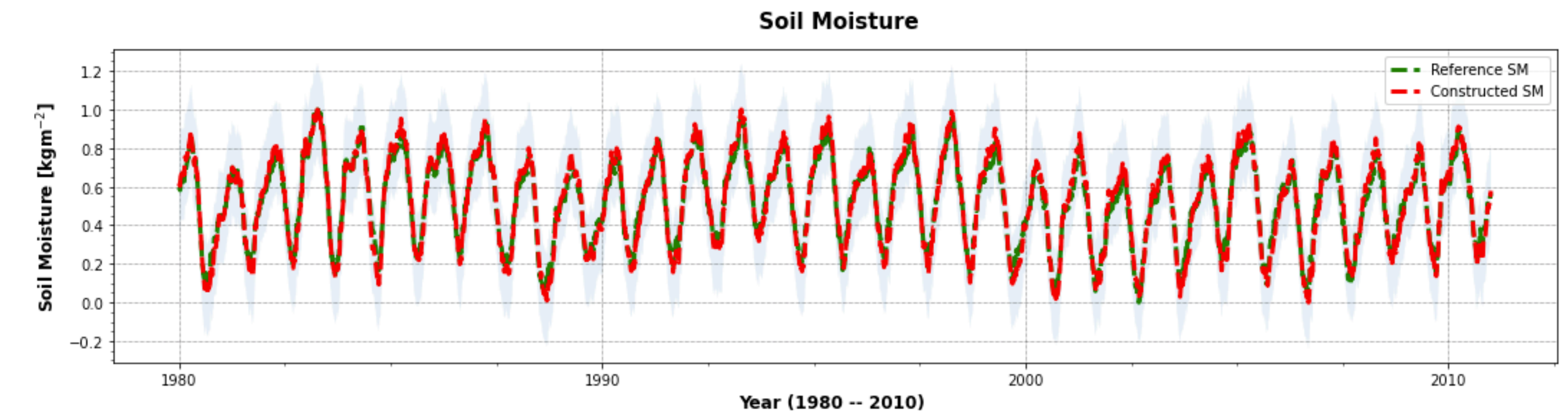

Fig.5. comparing the constructed soil moisture with the actual soil moisture dataset from the Climate Land Model. The constructed soil moisture is used in the output layer of the neural network to be predicted using the US.

Predicted Soil Moisture

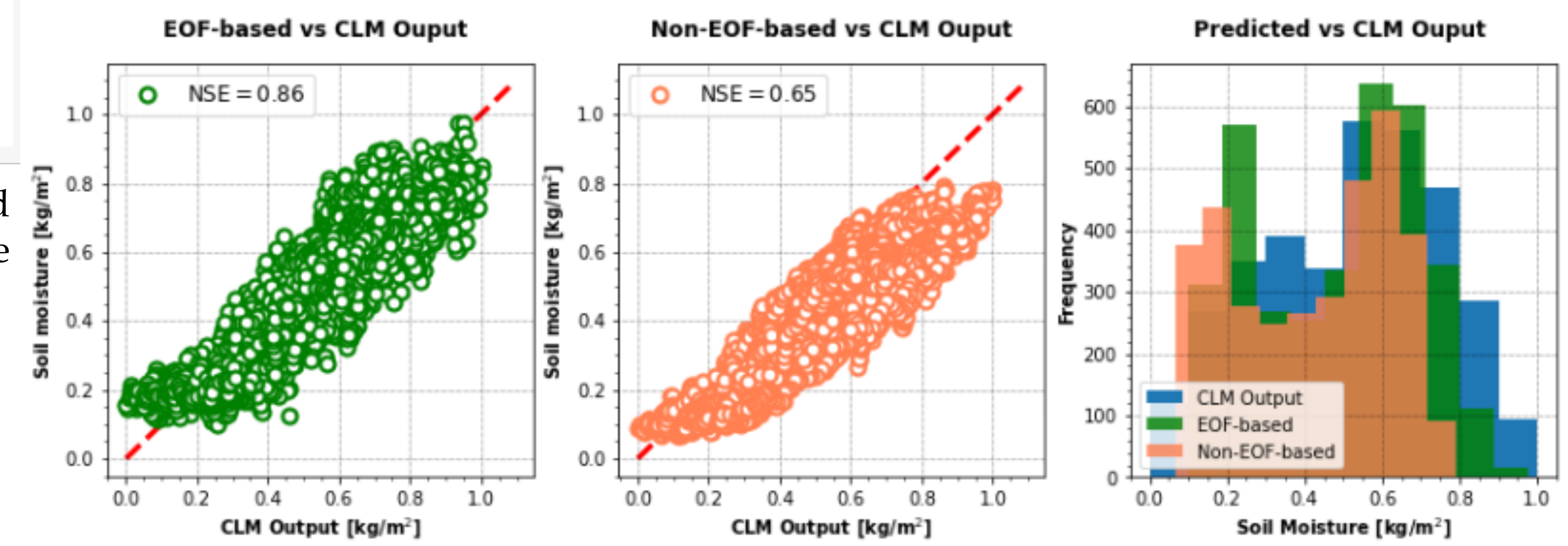

Fig.6. Scatter and distribution plot Comparing soil moisture prediction using the ML model with the actual CLM output over the US compared ton the EOF-based method.

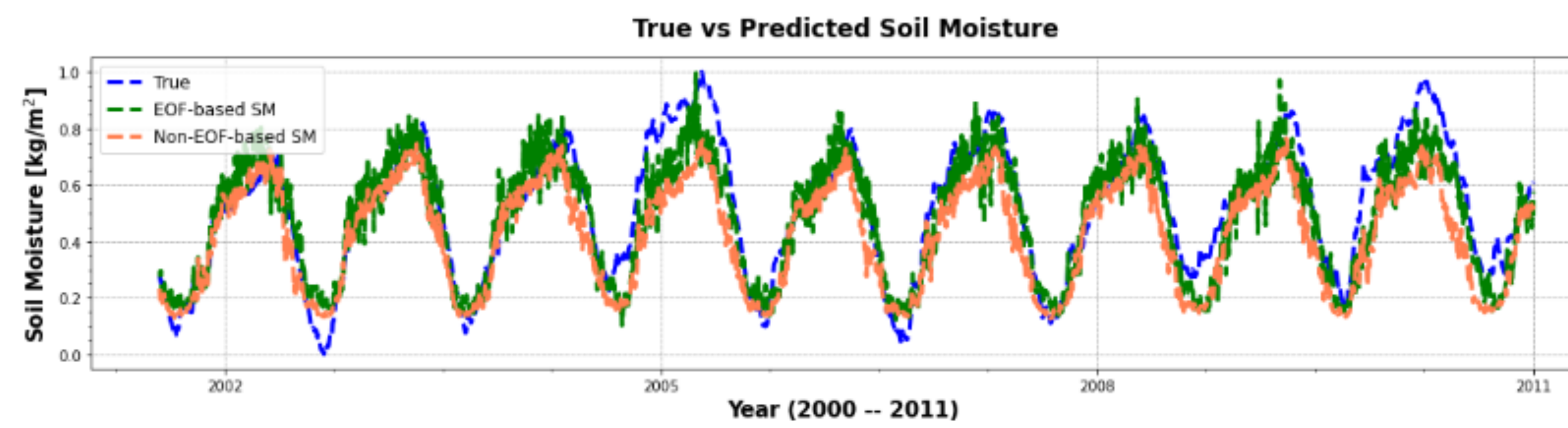

Fig.7 shows the Neural Network's Performance in predicting soil moisture in green color compared to the actual soil moisture output from the Climate Land Model in black color.

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

1. Wu, Y., Y. Chen, and Y. Tian (2022). Incorporating empirical orthogonal function analysis into machine learning models for streamflow prediction. Sustainability 14 (11), 6612.
2. Wang, T., T. E. Franz, R. Li, J. You, M. D. Shulski, and C. Ray (2017). Evaluating climate and soil effects on regional soil moisture spatial variability using eofs. Water Resources Research 53 (5), 4022–4035.
3. Hannachi, A., I. T. Jolliffe, and D. B. Stephenson (2007). Empirical orthogonal functions and related techniques in atmospheric science: A review. International Journal of Climatology: A Journal of the Royal Meteorological Society 27 (9), 1119–1152.

4. https://stats.stackexchange.com/questions/288949/what-exactly-is-a-principal-component-and-empirical-orthogonal-function