A Machine Learning Framework for Predicting and Understanding the Canadian Drought Monitor

Article *in* Water Resources Research · August 2023

DOI: 10.1029/2022WR033847

4 authors, including:

Jacob Mardian
University of Guelph
4 PUBLICATIONS **19** CITATIONS
SEE PROFILE

Aaron Berg
University of Guelph
240 PUBLICATIONS **8,737** CITATIONS
SEE PROFILE

AGU ADVANCING EARTH AND SPACE SCIENCES

Water Resources Research

RESEARCH ARTICLE
10.1029/2022WR033847

A Machine Learning Framework for Predicting and Understanding the Canadian Drought Monitor

Jacob Mardian[1,2], Catherine Champagne[2], Barrie Bonsal[3], and Aaron Berg[1]

[1]Department of Geography, Environment and Geomatics, University of Guelph, Guelph, ON, Canada, [2]AgroClimate, Geomatics and Earth Observation Division, Science and Technology Branch, Agriculture and Agri-Food Canada, Ottawa, ON, Canada, [3]Watershed Hydrology and Ecology Research Division, Environment and Climate Change Canada, Saskatoon, SK, Canada

Key Points:

- A machine learning model accurately predicted the Canadian Drought Monitor severity ratings
- Shapley Additive Explanation values are used to interpret the model and provide insight into drought dynamics in time and space
- This framework enables understanding of drought processes in near real-time

Correspondence to:

J. Mardian,
jmardian@uoguelph.ca

Citation:

Mardian, J., Champagne, C., Bonsal, B., & Berg, A. (2023). A machine learning framework for predicting and understanding the Canadian drought monitor. *Water Resources Research*, *59*, e2022WR033847. https://doi.org/10.1029/2022WR033847

Received 11 OCT 2022
Accepted 15 JUL 2023

Author Contributions:

Conceptualization: Jacob Mardian, Catherine Champagne, Aaron Berg
Data curation: Jacob Mardian, Catherine Champagne
Formal analysis: Jacob Mardian
Funding acquisition: Catherine Champagne, Aaron Berg
Investigation: Jacob Mardian
Methodology: Jacob Mardian
Project Administration: Catherine Champagne, Aaron Berg
Resources: Jacob Mardian, Catherine Champagne, Aaron Berg
Software: Jacob Mardian
Supervision: Catherine Champagne, Barrie Bonsal, Aaron Berg
Visualization: Jacob Mardian
Writing – original draft: Jacob Mardian
Writing – review & editing: Catherine Champagne, Barrie Bonsal, Aaron Berg

Abstract Drought is a costly natural disaster that impacts economies and ecosystems worldwide, so monitoring drought and communicating its impacts to individuals, communities, industry, and governments is important for mitigation, adaptation, and decision-making. This research describes a novel machine learning framework to predict and understand the Canadian Drought Monitor (CDM). This fully automated approach is trained on nearly two decades of expert analysis and would assist the comprehensive monitoring of drought impacts without the continued requirement of ground support, a benefit in many data-limited areas across the country. The framework also integrates the Shapley Additive Explanation (SHAP) variable importance metric to provide insight into drought dynamics in near real-time, demonstrating its usefulness for understanding the value of different data sets for drought assessments and dispelling the commonly held misconception that machine learning models are not useful for inference. The results demonstrate that the model can effectively predict the CDM maps and realistically capture the evolution of drought events over time. A SHAP analysis found that the Prairie drought of 2015 was related to a strong El Niño event that reduced water supply to a region already facing long-term water deficits, and the subsequent reduction in groundwater availability was detected by the Gravity Recovery and Climate Experiment satellite. Overall, this research shows strong potential to streamline the CDM methodology, integrate scientific insight into operations in near real-time using SHAP values, and provide an avenue to retrospectively extend the CDM for evaluating current and future drought events in a historical context.

1. Introduction

In recent years, a series of severe droughts have impacted North America, including the Great Plains drought of 2012 (Hoerling et al., 2014), the western Canada drought of 2015 (Szeto et al., 2016), and the extreme drought of 2021 that affected most of western North America. During these periods, rainfed crops and livestock are receiving inadequate water, soils are parched, groundwater levels are dangerously low, and irrigated regions may receive insufficient water allocations from municipal water agencies (Gutzmer, 2021). These water shortages have implications for water security, economic growth, and global food supply as North America produces a substantial proportion of the world's grain and livestock (Boyer et al., 2013). Compound extremes of drought and high temperatures may trigger extended heat waves and forest fires with devastating impacts on individuals and communities. For instance, dry conditions paired with record high temperatures in Canada fostered the conditions for uncontrolled wildfires from the interior of British Columbia to Northern Ontario in the summer of 2021 (NRCan, 2021).

Consequently, drought monitoring and early warning systems are necessary for mitigation, adaptation, and decision-making. Drought monitoring systems often utilize satellite and gridded climate data due to their continuous spatial coverage and frequent data acquisition that enables timely updates to drought status over large areas. Standardized drought indicators are frequently used to measure moisture anomalies (Liu et al., 2017; McKee et al., 1993; Vicente-Serrano et al., 2010), which may be derived from ground-based (Liu et al., 2015), satellite (Keikhosravi-Kiany et al., 2022), or model-driven weather observations (Spinoni et al., 2014). Another method uses satellite-based vegetation parameters to monitor plant response to drought conditions (Klisch & Atzberger, 2016; Liu et al., 2021), sometimes in combination with satellite-derived land surface temperatures (Karnieli et al., 2010; Son et al., 2012). Other effective earth observation techniques use the Gravity Recovery and Climate Experiment (GRACE) satellites to estimate terrestrial water storage anomalies (Houborg et al., 2012; K.

S. Kumar et al., 2021; Thomas et al., 2014; Tian et al., 2021) or satellites operating in the microwave spectrum to estimate soil moisture anomalies (Champagne et al., 2018; Narasimhan & Srinivasan, 2005; Oozeer et al., 2020; Tao et al., 2021).

A limitation of most existing drought indices is that they fail to represent multiple types of drought and their impacts, and instead tend to represent one or two types of droughts. For instance, meteorological droughts are precipitation deficits, agricultural droughts are soil moisture deficiencies that harm vegetation growth, and hydrological droughts are surface and subsurface water shortages that fail to meet the demand of humans and ecosystems. While these indices are useful for specific applications, they fail to provide a comprehensive evaluation of drought conditions and impacts.

The United States Drought Monitor (USDM) was developed in 2000 to provide a more comprehensive estimation of drought extent and severity for the USA and its territories (Svoboda et al., 2002). It blends a suite of in-situ, model-based, and satellite indicators along with expert interpretation to achieve a convergence of evidence on drought. It measures various types of drought, including meteorological, agricultural, and hydrological drought, their impacts (e.g., water resources, agriculture, forestry), and whether the impacts are short-term or long-term. Areas are classified as no drought (ND), abnormally dry (D0), moderate drought (D1), severe drought (D2), extreme drought (D3) or exceptional drought (D4). The Canadian Drought Monitor (CDM) was created in 2002 to monitor drought extent, duration and severity across Canada using a similar methodology (Lawrimore et al., 2002). This tool is useful for providing stakeholders such as farmers and water resource managers with the necessary information to enact drought adaptation and mitigation measures. However, many regions of Canada lack a ground support network required to adequately and consistently provide expert input into the CDM and could benefit from a fully automated approach similar to the methodology used for the European Drought Observatory (Cammalleri et al., 2021; Sepulcre-Canto et al., 2012). The development of an exclusively data-driven and automated approach trained on nearly two decades of CDM maps would enable comprehensive monitoring of drought impacts in many areas across the country lacking ground support. It would also provide a reproducible estimate of drought consistent with drought monitor categories that could be used to analyze historical droughts. Finally, the current CDM methodology is dependent on expert interpretation of drought, which can be subject to change as the climate evolves and impede our ability to accurately detect future shifts in drought frequency and severity. Operationalizing a data-driven approach trained on the historical database of CDM assessments ensures consistent definition of drought monitor categories over time. This advancement provides a reliable foundation for studying changes to drought dynamics under a non-stationary climate.

Machine learning is a valuable tool for this purpose due to three main reasons. First, it can learn the input-output relationship from training data, eliminating the need for prior knowledge of the system. In contrast, traditional statistical approaches rely on predefined models and assumptions. Second, it can effectively utilize big data to improve predictive accuracy, a necessary characteristic for drought monitoring given the multitude of variables to consider over large areas. Finally, it excels in capturing complex and non-linear relationships, a critical aspect in understanding the intricacies of drought phenomena. The utility of machine learning for drought monitoring was demonstrated by the Vegetation Drought Response Index (VegDRI), which uses decision trees to blend climate, satellite, and biophysical data into an agricultural drought monitoring index (Brown et al., 2008; Tadesse et al., 2017). Similarly, Han et al. (2021) established the Combined Drought Monitoring Index (CDMI) by blending satellite-based estimations of precipitation, vegetation health, and soil moisture using random forests.

A similar approach could be used to train a machine learning model on the CDM drought severity categories. The drought monitor is a widely trusted source of drought information in North America and is frequently used as a reference map in other studies (B. Li et al., 2019; Xu et al., 2020), but its multifaceted definition of drought and human input hinders interpretability. As a result, this is a promising approach for comprehensive monitoring of various types of drought and their impacts in an objective, reproducible, and transparent way.

A major criticism of machine learning is its relative lack of interpretability. However, recent advances in Explainable AI (XAI) techniques provide opportunities for not only making accurate predictions but understanding why these predictions are made. Global interpretability methods such as the Mean Decrease in Accuracy and Mean Decrease in Gini methods have long been used to assess the average importance of variables to model predictions but fail to provide observation level importance metrics for more granular interpretations. Recently, local interpretability methods have been developed to overcome this limitation, and may be applied to a variety of model architectures such as neural networks and random forests (Ribeiro et al., 2016a). For example, the Local

Interpretable Model-agnostic Explanation (LIME) method has been widely used, but exhibits high instability as interpretations vary substantially when the method is repeated (Ribeiro et al., 2016b). The Shapley Additive Explanation (SHAP) value approach has recently been proposed as a solution, using a strong theoretical foundation based in game theory to expand the LIME framework with greater mathematical accuracy and consistency (Lundberg & Lee, 2017; Molnar, 2022).

The use of SHAP values as a local variable importance metric in an automated CDM framework enables analysis of drought dynamics to understand the drivers of drought. Its use on an operational basis would allow practitioners to gain deeper insight into drought events as they unfold, integrating science into the operational model. Additionally, SHAP values can be used to weigh the reliability of predictions based on the most important variables and their values, a process that could be incorporated into the decision-making process. For example, if the model is predicting a severe drought because of long-term and sustained groundwater and soil moisture deficits, a drought practitioner may consider this to be strong evidence, while a severe drought prediction driven by slightly below average precipitation the previous month may be less reliable and likely does not constitute a severe drought.

Extreme Gradient Boosting (XGBoost) is an ideal machine learning model for this operational monitoring framework for three main reasons. First, it is scalable, efficient, and accurate, making it easy to implement (Chen & Guestrin, 2016). Second, XGBoost has only a few hyperparameters that require tuning and model performance is relatively insensitive to the choice of hyperparameters (Barnwal et al., 2022; Gao & Ding, 2020). This reduces the maintenance required by practitioners to update model hyperparameters over time. Finally, XGBoost integrates seamlessly with Tree SHAP allowing for calculation of SHAP values with significant speed advantages over the traditional SHAP computation approaches required for most models (Lundberg et al., 2018; K. Yang, 2021). These computational benefits are useful for understanding the key determinants of drought in real-time monitoring applications.

As a result, the objective of this study is twofold: (a) to build a high-resolution machine learning model with the ability to predict drought extent and severity in the Canadian Prairies from 2010 to 2019; and (b) to demonstrate the suitability of SHAP values for understanding the processes driving the CDM categories. These objectives will be achieved through the development of an XGBoost machine learning model that predicts historical CDM maps using satellite, climate, and ancillary data, and the use of SHAP values to understand the key determinants of the severe drought using the summer of 2015 as a case study.

2. Data and Methodology

2.1. Study Area

The study area and study period for this paper covers the agricultural region of the Canadian Prairies from January 2010 to December 2019. This region spans the provinces of Alberta, Saskatchewan, and Manitoba, and comprises a wide range of land covers, climates, soil types, and topographical landscapes (Figure 1). For example, while the area is dominated by agricultural land including grains, pastures, and vegetables (49.5%), it includes a substantial proportion of shrublands and natural grasslands (20.9%), forests (18.6%), wetlands (7.7%), urban areas (2.4%), and barren lands (0.9%). It is also the most drought-prone region of Canada, due to its semi-arid climate and its economic dependence on rainfed agriculture that is vulnerable to climate variability. This makes it an ideal location for developing and testing a drought monitoring system.

2.2. Data

A variety of data pertinent to drought in the prairies were used to predict the CDM drought categories (described below). This includes location-based predictors to capture spatial variations in drought, as well as variables that capture serial autocorrelation because drought is an inherently persistent phenomenon. Several drought indicators were used to capture moisture anomalies over different time periods as these are some of the most valuable ways to measure drought. Teleconnection indices were also used as model inputs because drought is known to be related to large-scale atmospheric circulation and sea surface temperature conditions. Finally, several satellite-based data sets related to soil moisture, vegetation, evapotranspiration (ET), and terrestrial water storage were included as predictors to measure the land surface response to drought conditions.

2.2.1. Canadian Drought Monitor

The monthly CDM categories were used as the response variable in this study. The CDM measures drought on an ordinal scale from no drought (ND) to abnormally dry (D0), moderate drought (D1), severe drought (D2),

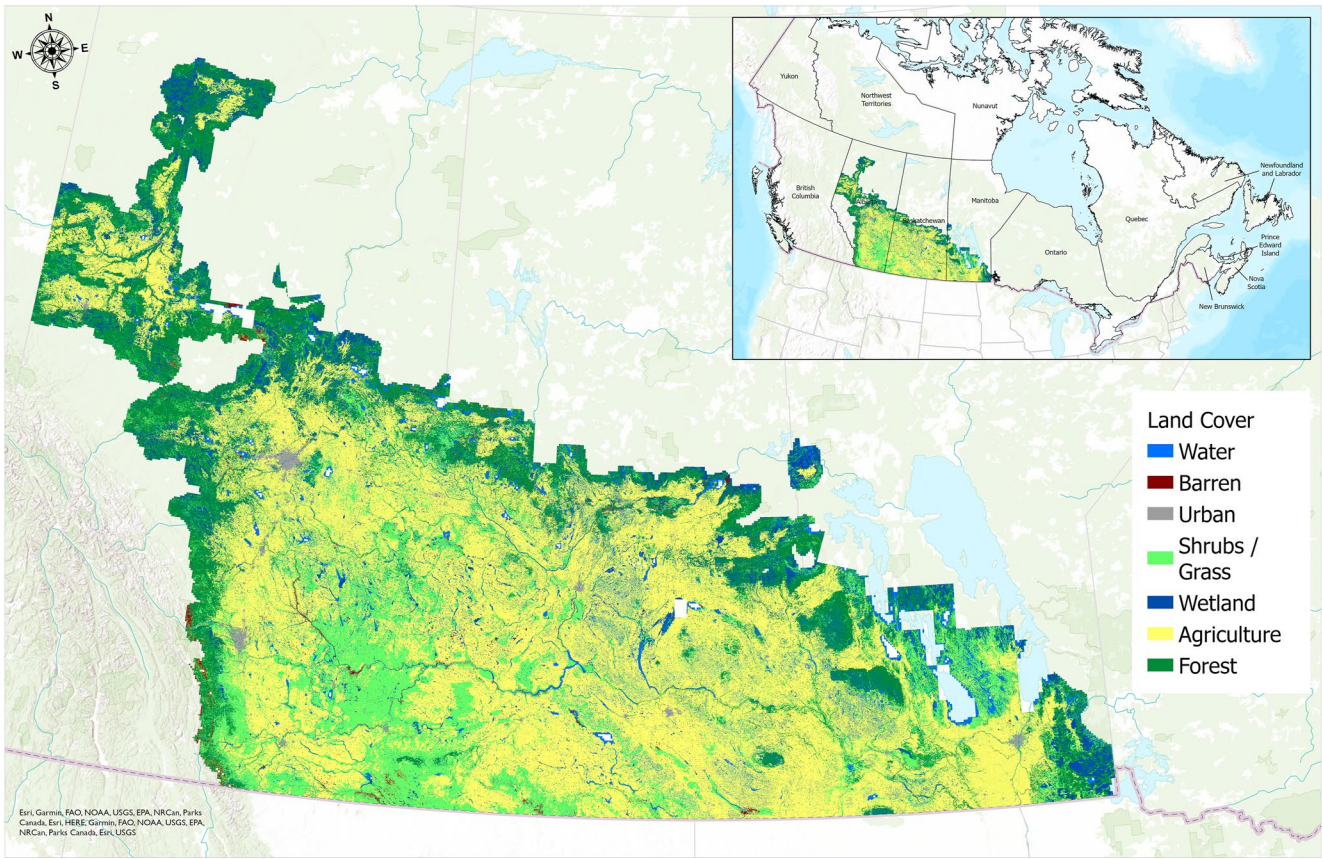

Figure 1. Canadian Prairies study area and its diverse land cover (AAFC, 2015).

extreme drought (D3), and exceptional drought (D4). These drought classes correspond to probability percentiles of drought occurrence, similar to the approach used by the USDM (Svoboda et al., 2002). For example, the D4 category represents a 2% or lower chance of drought occurrence. In this study, the D3 and D4 drought classes (extreme and exceptional droughts) were merged into one class (from now on D3/D4) as the low sample size of exceptional droughts presented issues for model training.

2.2.2. Temporal and Spatial Predictors

Ecoregions were used to define areas with similar land and climate characteristics using the Ecological Land Classification of Canada (Wilken, 1986). For each grid cell, the proportion of each drought category that occurred in its Ecoregion the previous month was included as a predictor. For example, a grid cell in the Mixed Grasslands ecoregion will include the proportion of ND, D0, D1, D2, and D3/D4 drought that occurred in the Mixed Grassland ecoregion the previous month (from now on NDProp, D0Prop etc.). This provides critical information about serial autocorrelation that is important due to the persistence of drought. It was structured by ecoregion to group together areas with similar flora, climate, hydrology, and soils, with the assumption that drought severity and impacts are related to these environmental characteristics. These variables were derived from the previous month's CDM maps and included as temporal predictors in the model.

Spatial predictors include the mean latitude (Lat) and longitude (Long) of each grid cell, allowing the model to vary its representation of drought processes across space. These are important predictors of drought given the spatial dependence of weather and climate phenomena.

2.2.3. Drought Indicators

Drought indicators are moisture anomalies derived from climate data. The Standardized Precipitation Evapotranspiration Index (SPEI) is a standardized drought indicator that estimates moisture anomalies by measuring precipitation against potential ET compared to normal conditions at a particular time of year and location

(Vicente-Serrano et al., 2010). SPEI was obtained on time scales of 1, 3, 6, 9, 12, and 24 months (SPEI1, SPEI3, etc.), calculated from Environment and Climate Change Canada's Meteorological Service weather stations and interpolated over the agricultural extent of Canada where station density is high. The Palmer Drought Severity Index (PDSI) is another drought indicator that measures relative aridity by calculating water balance terms for a two-layer soil model (Palmer, 1965). We apply a modified version of the PDSI, called the Palmer Drought Index (PDI), that was developed for operational drought monitoring in the Canadian Prairies with a modified water balance estimation (Akinremi et al., 1996; Chipanshi et al., 2013).

2.2.4. Teleconnections

Several monthly teleconnection indices that have shown to influence prairie climate (Bonsal & Shabbar, 2011), with a lag effect in many cases (Bonsal & Shabbar, 2008; Shabbar et al., 2011). As a result, each teleconnection index was included as a predictor for the current month and with a 3 and 6 months lag.

The Multivariate El Niño/Southern Oscillation (ENSO) Index version 2 (MEI) is used to represent the ENSO using a combination of sea level pressure, SSTs, zonal winds, meridional winds, and outgoing longwave radiation in the tropical Pacific. The Arctic Oscillation (AO) index describes the strength of circumpolar winds in the northern hemisphere. The North Atlantic Oscillation index represents the strength of the pressure gradient from the Icelandic Low to the Azores High in the North Atlantic and is highly correlated to the AO. The Atlantic Multidecadal Oscillation index describes a low frequency pattern in North Atlantic SSTs. The Pacific Decadal Oscillation index describes a low frequency pattern of SSTs and sea level pressure in the North Pacific. Finally, the Pacific North America index describes mid-tropospheric atmospheric circulation over the North Pacific Ocean and North America. All teleconnection values were retrieved from NOAA's Physical Sciences Laboratory (NOAA, 2023).

2.2.5. Satellite Indicators

The European Space Agency Climate Change Initiative satellite soil moisture data set (version 4.5) was obtained over the Canadian Prairies. This data set includes a blend of active and passive soil moisture retrievals from the Special Sensor Microwave Imager (SSM/I), the Advanced Microwave Scanning Radiometer (AMSR2 and AMSR-E), Windsat, the Metop Advanced Scatterometer (ASCAT), and the Soil Moisture and Ocean Salinity mission rescaled to the NASA Global Land Data Assimilation System for temporal consistency (Dorigo et al., 2017; Gruber et al., 2019). For this study, moving averages were calculated on 1-, 2-, 4-, and 8-month time scales (SM1, SM2 etc.). However, soil moisture data are not valid over Canada due to frozen soil conditions from December to March, so these indicators measure the last months of valid observations. For example, a 2-month soil moisture product in April includes an average of soil moisture in April and the preceding November, with the assumption that there is a high correlation between soil moisture after winter freeze-up and before spring thaw (J. Yang & Wang, 2019). This also ensures a consistent number of months are used in the soil moisture averages.

The Evaporative Stress Index (ESI) was obtained in 4-week running averages at a 1 km resolution (Anderson et al., 2011). The ESI represents a relative index of temporal anomalies, with the actual ET estimated from the thermal and optical sensors of the Moderate Resolution Imaging Spectrometer (MODIS) and the reference ET using a Penman Monteith approach. Monthly averages were computed from weekly data available from NASA (NASA SERVIR, 2023).

The GRACE and GRACE-FO total water storage product can track changes in water storage on the land surface based on gravitational fluctuations from water mass variations over time. This product is then vertically partitioned into surface soil moisture, root zone soil moisture, and shallow groundwater, spatially downscaled to 0.125° spatial resolution and temporally downscaled to the weekly time step using the using the Catchment Land Surface Model and an ensemble Kalman filter approach (Houborg et al., 2012; S. V. Kumar et al., 2016). This study uses the weekly groundwater percentiles resampled to the 5 km grid to compute groundwater indices on time scales of 1, 3, 6, 9, and 12 months (GRACE1, GRACE3 etc.). The weekly GRACE-based shallow groundwater drought indicator is available from the University of Nebraska-Lincoln (NASA, 2023).

The Canadian Ag-Land Monitoring System produces a weekly maximum Normalized Difference Vegetation Index (NDVI) composite from MODIS (AAFC, 2023d). This product was upscaled from 250 m to the modeling grid of 5 km, retaining both the mean and standard deviation statistics.

2.3. Modeling

2.3.1. Model Overview

Gradient Boosted Decision Trees (GBDTs) were presented as a decision tree (DT) ensemble framework to improve learning (Friedman, 2001). The technique is based on gradient boosting, where DTs are sequentially built with each DT learning from the errors of the previous DT. The model improves with each iteration using gradient descent to minimize a loss function, which is often log likelihood loss for multiclassification problems. It does this using the residuals to calculate the gradient (i.e., the partial derivative of the loss function). Given a data set $D = \{(x_i, y_i)\}$ $(i = 1, \ldots N)$, the first step is to initialize the model with a constant value:

$$f_o(x) = \text{argmin}_\gamma \sum_{i=1}^{N} L(y_i, \gamma)$$

where y_i is the observed value, γ is the predicted value (a constant), and L is the loss function. For each subsequent iteration $t = 1, \ldots, T$, calculate the steepest descent direction (gradient) of the loss function:

$$-G_t(x_i) = -\left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}\right]_{f(x)=f_{t-1}(x)}$$

Next, fit the model to predict the residuals of the previous tree:

$$\gamma_t = \underset{\gamma}{\text{argmin}} \sum_{i=1}^{N} L(y_i, f_{t-1}(x_i) + \gamma h_t(x_i))$$

where $h_t(x_i)$ is the DT made on the residuals. Finally, update the model:

$$f_t(x) = f_{t-1}(x) + \gamma_t h_t(x)$$

This framework is extended in the eXtreme Gradient Boost (XGBoost) model by introducing a regularization term to the objective function to reduce overfitting (Chen & Guestrin, 2016):

$$\text{Obj} = \sum_{i=1}^{N} L(y_i, \gamma) + \sum_{k=1}^{K} \Omega(f_t)$$

where $\Omega(f_t)$ is the regularization term at the tth time step, expressed as:

$$\Omega(f_k) = \alpha H + \frac{1}{2}\lambda \sum_{j=1}^{H} w_j^2$$

where α is the leaf complexity, H is the number of leaves, λ is the penalty, and w_j is the weight of each leaf node. Therefore, XGBoost can optimize tree structures more effectively by penalizing model complexity, which in turn reduces overfitting and improves generalization.

For classification problems, the GBDT and XGBoost models use majority voting to predict class labels. This means that each DT constructed (i.e., $t = 1, \ldots, T$) classifies the observation, and the most predicted class is the model output.

2.3.2. Model Training

In this paper, an XGBoost model is used for multiclassification to predict monthly CDM drought severity categories in the Canadian Prairies. The model treats each observation independently on a pixel-by-pixel basis and ensures spatial scale coherence by applying a 19×19 spatial filter to the model output.

Before making predictions, the model requires tuning of several hyperparameters, as outlined by Chen et al. (2021). These hyperparameters determine the structure and function of the algorithm, including the number of trees, maximum tree depth, rate of learning, subsampling, variable subsetting, and regularization parameters to prevent overfitting. The hyperparameters were chosen using a cartesian grid search that spans the typical range of values for each parameter, and the model with best combination of hyperparameters from the training period

of 2005–2009 was chosen to evaluate the model from 2010 to 2019. The accuracy of each combination of hyperparameters was assessed using a time series cross-validation technique that utilizes all available historical data to predict each month (Kuhn, 2019). This technique emulates real-time drought monitoring and ensures that future data is not used to predict historical data.

Additionally, the model faces an imbalanced classification problem due to the larger sample size of the ND class compared to the other drought classes. The proportions of each drought intensity class (D0, D1, D2, D3/D4) decreases successively, with ND accounting for 66% of all samples and D3/D4 comprising just 1.6%. Consequently, the model naturally places greater importance on learning the ND category to maximize overall classification accuracy, which is counterproductive considering the objective of predicting drought. To address this issue, the proportion of observed classes was supplied to the XGBoost weights parameter. This adjustment increases the learning rate for rare classes, thereby improving their performance.

2.3.3. Model Evaluation

The model with the optimal parameters from the training phase was used to evaluate predictions from 2010 to 2019. This evaluation followed the same sampling scheme, which exclusively employed historical data for each month. These predictions serve to assess the model's effectiveness for real-time operational drought monitoring, providing realistic metrics.

The accuracy metric quantifies the number of correctly labeled samples as a proportion of the total number of samples:

$$\text{Accuracy} = \frac{\hat{y}_i \cap y_i}{N} \cdot 100 \quad (i = 1, \ldots N),$$

where y_i are the observed class labels, $\hat{y}_i$ are the predicted class labels, and N is the sample size.

However, accuracy is a misleading evaluation metric for both imbalanced and ordinal multiclassification problems for three main reasons. First, the ND class contains significantly more samples than the other classes, meaning that maximizing overall accuracy places greater emphasis on accurately predicting the ND class. A skillful model should possess the capability to predict both the presence and absence of drought. Second, it uses a binary interpretation of classification accuracy (i.e., correct or incorrect), while the difference between the predicted and observed categories are also important. Finally, the accuracy metric is challenging to interpret for imbalanced classifiers because it is difficult to determine whether the model's accuracy is statistically significant or merely due to chance. For instance, while an accuracy above 50% in predicting coin flips indicates skill, determining the threshold for skillful accuracy in predicting weighted dice is non-intuitive.

To overcome the limitations of the overall accuracy metric, the weighted Cohen's Kappa (κ_w) metric is used to select the best model during the hyperparameter tuning stage and assess model performance (Cohen, 1968):

$$\kappa_w = \frac{p_{a(w)} - p_{e(w)}}{1 - p_{e(w)}}$$

where p_a is the proportion of agreements between the XGBoost model and the CDM drought categories (i.e., accuracy), and p_e is the proportion of agreements that would occur by chance:

$$p_e = \sum_{j=1}^{K} \frac{n_{j(\text{cdm})}}{N} \frac{n_{j(\text{xgb})}}{N}$$

where $n_{j(\text{cdm})}$ is the sample size of the jth class in the CDM, $n_{j(\text{xgb})}$ is the sample size of the jth class in the XGBoost model, and K is the number of drought classes. p_e will also be presented in the results section to provide context to accuracy figures. Finally, w is a matrix describing the degree of agreement between any combination of drought classes, where 1 indicates perfect agreement and 0 indicates perfect disagreement. A quadratic weighting scheme is used to place larger penalties on distant misclassifications. Note that the weighted kappa (κ_w) proposed in Cohen (1968) was used instead of the unweighted kappa (κ) as proposed in Cohen (1960). It is standard practice to use κ_w for ordinal scales because disagreements have varying levels of importance, while κ is used for nominal scales where all disagreements are considered equal.

This yields a coefficient where $\kappa_w < 0$ indicates the model performed worse than chance, $\kappa_w = 0$ indicates that the model performed equal to chance, and $\kappa_w = 1$ indicates perfect agreement between the model and CDM.

The coefficient can be interpreted using the following scale: 0–0.2 (slight agreement), 0.2–0.4 (fair agreement), 0.4–0.6 (moderate agreement), 0.6–0.8 (substantial agreement), and 0.8–1 (nearly perfect agreement) (Viera & Garrett, 2005).

To interpret model performance for each drought category (as opposed to overall performance), two additional metrics are presented. First, p_j is the precision (also called User's Accuracy) for the jth class:

$$p_j = \frac{\hat{y}_i \cap y_i}{n_{j(\text{xgb})}} \cdot 100 \quad \left(i = 1, \ldots n_{j(\text{xgb})}\right),$$

and r_j is the recall (also called Producer's Accuracy) for the jth class:

$$r_j = \frac{\hat{y}_i \cap y_i}{n_{j(\text{cdm})}} \cdot 100 \; (i = 1, \ldots n_{j(\text{cdm})}),$$

Put simply, precision is the proportion of correctly classified instances among the total number of predicted instances for a particular class. It quantifies the probability that a drought predicted by the model was observed in the CDM, providing insights into the reliability of the model's predictions. Conversely, recall is the proportion of correctly classified instances among the total number of observed instances for a specific class. It captures the probability of the model correctly predicting an observed drought, thereby reflecting the model's sensitivity.

2.4. Variable Importance

Shapley values were first introduced in game theory to determine the contributions made by each player in a cooperative game, with the idea that each player will receive a payout proportional to their contributions (Shapley, 1953). Recently, the SHAP method, based on the theoretical foundations of Shapley values, was developed to explain the impact of each covariate on model predictions (Lundberg & Lee, 2017):

$$\phi_m(v) = \sum_{S \subseteq N \backslash \{m\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} (v(S \cup \{m\}) - v(S))$$

where $\phi_m(v)$ is the contribution of covariate m, N is the set of all covariates, S is a subset of N, and $v(S)$ is the value of that subset. To elaborate, $S \subseteq N\backslash\{m\}$ indicates that we take every unique combination of covariates that excludes m. Given that $v(S)$ is the value for subset S, then $(v(S \cup \{m\}) - v(S))$ is the additional value when covariate i is included in S, otherwise known as the marginal contribution of m. The term $\frac{|S|!(|N|-|S|-1)!}{|N|!}$ is a scaling factor that is included to remove the effects of other covariates. Therefore, the formula quantifies the average marginal contribution of m to model predictions.

When ϕ_m is calculated for each covariate, the explanation model, $g(x')$, is complete and can be used to understand the original output, $f(x)$, in the form of a linear model:

$$f(x) = g\left(x'\right) = \phi_0 + \sum_{m=1}^{M} \phi_m z'_m$$

where $z' \in \{0,1\}^M$ and M is the number of covariates. ϕ_0 is the null model output (i.e., with no covariates), which is the expected probability of a CDM category occurring. This means that the prediction made by a null model plus the marginal contributions of each covariate must equal the model prediction, resembling linear regression where the sum of the intercept plus effects equals the prediction.

Finally, SHAP values inherit the units from the model they interpret, which is log-odds for XGBoost. We use a simple linear scaling to approximate probabilities from log-odds for enhanced interpretation:

$$\varphi_m \cong \rho \phi_m$$

where φ_m is the SHAP value in probability units, ϕ_m is the SHAP value in log-odds, and ρ is a scaling coefficient:

$$\rho = \frac{\sum_{m=1}^{M} \phi_m}{\hat{y} - \omega}$$

where $\hat{y}$ is the predicted class probability and ω is the empirical class proportion in the training data.

The use of SHAP values has increased substantially in recent years, including in the environmental sciences. Nandlall and Millard (2020) provide an excellent background of SHAP values and an overview of its application to remote sensing classifiers. Padarian et al. (2020) provide an example for interpreting digital soil maps and introduce the concept of mapping SHAP values to understand the spatial distribution of variable importance. It has also been used to interpret machine learning models for remote sensing NO_2 estimation (Ghahremanloo et al., 2021), remote sensing poverty mapping (Ayush et al., 2021), and urban vegetation mapping using aerial imagery (Abdollahi & Pradhan, 2021).

In this paper, SHAP values are calculated for all predictions in the summer drought of 2015, a suitable case study due to a severe drought permeating the prairies. The mean absolute SHAP values are then calculated for each category to provide global variable importance metrics and are clustered into groups of variables to elucidate the optimal types of data sets for drought monitoring. The average absolute SHAP value for each grid cell is calculated and presented as a map to illustrate the spatial variability of variable importance.

3. Results and Discussion

3.1. Classification Accuracy

The time series cross-validation technique used to evaluate model accuracy mimics real-time operational monitoring. It leverages all available historical data at each time step for training, enabling the model to predict the entire study area for the subsequent time step. It is important to note that this evaluation specifically focuses on the model's ability to predict drought in the Prairies for future months, and its performance in predicting other geographic locations is not examined in this study.

The XGBoost model had an overall accuracy of 71.3% (p_e = 47.6%), indicating that the correct drought class was identified 71.3% of the time, surpassing the 47.6% performance expected by chance alone. The κ_w score is 0.76, defined as substantial agreement with the CDM and approaching near perfect agreement (0.8–1.0). Both evaluation metrics demonstrate the model was skillful in predicting the CDM drought categories. The confusion matrix shows a more detailed account of model performance for each drought class (Figure 2). The precision and recall for each category are also summarized in Table 1. Notably, the model exhibits minimal systematic bias, as it predicts each category with roughly equal frequency as it occurs in the CDM. Furthermore, due to the subjective nature of the CDM, disagreements between the model and the CDM do not necessarily imply incorrect predictions but might reflect different interpretations of the drought phenomenon.

The ND (absence of drought) category has a recall of 85.1%, indicating that 85.1% of all ND observations in the CDM were correctly predicted. Among the incorrect predictions, the majority were misclassified as D0 (11.7%), while ND was seldom misclassified as D1, D2, or D3/D4 (3.1%). Conversely, the precision for the ND category is 88.9%, meaning that 88.9% of the ND predictions made by the model aligned with the ND category in the CDM. Among the cells incorrectly predicted as ND, a large portion belonged to the D0 category (9.6%). These results indicate that the ND category is predicted with both high sensitivity and reliability.

For the D0 (abnormally dry) category, the recall is 43.1% (p_e = 3.4%), and the omission error is 56.9%. The ND category was responsible for most of the omission errors (33.5%). The precision was 41.1% (p_e = 3.4%), with the commission errors mostly belonging to the ND category as well (40.7%). Evidently, distinguishing between ND and D0 is a challenge. This arises from the fact that neither class represents drought conditions, with D0 denoting short-term dryness typically observed during the onset or recovery of drought, and statistically only indicating a minor deviation from normal conditions (Svoboda et al., 2002). As a result, the accuracy of this class can be conceptualized as the ability of the model to respond to transitions in drought/ND status. Given the relatively infrequent occurrence of D0 (18.1% of all samples) compared to ND (66% of all samples), the small deviation in aridity was a challenge to detect.

The D1 (moderate drought) category has a recall of 42.6% (p_e = 1.1%). The omission error of 57.4% is primarily distributed between the D0 (29.7%) and D2 (17.1%) classes. The precision is 40.3% (p_e = 1.1%), with the commission errors of 59.7% distributed between the D0 (30.7%), ND (16.4%) and D2 (11.2%) classes.

The D2 (severe drought) class has a recall of 49% (p_e = 0.2%) with most of the omissions belonging to the D1 class (31%). The precision is 37.1% (p_e = 0.2%), with most of the commission errors representing the D1 class

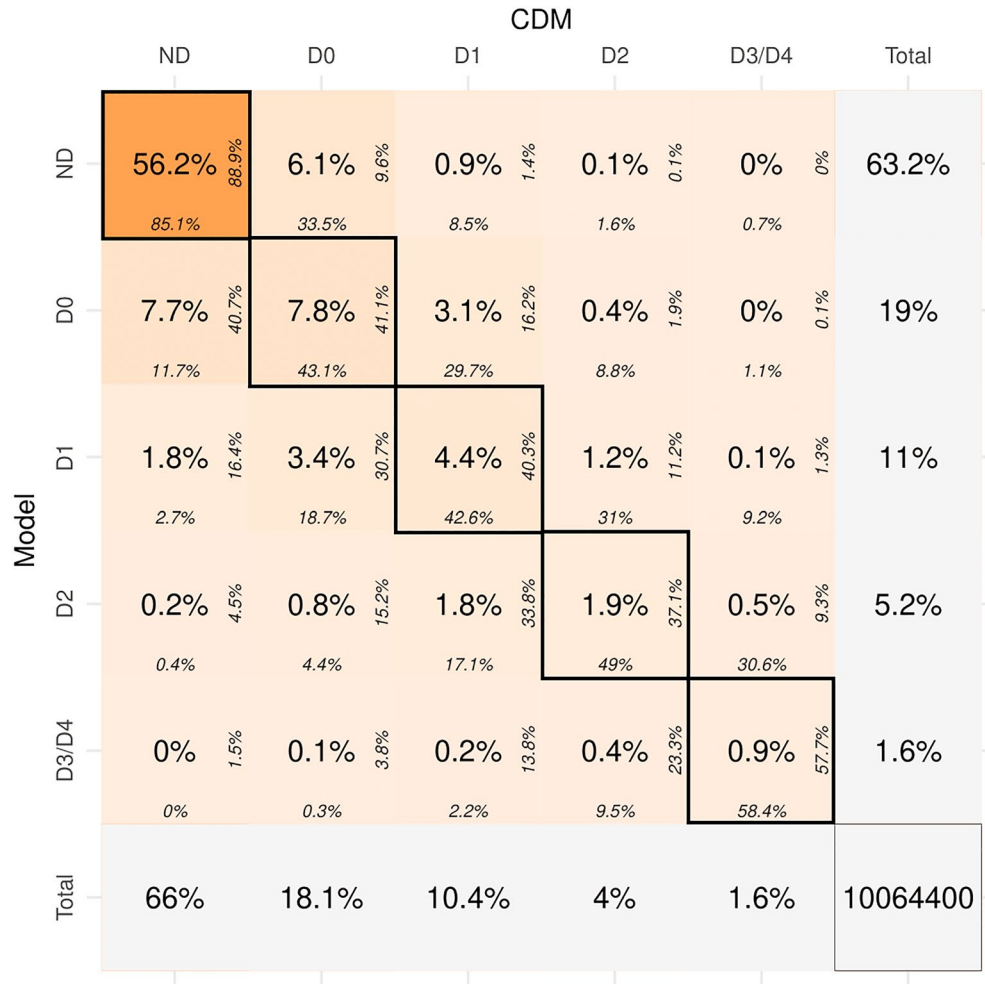

Figure 2. Confusion matrix of XGBoost model predictions versus observed Canadian Drought Monitor (CDM) categories. In each cell, the percentage of all samples falling into that category are shown in the center (also displayed by color). The diagonal cells (outlined in black) show correct predictions, with the recall shown at the bottom and the precision shown on the right side of the cell. For example, the top left cell indicates that 56.2% of all observations were no drought (ND) in both the CDM and the model, the recall indicates that 85.1% of all ND observations in the CDM were predicted as ND by the model, and the precision indicates that 88.9% of the ND predictions were ND in the CDM. The percentage of omission (commission) errors that fall into each category can be observed on the bottom (right-side) of the off-diagonal cells. For example, the cell in the first column and second row indicates that 11.7% of ND observations in the CDM were misclassified as D0 by the model.

Table 1
The Recall and Precision Metrics for Each Drought Category

Class	Recall (%)	Precision (%)	p_e (%)
ND	85.1	88.9	42.9
D0	43.1	41.1	3.4
D1	42.6	40.3	1.1
D2	49.0	37.1	0.2
D3/D4	58.4	57.7	0.03

Note. The omission error can be retrieved using the complement of the recall while the commission error can be retrieved using the complement of the precision. P_e is the recall and precision that would occur by chance alone and is provided for context.

(33.8%). The high recall compared to precision demonstrates that the model slightly overpredicts this category, which is further supported by its proportion of all predictions in the model (5.2%) compared to the CDM (4%).

Finally, D3/D4 has a recall of 58.4% (p_e = 0.03%), with the omission error of 41.6% primarily falling into the D2 (30.6%) categories. The precision is 57.7% (p_e = 0.03%), and the commission error of 52.3% comprises the D2 (23.3%) and D1 (13.8%) classes. The accuracy of this class surpasses all other classes except ND, demonstrating that the model can successfully identify the highest intensity droughts.

In general, the accuracy metrics for drought classes (D1, D2, D3/D4) are lower than the ND classes (ND, D0) because they are rare events that are challenging to identify. However, model misclassifications are often between drought classes and ND classes. For example, D3/D4 is classified as the D1, D2, or D3/D4 categories 98.2% of the time. Similarly, D2 (89.5%) and D1

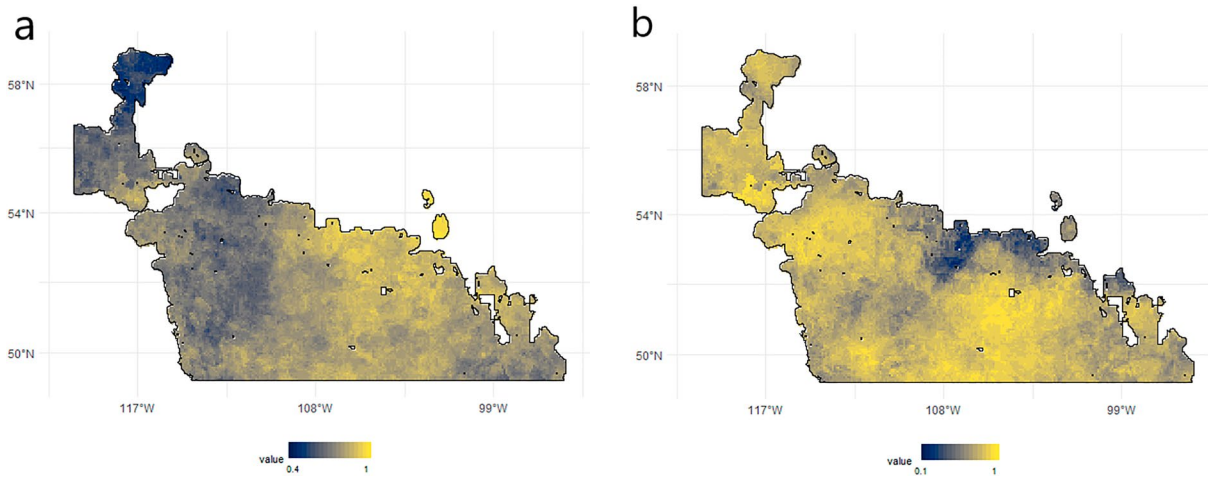

Figure 3. Pixelwise accuracy (a) and κ_w scores (b). Higher values, shown in yellow, indicate better model performance.

(61.9%) are frequently predicted as one of the three drought categories. This indicates that while the model can successfully detect drought, distinguishing the boundaries between adjacent drought classes poses difficulties. This challenge may be partially attributed to the subjective nature of human interpretation in the CDM, which incorporates aspects not captured solely by the biophysical data provided.

3.1.1. Spatial Patterns of Classification Accuracy

The performance of the model exhibited spatial heterogeneity, as demonstrated by the pixelwise accuracy and κ_w scores (Figure 3). In the north-central prairies (107°W, 53°N), a cluster of high accuracy scores of 80% or higher was observed. However, this region displayed the lowest κ_w values ranging from 0 to 0.2, indicating poor agreement only slightly above chance. Upon closer examination, it was found that these areas were primarily devoid of drought conditions, yet the model occasionally predicted abnormally dry conditions, resulting in a low κ_w score.

The κ_w map reveals that the model exhibited the highest skill in predicting drought in the south-central prairies (105°W, 51°N). This area has κ_w scores of 0.9 or higher, indicating nearly perfect agreement between the model and CDM. Similarly, notable clusters of high κ_w scores above 0.9 were observed in the western prairies, particularly in the vicinity of the Grand Prairie (117°W, 54°N), Edmonton (114°W, 53°N), and Lethbridge-Medicine Hat (112°W, 50°N) areas. These regions, which correspond to heavily populated areas and areas of intensive agriculture, showed the highest model performance. This outcome is likely attributable to the reliability and consistency of the CDM itself, as these areas benefit from a greater number of meteorological stations and local experts who provide ground-truth information.

3.2. Prediction Maps

The prediction skill of the model can also be visually assessed by mapping the predicted, observed, and difference maps each month. Figures 5–8 present the spatial patterns of drought for select seasons. The summer of 2012 was relatively drought free with a few exceptions (Figure 4). During June–August, the model had over 90% agreement with the CDM, accurately capturing the presence of aridity in most areas, including the small pockets in southern Manitoba (96°W, 50°N) and the Peace River region (118°W, 56°N). However, there were slight discrepancies in June, where the model predicted D1 instead of D0, and in August, where it underestimated drought by predicting D0 instead of a D1 event.

The following months of Fall 2012 had a rapid intensification and dissipation of drought that are ideal for examining the limits of model performance due to their challenging predictability (Figure 5). In September, there was widespread agreement on drought onset and intensification, although the model slightly underestimated the intensity of drought in the southern prairies (99–110°W, 50°N). In October, there was an improvement in drought conditions observed in the Peace River region (118°W, 56°N), which the model failed to capture, and it still generally underestimated drought in the southern prairies. However, the spatial patterns of drought were still

10.1029/2022WR033847

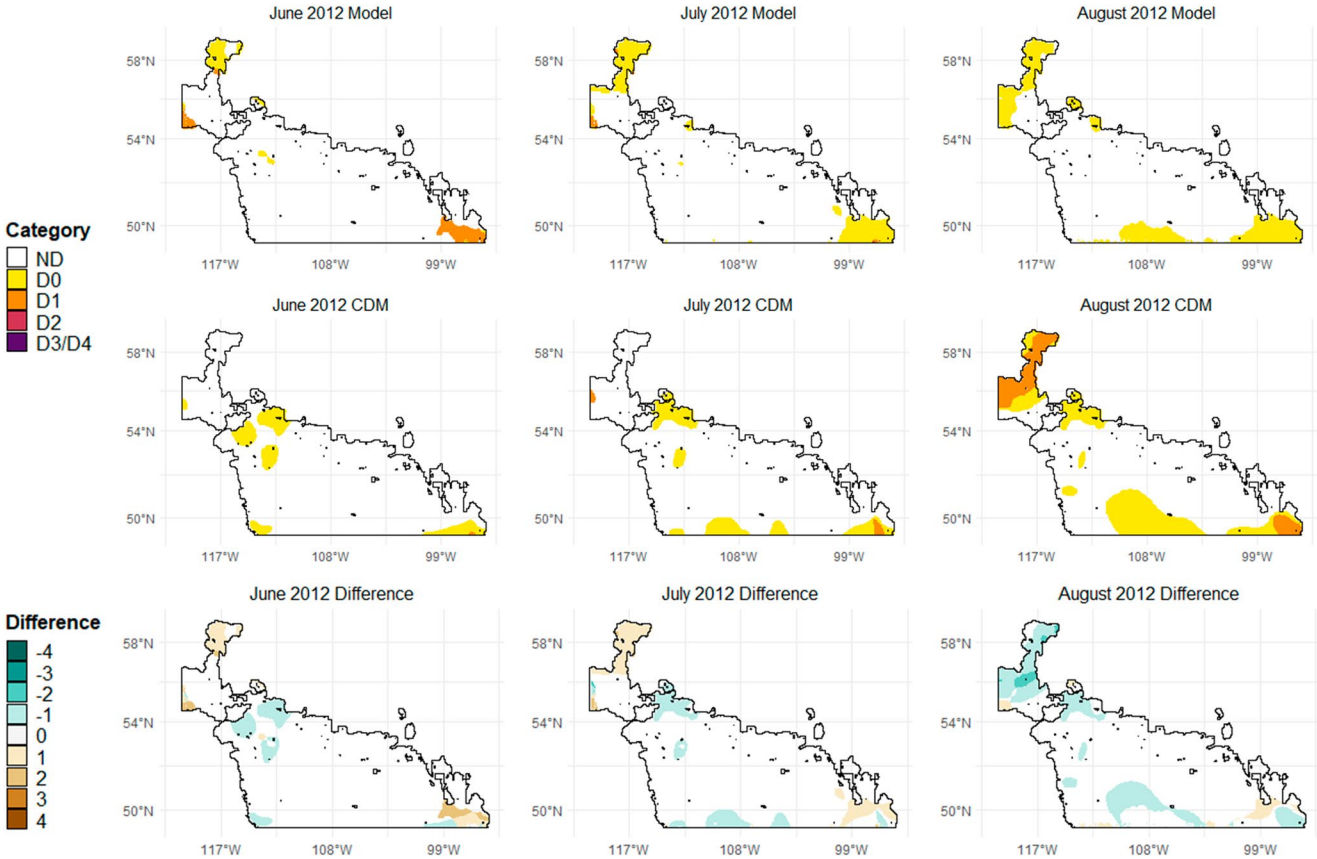

Figure 4. The model predictions versus Canadian Drought Monitor (CDM) drought categories in the summer of 2012. The bottom row displays the magnitude of misclassifications (Difference = Model − CDM).

captured, except for some D0 predictions in Alberta. Finally, in November, there was a rapid improvement and dissipation of drought conditions, which the model and CDM agreed on, while also detecting the persistence of D0 and D1 in the Peace River region.

The summer of 2015 was characterized by the most severe and widespread drought (Figure 6). The intensification of drought began in May (not shown), primarily in the form of D0 and D1 categories with two small pockets of D2. In June, there was a rapid and extensive intensification, with a large portion of the prairies experiencing drought severity ranging from D1 to D3/D4. While the model detected the onset and worsening of drought, it underestimated its severity in many areas, particularly the D2 and D3/D4 event in south-central Alberta and Saskatchewan (112°W, 51°N). In July, the western prairies (west of 108°W) witnessed further intensification, while improvement and dissipation occurred in the eastern prairies (east of 108°W) according to both the model and the CDM. However, a considerable portion of the D3/D4 drought was predicted as a D2. In August, drought persisted in the western prairies, albeit with a different spatial pattern. This pattern was reflected in both the model and CDM maps.

Next, the winter of 2017–2018 was marked by a widespread occurrence of winter drought, which continued from the previous season. The model showed exceptional congruence with the CDM maps predicting drought correctly or within one category throughout most of the study area (Figure 7). There was agreement on a localized D3/D4 drought event near Regina in December and January. However, in February, the model incorrectly predicted an improvement in drought severity from D3/D4 to D2 in south Saskatchewan (105°W, 51°N). The prediction maps for the winter of 2017–2018 highlight the model's ability to forecast drought events not only during the summer but also in the winter and shoulder seasons.

Finally, the spring drought of 2019 generally showed agreement between the model and CDM but exposed some model limitations (Figure 8). In April, there was a notable intensification of drought in central Saskatchewan

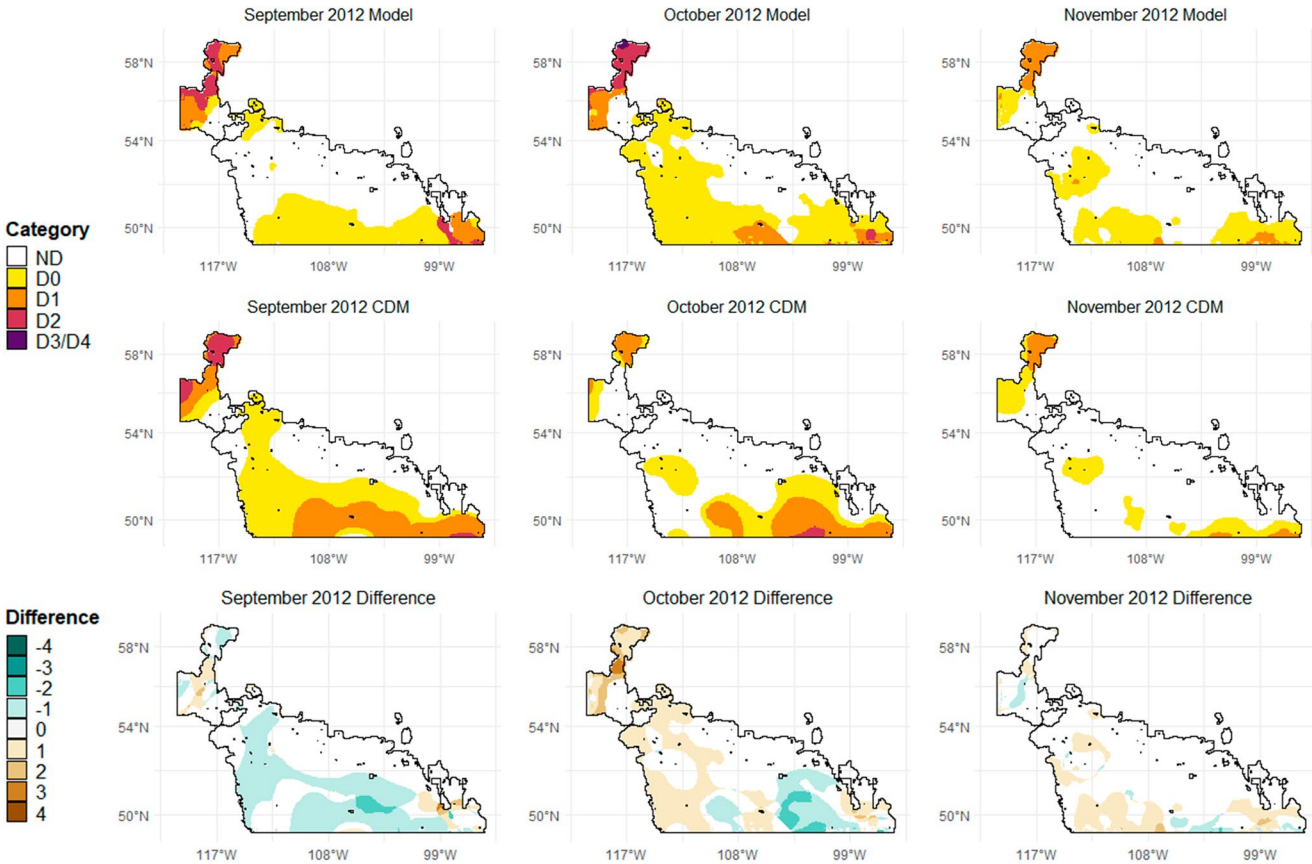

Figure 5. The model predictions versus Canadian Drought Monitor (CDM) drought categories in the fall of 2012. The bottom row displays the magnitude of misclassifications (Difference = Model − CDM).

(108–112°W, 52°N). Although the model predicted drought in this region, its severity was underestimated, while in other areas of the prairies, drought severity was overestimated. Subsequently, in May, there was a further intensification of drought that the model captured. Nevertheless, there was an overall overestimation of drought across many parts of the study area. These findings indicate a limitation of the model in accurately predicting flash drought conditions characterized by rapid onset and intensification.

Overall, the model exhibits a strong ability to capture the spatial structure and temporal evolution of drought events, as evidenced by its performance during various prairie droughts across all four seasons. However, the drought severity predictions were not as consistent. This reinforces the idea that the model is capable of distinguishing between drought and ND conditions, but the precise boundaries between adjacent drought classes remain ambiguous. In addition, the congruence between the model and CDM drought maps suffered in flash drought conditions, where rapid onset, intensification, and termination were underestimated.

3.3. Interannual Variability

The model performance, as measured by overall accuracy and κ_w, exhibits interannual variability (Figure 9). Generally, years characterized by higher drought coverage and severity (e.g., 2012, 2015, 2017–2019) exhibit lower overall accuracy scores, while years with less drought (e.g., 2010–2011, 2013–2014) show higher scores. This is attributed to the challenging nature of predicting infrequent drought events.

The κ_w metric provides a more reliable indicator of overall agreement between the predicted and observed drought categories, as it is less affected by the relative frequencies of each category. It exhibits substantial interannual variability with no clear trend (Figure 9). In the first year of the test period (2010), the κ_w value was nearly 0.9, indicating near perfect agreement between the model and CDM categories. The agreement was high due to the

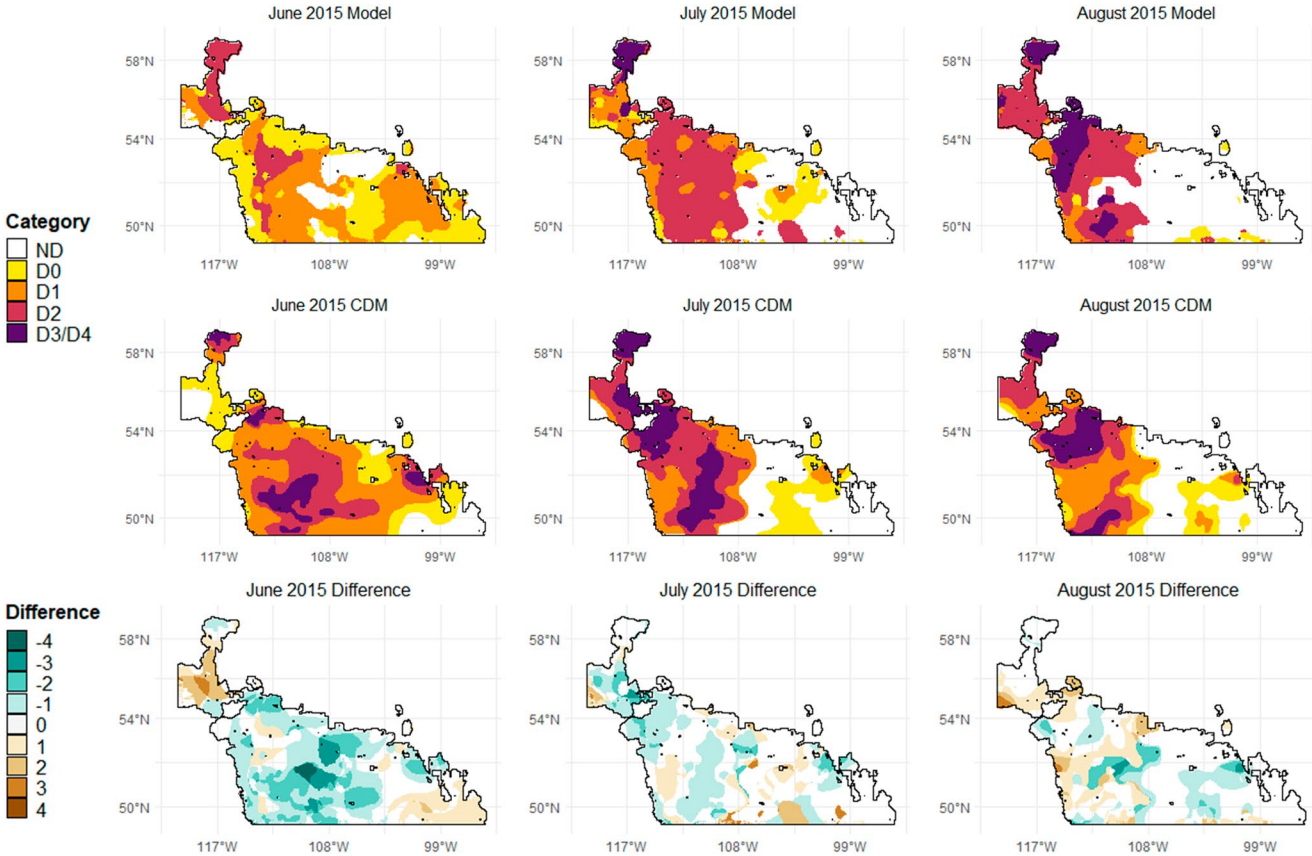

Figure 6. The model predictions versus Canadian Drought Monitor (CDM) drought categories in the summer of 2015. The bottom row displays the magnitude of misclassifications (Difference = Model − CDM).

absence of drought throughout most of the study area, except for a persistent drought in the Peace River region that was relatively simple to predict.

However, from 2011 to 2013, there was a decrease in the κ_w metric each year, reaching a minimum in 2013 at 0.5 (moderate agreement). These years generally saw an absence of drought, with localized pockets of rapidly changing drought conditions that were challenging to predict. This includes several periods of abnormally dry conditions that developed and dissipated quickly and were not captured well (not shown).

It is interesting to note that 2015 and 2017 witnessed the most extensive and severe droughts during the study period, resulting in a significant decline in overall accuracy. However, the κ_w scores of 0.75–0.80 indicate substantial agreement between the model and CDM. This suggests that although the exact classifications were not always correct, the misclassifications predominantly occurred between adjacent categories, and the overall evolution and spatial structure of drought was captured well.

In contrast, 2018 and 2019 also experienced widespread drought, but had slightly below average κ_w scores (0.6–0.7). This lower model performance compared to 2015 and 2017 suggests different drought characteristics that altered predictability. This could be attributed to the rapid intensification in August 2018 and rapid dissipation in July 2019. These flash droughts may be driven by different physical characteristics than conventional droughts that were not adequately represented in the training data. Alternatively, the limitations in capturing flash droughts could be attributed to the monthly time step used in the model, as these events typically develop over several weeks (Otkin et al., 2022).

The variability between the accuracy of each drought category decreases over time (Figure 10). Prior to 2015, the model performance of ND and extreme drought far exceeds the other three categories. However, the accuracy of each category is similar by the final years in the study period. There are two reasons for this. First, the progressively longer time series available for model training improves predictability of D0, D1, and D2 drought

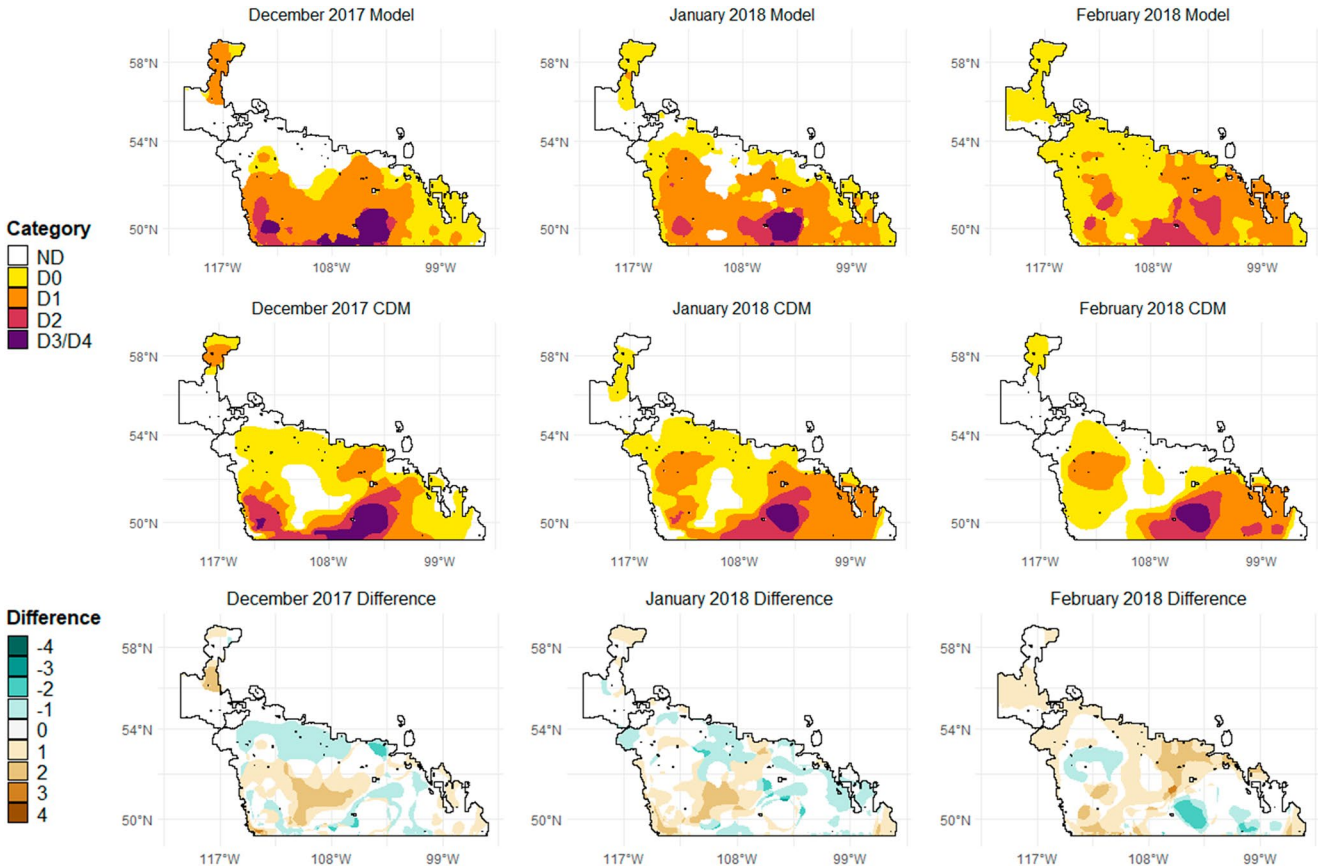

Figure 7. The model predictions versus Canadian Drought Monitor (CDM) drought categories in the winter of 2017/2018. The bottom row displays the magnitude of misclassifications (Difference = Model − CDM).

as the model learns to differentiate the subtle variations in the data that distinguish each category. Therefore, it is expected that the model will continue to become more reliable and consistent over time. At the same time, the high incidence of drought during the later years leads to a decrease in the accuracy of predicting ND conditions. As widespread drought becomes more prevalent, it becomes increasingly challenging for the model to accurately identify pockets of ND amidst these extensive drought conditions.

3.4. Variable Importance Case Study

3.4.1. Analysis of an Individual Prediction

SHAP values were calculated for each observation (grid cell) during the summer of 2015 to gain insights into the underlying factors driving the occurrence of the severe and widespread drought event in the study area (see Figure 6). The results for a randomly selected D3/D4 grid cell in July of 2015 are shown in Figure 11. The baseline probability of experiencing a D3/D4 drought is 1.46%, based on its frequency in the training data. However, the model predicts a probability of 48% for this grid cell. This difference is attributable to the values of the model predictors. The SHAP values provide an understanding of how each predictor variable contributes to the prediction. By analyzing these values, it becomes possible to discern the relative importance of different factors in driving the drought event.

Among the variables, the D3/D4Prop variable has the highest impact on increasing the probability of the D3/D4 event. Its marginal probability of nearly 30%, suggests that the proportion of D3/D4 drought in grid cell's surrounding area during the previous month significantly influences the occurrence of a D3/D4 drought in that specific location. With a proportion of 9%, indicating a higher prevalence of D3/D4 drought than usual, this variable contributes significantly to the increased probability of the event. GRACE groundwater percentiles averaged over the previous 9 months also play a substantial role. With an average of 12.59%, indicating lower groundwater

10.1029/2022WR033847

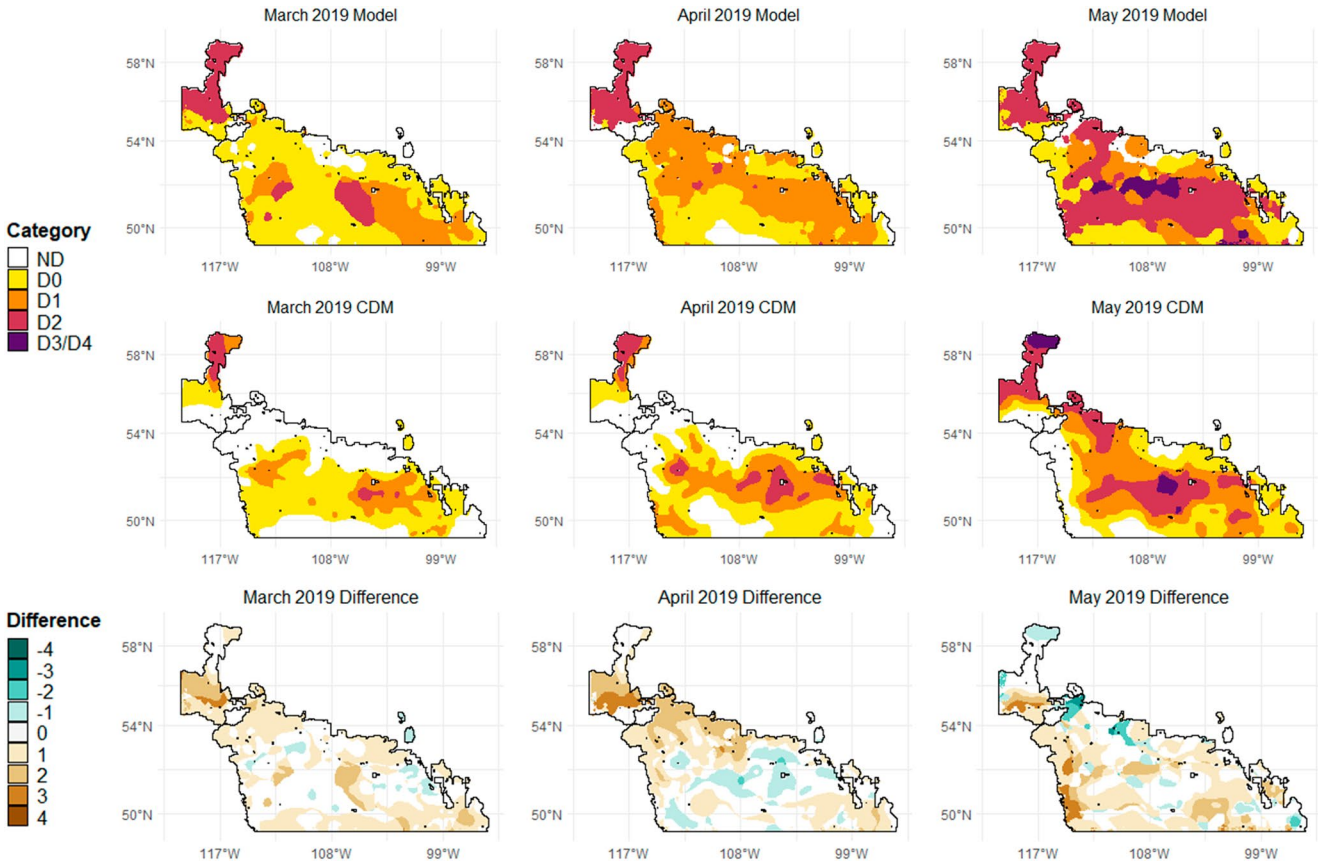

Figure 8. The model predictions versus Canadian Drought Monitor (CDM) drought categories in the spring of 2019. The bottom row displays the magnitude of misclassifications (Difference = Model − CDM).

levels, this variable increases the probability of a D3/D4 event by approximately 17%. The SPEI12 drought indicator, reflecting the meteorological water balance over the previous 12 months, was −2.01 standard deviations below the average at the selected location. This contributes to an increased probability of a D3/D4 drought by about 11%. Similarly, PDI was −2.84 at that location, increasing the probability of a D3/D4 drought by about 7%. The MEI value of 1.73 is indicative of an El Niño event, raising the probability of a D3/D4 drought by about 10%.

3.4.2. Most Important Types of Variables

Furthermore, the absolute mean SHAP value for different groups of variables were calculated to understand the types of data with the highest impact on predicting the summer drought of 2015 (Figure 12). The results demonstrate that the most important groups of variables were drought indicators and temporal variables. With values exceeding 15% for each drought category, this indicates that climatological anomalies and drought characteristics the previous month were key predictors of every drought category. For example, the mean absolute SHAP value of drought indicators for D2 drought is about 25%, meaning that the SPEI and PDI variables together changed the probability of the model predicting a D2 drought by 25% on average.

Other impactful data sets include GRACE, spatial variables, and teleconnection indices, demonstrating the importance of location, groundwater levels, and large-scale ocean-atmosphere processes in predicting drought in the Canadian Prairies. The satellite-based NDVI, ESI, and satellite soil moisture variables were not very impactful for predicting the summer 2015 drought, although these data sets have demonstrated success in monitoring and predicting drought in the past (Anderson et al., 2016; Cammalleri et al., 2021; Champagne et al., 2012; Klisch & Atzberger, 2016).

Drought indicators were the most impactful variables for predicting ND with a marginal probability of 43% (Figure 12), suggesting that simple climatological water balance approaches were sufficient for delineating drought and ND areas in the prairies. These indicators were important for predicting the other categories as well,

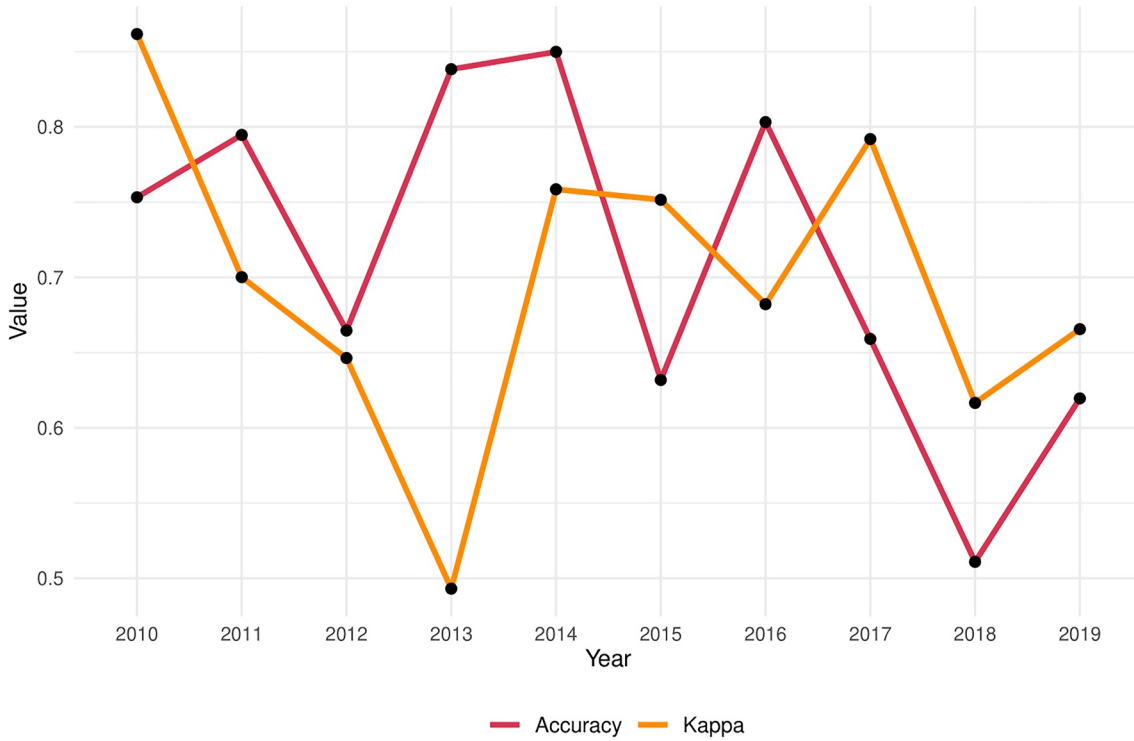

Figure 9. Line chart showing annual accuracy (orange) and κ_w scores (red) averaged over the entire study area.

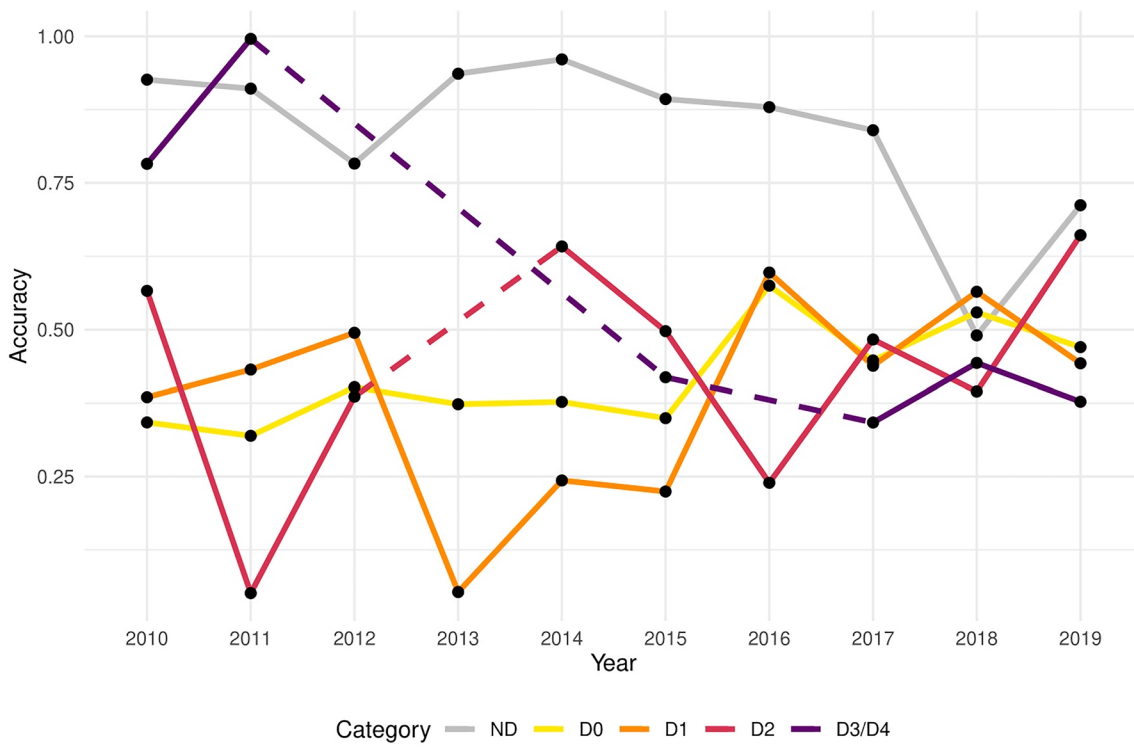

Figure 10. Annual accuracy by drought category, where accuracy is the average recall and precision of the corresponding drought category. Dashed lines indicate missing data.

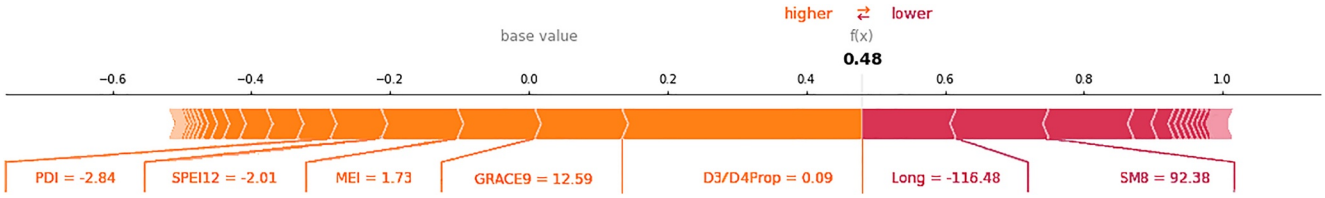

Figure 11. Force plot showing the most impactful predictors of a D3/D4 drought observation of the Peace River region in July 2015. The baseline probability (base value) of a D3/D4 drought is 0.0146, while the predicted probability for this observation is 0.48. The marginal contribution of each variable to this prediction is demonstrated by the size of the bars. Variables in orange increase the probability of a D3/D4 drought, while variables in red decrease the probability. The original feature values are displayed below the bars.

although further complementary information is required to predict the severity of drought. For instance, for the D3/D4 category, the impact of drought indicators was relatively lower compared to other categories, while the influence of GRACE groundwater percentiles and spatial variables (latitude/longitude) was considerably higher. This indicates that D3/D4 droughts are characterized by groundwater shortages, and location is a crucial predictor of drought severity.

More specifically, the absolute mean SHAP value was calculated for each variable for a more detailed account of variable importance (Figure 13). The most impactful variables were D2Prop and D3/D4Prop, with NDProp, D0Prop, and D1Prop also amongst the most important variables. These temporal indicators provide valuable information about the frequency and severity of drought that occurred in the surrounding area the previous month, serving as a baseline probability for drought occurrence. Surprisingly, the most impactful variable for abnormally dry (D0) predictions is the proportion of D2 drought the previous month (shown by the size of the yellow bars in Figure 12). This suggests that months with a high proportion of severe drought are more likely to experience below-average precipitation, even in areas where the impacts are not severe enough to be classified as drought. This observation aligns with the drought patterns depicted in Figures 5–9, where areas of drought often exhibit radial patterns that are frequently surrounded by D0 conditions.

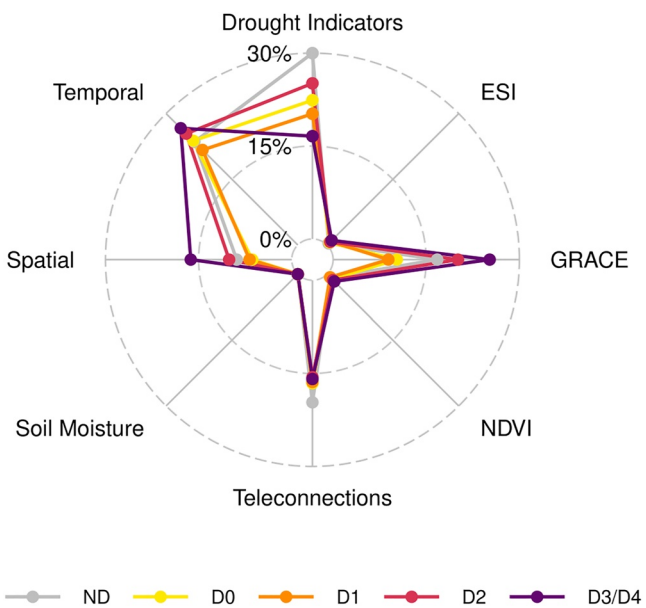

Figure 12. Radar chart showing the most impactful groups of variables for each category in the summer of 2015 (JJA). Calculated by taking the sum of the mean absolute Shapley Additive Explanation values of all variables within that group. Drought indicators include Palmer Drought Index and Standardized Precipitation Evapotranspiration Index, teleconnections include Atlantic Multidecadal Oscillation, Arctic Oscillation, Multivariate ENSO Index, North Atlantic Oscillation, Pacific Decadal Oscillation, and Pacific North America, soil moisture refers to satellite near surface soil moisture, spatial predictors are latitude, longitude, and ecoregion, and temporal predictors are the proportion of each drought category observed the previous month. Note that drought indicators were 43% for no drought but scaled to 30% for this figure.

PDI was the third most impactful variable overall, primarily due to its high impact in predicting the ND category where it has an absolute mean SHAP value of 17% (Figure 13). This means that PDI alone changed the probability of the model predicting ND by 17% on average. Other impactful variables for ND prediction are the longitude, SPEI3, NDProp, and GRACE1. This demonstrates that in this case, ND can be predicted using the amount of drought the previous month, short to medium term climatological anomalies, and short-term groundwater levels.

There is a strong association between drought severity and ENSO in the Prairies (Figure 13). The MEI values of 1.85, 1.73, and 1.92 for June, July, and August of 2015 are indicative of a strong El Niño event that rapidly intensified the previous spring, offering early warning potential for the 2015 drought event. El Niño is particularly important for predicting high severity drought events, as MEI has a mean absolute SHAP value of 6% for the D3/D4 category. This is supported by Z. Li et al. (2018) who found that the summer 2015 prairie drought and other historical droughts are related to positive ENSO phases.

In addition, the impact of MEI was spatially heterogeneous. Figure 14 shows the MEI impact (SHAP values) on D2 drought prediction in the summer of 2015. MEI was the most important in southern Alberta and Saskatchewan, particularly near Medicine Hat, Alberta (110°W, 50°N). MEI was also impactful in the Rocky Mountain foothills west of Edmonton (117°W, 53°N) and other pockets in Saskatchewan and southern Manitoba. This suggests a

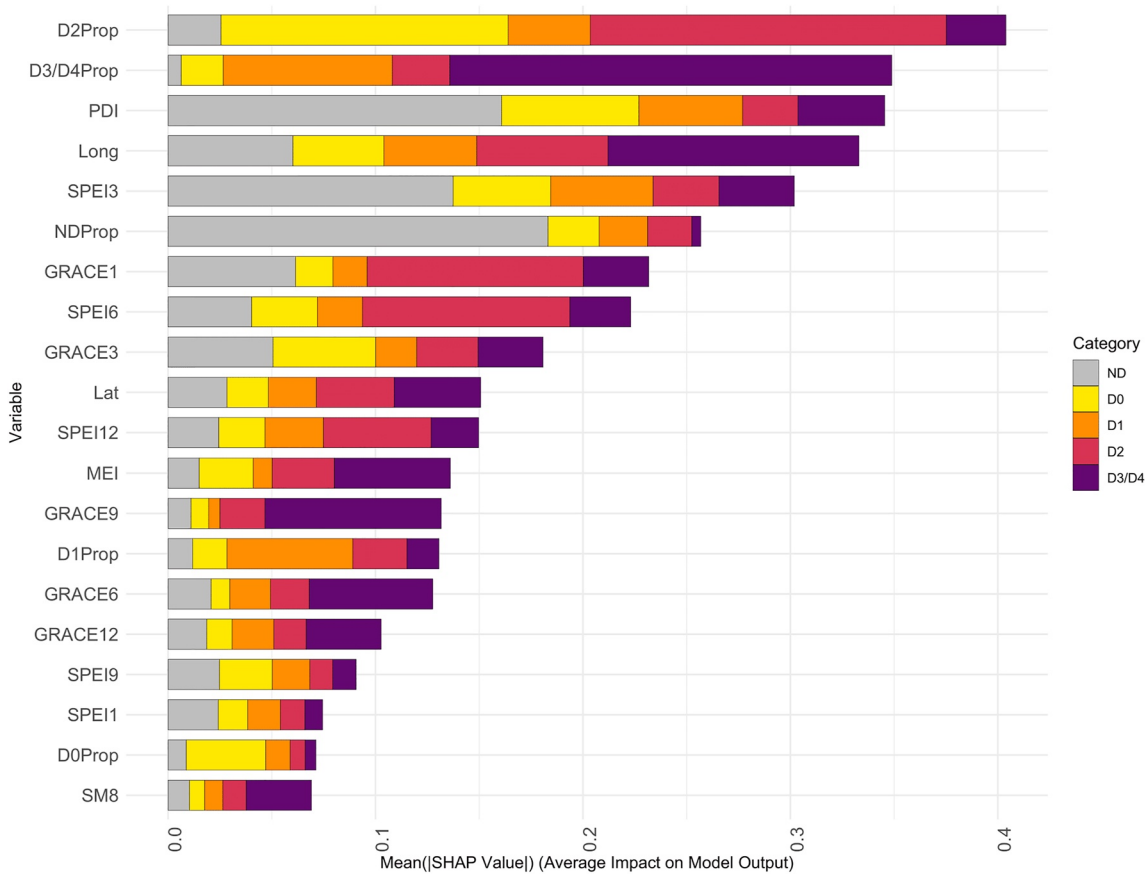

Figure 13. Stacked barplot showing the top 20 most impactful variables overall and for each category in the summer of 2015 (JJA). Calculated by taking the mean absolute Shapley Additive Explanation values for each variable by drought category.

teleconnection exists between ENSO and local weather in these areas in the summer of 2015. More research is required to understand if these spatial relationships remain for other drought events.

3.5. Limitations and Future Research

A key limitation to machine learning variable importance approaches is their sensitivity to model selection, as different models can reach the same conclusions through different pathways, a concept known as equifinality. While this sensitivity has been demonstrated for global variable importance metrics (Schmidt et al., 2020), the stability of SHAP values across different algorithms requires further investigation. This research is a critical first step toward understanding the key determinants of this drought, but future research may benefit from using a multi-model approach to quantify the contributions of model uncertainty in SHAP values.

This paper provides a proof of concept for an operational framework that monitors drought and assesses its underlying processes in real-time. The key drivers of the 2015 Prairie drought are identified, but a more comprehensive examination of SHAP variable importance across a wide range of drought conditions would provide further insight into drought dynamics in the Prairies and improve early warning systems to anticipate, adapt, and mitigate future droughts in the region.

4. Conclusions

Drought is a difficult phenomenon to predict given the complex and non-linear dynamical drivers and its varied impacts across multiple sectors. This research produced a machine learning model that monitors drought in the Canadian Prairies at 5-km resolution from 2010 to 2019. The use of XGBoost in conjunction with the κ_w evaluation metric overcame several challenges of categorical drought monitoring introduced by imbalanced and

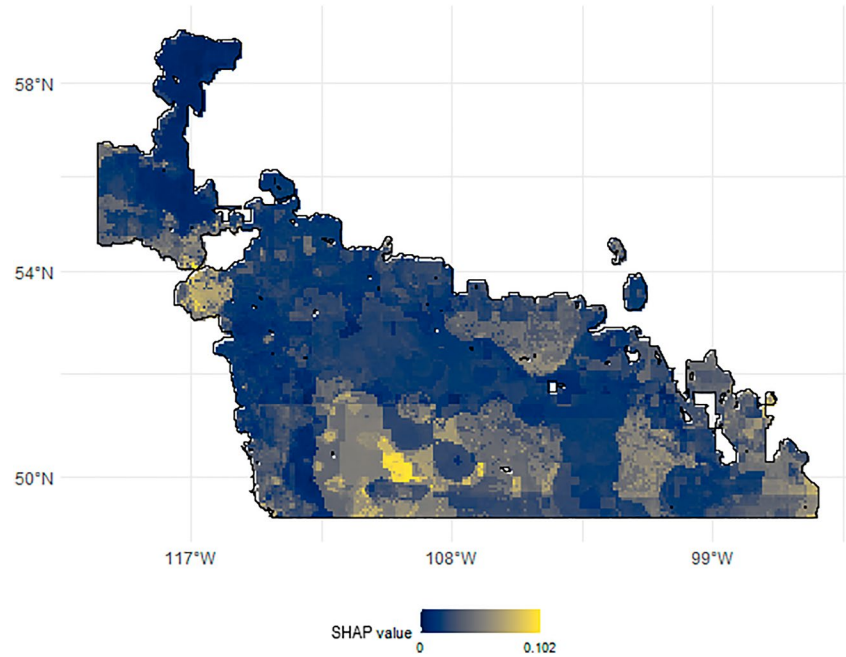

Figure 14. Map illustrating the spatial variability of Multivariate ENSO Index (MEI) impact on D2 prediction in the summer of 2015. High Shapley Additive Explanation values in yellow show areas where MEI had higher importance on drought predictability in the summer of 2015.

ordinal classification. It demonstrated substantial agreement with the CDM, capturing the evolution of historical droughts in a timely manner and describing their spatial structure well. This includes reliable performance in the absence of drought, and a sensitivity to a variety of drought conditions, including high severity droughts covering large areas. Given its high resolution, the model has the potential to capture local variations in drought conditions, although a more granular validation data set is required to evaluate its performance at this scale.

The automated approach to the CDM presented in this paper could streamline the CDM convergence of evidence methodology by eliminating or reducing the need for expert input, enabling a more rapid assessment of drought at high spatial resolutions and improving temporal consistency of the CDM database. This objective and reproducible model could also be used for retrospective drought analysis to better understand current droughts in a historical context. In addition, the trade-off between accuracy and interpretability has presented challenges for machine learning applications in climatology, especially where transparency of methods are needed for interpretation. We suggest that this trade-off can be mitigated using advances in XAI. This was demonstrated using a case study of the Prairie drought of 2015, where it was found that spatial and temporal predictors, drought indicators, GRACE groundwater percentiles, and teleconnection indices were all important model variables, providing insight into drivers of the drought event. Running this model operationally in near real time enables understanding of drought processes on-the-fly, offering the prospect of integrating science and operations for continuous improvement. Understanding why predictions were made (i.e., the physical processes driving drought) may also help practitioners weigh the reliability of predictions.

Overall, this research provides a high-resolution drought monitoring and interpretation framework using state-of-the-art machine learning and XAI techniques. It also demonstrates the applicability of this framework to operational drought monitoring systems over large areas, allowing stakeholders to not only mitigate, adapt, and make decisions about current droughts, but understand early warning signs for future droughts.

Data Availability Statement

All data analyzed in this study are publicly available. Many products can be downloaded from AAFC, including the Canadian Drought Monitor maps (AAFC, 2023a), Ecoregions (Ecological Stratification Working Group, 1995), SPEI (AAFC, 2023c), PDI (AAFC, 2023b), and NDVI (AAFC, 2023d). All teleconnection indices

can be downloaded from NOAA (NOAA, 2023), the soil moisture data from ESA (ESA, 2023), ESI from NASA SERVIR (NASA SERVIR, 2023), and NASA's model-driven estimates of groundwater percentiles from GRACE and GRACE-FO from the University of Nebraska-Lincoln (NASA, 2023).

Acknowledgments
This research was funded in part by Agriculture and Agri-Food Canada, the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Canada First Research Excellence Fund: Food from Thought. The authors wish to acknowledge Trevor Hadwen, Richard Warren, and Tyler Black for assistance with data preparation.

References

AAFC. (2015). Annual Crop Inventory [Dataset]. Retrieved from https://open.canada.ca/data/en/dataset/ba2645d5-4458-414d-b196-6303ac06c1c9

AAFC. (2023a). Canadian drought monitor (CDM) [Dataset]. Retrieved from https://agriculture.canada.ca/en/agricultural-production/weather/canadian-drought-monitor

AAFC. (2023b). Palmer Modified Drought Index [Dataset]. Retrieved from https://agriculture.canada.ca/atlas/data_donnees/agClimate/data_donnees/tif/pmdi/

AAFC. (2023c). Standardized Precipitation Evapotranspiration Index (SPEI) [Dataset]. Retrieved from https://agriculture.canada.ca/atlas/data_donnees/agClimate/data_donnees/tif/spei/

AAFC. (2023d). The Canadian Ag-land monitoring system (CALMS) [Dataset]. Retrieved from https://agriculture.canada.ca/atlas/data_donnees/calms/

Abdollahi, A., & Pradhan, B. (2021). Urban vegetation mapping from aerial imagery using explainable AI (XAI). *Sensors*, *21*(14), 4738. https://doi.org/10.3390/s21144738

Akinremi, O. O., McGinn, S. M., & Barr, A. G. (1996). Evaluation of the Palmer Drought Index on the Canadian Prairies. *Journal of Climate*, *9*(5), 897–905. https://doi.org/10.1175/1520-0442(1996)009<0897:eotpdi>2.0.co;2

Anderson, M. C., Kustas, W. P., Norman, J. M., Hain, C. R., Mecikalski, J. R., Schultz, L., et al. (2011). Mapping daily evapotranspiration at field to continental scales using geostationary and polar orbiting satellite imagery. *Hydrology and Earth System Sciences*, *15*(1), 223–239. https://doi.org/10.5194/hess-15-223-2011

Anderson, M. C., Zolin, C. A., Sentelhas, P. C., Hain, C. R., Semmens, K., Tugrul Yilmaz, M., et al. (2016). The Evaporative Stress Index as an indicator of agricultural drought in Brazil: An assessment based on crop yield impacts. *Remote Sensing of Environment*, *174*, 82–99. https://doi.org/10.1016/j.rse.2015.11.034

Ayush, K., Uzkent, B., Tanmay, K., Burke, M., Lobell, D., & Ermon, S. (2021). Efficient poverty mapping from high resolution remote sensing images. *Proceedings of the AAAI Conference on Artificial Intelligence*, *35*(1), 12–20. https://doi.org/10.1609/aaai.v35i1.16072

Barnwal, A., Cho, H., & Hocking, T. (2022). Survival regression with accelerated failure time model in XGBoost. *Journal of Computational and Graphical Statistics*, *31*(4), 1292–1302. https://doi.org/10.1080/10618600.2022.2067548

Bonsal, B., & Shabbar, A. (2008). Impacts of large-scale circulation variability on low streamflows over Canada: A review. *Canadian Water Resources Journal*, *33*(2), 137–154. https://doi.org/10.4296/cwrj3302137

Bonsal, B., & Shabbar, A. (2011). Large-scale climate oscillations influencing Canada, 1900-2008. Technical Thematic Report No. 4. In *Canadian biodiversity: Ecosystem status and trends 2010* (pp. 1–15). Canadian Councils of Resource Ministers.

Boyer, J. S., Byrne, P., Cassman, K. G., Cooper, M., Delmer, D., Greene, T., et al. (2013). The U.S. drought of 2012 in perspective: A call to action. *Global Food Security*, *2*(3), 139–143. https://doi.org/10.1016/j.gfs.2013.08.002

Brown, J. F., Wardlow, B. D., Tadesse, T., Hayes, M. J., & Reed, B. C. (2008). The Vegetation Drought Response Index (VegDRI): A new integrated approach for monitoring drought stress in vegetation. *GIScience and Remote Sensing*, *45*(1), 16–46. https://doi.org/10.2747/1548-1603.45.1.16

Cammalleri, C., Arias-Munõz, C., Barbosa, P., De Jager, A., Magni, D., Masante, D., et al. (2021). A revision of the combined drought indicator (CDI) used in the European drought observatory (EDO). *Natural Hazards and Earth System Sciences*, *21*(2), 481–495. https://doi.org/10.5194/nhess-21-481-2021

Champagne, C., Berg, A. A., McNairn, H., Drewitt, G., & Huffman, T. (2012). Evaluation of soil moisture extremes for agricultural productivity in the Canadian prairies. *Agricultural and Forest Meteorology*, *165*, 1–11. https://doi.org/10.1016/j.agrformet.2012.06.003

Champagne, C., Zhang, Y., Cherneski, P., & Hadwen, T. (2018). Estimating regional scale hydroclimatic risk conditions from the soil moisture active-passive (SMAP) satellite. *Geosciences (Switzerland)*, *8*(4), 1–18. https://doi.org/10.3390/geosciences8040127

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. San Francisco, California*.

Chen, T., He, T., Benesty, B., Khotilovich, V., Tang, Y., Cho, H., et al. (2021). xgboost: Extreme Gradient Boosting. R package version 1.4.1.1 [Software]. Retrieved from https://CRAN.R-project.org/package=xgboost

Chipanshi, A. C., Warren, R. T., L'Heureux, J., Waldner, D., McLean, H., & Qi, D. (2013). Use of the national drought model (NDM) in monitoring selected agroclimatic risks across the agricultural landscape of Canada. *Atmosphere-Ocean*, *51*(5), 471–488. https://doi.org/10.1080/07055900.2013.835253

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, *70*(4), 213–220. https://doi.org/10.1037/h0026256

Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., et al. (2017). ESA CCI soil moisture for improved Earth system understanding: State-of-the art and future directions. *Remote Sensing of Environment*, *203*, 185–215. https://doi.org/10.1016/j.rse.2017.07.001

Ecological Stratification Working Group. (1995). A national ecological framework for Canada [Dataset]. Agriculture and Agri-Food Canada. Retrieved from https://agriculture.canada.ca/atlas/data_donnees/nationalEcologicalFramework/data_donnees/

ESA. (2023). ESA CCI soil moisture [Dataset]. Retrieved from https://www.esa-soilmoisture-cci.org/data

Friedman, J. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, *29*(5), 1189–1232. Retrieved from https://www.jstor.org/stable/2699986. https://doi.org/10.1214/aos/1013203451

Gao, L., & Ding, Y. (2020). Disease prediction via Bayesian hyperparameter optimization and ensemble learning. *BMC Research Notes*, *13*(1), 1–6. https://doi.org/10.1186/s13104-020-05050-0

Ghahremanloo, M., Lops, Y., Choi, Y., & Yeganeh, B. (2021). Deep learning estimation of daily ground-level NO_2 concentrations from remote sensing data. *Journal of Geophysical Research: Atmospheres*, *126*(21), 1–18. https://doi.org/10.1029/2021JD034925

Gruber, A., Scanlon, T., Van Der Schalie, R., Wagner, W., & Dorigo, W. (2019). Evolution of the ESA CCI soil moisture climate data records and their underlying merging methodology. *Earth System Science Data*, *11*(2), 717–739. https://doi.org/10.5194/essd-11-717-2019

Gutzmer, D. (2021). *Impact summary June 2021*. National Drought Mitigation Center. Retrieved from https://storymaps.arcgis.com/stories/4b9f22a9781141a690f6454442bdb559

Han, H., Bai, J., Yan, J., Yang, H., & Ma, G. (2021). A combined drought monitoring index based on multi-sensor remote sensing data and machine learning. *Geocarto International*, *36*(10), 1161–1177. https://doi.org/10.1080/10106049.2019.1633423

Hoerling, M., Eischeid, J., Kumar, A., Leung, R., Mariotti, A., Mo, K., et al. (2014). Causes and predictability of the 2012 Great Plains drought. *Bulletin of the American Meteorological Society*, *95*(2), 269–282. https://doi.org/10.1175/BAMS-D-13-00055.1

Houborg, R., Rodell, M., Li, B., Reichle, R., & Zaitchik, B. F. (2012). Drought indicators based on model-assimilated gravity recovery and climate experiment (GRACE) terrestrial water storage observations. *Water Resources Research*, *48*(7), W07525. https://doi.org/10.1029/2011WR011291

Karnieli, A., Agam, N., Pinker, R. T., Anderson, M., Imhoff, M. L., Gutman, G. G., et al. (2010). Use of NDVI and land surface temperature for drought assessment: Merits and limitations. *Journal of Climate*, *23*(3), 618–633. https://doi.org/10.1175/2009JCLI2900.1

Keikhosravi-Kiany, M. S., Masoodian, S. A., Balling, R. C., & Darand, M. (2022). Evaluation of Tropical Rainfall Measuring Mission, Integrated Multi-satellite Retrievals for GPM, Climate Hazards Centre InfraRed Precipitation with Station data, and European Centre for Medium-Range Weather Forecasts Reanalysis v5 data in estimating prec. *International Journal of Climatology*, *42*(4), 2039–2064. https://doi.org/10.1002/joc.7351

Klisch, A., & Atzberger, C. (2016). Operational drought monitoring in Kenya using MODIS NDVI time series. *Remote Sensing*, *8*(4), 267. https://doi.org/10.3390/rs8040267

Kuhn, M. (2019). The caret package. R package version 6.0. Retrieved from https://topepo.github.io/caret/data-splitting.html#time

Kumar, K. S., Anandraj, P., Sreelatha, K., Bisht, D. S., & Sridhar, V. (2021). Monthly and seasonal drought characterization using GRACE-based groundwater drought index and its link to teleconnections across South Indian River basins. *Climate*, *9*(4), 56. https://doi.org/10.3390/cli9040056

Kumar, S. V., Zaitchik, B. F., Peters-Lidard, C. D., Rodell, M., Reichle, R., Li, B., et al. (2016). Assimilation of gridded GRACE terrestrial water storage estimates in the North American land data assimilation system. *Journal of Hydrometeorology*, *17*(7), 1951–1972. https://doi.org/10.1175/JHM-D-15-0157.1

Lawrimore, J., Heim, R. R., Svoboda, M., Swail, V., & Englehart, P. J. (2002). Beginning a new era of drought monitoring across North America. *Bulletin of the American Meteorological Society*, *83*(8), 1191–1192. https://doi.org/10.1175/1520-0477(2002)083<1191:BANEOD>2.3.CO;2

Li, B., Rodell, M., Kumar, S., Beaudoing, H. K., Getirana, A., Zaitchik, B. F., et al. (2019). Global GRACE data assimilation for groundwater and drought monitoring: Advances and challenges. *Water Resources Research*, *55*(9), 7564–7586. https://doi.org/10.1029/2018WR024618

Li, Z., Li, Y., Bonsal, B., Manson, A. H., & Scaff, L. (2018). Combined impacts of ENSO and MJO on the 2015 growing season drought on the Canadian Prairies. *Hydrology and Earth System Sciences*, *22*(10), 5057–5067. https://doi.org/10.5194/hess-22-5057-2018

Liu, Y., Dang, C., Yue, H., Lyu, C., & Dang, X. (2021). Enhanced drought detection and monitoring using sun-induced chlorophyll fluorescence over Hulun Buir Grassland, China. *Science of the Total Environment*, *770*, 145271. https://doi.org/10.1016/j.scitotenv.2021.145271

Liu, Y., Yang, X., Ren, L., Yuan, F., Jiang, S., & Ma, M. (2015). A new physically based self-calibrating Palmer Drought Severity Index and its performance evaluation. *Water Resources Management*, *29*(13), 4833–4847. https://doi.org/10.1007/s11269-015-1093-9

Liu, Y., Zhu, Y., Ren, L., Singh, V. P., Yang, X., & Yuan, F. (2017). A multiscalar Palmer drought severity index. *Geophysical Research Letters*, *44*(13), 6850–6858. https://doi.org/10.1002/2017GL073871

Lundberg, S. M., Erion, G. G., & Lee, S.-I. (2018). Consistent individualized feature attribution for tree ensembles. ArXiv Preprint, arXiv:1802. Retrieved from http://arxiv.org/abs/1802.03888

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Proceedings of the advances in neural information processing systems (NIPS)*, Long Beach, CA, USA, 4–9 December 2017 (pp. 4765–4774). https://doi.org/10.48550/arXiv.1705.07874

McKee, T., Doesken, N., & Kleist, J. (1993). The relationship of drought frequency and duration to time scales. In *Eighth conference on applied climatology, 17–22. Anaheim, California*. https://doi.org/10.1002/jso.23002

Molnar, C. (2022). Interpretable machine learning: A guide for making black box models explainable (2nd ed.). Retrieved from https://christophm.github.io/interpretable-ml-book

Nandlall, S. D., & Millard, K. (2020). Quantifying the relative importance of variables and groups of variables in remote sensing classifiers using Shapley values and game theory. *IEEE Geoscience and Remote Sensing Letters*, *17*(1), 42–46. https://doi.org/10.1109/LGRS.2019.2914374

Narasimhan, B., & Srinivasan, R. (2005). Development and evaluation of Soil Moisture Deficit Index (SMDI) and Evapotranspiration Deficit Index (ETDI) for agricultural drought monitoring. *Agricultural and Forest Meteorology*, *133*(1–4), 69–88. https://doi.org/10.1016/j.agrformet.2005.07.012

NASA. (2023). GRACE-based shallow groundwater drought indicator [Dataset]. Retrieved from https://nasagrace.unl.edu/

NASA SERVIR. (2023). Evaporative Stress Index [Dataset]. Retrieved from https://servirglobal.net/Global/Evaporative-Stress-Index-4-weeks

NOAA. (2023). *Climate Indices: Monthly atmospheric and ocean time series*. Physical Sciences Laboratory. Retrieved from https://psl.noaa.gov/data/climateindices/list/

NRCan. (2021). Canadian Wildland Fire Information System. Retrieved from https://cwfis.cfs.nrcan.gc.ca/maps/fm3?type=tri

Oozeer, Y., Fletcher, C. G., & Champagne, C. (2020). Evaluation of satellite-derived surface soil moisture products over agricultural regions of Canada. *Remote Sensing*, *12*(9), 1–17. https://doi.org/10.3390/RS12091455

Otkin, J. A., Woloszyn, M., Wang, H., Svoboda, M., Skumanich, M., Pulwarty, R., et al. (2022). Getting ahead of flash drought: From early warning to early action. *Bulletin of the American Meteorological Society*, *E2188*(10), 2188–2202. https://doi.org/10.1175/BAMS-D-21-0288.1

Padarian, J., McBratney, A. B., & Minasny, B. (2020). Game theory interpretation of digital soil mapping convolutional neural networks. *Soil*, *6*(2), 389–397. https://doi.org/10.5194/soil-6-389-2020

Palmer, W. C. (1965). Meteorological drought. In *U.S. Weather Bureau, Res. Pap. No. 45*. Retrieved from https://www.ncdc.noaa.gov/temp-and-precip/drought/docs/palmer.pdf

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016a). Model-agnostic interpretability of machine learning. In *ICML workshop on human interpretability in machine learning*. Retrieved from http://arxiv.org/abs/1606.05386

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016b). "Why should I trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135–1144). https://doi.org/10.1145/2939672.2939778816-3020

Schmidt, L., Heße, F., Attinger, S., & Kumar, R. (2020). Challenges in applying machine learning models for hydrological inference: A case study for flooding events across Germany. *Water Resources Research*, *56*(5), e2019WR025924. https://doi.org/10.1029/2019WR025924

Sepulcre-Canto, G., Horion, S., Singleton, A., Carrao, H., & Vogt, J. (2012). Development of a combined drought indicator to detect agricultural drought in Europe. *Natural Hazards and Earth System Science*, *12*(11), 3519–3531. https://doi.org/10.5194/nhess-12-3519-2012

Shabbar, A., Bonsal, B. R., & Szeto, K. (2011). Atmospheric and oceanic variability associated with growing season droughts and pluvials on the Canadian prairies. *Atmosphere-Ocean*, *49*(4), 339–355. https://doi.org/10.1080/07055900.2011.564908

Shapley, L. (1953). A value for n-person games. In H. W. Kuhn & A. W. Tucker (Eds.), *Contributions to the theory of games (AM-28)* (Vol. II, pp. 307–317). Princeton University Press. https://doi.org/10.1515/9781400881970-018

Son, N. T., Chen, C. F., Chen, C. R., Chang, L. Y., & Minh, V. Q. (2012). Monitoring agricultural drought in the lower Mekong basin using MODIS NDVI and land surface temperature data. *International Journal of Applied Earth Observation and Geoinformation*, *18*(1), 417–427. https://doi.org/10.1016/j.jag.2012.03.014

Spinoni, J., Naumann, G., Carrao, H., Barbosa, P., & Vogt, J. (2014). World drought frequency, duration, and severity for 1951-2010. *International Journal of Climatology*, *34*(8), 2792–2804. https://doi.org/10.1002/joc.3875

Svoboda, M. D., LeComte, D., Hayes, M., Heim, R., Gleason, K., Angel, J., et al. (2002). The drought monitor. *Bulletin of the American Meteorological Society*, *83*(8), 1181–1190. https://doi.org/10.1175/1520-0477-83.8.1181

Szeto, K., Zhang, X., White, R. E., & Brimelow, J. (2016). The 2015 extreme drought in Western Canada. *Bulletin of the American Meteorological Society*, *97*(12), S42–S46. https://doi.org/10.1175/BAMS-D-16-0147.1

Tadesse, T., Champagne, C., Wardlow, B. D., Hadwen, T. A., Brown, J. F., Demisse, G. B., et al. (2017). Building the Vegetation Drought Response Index for Canada (VegDRI-Canada) to monitor agricultural drought: First results. *GIScience and Remote Sensing*, *54*(2), 230–257. https://doi.org/10.1080/15481603.2017.1286728

Tao, L., Ryu, D., Western, A., & Boyd, D. (2021). A New Drought Index for soil moisture monitoring based on MPDI-NDVI trapezoid space using MODIS data. *Remote Sensing*, *13*(1), 122. https://doi.org/10.3390/rs13010122

Thomas, A. C., Reager, J. T., Famiglietti, J. S., & Rodell, M. (2014). A GRACE-based water storage deficit approach for hydrological drought characterization. *Geophysical Research Letters*, *41*(5), 1537–1545. https://doi.org/10.1002/2014GL061184.Received

Tian, K., Wang, Z., Li, F., Gao, Y., Xiao, Y., & Liu, C. (2021). Drought events over the Amazon River Basin (1993 – 2019) as detected by the climate-driven total water storage change. *Remote Sensing*, *13*(6), 1124. https://doi.org/10.3390/rs13061124

Vicente-Serrano, S. M., Beguería, S., & Lopez-Moreno, J. I. (2010). A Multi-Scalar Drought Index sensitive to global warming: The Standardized Precipitation Evapotranspiration Index – SPEI. *Journal of Climate*, *23*(7), 1696–1718. https://doi.org/10.1175/2009jcli2909.1

Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: The Kappa statistic. *Family Medicine*, *37*(5), 360–363. https://doi.org/10.1001/jama.268.18.2513

Wilken, E. B. (1986). Terrestrial ecozones of Canada. In *Ecological land classification, Series No. 19* (p. 26). Environment and Climate Change Canada

Xu, L., Abbaszadeh, P., Moradkhani, H., Chen, N., & Zhang, X. (2020). Continental drought monitoring using satellite soil moisture, data assimilation and an integrated drought index. *Remote Sensing of Environment*, *250*, 112028. https://doi.org/10.1016/j.rse.2020.112028

Yang, J. (2021). Fast TreeSHAP: Accelerating SHAP value computation for trees. ArXiv Preprint, arXiv:2109. Retrieved from http://arxiv.org/abs/2109.09847

Yang, K., & Wang, C. (2019). Seasonal persistence of soil moisture anomalies related to freeze–thaw over the Tibetan Plateau and prediction signal of summer precipitation in eastern China. *Climate Dynamics*, *53*(3–4), 2411–2424. https://doi.org/10.1007/s00382-019-04867-1