An Interpretable Machine Learning Model to Explore Relationships between Drought Indices and Ecological Drought Impacts in the Cheyenne River Basin, USA

by
Anne Britton

A thesis submitted to Johns Hopkins University in conformity with the requirements for the degree of a Master of Science in Environmental Sciences and Policy

Baltimore, Maryland
May 2023

Executive Summary

This research aims to assess the feasibility of using machine learning (ML) and explainable artificial intelligence (XAI) to identify the most effective drought indices for predicting changes in vegetation health. By doing so, the researcher intends to provide actionable results relevant to monitoring ecological drought and its impacts. The author, who has prior experience in conservation-related drought research focused on vegetation health and hydrological impacts, wanted to apply underutilized XAI techniques in the ecological drought space to a specific region to both determine the effectiveness of this approach and provide actionable results for drought monitoring in the study area.

The author's experience performing remote sensing-based drought research, as well as technical and scientific skills developed during her Master of Science in Environmental Sciences and Policy at Johns Hopkins University, form the foundation of this project. The author has acquired expertise in courses such as *Landscape Ecology*, *Hydrology & Water Resources*, *Programming and Data Management*, *Environmental Applications of GIS*, and *Analysis of Environmental & Ecological Data*. Additionally, the author's proficiency in Python, a necessary skill for this project, was first developed during her experience as a researcher in the NASA DEVELOP National Program and was honed during her work as a research assistant in Benjamin Zaitchik's Hydroclimate Research Group in the Johns Hopkins University Department of Earth & Planetary Sciences. Overall, this capstone project synthesizes the author's learnings both inside and outside of the classroom during her time at Johns Hopkins University and contributes to the field of ecological drought research through the use of XAI methods.

Abstract

Rangeland ecosystems across the United States have significant biological, economic, and cultural value. However, the increasing frequency and severity of droughts across the country may lead to unforeseen impacts on these ecosystems. To address this challenge, this study aimed to identify relationships between drought indices and vegetation health in the Cheyenne River Basin, USA, using machine learning (ML) and explainable artificial intelligence (XAI) methods. Using Terra Moderate Resolution Imaging Spectroradiometers (MODIS), University of Idaho Gridded Surface Meteorological Dataset (gridMET), and Daymet data, the study employed XGBoost Regressor and Extra Trees Regressor models in unison with SHapley Additive exPlanations (SHAP) to evaluate predictive performance and the connections between drought indices, environmental variables, and the Normalized Difference Vegetation Index (NDVI). Tests of model performance demonstrated that the XGBoost model performed moderately well at predicting NDVI and was therefore useful for further XAI analysis with SHAP. SHAP explainer results showed that the Palmer Drought Severity Index (PDSI), the 90-day Standardized Precipitation Index (SPI), and snow water equivalent (SWE), were the most important predictors of NDVI values and are therefore closely associated with vegetation health in the study area. The findings of this study first demonstrate the feasibility and usefulness of applying XAI, an underutilized method in the drought space, to study ecological drought indicators. Secondly, results provide an understanding of which commonly used drought indices correlate with effects on vegetation health in the study area, as well as the specific directionality of these relationships. These results can be used to inform drought research and monitoring practices and anticipate ecological drought impacts in the Cheyenne River Basin.

Advisor and Primary Reader: Dr. Garrett Graham

Secondary Reader: Dr. Michael Schwebel

Acknowledgments

First, I would like to acknowledge my capstone advisor, Dr. Garrett Graham, a Research Associate with the North Carolina Institute for Climate Studies and a Cooperative Institute Affiliate with NOAA National Centers for Environmental Information, for providing extensive resources, knowledge, and advice to ground this project in proven and reproducible machine learning and XAI methods. I would also like to thank Molly Woloszyn from NOAA's National Integrated Drought Information System for providing me with support and drought expertise throughout the proposal and production of this research. Additionally, I would like to thank Dr. Michael Schwebel for being my capstone instructor and secondary reader, and Dr. Cassandra Hansen for being my academic advisor throughout my Masters.

Lastly, I would like to thank Dr. Grace Brush, Professor Emerita of Environmental Health and Engineering at Johns Hopkins University, who agreed to take on an undergraduate Anthropology major as a research assistant, fostered my love for ecology, and entirely changed my perspective on my own abilities as a scientist. I am incredibly lucky to have learned from and worked under Grace and endlessly look forward to our emails, phone calls, book exchanges, and discussions over lunch.

Table of Contents

List of Tables

List of Figures

Glossary

AOI – Area of Interest, a geographic region or location that is the focus of analysis (i.e., the Cheyenne River Basin).

API – Application Programming Interface, a set of protocols and tools for building software applications.

Daymet – A gridded dataset of daily weather and climate variables for North America, including temperature, precipitation, humidity, and radiation.

Drought index – An index used to assess drought conditions, typically based on meteorological, hydrological, or agricultural data.

Earth observations – The collection and analysis of data from satellites, airborne sensors, and ground-based measurements to better understand the Earth's systems.

Ecological drought – A type of drought that affects ecosystems, typically characterized by reduced water availability and changes in vegetation and wildlife.

EDDI – Evaporative Demand Drought Index, an index that quantifies the atmospheric demand for moisture.

Environmental variable – A variable that describes a characteristic of the environment, such as temperature or precipitation.

Feature Importance – A measure of the importance of each feature in a machine learning model for making predictions.

GEE – Google Earth Engine, a cloud-based platform for analyzing and processing geospatial data.

gridMET – University of Idaho Gridded Surface Meteorological Dataset, a high-resolution meteorological dataset derived from various meteorological observation sources.

ML – Machine Learning, a subfield of artificial intelligence that focuses on developing algorithms that can learn from and make predictions on data.

MODIS – Moderate Resolution Imaging Spectroradiometer, a NASA Earth Observing System instrument that provides global coverage of land, ocean, and atmosphere.

MSE – Mean Squared Error, a common metric for evaluating the performance of regression models.

NDVI – Normalized Difference Vegetation Index, a remotely sensed index that measures the greenness of vegetation.

NNs – Neural Networks, a type of machine learning algorithm that models the structure and function of the human brain.

Palmer Z – A drought index that incorporates both precipitation and temperature data.

PDSI – Palmer Drought Severity Index, a commonly used index for assessing drought conditions.

Pentad – A period of five consecutive days, often used in climate and meteorological analyses.

R-squared – A statistical measure that represents the proportion of variance in a dependent variable that can be explained by an independent variable.

RMSE – Root Mean Squared Error, a measure of the difference between predicted and observed values in a regression model.

SHAP – SHapley Additive exPlanations, a method for interpreting machine learning models and understanding feature importance.

SPEI – Standardized Precipitation Evapotranspiration Index, a drought index that considers both precipitation and evapotranspiration.

SPI – Standardized Precipitation Index, a meteorological drought index that quantifies precipitation deficits.

SVR – Support Vector Regression, a type of machine learning algorithm used for regression problems.

SWE – Snow Water Equivalent, a measurement of the amount of water contained within snowpack.

Tree-based methods – Machine learning algorithms that use decision trees as the primary model structure, including Random Forest and Gradient Boosting.

XAI – Explainable Artificial Intelligence, a field that seeks to make artificial intelligence transparent and understandable to humans.

XGBoost – Extreme Gradient Boosting, a popular machine learning algorithm for regression and classification problems.

1. Introduction

Ecological drought is increasingly cited as a result of a warming climate, shifting oceanic and atmospheric processes, and increased human water usage. As ecosystems become more vulnerable to these effects, processes driving ecological drought may result in ecosystem transformations and subsequent effects on ecosystem services (Crausbay et al., 2017). Crausbay et al. best define ecological drought in their foundational paper, *Defining Ecological Drought for the Twenty-First Century*, as an "episodic deficit in water availability that drives ecosystems beyond thresholds of vulnerability, impacts ecosystem services, and triggers feedbacks in natural and/or human systems," (Crausbay et al., 2017). When exactly various ecosystems are driven beyond these thresholds, to what extent ecosystem services may be impacted, and how to properly measure and predict these variables, are each questions that vary by impact, ecosystem, and scale.

While accurate drought characterization can be difficult, there are many established ways to measure and forecast hydrological and meteorological drought (Hao et al., 2017). For instance, the U.S. Drought Monitor (USDM) incorporates physical drought indicators, such as the Palmer Drought Severity Index (PDSI), the Standardized Precipitation Index (SPI), the Keetch-Byram Drought Index for fire, vegetation health and soil moisture data, and hydrologic data into one map to provide the current depiction of drought (Hao et al., 2017; U.S. Drought Monitor, 2022). However, links between many of these traditional drought indices and ecological drought impacts, such as declines in vegetation health, are not always well established at a variety of ecosystems and scales, particularly at scales useful to natural resource management (Bradford et al., 2020; Crausbay et al., 2020; Wiens & Bachelet, 2010).

To bridge this gap, this research employed a combined approach of machine learning (ML) models and SHapley Additive exPlanations (SHAP), an eXplainable AI (XAI) method, to evaluate connections between drought indices, environmental variables, and vegetation health. The purpose of this research was to determine which features, as indicated by SHAP values obtained from a trained ML model, contribute most towards predicting Normalized Difference Vegetation Index (NDVI) values in the Cheyenne River Basin, a subbasin within Missouri River Basin. The Cheyenne River Basin was selected due to the presence of non-agricultural areas that allowed for a more consistent assessment of vegetation health, as measured by Terra MODIS-derived NDVI, and the occurrence of multiple drought events across the basin over the past two decades. For this research, two alternative hypotheses were proposed:

1. **(H1)** An ML regression model that incorporates drought indices and environmental variables can provide a reliable prediction of NDVI with at least 50% of the variance explained.
2. **(H2)** In a reliably predictive ML regression model that includes drought indices and environmental variables features, drought indices will have higher absolute mean SHAP values than environmental variables, indicating that drought indices have a stronger association with NDVI in the Cheyenne River Basin.

This research project makes progress towards one of the National Oceanic and Atmospheric Administration's (NOAA) National Integrated Drought Information System (NIDIS) key priorities in the 2021-2023 Missouri River Basin Drought Early Warning System (DEWS) Strategic Action Plan (NOAA/NIDIS, 2020). This plan, which was developed in consultation with partners and stakeholders throughout the Missouri River Basin, recognizes the need to build a

comprehensive understanding of drought indicators and their application within the Basin, with a specific focus on identifying drought indicators most useful for monitoring ecological drought. By identifying the most significant predictors of NDVI in the basin using SHAP values, this research will contribute to the overall understanding of drought impacts on vegetation health in the region and provide valuable information for decision-making and resource management, including the identification of promising ecological drought indicators for the region.

2. Literature Review

2.1 Machine Learning and Explainable AI in Drought Research

As the amount of available Earth system data continues to increase, a key research challenge comes with extracting useful and actionable information from these data (Reichstein et al., 2019). One of the main issues that research must address is how to create models that can learn as much from data as possible, while still providing actionable and directed insights. As such, ML has become a key approach in geoscience research (Reichstein et al., 2019). In the arena of drought research, ML methods are beneficial since they are less time-consuming, typically require fewer inputs, and are generally less complex than complete physical models (Mokhtar et al., 2021; Sundararajan et al., 2021).

ML techniques such as random forest, boosted regression trees, support vector regression (SVR), and neural networks (NNs) have increasingly been used to successfully model and predict drought conditions (Belayneh & Adamowski, 2013; Dikshit et al., 2022; Park et al., 2016; Shamshirband et al., 2020; Sundararajan et al., 2021). However, as published research on ecological drought increases (Figure 1), there is still a relative lack of research specifically examining ecological drought using ML methodologies (Clarivate Web of Science, 2023).

Therefore, this area is ripe for exploration using ML techniques that have been used to successfully predict other types of drought.

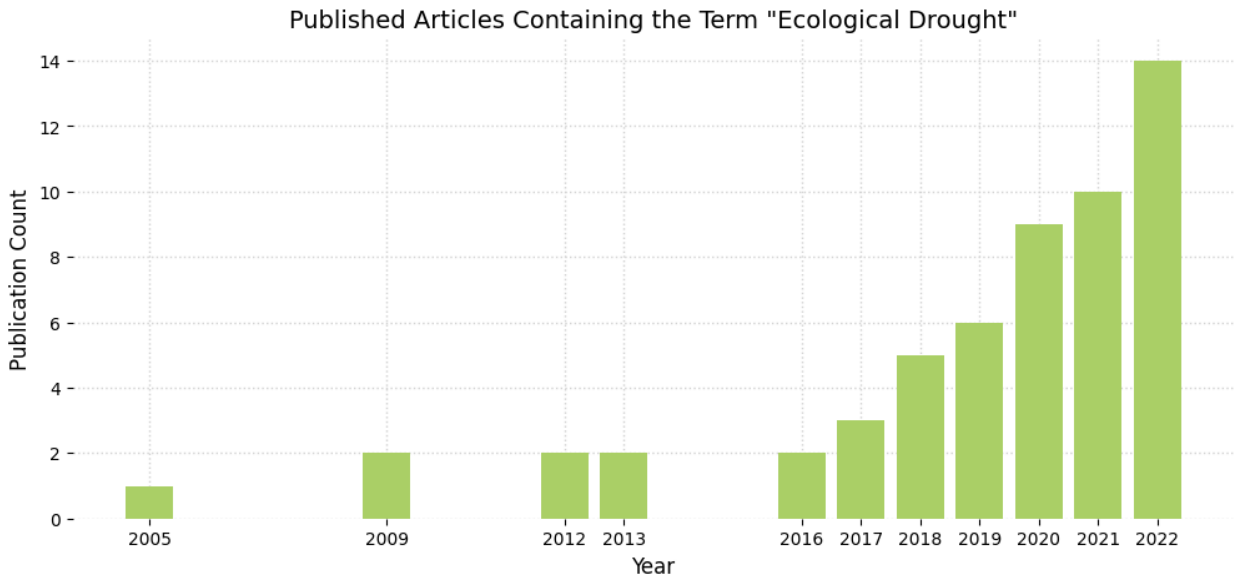


Figure 1. Number of published articles by year containing the term "ecological drought," from Clarivate Web of Science, 2023.

Extra Trees Regressor and XGBoost, the two ML models selected for this project, are powerful algorithms widely used in Earth science and environmental sciences (Chen & Guestrin, 2016; Geurts et al., 2006; Liaw & Wiener, 2002). Extra Trees Regressor, a variant of decision tree-based ensemble methods, is an ML technique introduced by Geurts et al. (2006) as an extension of the Random Forest algorithm. It works by constructing multiple decision trees with randomized splitting rules and aggregating their predictions to obtain a final output. The Extra Trees Regressor is shown to be less prone to overfitting when compared with other tree-based ensemble methods, as the randomization of splitting rules reduces the variance of the predictions (Geurts et al., 2006).

Extreme Gradient Boosting (XGBoost) is another popular variant of decision tree-based ensemble methods, similar to Extra Trees Regressor. XGBoost was introduced in 2014 as a

gradient tree boosting system and is available as an open-source software library. Like Extra Trees Regressor, XGBoost constructs multiple decision trees, but instead of using randomized splitting rules, it utilizes gradient boosting to iteratively improve the accuracy of the predictions (Chen & Guestrin, 2016). XGBoost has been shown to be particularly effective in solving a wide range of machine learning problems and is scalable, running "more than ten times faster than existing popular solutions on a single machine," making it particularly suitable for a project of this scope (Chen & Guestrin, 2016).

Despite the significant benefits of ML methods, Balti et al. (2020) highlight that ML approaches for drought monitoring have faced recent criticism for their poor capacity in reasoning and diagnosing the logic behind generated decisions, a concern also echoed by Samek et al. (2017) (Balti et al., 2020; Samek et al., 2017). This limitation is often attributed to the limited interpretability of ML models. Therefore, in this project, SHAP will be used to interpret the relationships between features and results. SHAP was first introduced as a solution concept derived from Shapley values in game theory to determine the contributions of individual players in a cooperative game (Shapley, 2016). It was developed into a method for interpreting complex machine learning models by quantifying the contribution of each feature to the model's output (Lundberg & Lee, 2017; Shapley, 2016). SHAP measures how much a feature adds to or subtracts from the prediction compared to its absence, which allows for a more nuanced understanding of how a model is making its predictions than methods such as feature importance scoring. Research has shown that using SHAP for drought prediction can significantly inform resulting decision-making practices (Dikshit & Pradhan, 2021a, 2021b; Lundberg & Lee, 2017).

2.2 Machine Learner Target

This project used the Normalized Difference Vegetation Index (NDVI) as a proxy for ecological drought and the target data for the machine learner. NDVI is a widely used remote sensing index that quantifies the greenness of vegetation on the Earth's surface based on the reflectance of near-infrared and red light by plants. NDVI can be derived from satellite imagery using the reflectance ratio between red and near-infrared bands (Pettorelli et al., 2005). This formula (below) generates values on a scale of -1 to +1. Negative values of NDVI typically indicate bare soil or little vegetative cover, while low positive values suggest unhealthy vegetative cover and high positive values indicate a high degree of healthy vegetative cover.

$$NDVI = \frac{NIR - RED}{NIR + RED}$$

Studies have shown that NDVI is a valuable tool to monitor drought conditions and their impacts on vegetation health (Anyamba & Tucker, 2012; Tucker et al., 1986). For example, studies have found that declining NDVI values are indicative of reduced vegetation activity and productivity during drought events (Park et al., 2016; Pettorelli et al., 2005; Phillips et al., 2008). According to Pettorelli et al. (2005), "NDVI has shown consistent correlation with vegetation biomass and dynamics in various ecosystems worldwide." NDVI is therefore a valuable proxy for the impacts of ecological drought on vegetation and has been successfully predicted in the past using ML methods (Li et al., 2021; Roy, 2021). However, this project was distinguished from past predictive NDVI research due to the goal of determining which commonly used and accessible drought indices and environmental variables contribute most towards predicting NDVI, thereby working towards identifying drought indicators for ecological drought in the study area.

2.3 Machine Learner Features

Several drought indices and environmental variables were selected as features for the ML model. These features encompassed different aspects of drought dynamics, including short and long-term soil moisture status, atmospheric evaporative demand, precipitation anomalies, temperature, and snow cover dynamics. Features were selected by considering relationships between drought and vegetation dynamics and the inclusion of commonly used, easily reproducible indices that are accessible to a variety of stakeholders. The inclusion of multiple drought indices and environmental variables provided a more holistic approach to capturing the complex interactions between drought and vegetation dynamics, which can enhance the accuracy and robustness of a predictive model.

Drought Indices

- Palmer Drought Severity Index (PDSI): A widely used drought index that quantifies drought conditions based on precipitation and temperature data while taking into account the water-holding capacity of soils (Palmer, 1965). The index provides information on the long-term moisture status of an area with a timescale of approximately nine months (Svoboda & Fuchs, 2016).
- Palmer Z Score: A derived score from PDSI that responds better to short-term drought conditions and can be used to identify developing and regressing drought conditions (Palmer, 1965; Svoboda & Fuchs, 2016).
- Standardized Precipitation Index (SPI): An index that represents the likelihood of precipitation as calculated at any number of timescales, from 1 month to 48 months or longer, using historical precipitation records (McKee et al., 1993). The simplicity of the

index and the ability to calculate SPI at varying timescales allows for use in various applications, however, it does not account for a temperature component (Svoboda & Fuchs, 2016).

- Standardized Precipitation Evapotranspiration Index (SPEI): A drought index that incorporates both precipitation and potential evapotranspiration (derived from temperature data) to determine drought (Svoboda & Fuchs, 2016). The inclusion of temperature data may help to account for a wider variety of drought impacts. Like SPI, SPEI is calculated over different time scales (Vicente-Serrano et al., 2010).
- Evaporative Demand Drought Index (EDDI): An experimental drought index that indicates how anomalous atmospheric evaporative demand is across a variety of timescales, making it a relevant feature for capturing the atmospheric moisture stress on vegetation (Hobbins et al., 2016). EDDI excels at catching the early warning signs of water stress on weekly to monthly timeframes (Hobbins et al., 2016).

Environmental Variables

- Daily Maximum and Minimum Temperature: An important factor in determining drought impacts as well as plant growth and development. Maximum temperature refers to the highest temperature recorded during a 24-hour period, while minimum temperature refers to the lowest temperature recorded during the same period (Abatzoglou, 2013).
- Daily Precipitation: Refers to moisture in the form of rain, snow, sleet, or hail. Precipitation plays a critical role in determining water availability in ecosystems (Abatzoglou, 2013).

- Snow Water Equivalent (SWE): The amount of water that would be equivalent to the depth of snow cover if it melted. Snow cover can impact NDVI as it affects the amount of snowmelt and therefore the timing, amount, and duration of water available for the growing season (US Department of Commerce & NOAA's Weather Service, n.d.).

3. Methods

This section presents the methods used in this study. It begins with an overview of the study area (*3.1*), and the subsequent subsections detail the data acquisition process, data preprocessing steps, model selection, and hyperparameter optimization techniques employed. Specifically, subsection *3.2* outlines the data sources used in this study, while subsection *3.3* describes the steps taken to clean, preprocess, and transform the data into a format suitable for analysis. Subsection *3.4* discusses the selection methods and criteria used to choose the models for the analysis. Finally, subsection *3.5* provides a detailed explanation of the hyperparameter optimization techniques used to fine-tune the models. Together, these subsections provide a comprehensive account of the methods employed in this study.

3.1 Area of Study

The Cheyenne River Basin spanning South Dakota, Wyoming, and Nebraska was chosen for this study for its extensive rangeland ecosystem and history of drought (Figure 2). The study area is characterized by a range of physical features, including rolling hills, plains, badlands, and plateaus (Culler et al., 1961; Ehlert, 2022). The Cheyenne River runs through the center of the region, originating in Wyoming and flowing eastward through South Dakota before eventually joining the Missouri River. The geology of the region is primarily defined by marine sediments from the Mesozoic and Cenozoic eras, but also includes igneous and metamorphic rocks (Culler

et al., 1961; Ehlert, 2022). This region's climate is characterized by its semi-arid to arid conditions, with limited precipitation and high variability in temperature (Ehlert, 2022). The annual precipitation in the basin ranges from 10 to 20 inches, with the majority falling as rain during the spring and summer months (Culler et al., 1961; United States Bureau of Reclamation, 2019).

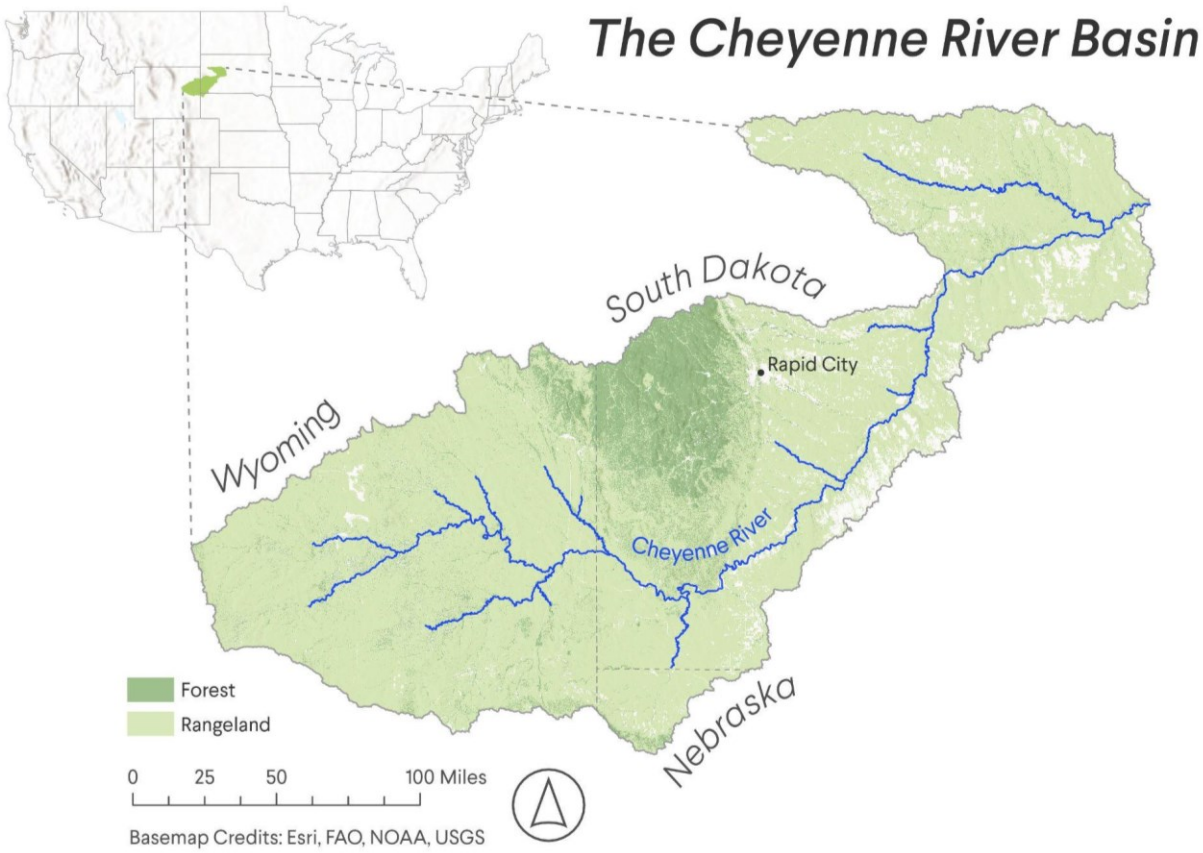


Figure 2. The extent of the Cheyenne River Basin, spanning Wyoming, Nebraska, and South Dakota.

The basin is predominantly made up of rangelands, with forests occurring in the north-central portion of the study area, and a minimal amount of developed and agricultural land interspersed throughout (Figure 2). Rangelands are a key ecosystem in this region, with native rangeland accounting for 46% of Nebraska's land mass, over 50% of South Dakota's land mass,

and over 85% of Wyoming's land mass (Boden, 2023; Ehlert, 2023; "Introduction to Wyoming Rangelands," 2023). Rangelands consist primarily of native grasses, forbs, and shrubs, and may also include woodlands with open canopies and a substantial understory (Boden, 2023; Ehlert, 2023; "Introduction to Wyoming Rangelands," 2023). The definition does not specify a specific land use, emphasizing that rangelands are defined by the ecosystems they sustain, rather than how they are utilized. This means that resource management strategies must be designed to be implemented across private, state, and federal lands, and to consider the relationship between climatic, environmental, and sociological factors. According to Krista Ehlert, Assistant Professor and Range Specialist at South Dakota State University, "Land managers need tools and techniques to help them monitor rangeland condition, improve utilization of rangeland resources, control invasive species, and develop management plans to respond to challenges resulting from drought and other natural disasters," (Ehlert, 2023). Due to the complex nature of these ecosystems and their uses, the effects of ecological drought on rangelands could lead to ecosystem transformations and impacts on a variety of ecosystem services.

3.2 Data Acquisition

The data acquisition process for this project involved acquiring data using the Google Earth Engine (GEE) Python API in a Jupyter Notebook hosted on Google Colaboratory.[1] GEE provides open access to a diverse collection of satellite and remote sensing datasets. All data are available to any GEE user, improving reproducibility. The following Earth observation data were acquired as image collections for the date range of February 24, 2000 (Terra MODIS's start

[1] In addition to the use of Google Colaboratory, Github was used for version control and code storage. See Appendix I for a link to the Github repository for this project.

date), and December 31, 2021: Terra Moderate Resolution Imaging Spectroradiometer (MODIS) surface reflectance data, Gridded Surface Meteorological (gridMET) Dataset Palmer Drought Severity Index (PDSI), Evaporative Demand Drought Index (EDDI), Standardized Precipitation Index (SPEI), Standardized Precipitation Evapotranspiration Index (SPEI), maximum and minimum temperature, and precipitation data, and Daymet snow water equivalent (SWE) data (Table 1).

Table 1. Earth observations acquired for the ML models.

Data Type	Data Platform	Variable(s)	Spatial & Temporal Resolution	Time Period
Drought Impact (Target)	MOD09GA v006: MODIS/Terra Surface Reflectance Daily L2G Global 1 km SIN Grid	Normalized Difference Vegetation Index (NDVI) - Derived	1 km, Daily	2000 – Present
Drought Indices (Features)	Gridded Surface Meteorological (gridMET) Dataset	Palmer Drought Severity Index (PDSI) Palmer Z Score Evaporative Demand Drought Index (EDDI) – 30, 90, 180-day Standardized Precipitation Index (SPI) – 30, 90, 180-day Standardized Precipitation Evapotranspiration Index (SPEI) – 30, 90, 180-day	4 km, Pentads (5-day)	1979 – Present
Environmental Variables (Features)	Gridded Surface Meteorological (gridMET)	Maximum & Minimum Temperature	4 km, Daily	1979 – Present

	Dataset	Precipitation		
	Daymet	Snow Water Equivalent (SWE)	1 km, Daily	1980 – Present

3.3 Data Preprocessing

To calculate an NDVI band from the MODIS surface reflectance data, the image collection obtained through GEE was first quality controlled. Using bits 10 (the internal cloud algorithm flag bit) and 15 (the internal snow mask bit) from the QA band, cloudy and snow pixels were masked from the images. Next, a new NDVI band was computed for each image in the filtered MODIS collection using the normalized difference of bands *sur_refl_b02* and *sur_refl_b01*. The select function was then applied to retain only the NDVI band in the image collection. This process produced a new image collection that contained NDVI data for the selected date range with cloudy and snowy pixels masked out.

After completing the masking and NDVI calculations on the MODIS data, GEE was used to calculate the spatially averaged value for each variable in Table 1 across the study area over time. First, the study area shapefile (AOI) was imported into Colab and converted to a GEE object. Next, a function was built to extract necessary parameters such as the image collection, variable, AOI, and spatial scale from an input list. The function then calculated the mean value of the variable across the AOI for every image in the collection. This process produced a list of lists where each image had a two element list that contained the image's date and mean value. The nested list was then converted to a DataFrame using the pandas.DataFrame constructor. Output DataFrames for each variable were concatenated into a single DataFrame containing

spatially averaged values for the AOI for all variables between February 24, 2000, and December 31, 2021.

The next step was to calculate anomalies. This was done by creating a function to group the data by day of the year and calculate the daily mean for each variable. The function then mapped each day of data to the average value for that day of year and calculated the daily anomaly as the difference between the original value and the average value for that day of the year. The function was applied to each column of the DataFrame using a for-loop, resulting in a DataFrame that only contained the anomaly data.

The next preprocessing step was to resample the environmental variable data to match the temporal frequency of the gridMET drought indices. Five-day periods - pentads - are used to calculate the drought indices acquired from gridMET. To both resample and smooth the data for the machine learner, a 30-day rolling average was calculated for every pentad date across all variables, ensuring that all data was resampled to a five-day temporal resolution. Finally, data were standardized using the following formula (Figure 3):

$$Standardized\ Data = (Data - Mean(Data)) \div Standard\ Deviation(Data)$$

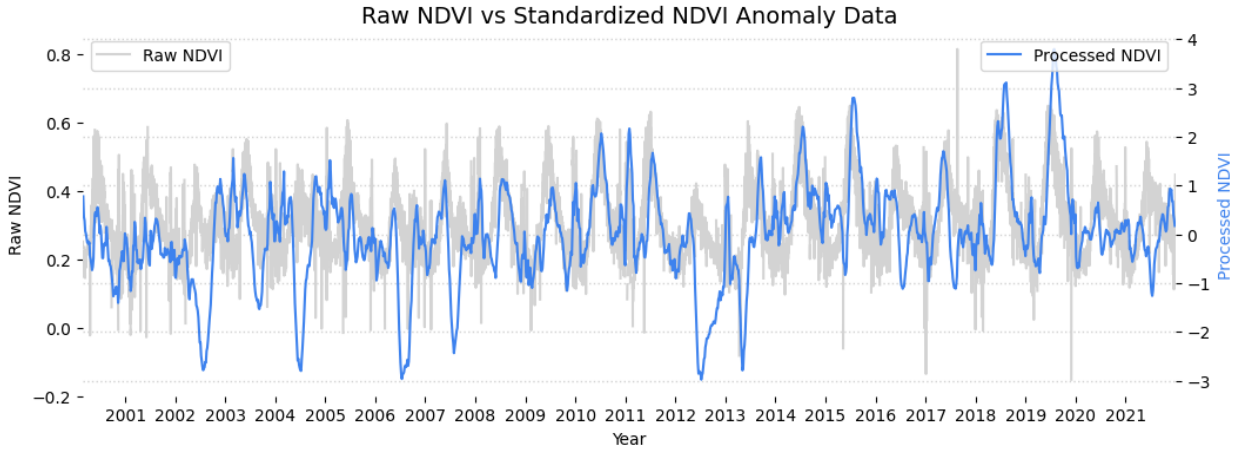


Figure 3. Raw NDVI versus standardized NDVI anomaly data as an example of the difference between the raw and preprocessed data.

3.4 Model Selection

After finalizing data preprocessing, the project next turned toward selecting an appropriate ML model. Using the .values() function, values in the DataFrame were converted to NumPy arrays. NDVI assumed the y value as the target for the learner, while all other variables served as features in the X set. Using the scikit-learn Python machine learning library, the data were first run through an untuned Decision Tree regressor to ensure that the data were formatted and preprocessed correctly.

Next, the Lazy Predict package was used to explore the performance of a variety of regression models on the data. Lazy Predict is beneficial because it runs the data through 40+ regression models without having to code each model individually (Pandala, 2022). It provides several statistics for each model such as adjusted R-squared, R-squared, Root Mean Squared Error (RMSE), and time to run. To cross-validate these statistics across the data, a function was created to run Lazy Predict across five folds of shuffled data for each untuned model and take the mean of the folds' performances. The results from Lazy Predict gave a preliminary idea of each model's untuned performance on the data (Table 2).

Table 2. Top five performing models from shuffled k-fold cross-validation of Lazy Predict.

Model	Adjusted R-Squared	R-Squared	RMSE	Time Taken
Extra Trees Regressor	0.93	0.93	0.26	0.65
Gaussian Process Regressor	0.92	0.93	0.27	0.23
LGBM Regressor	0.89	0.90	0.32	0.20
Hist Gradient Boosting Regressor	0.89	0.89	0.32	0.57
XGBoost Regressor	0.89	0.89	0.33	0.57

Based on the performance of the models in Lazy Predict, the Extra Trees Regressor was selected as the primary model for the project, while the XGBoost Regressor was selected as a comparison model. These models were chosen due to their Adjusted R-Squared and R-Squared values, RMSE, and the moderate amount of time they take to run. Additionally, the Extra Trees Regressor and XGBoost Regressor models are both tree-based models, which are generally considered to be more interpretable than Gaussian Process Regressors. Tree-based models can provide clear insights into feature importance and decision rules, which was key to this project's goal of examining the SHAP values of each model's features (Molnar, 2023).

3.5 Hyperparameter Optimization

Once models were selected, the data were split into training and test sets using scikit's train_test_split function, reserving the last 20% of data for testing. This ensured that no information leakage would occur between the test and training sets when performing hyperparameter optimization. Hyperparameter optimization is an essential step in machine learning model development as it helps tune the model's performance and reduce overfitting by finding the best combination of hyperparameter values (Müller & Guido, 2016). In this study, the Extra Trees Regressor and the XGBoost Regressor were optimized using two steps of hyperparameter tuning on training data - random search and grid search - in combination with k-fold cross-validation (Figure 4). Cross-validation ensures that a variety of data are represented across folds and gives a better estimate of how model performance will generalize, leading to a more robust model (Müller & Guido, 2016).

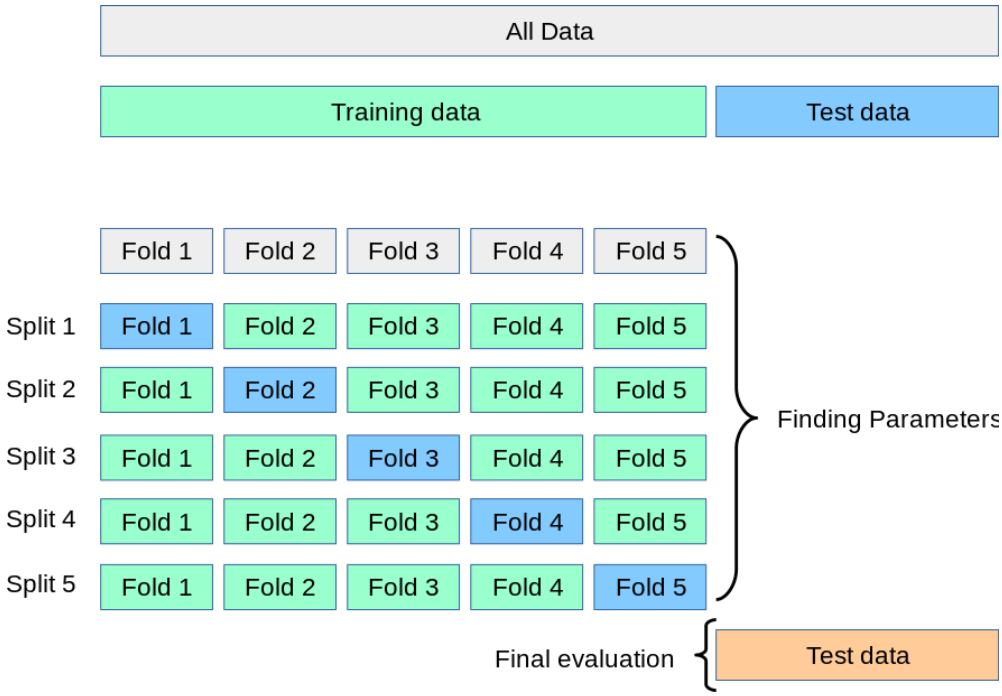


Figure 4. "Cross-validation: evaluating estimator performance" from the [scikit-learn documentation.]

In the first step of hyperparameter tuning, random search was performed on both models using the RandomizedSearchCV function from the scikit-learn model_selection module. This function samples hyperparameter values randomly based on given ranges, taking the regression model, the parameter ranges to search over, the number of iterations, and the cross-validation object as inputs. Parameter ranges were defined using a dictionary, which specified the ranges of hyperparameter values to search over. Five-fold cross-validation was performed across 100 iterations of each regression model object. The best hyperparameters and their corresponding mean cross-validated R-Squared values were returned using the best_params_ and best_score_ attributes of the RandomizedSearchCV object, respectively. The best hyperparameters were selected based on the highest mean cross-validated R-Squared value. Six rounds of random search were performed to narrow the hyperparameter ranges in preparation for grid search.

In the second step of hyperparameter tuning, a grid search was performed on both models based on the best parameter ranges from random search to further fine-tune the hyperparameters of the models. Unlike random search, grid search exhaustively searches through all possible combinations of hyperparameter values within a predefined range. Similar to the RandomizedSearchCV function, the GridSearchCV function from scikit-learn takes in the regression model object, the parameter grid to search over, and the cross-validation object as inputs. After three rounds of grid search, the final hyperparameters were selected for each model (Table 3).

Table 3. Best hyperparameters for the Extra Trees Regressor and XGBoost Regressor models.

Model	Best Hyperparameters	
Extra Trees Regressor	bootstrap = False max_depth = 18 max_features = 1.0	min_samples_leaf = 1 min_samples_split = 2 n_estimators = 900
XGBoost Regressor	colsample_bytree = 0.75 max_depth = 7 learning_rate = 0.031	min_child_weight = 4 subsample = 0.62 n_estimators = 950

4. Analysis

4.1 Model Performance (H1)

Using the parameters outlined above, scikit-learn's fit() function was used to train each model on the training data. Once both models were trained, they were saved using Joblib's dump() function to ensure reproducibility. Scikit-learn's predict() function was used on the test data for each model, and Mean Squared Error (MSE), RMSE, and R-squared scores were computed for the test set's predictions to assess how well the models performed. MSE and RMSE are both measures of the accuracy of a regression model, with lower values indicating

better performance, as they reflect smaller prediction errors between the model's predictions and the actual values of the target variable (Chicco et al., 2021). The R-squared value represents the proportion of variance in the target variable explained by the model and is arguably a more informative and truthful metric to evaluate regression analyses (Chicco et al., 2021). Actual and predicted NDVI values were plotted on line charts using the Matplotlib library.

4.2 Feature Importance and SHAP Values (H2)

To determine the importance of features in each model, feature importance scores were calculated using the feature_importances_ attribute of each model. Feature importance scores are calculated based on the model's internal computations during the training process and provide a relative measure of how much each feature contributes to the model's predictive performance (Müller & Guido, 2016). These scores were visualized in horizontal bar charts using Matplotlib.

Additionally, the SHAP library was used to explain the results of each model. SHAP values can be used to explain not only the magnitude but also the direction of the effect of each feature on a specific prediction (Lundberg & Lee, 2017). To compute SHAP values for each model, an explainer object was created using the Explainer class from the SHAP library, and the trained models were each passed as an argument. The explainer object was used to calculate the SHAP values for the test data in both models using the explainer() function with x_test as the input feature. To visualize the SHAP values, summary plots were created to display the feature importance values in descending order. Bar plots and decision plots of SHAP values were also created for each model as additional visualizations.

4.3 Adjusted V2 Model (H1 & H2)

Once analyses were performed on the models, additional research questions were formulated surrounding the time ranges of the selected drought indices. To look at the potential influence of longer-term drought indices on NDVI predictions, an additional XGBoost Regressor model (referred to as XGBoost Regressor V2) was created with adjusted features outlined in Table 4. Original features that had low feature importance scores and SHAP values, such as minimum and maximum temperature, were removed, while several indices with longer accumulation periods, up to five years, were added. All data were preprocessed using the same methodology as the data in the original models.

Table 4. Earth observations used for the XGBoost Regressor V2 model.

Data Type	Variable(s)
Drought Impact (Target)	Normalized Difference Vegetation Index (NDVI)
Drought Indices (Features)	Palmer Drought Severity Index (PDSI) Palmer Z Score Evaporative Demand Drought Index (EDDI) – 1, 5-year Standardized Precipitation Index (SPI) – 90, 180, 270-day; 1, 2, 5-year Standardized Precipitation Evapotranspiration Index (SPEI) – 270-day; 1, 2, 5-year
Environmental Variables (Features)	Snow Water Equivalent

The XGBoost Regressor V2 model underwent hyperparameter optimization in the same manner as the original, with several rounds of random search followed by grid searches to select the optimal parameters for the new model. Using the best parameters from grid search, the model was trained and saved. MSE, RMSE, and R-squared scores were computed for the model's test set's predictions, and feature importance scores were again calculated. Lastly, SHAP values were produced and visualized to explain model predictions.

5. Results

5.1 Model Performance (H1)

The Extra Trees Regressor model returned an MSE of 0.54, an RMSE of 0.74, and an R-squared of 0.48 (Table 5). These metrics suggest that the Extra Trees Regressor model performed moderately well in predicting NDVI values based on the input features. However, the XGBoost model demonstrated better performance than the Extra Trees Regressor model, with a lower MSE of 0.48 and RMSE of 0.69 (Table 5). The model's R-squared value was higher at 0.54, suggesting that the XGBoost model explained a larger proportion of the variance in the NDVI values, making it a more accurate predictor.

Table 5. Test set performance post-hyperparameter tuning.

Model	MSE	RMSE	R-Squared
Extra Trees Regressor	0.54	0.74	0.48
XGBoost Regressor	0.48	0.69	0.54

Model performance is visually represented in Figure 5, which plots the true test values and predicted values from the Extra Trees and XGBoost Regressor models. Model predictions generally tended to be directionally correct, with much of the error coming from the underestimation of large NDVI spikes around indices 75 and 150 and the overall depression of NDVI values across the predicted set. From index 200 onward in both models, the predicted data are relatively accurate. Some lead and lag effects can be seen throughout.

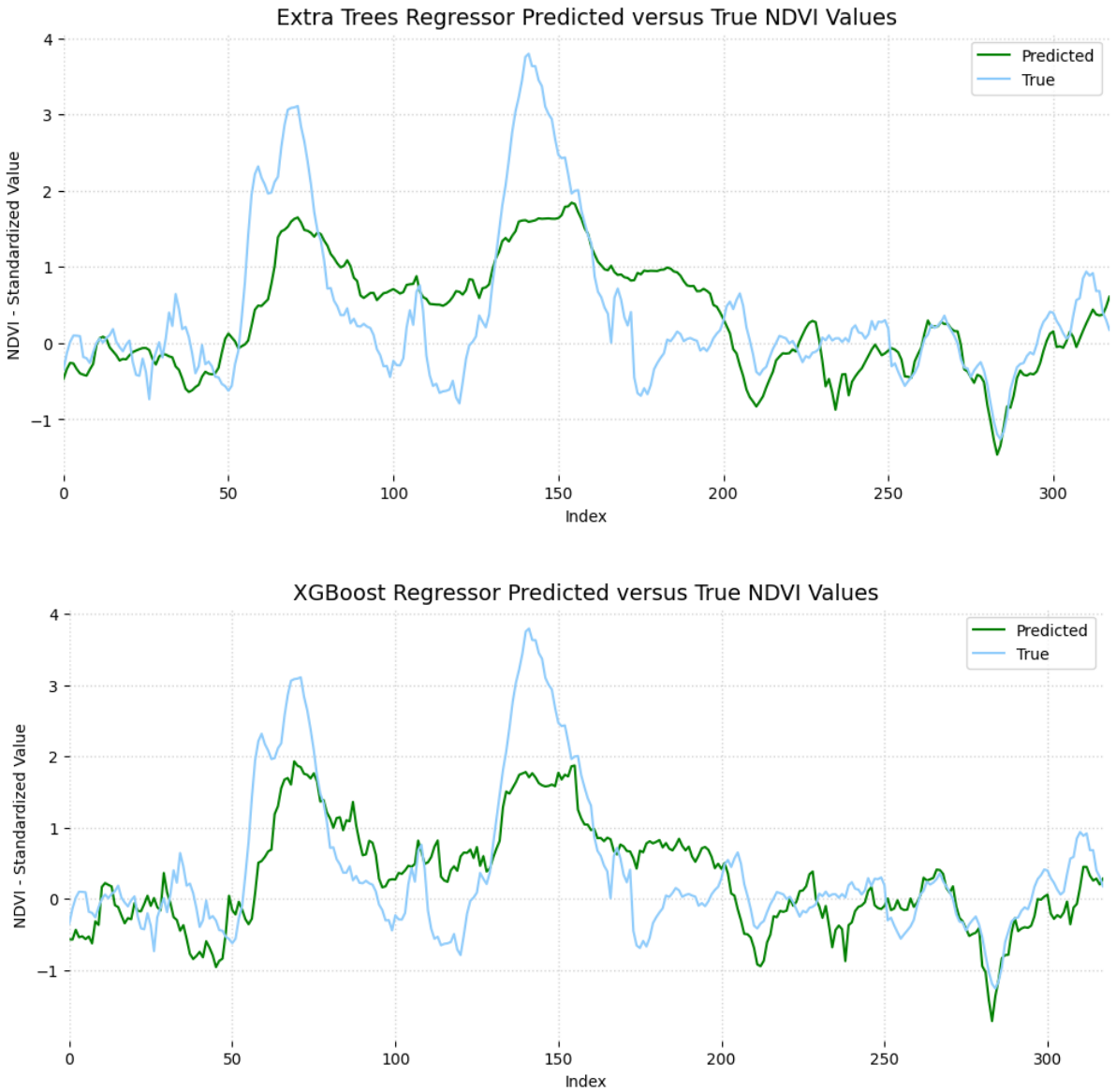


Figure 5. True and predicted NDVI values from the Extra Trees (top) and XGBoost (bottom) Regressors.

5.2 Feature Importance and SHAP Values (H2)

Based on the performance evaluation of both models, the XGBoost model was selected for further analysis.[2] The feature importance scores computed for this model were ranked

[2] Since the Extra Trees Regressor was not selected as the optimal model, its feature importance scores and SHAP summary plots have been included in the Appendix for brevity.

based on their contribution to the model's ability to predict NDVI values over time in the study area (Figure 6, left). The higher the importance score, the more influential the feature was in determining predicted NDVI values. According to the feature importance ranking scores, PDSI had the highest importance with a score of 0.26, followed by SPI 90-day with a score of 0.17, and SWE with a score of 0.11. Other important features included SPI 180-day with a score of 0.09, Palmer Z with a score of 0.09, and SPEI 180-day with a score of 0.06. Features with relatively lower importance included SPEI 90-day, EDDI 90-day, EDDI 180-day, precipitation, EDDI 30-day, SPEI 30-day, SPI 30-day, minimum temperature, and maximum temperature. These features collectively contribute to the XGBoost model's ability to predict NDVI values in the study area, with PDSI and SPI 90-day being the most influential.

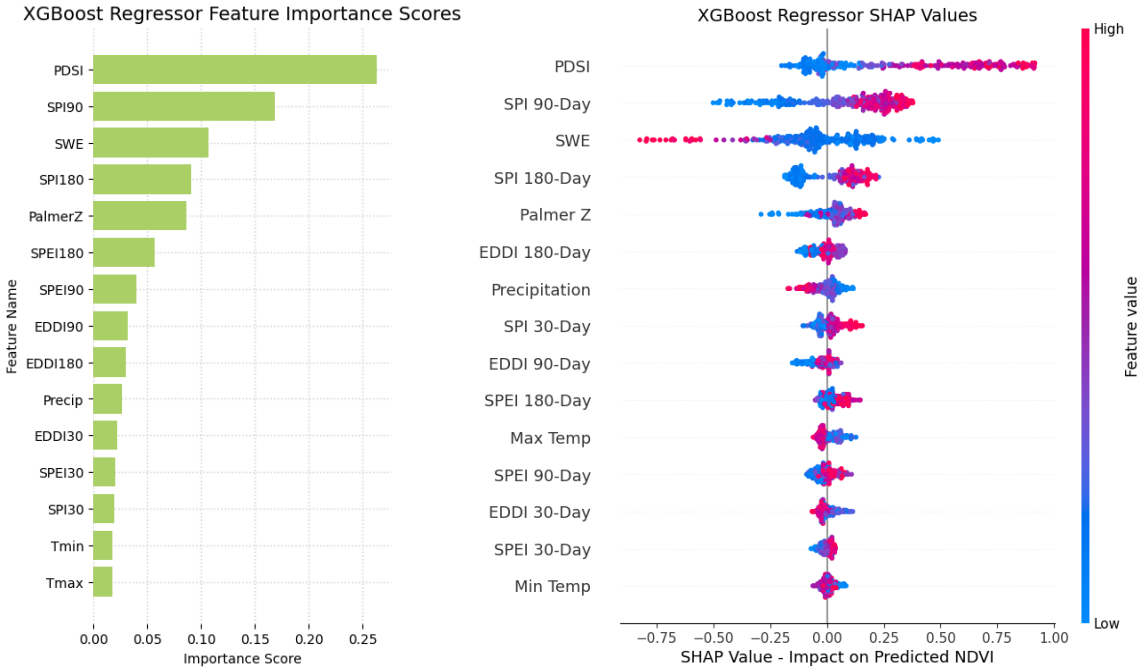


Figure 6. Feature importance scores (left) and SHAP values (right) from the XGBoost Regressor.

While feature importance scores provided a general indication of the relative importance of different features in the model, SHAP values offered a more detailed and

directional explanation of the contribution of each feature for a specific predicted NDVI value. The right plot in Figure 6 illustrates the SHAP values for each feature as individual points grouped by feature name. Shades of pink denote higher feature values while shades of blue indicate lower feature values. Negative impacts on NDVI are to the left of the vertical axis, while positive impacts are to the right. The plot shows that PDSI had the most significant overall contribution to NDVI prediction, with a direct relationship between higher PDSI values and higher SHAP values, and vice versa. The SPI 90-day index ranked second, with a relatively balanced impact on positive and negative SHAP values compared to PDSI, which had positively skewed SHAP values overall. SWE had an inverse relationship with resulting SHAP values and a wider distribution than SPI 90-day. Additionally, SPI 180-day and Palmer Z index had relatively high SHAP scores and mirrored the direct relationships observed for PDSI and SPI 90-day.

5.3 Adjusted V2 Model (H1 & H2)

After model features were adjusted to include indices with longer accumulation periods and remove unimportant features, the XGBoost Regressor V2 model resulted in an MSE of 0.59, an RMSE of 0.77, and an R-squared of 0.43, underperforming the original XGBoost Regressor model (R-squared: 0.54). Results from the feature importance ranking scores showed that SPEI 1-year had the highest importance with a score of 0.22, followed by PDSI with a score of 0.14, SPEI 2-year with a score of 0.10, SPI 90-day with a score of 0.08, SPI 5-year with a score of 0.08, and SWE with a score of 0.08 (Figure 7, left). In descending order, the rest of the features were SPI 180-day, SPEI 5-year, SPEI 270-day, SPI 270-day, Palmer Z, SPI 1-year, EDDI 1-year, EDDI 5-year, and SPI 2-year.

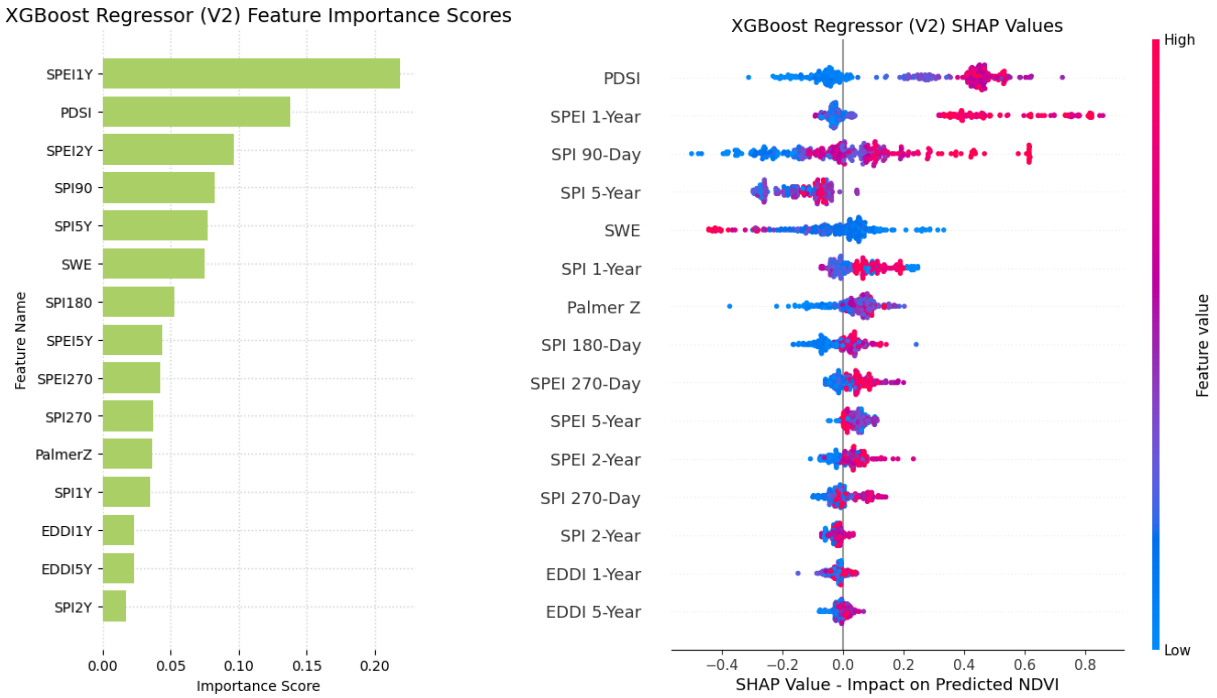


Figure 7. Feature importance scores (left) and SHAP values (right) from Version 2 of the XGBoost Regressor.

According to the SHAP values calculated from the XGBoost V2 model (Figure 7, right), PDSI remained the most significant factor in predicting NDVI, just as in the original XGBoost V1 model. Higher PDSI values consistently resulted in higher SHAP values, and vice versa, although the distribution of PDSI SHAP values in this model was only slightly positively skewed. The second most important factor was SPEI 1-year, which exhibited an overall positive skew in its SHAP values. SPI 90-day came in third place, with a relatively even distribution and a direct relationship with SHAP values. SPI 5-year was the next most important feature, but due to the distribution of its feature values across the x-axis, it was difficult to determine its directionality of impact. Finally, SWE was again among the top five features, demonstrating a similar relationship and distribution as in the first XGBoost model.

6. Discussion

6.1 Model Performance (H1)

Model performance results show that the XGBoost model was more effective than the Extra Trees Regressor for this prediction task, and support the acceptance of alternative hypothesis one (H1), as the XGBoost Regressor's R-squared value was greater than 0.50. The acceptance of H1 demonstrates that the performance of the XGBoost model was deemed sufficient for the prediction task. This conclusion is important because it establishes the credibility and reliability of the subsequent analysis using SHAP.

XGBoost has been the chosen method for several other drought and NDVI prediction studies, including Li et al.'s 2021 paper. In this study, the authors achieved an R-squared of 0.83, significantly higher than the R-squared value achieved here, using historical NDVI values along with six environmental variables as model features (Li et al., 2021). While Li et al. developed a model that better explains the variance in their NDVI data, it did not incorporate drought indices and therefore addressed a different goal than this research. In comparison, this research aimed for a sufficiently predictive model to inform the relationship of commonly used drought indices to vegetation health. This is an important distinction between past predictive work and this research, as stakeholders throughout the wider Missouri River Basin have expressed the need to develop a more comprehensive understanding of drought indicators and their specific relationship to ecological drought (NOAA/NIDIS, 2020).

In addition, it is worth discussing why XGBoost may have outperformed Extra Trees Regressor. One reason is the more advanced regularization techniques employed by XGBoost, such as L1 and L2 regularization and tree pruning, which help to prevent overfitting and improve model generalization (Chen & Guestrin, 2016). Moreover, while Extra Trees Regressor uses randomized methods to enhance model accuracy, XGBoost leverages gradient boosting

and adaptive boosting, resulting in a more sophisticated learning algorithm that improves model performance (Chen & Guestrin, 2016). However, it is important to note that the efficacy of these algorithms depends on the specific research question and data being analyzed, and XGBoost may not always outperform Extra Trees Regressor. It is therefore essential to test various models using tools like Lazy Predict to determine the best approach for a given task

6.2 Feature Importance and SHAP Values (H2)

The feature importance scores and SHAP values computed from both versions of the XGBoost model demonstrated that drought indices were among the most important features in predicting NDVI values in the Cheyenne River Basin. These findings support the acceptance of alternative hypothesis two (H2) that in a reliably predictive ML model that includes both drought indices and environmental variables features, drought indices will have higher absolute SHAP values than environmental variables, meaning that changes in the values of drought indices are more strongly linked to changes in the predicted NDVI values. Notably, for the V1 and V2 XGBoost models, PDSI scored first and second respectively in computed feature importance scores and had the highest absolute mean SHAP value in both models. This reinforces that PDSI was the most important factor in predicting NDVI across models, demonstrating a direct relationship between PDSI values and NDVI, and highlights the feasibility of using SHAP to identify relationships between drought indices and vegetation health.

A key takeaway from the SHAP results across models is the importance of temporal scale. Top features in the feature importance and SHAP analyses capture drought conditions at different temporal scales (90 days, 180 days, 9 months - 1 year), which allows the model to account for both short-term and long-term drought impacts on NDVI. Drought conditions at

longer temporal scales (e.g., PDSI) may have a cumulative effect on vegetation health, while shorter-term drought conditions (e.g., SPI 90-day) may have a more immediate impact. As PDSI incorporates precipitation and temperature data, and, by extension, water balance and potential evapotranspiration (PET), it is typically a good indicator of soil moisture at a timescale of 9-12 months (McEvoy et al., 2019; Vicente-Serrano et al., 2010). This may be one reason why SPI 90-day co-occurs with PDSI as a shorter-range precipitation-focused index that may complement the data provided by PDSI. This multi-scale information may be crucial for monitoring vegetation health over time in the study area.

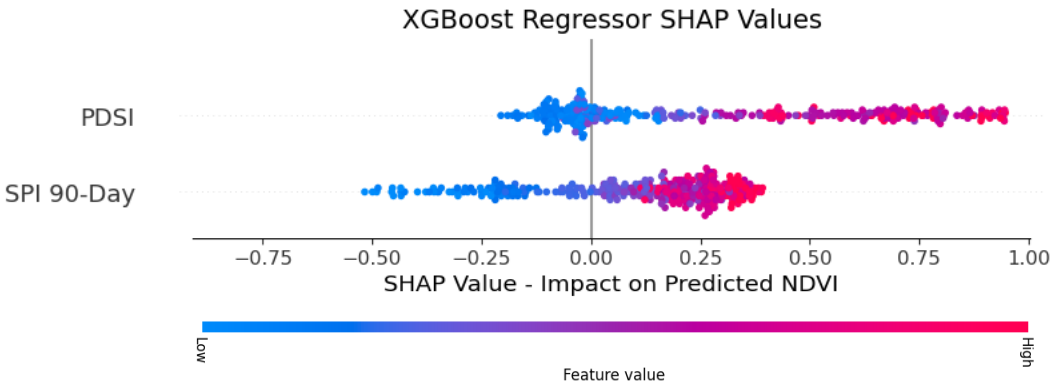


Figure 8. Distribution of SHAP values for PDSI and SPI 90-day. SPI 90-day is more evenly distributed across negative and positive SHAP values than PDSI, which skews positive.

An additional explanation for the co-occurrence of SPI 90-day and PDSI across models may be due to the distribution of the features' SHAP values (Figure 8). SPI 90-day showed a more even distribution than PDSI across negative and positive SHAP values, while PDSI skewed positively overall. This implies that PDSI was stronger at predicting high NDVI values, and weaker at predicting low NDVI (Figure 9). Due to the more even distribution of SPI 90-day across both the V1 and V2 XGBoost models, it is possible that the relative importance of this index also has to do with its effectiveness at predicting lower NDVI values, and by extension negative drought impacts on vegetation health.

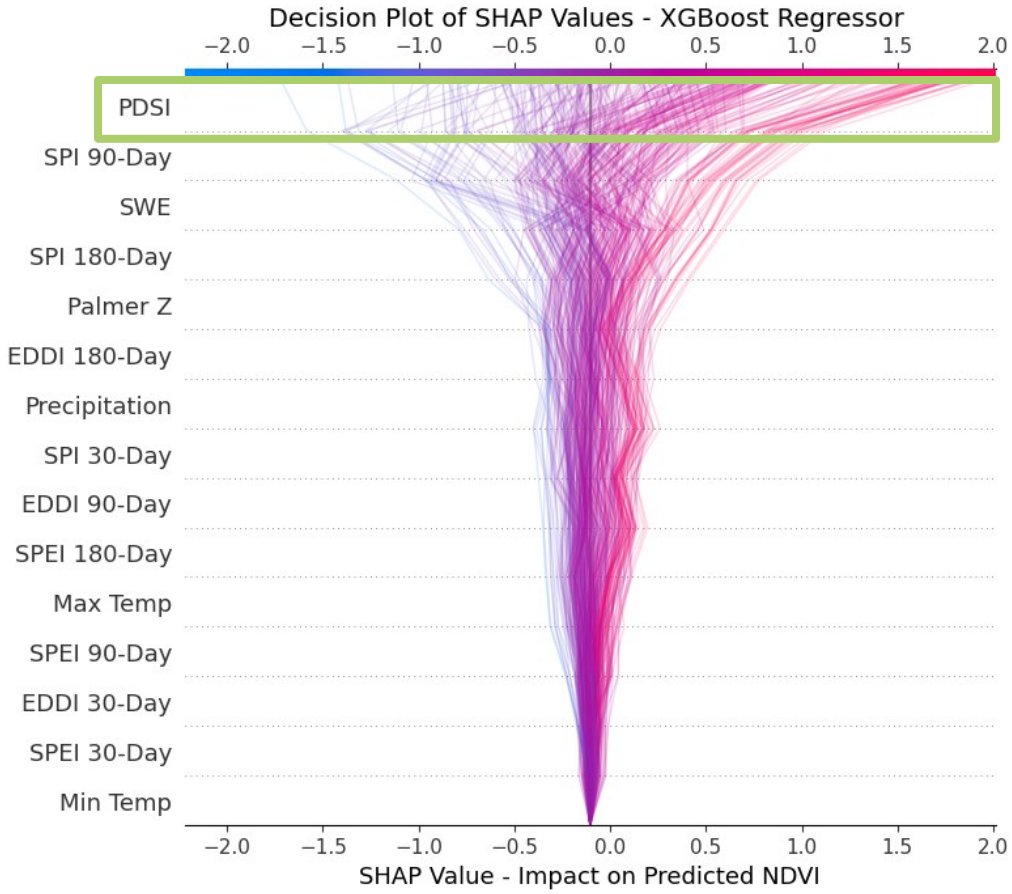


Figure 9. A decision plot of SHAP values and model features, highlighting PDSI. Within the PDSI feature row, lines skew positively towards higher SHAP values.

SPEI 1-year also co-occurred with PDSI in the V2 model. As previously described, PDSI is based on the assumption that the availability of soil moisture is a key factor in determining whether a region is experiencing drought or not, while SPEI takes into account both precipitation and evapotranspiration directly (Palmer, 1965; Vicente-Serrano et al., 2010). In this way, SPEI accounts for a different way of measuring water availability than PDSI. The differences in the underlying index assumptions may provide the model with an additional way of predicting drought impacts on vegetation health and lead to the indices' co-occurrence with high SHAP values. Fundamentally, the overall performance of PDSI across models demonstrates that it is a critical index for land managers and decision-makers to monitor when considering

the impacts of ecological drought on vegetation health in the Cheyenne River Basin, and is even more powerful when used in concert with complementary features such as shorter range indices (SPI 90-Day) or those with different underlying water availability assumptions (SWE).

In this study, SWE was the one environmental variable to consistently outperform the absolute mean SHAP values of drought indices. However, in contrast to other model features, high SWE values were associated with low NDVI values, while low SWE values were associated with both high and low NDVI values (Figure 10). This relationship between SWE and NDVI values may seem counterintuitive as water availability typically leads to increased vegetation health. In this case, it is important to note that snow accumulation in the winter months suppresses vegetation growth, resulting in lower NDVI values (Grippa et al., 2005; T. Wang et al., 2013; Y. Wang et al., 2022). However, once the snow begins to melt in the spring, water availability for vegetation has been shown to increase, leading to higher NDVI values (Matongera et al., 2021; Paudel & Andersen, 2013). Thus, SWE can have an inverse relationship with NDVI values during the winter months and a positive relationship during the following spring and summer months, creating a lag effect that serves as an indicator of future snowmelt and water availability. In years with little snowpack, this mechanism may be weaker, as less snowpack leads to worsened vegetation health in the growing months. Therefore, the underlying mechanism driving the association of NDVI and SWE may explain the distribution of SHAP values across the x-axis. Nevertheless, the high SHAP values of SWE suggest that it is a critical variable in predicting water availability in the study area.

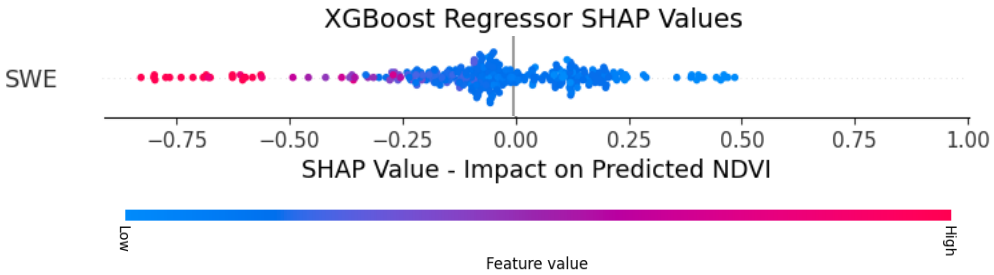


Figure 10. Distribution of SHAP values for SWE. High SWE values were associated with low NDVI values, while low SWE values were associated with both high and low NDVI values.

SHAP results suggest that vegetation health monitoring efforts in the Cheyenne River Basin should focus on using PDSI, SPI 90-day, SPEI 1-year, and SWE, which were consistently identified as important predictors of NDVI values in this study. Further, the findings surrounding the importance of the temporal scale of the data provide clues as to which drought-related climate effects may affect vegetation health at different timescales. For example, if a shorter-term precipitation index like SPI 90-day is found to be important in predicting vegetation health, this may indicate that vegetation types impacted by precipitation availability may be more susceptible to short-term droughts. If a longer-term index such as PDSI is found to be important, it may indicate that vegetation types impacted by trends of increased evapotranspiration may be susceptible to longer-term droughts. In summary, understanding the importance of different time scales of data can help reveal the mechanisms underlying drought impacts on vegetation health and can inform better decision-making in managing these impacts.

These results also support the findings of the few previous studies that have used XAI techniques in the domain of drought research. In particular, this study's successful use of SHAP aligns with findings in Dikshit and Pradhan's paper *Explainable AI in Drought Forecasting*, which suggested that SHAP is useful to understand the impact of variables within drought-related

models. The utilization of SHAP in this study not only reinforces the findings presented by Dikshit and Pradhan but also emphasizes the practicality of using SHAP to assess the influence of various variables within drought-related models. These collective findings contribute to the growing consensus on the effectiveness of XAI methodologies in enhancing the understanding of complex phenomena like drought and pave the way for further advancements in the field.

6.3 Limitations

The main limitations of this research are four-fold. First, the MODIS-derived NDVI data have limitations that could affect their accurate representation of vegetation health. For instance, factors such as cloud cover, atmospheric conditions, and solar angle can affect the accuracy and consistency of NDVI. Additionally, NDVI may not capture changes in vegetation density or structure, which can also be a resultant impact of ecological drought conditions. Second, the pentad drought indices are limited in capturing daily drought conditions, as they represent only five-day intervals and may not capture spatial variability or lagged effects of drought on vegetation health. Daily data were resampled to this temporal resolution to avoid having to interpolate a significant portion of the data for many features used by the ML model.

Next, XGBoost, Extra Trees, and ML regression algorithms in general have limitations that can impact their ability to predict targets accurately. Most notably, the quality and representativeness of the training data play a critical role in the performance of a machine learner. Even with careful preprocessing, remotely sensed data contain some amount of error and uncertainty, leading to potential loss in the model. For this research specifically, results are also limited by the moderate R-squared values of both models. Additionally, caution should be taken when interpreting and generalizing results from ML models, as they may not capture all

complex relationships between features and targets and may be influenced by region-specific factors. Lastly, while SHAP values can provide valuable insight into the importance of different features, they can be computationally expensive and may not provide accurate or meaningful insights if the model is poorly constructed or trained on biased data. SHAP values assume that input features are independent, which may not be true in all cases and may not fully capture joint interactions between features(Lundberg & Lee, 2017). Therefore, SHAP values should be used with caution and other interpretability techniques to understand the relationships between features and NDVI values comprehensively.

7. Future Work

The field of ML research related to ecological drought is currently limited, and there is a clear need for future studies to be conducted across various regions and ecosystems. This will help to determine whether the relative importance of drought indices change significantly based on the region or vegetation type, ultimately producing results at scales useful to natural resource management. When conducting these studies, a focus on improving model performance is advised, through the incorporation of additional data sources, such as soil moisture products, streamflow data, or the Vegetation Drought Response Index (VegDRI), or through the use of more sophisticated, proven, modeling techniques such as Long Short Term Memory (LSTM) models (Dikshit & Pradhan, 2021a). Additionally, while computing daily drought indices manually requires more time and effort, higher temporal resolution data would provide the machine learner with more points to train and test on. Another avenue worth exploring is using ML for spatial analysis to predict spatial patterns of NDVI based on drought indices across regions. Finally, it is clear that ensuring the interpretability of models in future

ML research related to ecological drought is key for developing actionable results. Continued exploration and integration of XAI techniques into drought research hold tremendous potential for refining models, improving accuracy, and better understanding variable importance and directionality. In summary, conducting studies across various regions and ecosystems, exploring other data sources, analyzing spatial patterns, and using XAI for interpreting ML models are each important areas of research that could contribute to better future understanding of ecological drought and provide relevant and actionable information to resource managers.

8. Conclusion

This research project stands apart from previous predictive NDVI research due to its focus on identifying the most influential drought indices and environmental variables for predicting NDVI, thereby uncovering indicators of vegetation stress in the study area. While relatively accurate NDVI prediction has been achieved in the past using machine learning methods (Li et al., 2021; Roy, 2021), the significance of this project lies in both its introduction of XAI methods into the ecological drought field and its identification of ecological drought indicators in the study area. By leveraging SHAP, this research not only provides insights into the ML model's predictions but also empowers human users to scrutinize their intuitions and validate them against the model's interpretations. The use of XAI to interpret ML predictions represents a novel and valuable approach in the domain of ecological drought. Consequently, this research contributes to both the application of interpretable predictive modeling techniques in the domain and the development of tools for drought monitoring and management.

To conclude, this study will offer a practical application of these results. As previously discussed, drought monitoring is carried out using a variety of methods across the United States, including those used by the United States Drought Monitor (USDM). For the UDSM, a variety of drought indicators are synthesized by a map author through a convergence of evidence approach, who then work with local observers (e.g., state climate offices, state agencies, National Weather Service offices, and others) across the country to provide an on-the-ground view of the data and drought impacts for "ground truthing." This method of drought monitoring can be informed by the results of this study, as they provide context to know which commonly used drought indices correlate with vegetation health impacts. This work can help those monitoring drought on a weekly basis in the region, and potentially in similar climates and ecosystems, better account for ecological drought and ultimately may result in more effective responses to these impacts.

9. References

Anyamba, A., & Tucker, C. J. (2012). Historical Perspectives on AVHRR NDVI and Vegetation Drought Monitoring. In B. D. Wardlow, M. C. Anderson, & J. P. Verdin (Eds.), *Remote Sensing of Drought* (0 ed., pp. 23–49). CRC Press. https://doi.org/10.1201/b11863-9

Balti, H., Ben Abbes, A., Mellouli, N., Farah, I. R., Sang, Y., & Lamolle, M. (2020). A review of drought monitoring with big data: Issues, methods, challenges and research directions. *Ecological Informatics*, *60*, 101136. https://doi.org/10.1016/j.ecoinf.2020.101136

Belayneh, A., & Adamowski, J. (2013). Drought forecasting using new machine learning methods. *Journal of Water and Land Development*, *18*(9), 3–12. https://doi.org/10.2478/jwld-2013-0001

Boden, D. (2023). *Nebraska Rangelands: Rangelands*. https://unl.libguides.com/c.php?g=51759&p=333380

Bradford, J. B., Schlaepfer, D. R., Lauenroth, W. K., & Palmquist, K. A. (2020). Robust ecological drought projections for drylands in the 21st century. *Global Change Biology*, *26*(7), 3906–3919. https://doi.org/10.1111/gcb.15075

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, *7*, e623. https://doi.org/10.7717/peerj-cs.623

Clarivate Web of Science. (2023). *"Ecological Drought" Analyze Results*. Clarivate Web of Science. https://www-webofscience-com.proxy1.library.jhu.edu/wos/woscc/analyze-results/c64bd0be-48a8-48ee-842a-621365883054-831d4de7

Crausbay, S. D., Betancourt, J., Bradford, J., Cartwright, J., Dennison, W. C., Dunham, J., Enquist, C. A. F., Frazier, A. G., Hall, K. R., Littell, J. S., Luce, C. H., Palmer, R., Ramirez, A. R., Rangwala, I., Thompson, L., Walsh, B. M., & Carter, S. (2020). Unfamiliar Territory: Emerging Themes for Ecological Drought Research and Management. *One Earth*, *3*(3), 337–353. https://doi.org/10.1016/j.oneear.2020.08.019

Crausbay, S. D., Ramirez, A. R., Carter, S. L., Cross, M. S., Hall, K. R., Bathke, D. J., Betancourt, J. L., Colt, S., Cravens, A. E., Dalton, M. S., Dunham, J. B., Hay, L. E., Hayes, M. J., McEvoy,

J., McNutt, C. A., Moritz, M. A., Nislow, K. H., Raheem, N., & Sanford, T. (2017). Defining Ecological Drought for the Twenty-First Century. *Bulletin of the American Meteorological Society*, *98*(12), 2543–2550. https://doi.org/10.1175/BAMS-D-16-0292.1

Culler, R. C., Hadley, R. F., & Schumm, S. A. (1961). *Hydrology of the upper Cheyenne River basin: Part A. Hydrology of stock-water reservoirs in upper Cheyenne River basin; Part B. Sediment sources and drainage-basin characteristics in upper Cheyenne River basin*. https://doi.org/10.3133/wsp1531

Dikshit, A., & Pradhan, B. (2021a). Explainable AI in drought forecasting. *Machine Learning with Applications*, *6*, 100192. https://doi.org/10.1016/j.mlwa.2021.100192

Dikshit, A., & Pradhan, B. (2021b). Interpretable and explainable AI (XAI) model for spatial drought prediction. *Science of The Total Environment*, *801*, 149797. https://doi.org/10.1016/j.scitotenv.2021.149797

Dikshit, A., Pradhan, B., & Santosh, M. (2022). Artificial neural networks in drought prediction in the 21st century–A scientometric analysis. *Applied Soft Computing*, *114*, 108080. https://doi.org/10.1016/j.asoc.2021.108080

Ehlert, K. (2022). Landscape and Watershed Setting of the Northern Great Plains of Western South Dakota. *South Dakota State University*.

Ehlert, K. (2023). *Partner State Resources for South Dakota Rangelands*. https://rangelandsgateway.org/states/south-dakota

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, *63*(1), 3–42. https://doi.org/10.1007/s10994-006-6226-1

Grippa, M., Kergoat, L., Le Toan, T., Mognard, N. M., Delbart, N., L'Hermitte, J., & Vicente-Serrano, S. M. (2005). The impact of snow depth and snowmelt on the vegetation variability over central Siberia. *Geophysical Research Letters*, *32*(21), L21412. https://doi.org/10.1029/2005GL024286

Hao, Z., Yuan, X., Xia, Y., Hao, F., & Singh, V. P. (2017). An Overview of Drought Monitoring and Prediction Systems at Regional and Global Scales. *Bulletin of the American Meteorological Society*, *98*(9), 1879–1896. https://doi.org/10.1175/BAMS-D-15-00149.1

Hobbins, M. T., Wood, A., McEvoy, D. J., Huntington, J. L., Morton, C., Anderson, M., & Hain, C. (2016). The Evaporative Demand Drought Index. Part I: Linking Drought Evolution to Variations in Evaporative Demand. *Journal of Hydrometeorology*, *17*(6), 1745–1761.

https://doi.org/10.1175/JHM-D-15-0121.1

Introduction to Wyoming Rangelands. (2023). *Wyoming Rangelands*. https://uwyoextension.org/uwrange/

Li, X., Yuan, W., & Dong, W. (2021). A Machine Learning Method for Predicting Vegetation Indices in China. *Remote Sensing*, *13*(6), Article 6. https://doi.org/10.3390/rs13061147

Liaw, A., & Wiener, M. (2002). *Classification and Regression by randomForest*. *2*.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 4768–4777.

Matongera, T. N., Mutanga, O., Sibanda, M., & Odindi, J. (2021). Estimating and Monitoring Land Surface Phenology in Rangelands: A Review of Progress and Challenges. *Remote Sensing*, *13*(11), Article 11. https://doi.org/10.3390/rs13112060

McEvoy, D. J., Hobbins, M., Brown, T. J., VanderMolen, K., Wall, T., Huntington, J. L., & Svoboda, M. (2019). Establishing Relationships between Drought Indices and Wildfire Danger Outputs: A Test Case for the California-Nevada Drought Early Warning System. *Climate*, *7*(4), Article 4. https://doi.org/10.3390/cli7040052

McKee, T. B., Doesken, N. J., & Kleist, J. (1993). *The Relationship of Drought Frequency and Duration to Time Scales*.

Mokhtar, A., Jalali, M., He, H., Al-Ansari, N., Elbeltagi, A., Alsafadi, K., Abdo, H. G., Sammen, S. Sh., Gyasi-Agyei, Y., & Rodrigo-Comino, J. (2021). Estimation of SPEI Meteorological Drought Using Machine Learning Algorithms. *IEEE Access*, *9*, 65503–65523. https://doi.org/10.1109/ACCESS.2021.3074305

Molnar, C. (2023). *Interpretable Machine Learning*. https://christophm.github.io/interpretable-ml-book/#

Müller, A. C., & Guido, S. (2016). *Introduction to Machine Learning with Python: A Guide for Data Scientists*. O'Reilly Media, Inc.

NOAA/NIDIS. (2020, December 14). *2021-2023 Missouri River Basin Drought Early Warning System Strategic Action Plan*. Drought.Gov. https://www.drought.gov/documents/2021-2023-missouri-river-basin-drought-early-warning-system-strategic-action-plan

Palmer, W. (1965). Meteorological Drought. Research Paper No. 45, 1965, 58 p. *US Department*

of Commerce Weather Bureau, 1–65.

Pandala, S. R. (2022). *Lazy Predict* (0.2.12). https://github.com/shankarpandala/lazypredict

Park, S., Im, J., Jang, E., & Rhee, J. (2016). Drought assessment and monitoring through blending of multi-sensor indices using machine learning approaches for different climate regions. *Agricultural and Forest Meteorology*, *216*, 157–169. https://doi.org/10.1016/j.agrformet.2015.10.011

Paudel, K. P., & Andersen, P. (2013). Response of rangeland vegetation to snow cover dynamics in Nepal Trans Himalaya. *Climatic Change*, *117*(1), 149–162. https://doi.org/10.1007/s10584-012-0562-x

Pettorelli, N., Vik, J. O., Mysterud, A., Gaillard, J.-M., Tucker, C. J., & Stenseth, N. Chr. (2005). Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends in Ecology & Evolution*, *20*(9), 503–510. https://doi.org/10.1016/j.tree.2005.05.011

Phillips, L. B., Hansen, A. J., & Flather, C. H. (2008). Evaluating the species energy relationship with the newest measures of ecosystem energy: NDVI versus MODIS primary production. *Remote Sensing of Environment. 112: 4381-4392.*, *112*, 4381–4392. https://doi.org/10.1016/j.rse.2008.04.012

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), Article 7743. https://doi.org/10.1038/s41586-019-0912-1

Roy, B. (2021). Optimum machine learning algorithm selection for forecasting vegetation indices: MODIS NDVI & EVI. *Remote Sensing Applications: Society and Environment*, *23*, 100582. https://doi.org/10.1016/j.rsase.2021.100582

Samek, W., Wiegand, T., & Müller, K.-R. (2017). *Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models* (arXiv:1708.08296). arXiv. https://doi.org/10.48550/arXiv.1708.08296

Shamshirband, S., Hashemi, S., Salimi, H., Samadianfard, S., Asadi, E., Shadkani, S., Kargar, K., Mosavi, A., Nabipour, N., & Chau, K.-W. (2020). Predicting Standardized Streamflow index for hydrological drought using machine learning models. *Engineering Applications of Computational Fluid Mechanics*, *14*(1), 339–350. https://doi.org/10.1080/19942060.2020.1715844

Shapley, L. S. (2016). 17. A Value for n-Person Games. In *17. A Value for n-Person Games* (pp. 307–318). Princeton University Press. https://doi.org/10.1515/9781400881970-018

Sundararajan, K., Garg, L., Srinivasan, K., Bashir, A., Kaliappan, J., Ganapathy, G., Selvaraj, S. K., & Thiruvadi, M. (2021). A Contemporary Review on Drought Modeling Using Machine Learning Approaches. *Computer Modeling in Engineering and Sciences*, *128*, 447–487. https://doi.org/10.32604/cmes.2021.015528

Svoboda, M., & Fuchs, B. A. (2016). Handbook of Drought Indicators and Indices. *World Meteorological Organization (WMO) and Global Water Partnership (GWP)*. https://library.wmo.int/doc_num.php?explnum_id=3057

Tucker, C. J., Justice, C. O., & Prince, S. D. (1986). Monitoring the grasslands of the Sahel 1984-1985. *International Journal of Remote Sensing*, *7*(11), 1571–1581. https://doi.org/10.1080/01431168608948954

United States Bureau of Reclamation. (2019). *Basin Precipitation—Monthly and Cumulative*.

US Department of Commerce, & NOAA's Weather Service. (n.d.). *Snow Water Equivalent and Depth Information*. NOAA's National Weather Service. Retrieved April 13, 2023, from https://www.weather.gov/marfc/Snow

U.S. Drought Monitor. (2022). *What is the USDM*. U.S. Drought Monitor. https://droughtmonitor.unl.edu/About/WhatistheUSDM.aspx

Vicente-Serrano, S. M., Beguería, S., & López-Moreno, J. I. (2010). A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index. *Journal of Climate*, *23*(7), 1696–1718. https://doi.org/10.1175/2009JCLI2909.1

Wang, T., Peng, S., Lin, X., & Chang, J. (2013). Declining snow cover may affect spring phenological trend on the Tibetan Plateau. *Proceedings of the National Academy of Sciences*, *110*(31), E2854–E2855. https://doi.org/10.1073/pnas.1306157110

Wang, Y., Chen, Y., Li, P., Zhan, Y., Zou, R., Yuan, B., & Zhou, X. (2022). Effect of Snow Cover on Detecting Spring Phenology from Satellite-Derived Vegetation Indices in Alpine Grasslands. *Remote Sensing*, *14*(22), Article 22. https://doi.org/10.3390/rs14225725

Wiens, J., & Bachelet, D. (2010). Matching the Multiple Scales of Conservation with the Multiple Scales of Climate Change. *Conservation Biology : The Journal of the Society for Conservation Biology*, *24*, 51–62. https://doi.org/10.1111/j.1523-1739.2009.01409.x

10. Data & Software

Abatzoglou, J. T. (2013). Development of gridded surface meteorological data for ecological applications and modelling [Data set]. *International Journal of Climatology, 33*: 121–131. https://doi.org/10.1002/joc.3413

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System (1.7.5) [Software]. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Google Colaboratory (2.84.0) [Software]. (n.d.). https://colab.research.google.com/

Google Earth Engine API (0.1.276) [Software]. (2021). Google. https://github.com/google/earthengine-api

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy (1.22.4) [Software]. *Nature*, *585*(7825), Article 7825. https://doi.org/10.1038/s41586-020-2649-2

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment (3.7.1) [Software]. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions (0.41.0) [Software]. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 4768–4777.

McKinney, W., & others. (2010). Data structures for statistical computing in python (1.5.3) [Software]. *Proceedings of the 9th Python in Science Conference* (Vol. 445, pp. 51–56).

OpenAI. (2023). *ChatGPT* (Version GPT-3.5) [Software]. Accessed multiple times between January 2023 and April 2023 for coding assistance. https://chat.openai.com/chat

Pandala, S. R. (2022). *Lazy Predict* (0.2.12) [Software]. https://github.com/shankarpandala/lazypredict

Pedregosa et al. (2011). *Scikit-learn: Machine Learning in Python* (1.2.2) [Software]. JMLR 12, pp. 2825-2830, 2011. https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html

Thornton, M.M., Shrestha, R., Wei, Y., Thornton, P.E., Kao, S-C., & Wilson, B.E. (2022). *Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 4 R1* [Data set]. ORNL DAAC, Oak Ridge, Tennessee, USA. https://doi.org/10.3334/ORNLDAAC/2129

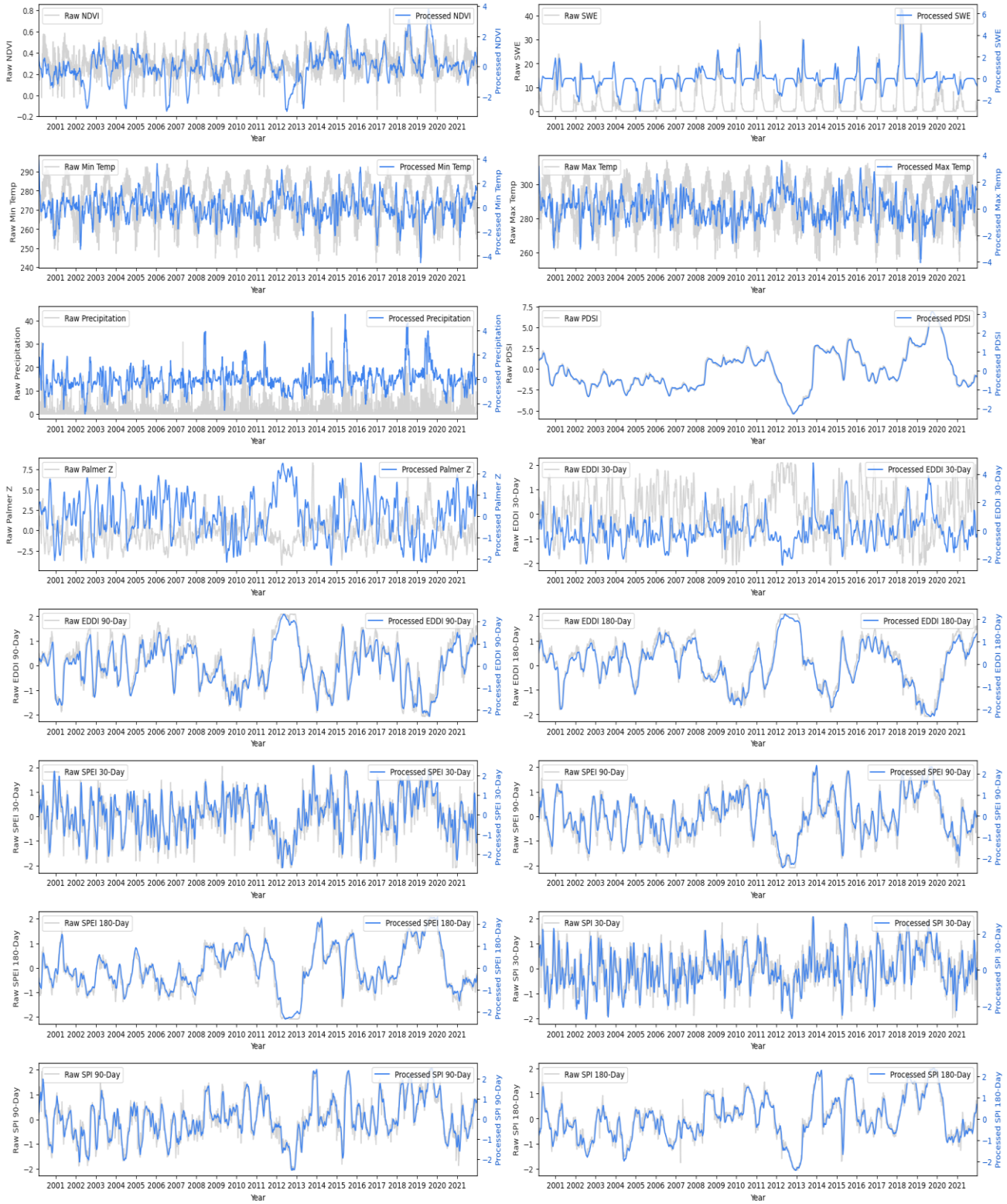
Vermote, E., Wolfe, R. (2015). *MOD09GA MODIS/Terra Surface Reflectance Daily L2G Global 1kmand 500m SIN Grid V006* [Data set]. NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/MODIS/MOD09GA.006

11. Appendices

Appendix I
Link to Github Repository: github.com/anniebritton/Ecological-Drought-ML-Modeling

Appendix II

Normalized anomaly data versus raw data displaying the data achieved through preprocessing.



Appendix III

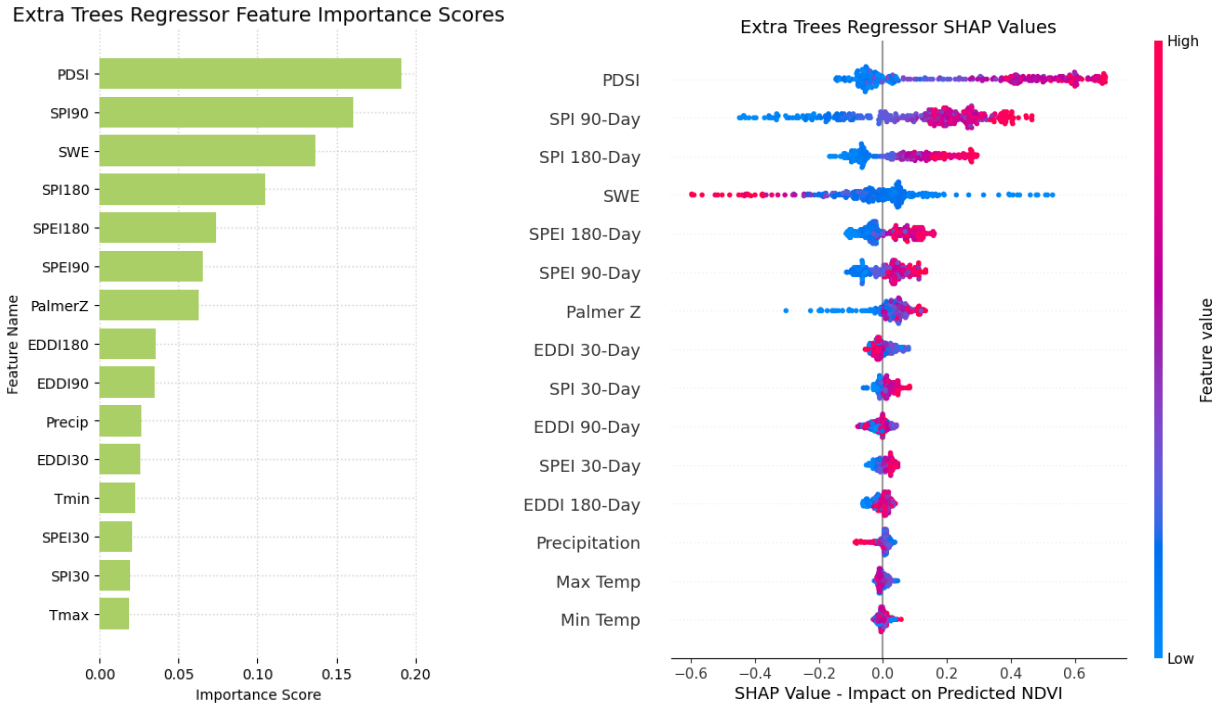
Model Performance from shuffled k-fold cross-validation of Lazy Predict.

Model	Adjusted R-Squared	R-Squared	RMSE	Time Taken
ExtraTreesRegressor	0.93	0.93	0.26	0.65
GaussianProcessRegressor	0.92	0.93	0.27	0.23
LGBMRegressor	0.89	0.90	0.32	0.20
HistGradientBoostingRegressor	0.89	0.89	0.32	0.57
XGBRegressor	0.89	0.89	0.33	0.57
RandomForestRegressor	0.88	0.88	0.34	1.67
BaggingRegressor	0.85	0.86	0.37	0.18
KNeighborsRegressor	0.81	0.82	0.42	0.04
GradientBoostingRegressor	0.78	0.79	0.45	0.78
MLPRegressor	0.77	0.78	0.46	2.93
SVR	0.77	0.78	0.47	0.17
NuSVR	0.76	0.77	0.47	0.23
DecisionTreeRegressor	0.73	0.74	0.51	0.08
ExtraTreeRegressor	0.68	0.70	0.55	0.03
AdaBoostRegressor	0.67	0.68	0.56	0.30
KernelRidge	0.48	0.51	0.70	0.16
Ridge	0.48	0.51	0.70	0.02
RidgeCV	0.48	0.51	0.70	0.02
LinearRegression	0.48	0.51	0.70	0.03
TransformedTargetRegressor	0.48	0.51	0.70	0.02
LassoLarsIC	0.48	0.51	0.70	0.04
BayesianRidge	0.48	0.50	0.70	0.02
SGDRegressor	0.48	0.50	0.70	0.02
HuberRegressor	0.48	0.50	0.70	0.10

ElasticNetCV	0.47	0.50	0.70	0.21
LassoCV	0.47	0.50	0.70	0.30
LassoLarsCV	0.47	0.49	0.71	0.10
LinearSVR	0.46	0.49	0.71	0.07
OrthogonalMatchingPursuitCV	0.45	0.48	0.72	0.04
LarsCV	0.45	0.47	0.72	0.07
TweedieRegressor	0.42	0.45	0.74	0.18
OrthogonalMatchingPursuit	0.32	0.36	0.80	0.02
ElasticNet	0.03	0.08	0.96	0.01
QuantileRegressor	-0.05	-0.00	1.00	52.36
DummyRegressor	-0.05	-0.00	1.00	0.02
Lasso	-0.05	-0.00	1.00	0.03
LassoLars	-0.05	-0.00	1.00	0.04
PassiveAggressiveRegressor	-0.06	-0.01	0.98	0.02
RANSACRegressor	-0.08	-0.03	1.00	0.30
Lars	-2.60	-2.43	1.64	0.04

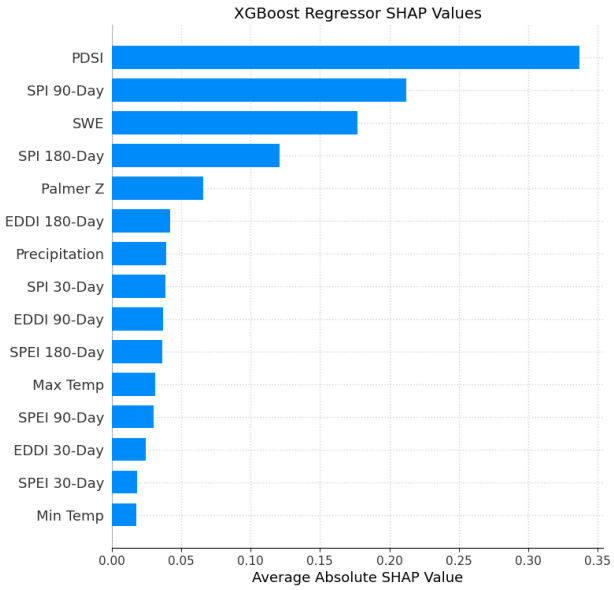
Appendix IV

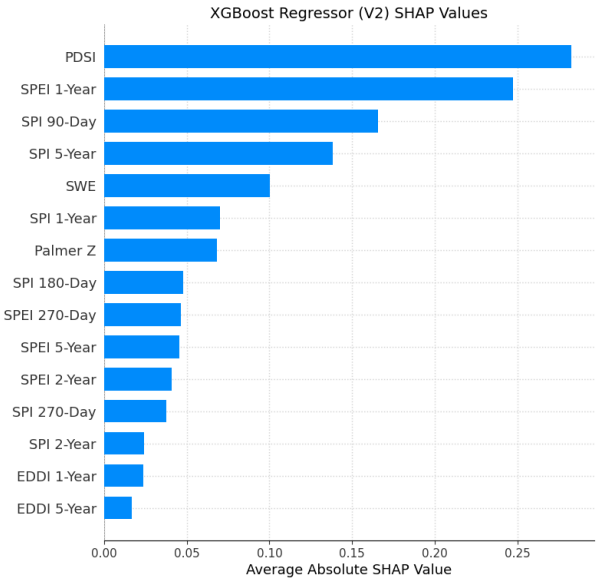
Feature importance scores (left) and SHAP values (right) from the Extra Trees Regressor.

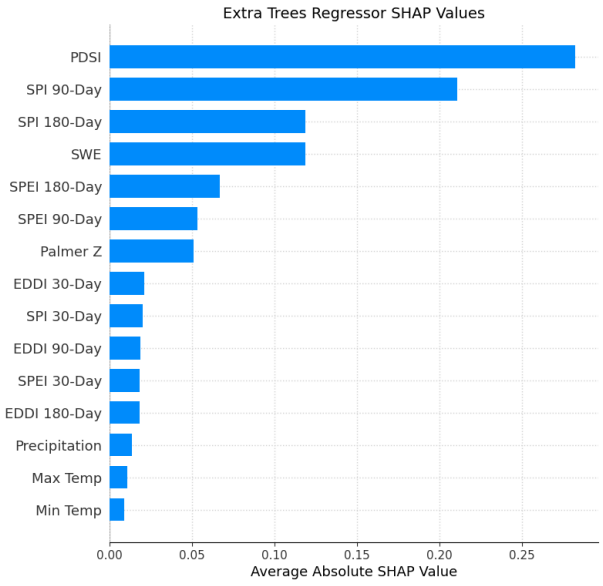


Appendix V

Average absolute SHAP values for the XGBoost Regressor, Extra Trees Regressor, and XGBoost V2 Regressor.

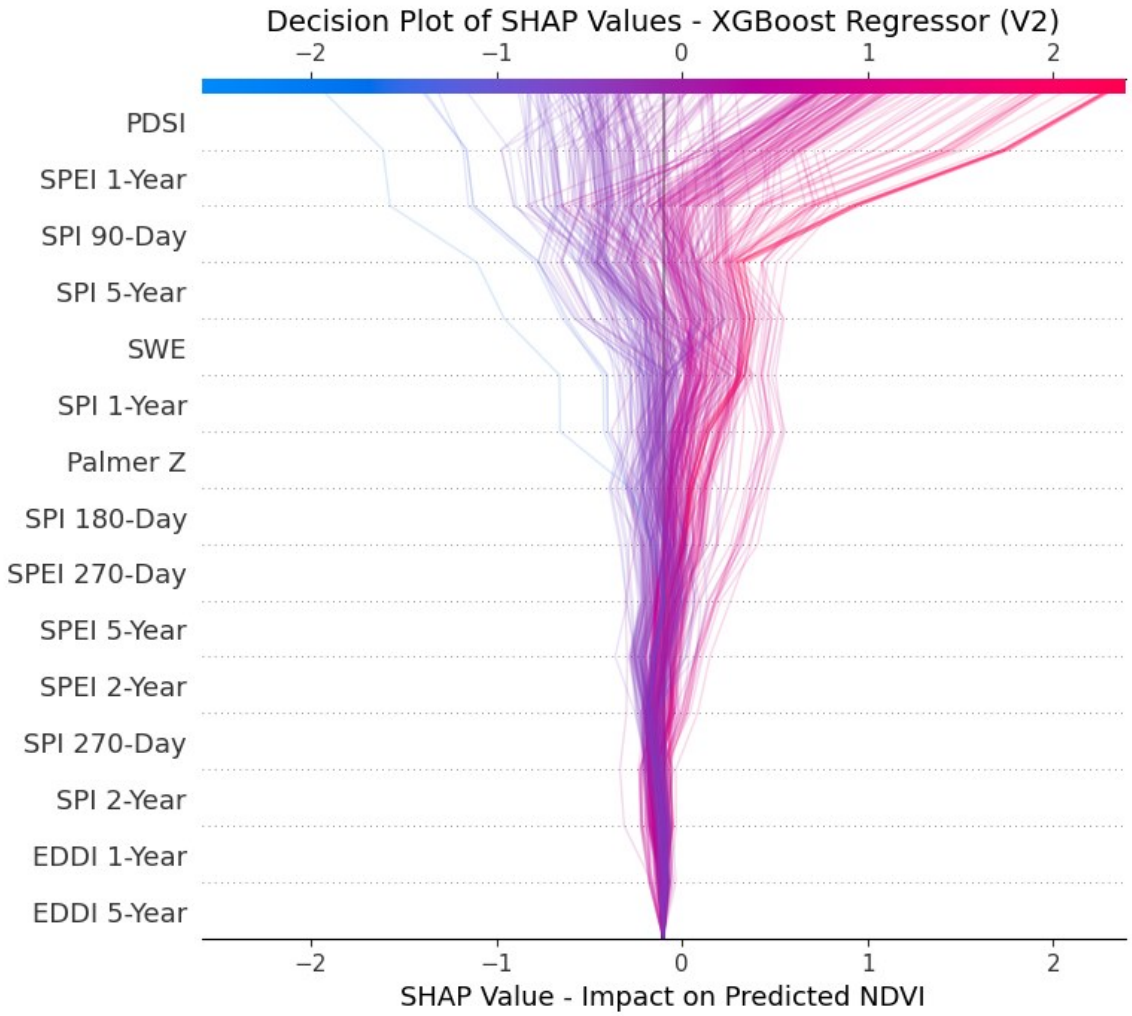


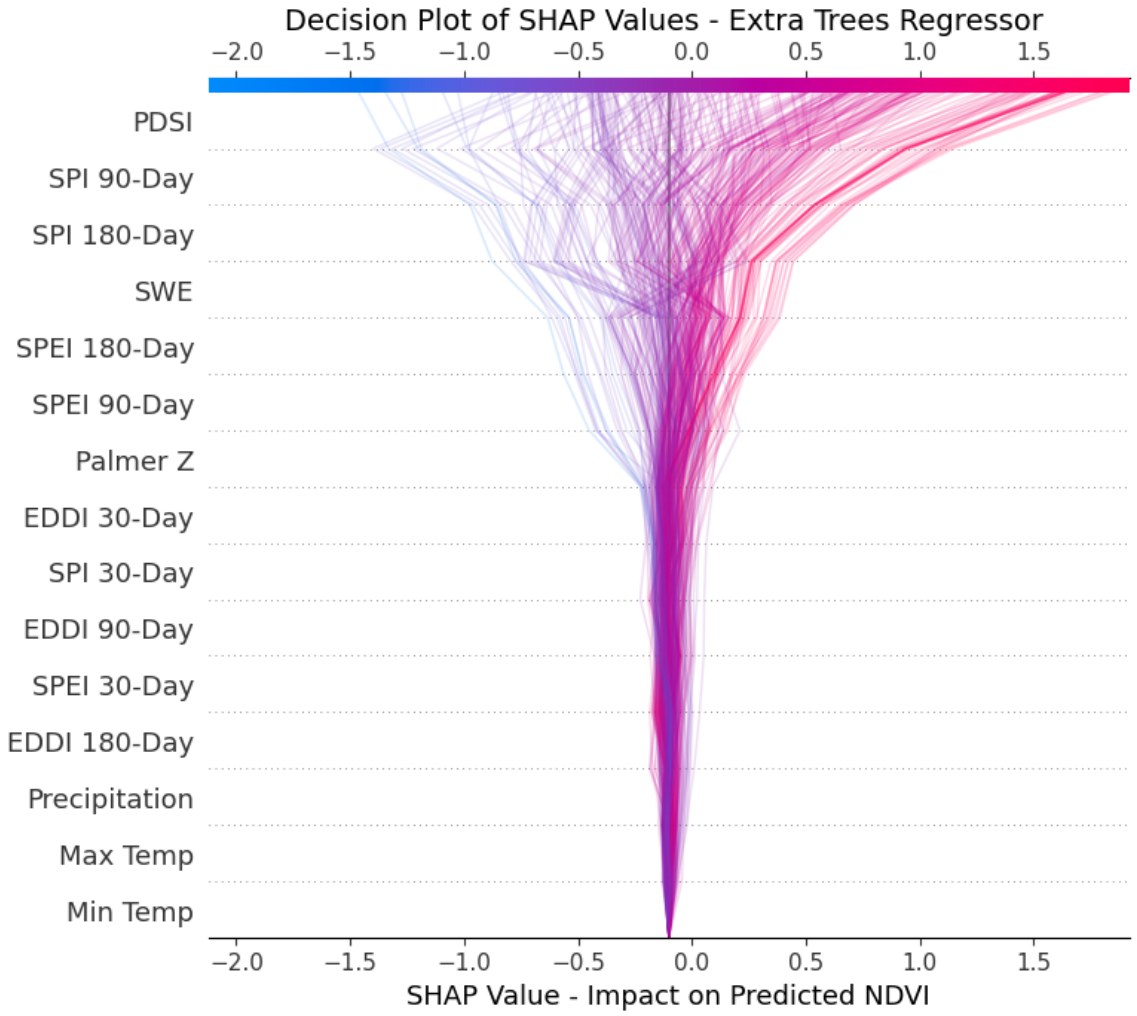




Appendix VI

Decision plots of SHAP values and model features showing the overall positive skew of predictions in the XGBoost V2 and Extra Trees models. The color of the top bar indicates whether the feature's value is high (red) or low (blue) relative to other instances in the dataset. The x-axis represents the impact of the feature on the model output, with positive values indicating that the feature increases the prediction and negative values indicating that it decreases the prediction.




Decision Plot of SHAP Values - Extra Trees Regressor
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
PDSI
SPI 90-Day
SPI 180-Day
SWE
SPEI 180-Day
SPEI 90-Day
Palmer Z
EDDI 30-Day
SPI 30-Day
EDDI 90-Day
SPEI 30-Day
EDDI 180-Day
Precipitation
Max Temp
Min Temp
SHAP Value - Impact on Predicted NDVI