

Evaluation of Flash Drought Identification with Machine Learning Techniques.

Part I: Traditional Machine Learning Algorithms

STUART EDRIS^a, AMY MCGOVERN^{a,b}, JEFFREY B. BASARA^c, JORDAN I. CHRISTIAN^d, JASON C. FURTADO^a,
HENRY OLAYIWOLA^a, AND XIANGMING XIAO^e

^a School of Meteorology, University of Oklahoma, Norman, Oklahoma

^b School of Computer Science, University of Oklahoma, Norman, Oklahoma

^c Department of Environmental, Earth, and Atmospheric Sciences, University of Massachusetts Lowell, Lowell, Massachusetts

^d Department of Atmospheric Sciences, University of North Dakota, Grand Forks, North Dakota

^e Department of Microbiology and Plant Biology, University of Oklahoma, Norman, Oklahoma

(Manuscript received 17 July 2024, in final form 3 May 2025, accepted 12 June 2025)

ABSTRACT: Drought is an extreme event that can have a number of impacts on water resources and agricultural productivity. To mitigate the impacts, many studies have focused on improving drought predictions. As part of this, machine learning (ML) has emerged as a useful tool for drought identification and prediction, which has been shown to skillfully predict seasonal-scale droughts. In addition, a subset of drought, termed flash drought, has gained increased attention due to its rapid development and difficulty to predict. Studies focused on flash droughts have found a number of variables that accurately characterize their rapidly evolving conditions. However, while much work has been conducted to investigate flash droughts, limited work has been completed using ML techniques. This study investigated the ability of ML algorithms to represent and predict flash drought events via two parts. The first part focused on standard ML techniques such as random forests, Ada boosting, and support vector machines, while the second part focused on neural-network-based models. The ML models were trained on a set of key flash drought variables (soil moisture, evaporation, potential evaporation, temperature, and precipitation). Results showed Ada boosting was the most effective at representing and predicting flash drought events and was able to learn the climatological hotspots and some of the seasonality for flash droughts. However, the ML models tended to over rely on a few variables and were biased toward those variables. Additionally, case studies showed that the ML models struggled to capture the precise timing of flash drought events, resulting in a low skill score for all the examined ML models.

KEYWORDS: Drought; Atmosphere-land interaction; Classification; Decision trees; Machine learning; Support vector machines

1. Introduction

Droughts are dry extremes in the hydrologic cycle and can have profound impacts on a society's and ecology's water resources, particularly depending on the level of societal preparedness or ability to mitigate drought impacts. For example, in regions where drought resistance is not well developed, drought events have been linked to recent food crises (Dikshit et al. 2020; Krishnamurthy R et al. 2022; Lesk et al. 2022). Even in regions where drought resistance is more developed, drought results in billions of dollars in damage (Otkin et al. 2016; Dikshit et al. 2020), exacerbated more by the fact that some regions, such as the contiguous United States, experience droughts almost on a yearly time scale (Peters-Lidard et al. 2021).

An important feature of reducing the impact of drought is the ability to identify and predict drought occurrence. Drought prediction has evolved into three main types: 1) dynamical methods, which use physical models and physical processes to estimate

patterns in the hydrologic cycle; 2) statistical methods, which use statistical correlations between drought indicators and climatological features to predict droughts and include statistical models such as machine learning; and 3) hybrid methods, which combine the previous two methods (Hao et al. 2018; AghaKouchak et al. 2022). In the past decade, machine learning (ML) has emerged as a viable approach to predict drought events due to various ML techniques' ability to capture and learn nonlinearities in the hydrologic and climate systems (Dikshit et al. 2022; Hsieh 2022). ML approaches in drought prediction have mostly focused on long-term (seasonal to interannual) drought events, connected to climatological indices and surface variables to predict drought-related indices (Dikshit et al. 2022; Ganguli and Reddy 2014; Rhee and Im 2017; Mohamadi et al. 2020; Dikshit and Pradhan 2021) such as the standardized precipitation index (SPI) (McKee et al. 1993, 1995), standardized precipitation evaporation index (SPEI) (Vicente-Serrano et al. 2010), and Palmer drought severity index (PDSI) (Palmer 1965; Wells et al. 2004). Further, ML approaches traditionally utilized for drought have included random forests (RFs), boosted tree algorithms, support vector machines (SVMs) and support vector regressors (SVRs), and different forms of artificial neural networks (ANNs) (Dikshit and Pradhan 2021; Dikshit et al. 2022; Zhang et al. 2022).

Another recent emergence in drought research has been that of rapidly developing drought events, termed flash droughts

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/AIES-D-24-0068.s1>.

Corresponding author: Stuart Edris, sgedris@ou.edu

(FDs). Whereas traditional, long-term drought events develop slowly over a period of months to years, FDs develop over a short time scale (~ 3 – 6 weeks depending on the definition used), following a period of rapid intensification of drought conditions (Otkin et al. 2018; Christian et al. 2021; Qing et al. 2022). FDs produce additional impacts such as decreased crop yield, ecosystem stress, and increased fire risk as a result of the desiccation of the landscape (Christian et al. 2021; Alencar and Paton 2022; Gavahi et al. 2022). However, FDs have proven difficult to identify in real time and to predict due to the inherent nonlinearity in the evaporative feedback with the land surface during the rapid drying period (Otkin et al. 2018; Christian et al. 2019a, 2021; Osman et al. 2021; Otkin et al. 2021, hereafter O21; Tyagi et al. 2022), and the ability of FDs to rapidly develop even from above-normal moisture environments (Basara et al. 2019). In addition, because of the rapid development of FDs, long-term drought indicators, such as climate indices and drought indices (e.g., SPI and PDSI), are not as effective as or as strongly correlated with FD events (Otkin et al. 2018; Lisonbee et al. 2022; Tyagi et al. 2022). As a result, many studies have sought to increase our understanding of FD events.

While the lack of precipitation is the first-order driver of drought, recent studies have found that quickly changing land surface and hydrologic variables are the best indicators and drivers of FDs. The most prominent hydrologic variables include soil moisture (SM) (Osman et al. 2021; O21; Alencar and Paton 2022), evapotranspiration (ET), and potential ET (PET) (Otkin et al. 2018; Christian et al. 2019b; Tyagi et al. 2022). In addition, air and surface temperature are also key drivers of FD (Tyagi et al. 2022). To aid the identification of FD events, numerous new indices have been created to capture the rapidly developing nature of FD events including the evaporative drought demand index (EDDI) (Hobbins et al. 2016), the standardized evaporative demand index (SEDI) (Li et al. 2020), the standardized evaporative stress ratio (SESr) (Christian et al. 2019b, 2021, 2022, 2023, hereafter C23), and the flash drought intensity index (FDII) (O21). Compared to traditional methods, these indices have been shown to capture the rapidly evolving conditions of FD when calculated at the pentad and weekly time scales, while also capturing deteriorating vegetative conditions that mark some of the main impacts of FD.

While there is no consensus on a definition of FD (Otkin et al. 2018; Lisonbee et al. 2022), the numerous definitions to quantify FD cluster around quantifying the rapid intensification period (a metric of choice decreases by a certain amount within a 3–6 week period) and the drought component (the metric of choice falls below the 20th percentile to quantify the drought) (e.g., see Ford and Labosier 2017; Christian et al. 2019b, 2022, C23; Yuan et al. 2019; Li et al. 2020; Liu et al. 2020, hereafter L20; Noguera et al. 2020, hereafter N20; Pendergrass et al. 2020, hereafter P20; Osman et al. 2021; O21; Qing et al. 2022). While these different FD definitions involve similar concepts and identify similar hotspots in FD occurrence, each examines FD development in different ways and can have comparatively different results (Osman et al. 2021; Alencar and Paton 2022; Tyagi et al. 2022). For example, Christian et al. (2019b) identify FD using evaporative stress, which places stronger emphasis on

FDs in regions where the land surface coupling (more precisely, the coupling between the temperature and ET) is strongest, and underemphasizes locations where the coupling is not as strong (Osman et al. 2021). Thus, not only are no two droughts or FDs the same, neither are two FD definitions the same.

Despite advances in FD studies, identifying FD in real time and predicting them remains a challenge. In addition, despite ML techniques' ability to represent drought events, the utilization of ML techniques to identify and predict FD events has been limited (Zhang et al. 2022; Foroumandi et al. 2024; Scott et al. 2024; Lorenz et al. 2024). Zhang et al. (2022) used multilinear regression, RFs, and a recurrent neural network (RNN) to distinguish rapid intensification from a set of drought events, finding that the RFs produced the most stable statistical results. Foroumandi et al. (2024) incorporated a generative adversarial network (GAN) to predict the standardized soil moisture index (SSI), which was then connected to FD. The GAN model was able to learn complex nonlinear features and sudden changes in the soil moisture, effectively learning SSI patterns. However, potential avenues for identifying FDs exist in previously tested algorithms such as 1) RFs and boosted trees (Tyagi et al. 2022) and 2) in neural network (NN) algorithms, which are still new to the FD community, though there have been a number of applications toward monitoring drought severity (e.g., Kaur and Sood 2020; Dhyani and Pandya 2021; Foroumandi et al. 2023; Liu et al. 2024). As such, this study seeks to conduct a comprehensive investigation of how various ML techniques can represent FD events. Because several ML techniques have been used to represent drought historically but have not been explored before for FD events, this study has been split into two parts to represent the full breadth of different ML algorithms in FD analysis. The contributions of this first part include use of traditional ML techniques (i.e., RFs, SVMs, and decision trees boosted by the Ada algorithm; Pedregosa et al. 2011), and Edris et al. (2025, hereafter Part II) investigates NN techniques (i.e., various neural network architectures). While NN generally performs better than simple ML models, exploring some of the simpler models can still prove insightful, especially given tree-based models have still had comparable performance to neural network models for FD classification and prediction of FD variables (e.g., Li et al. 2022; Prodhan et al. 2021; Zhang et al. 2022), and were shown to be comparable for FD predictions as well (Part II). In addition, these simpler models provide some level of interpretability to understand why the ML models make their predictions, and how to improve on them (Flora et al. 2024).

2. Datasets

a. North American Regional Reanalysis

Data for this project originated from the North American Regional Reanalysis (NARR) dataset (Mesinger et al. 2006). The NARR provides high-resolution data, $0.3^\circ \times 0.3^\circ$, or approximately $32 \text{ km} \times 32 \text{ km}$, over North America and for over 40 years of data (from 1979 to the present). Surface variables in the NARR are determined by the Noah land surface

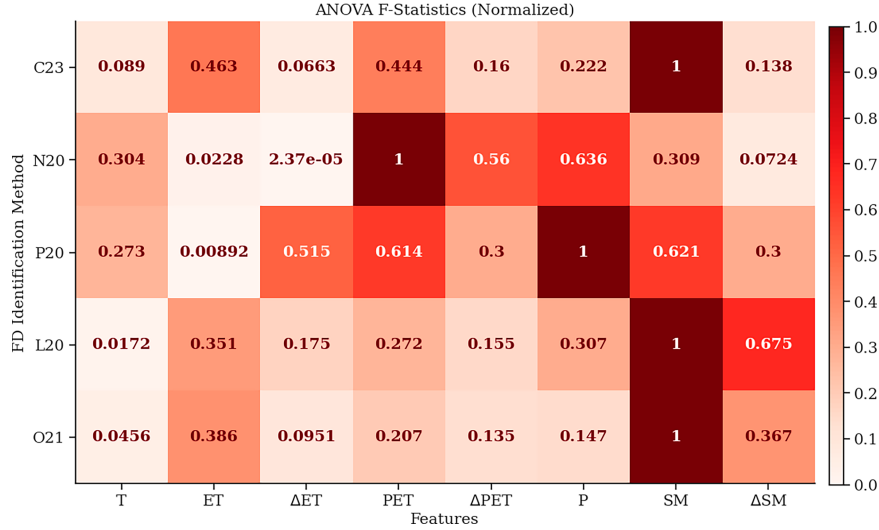


FIG. 1. Relative importance of each variable to identifying FD according to the normalized ANOVA F statistic. Bolded values denote statistical significance at the 95% confidence level.

model (Ek et al. 2003). The variables used for FD identification were SM (0–40 cm average), ET, PET, surface temperature T , and precipitation P . The dataset was also subsetting to focus on the contiguous United States (CONUS). The seasonality of FD favors spring and summer seasons in North America (Chen et al. 2019; Christian et al. 2019a, 2021). In addition, FDs have their greatest impact during the agricultural growing season (April–October in North America), and hence many FD studies focus on the growing season for FD research (Christian et al. 2019b; Shah et al. 2022). Thus, for this project, we chose the growing seasons for 1979–2021 and averaged the data into pentads (5-day, nonoverlapping mean). As such, each growing season contained 43 pentads. The pentad averaging was performed to retain the signal of FD predictors and minimize day-to-day variations. Last, this study tested and verified the ML models using K -fold cross validation (Hastie et al. 2009). For K -fold cross validation, each fold in the cross validation must be statistically independent from each other, which means that the dataset could not be divided randomly since some of the variables, such as temperature, PET, and especially soil moisture, are autocorrelated to various degrees. However, because there is a 5-month difference between each growing season, we assumed that the autocorrelation was negligible between one growing season and the next. Hence, each growing season is statistically independent and considered onefold. For each ML experiment, onefold was reserved for validation, onefold for testing, and the remaining folds for training. There were also 43 experiments for each ML model trained (and for each FD identification method explored) so that each fold was treated as a validation and test fold.

b. Input dataset

In addition to the five variables mentioned above, the changes in ET, PET, and SM from one pentad to the next were used to capture the rapid intensification component of

FD. For this study, the temporal changes in the three main drivers, SM, ET, and PET, were calculated as

$$\Delta x_p = x_p - x_{p-1}, \quad (1)$$

where x was either SM, ET, or PET; p was the pentad; and $p - 1$ was the previous pentad. Additionally, a 6-pentad leading moving average was applied to each change variable so that ΔSM , ΔET , and ΔPET all represented the mean 30-day change in their respective variable. Thus, the input parameters used to identify FD, in the order they are given in the ML algorithms, were T , P , ET, ΔET , PET, ΔPET , SM, and ΔSM . Moreover, prior to training, each variable was standardized for each grid point, relative to its respective dataset (training, validation, or testing). Last, the input data were reformatted to a space and time by input variable matrix so that each grid and pentad was treated as an example of FD identification for ML training. These variables were chosen due to their known correlation to FD and because they are known to describe the land surface processes that drive FD events (Otkin et al. 2018, 2021; Christian et al. 2019b; Tyagi et al. 2022). In addition, the change variables were designed to help the ML models learn the rapid change component that distinguishes FD from standard drought events and are also prominent drivers toward FD events (Fig. 1).

c. Output dataset

For this project, the FD identification problem was treated as a classification problem. Thus, the output dataset consisted of two values or labels:¹ 0 (i.e., there was no FD for that grid point and pentad) or 1 (i.e., there was FD for that grid point

¹ There was also a hidden, third class for any sea or ocean grid points whose class weight was set to 0 so that only land values are used in training.

TABLE 1. Overview of FD identification methods used in this study and their criteria for FD. Note the criteria are a summarized version, and the reader is referred to the original paper for full details (e.g., on full details for calculating obs and DRO_SUM for the O21 method).

Method	Metric	Criteria	Publication
C23	SESR (standardized ET/PET)	1. Rapid intensification period is 30+ days 2. SESR < 20th percentile at the end of rapid intensification 3. Pentad to pentad change in SESR < 25th percentile during rapid intensification	Christian et al. (2019b), revised in C23
N20	SPEI ($P - PET$ transformed to a standard normal distribution)	1. Rapid Intensification is 28+ days 2. SPEI decrease is at least 2 z units across the intensification period 3. SPEI < -1.28 z units at end of the intensification period	N20
P20	EDDI (PET transformed to a standard normal distribution)	1. EDDI increases by 50th percentiles in 3 weeks 2. EDDI does not decrease below the 50th-percentile increase for another 3 weeks	P20
L20	SM percentiles	1. During the rapid intensification period: mean change in SM percentiles $\geq$ 5th percentiles/pentad or maximum intensification rate of SM percentiles of 7.5th percentiles/pentad 2. SM percentiles $\leq$ 20 at the end of the rapid intensification period	L20
O21	FDII (determined from SM percentiles)	1. $FD_INT = (PER_BASE / (T_BASE)^{-1} \times \max(\text{obs})$ if $\max(\text{obs}) > PER_BASE / T_BASE$; 0 otherwise ($PER_BASE = 15$th percentiles, $T_BASE = 4$ pentads, and obs is the observed time-average change in SM percentiles up to 10 pentads) 2. $DRO_SEV = DRO_SUM / N_{pent}$ if $FD_INT > 0$; 0 otherwise ($DRO_SUM = \sum_i^{N_{pent}} (20 - SM_{percentile_i})$, N_{pent} can look up to 18 pentads forward) 3. $FDII = FD_INT \times DRO_SEV$	O21

and pentad). However, because there is no consensus on a definition for FD, five different FD identification methods were chosen for this project, and five separate ML methods (for each ML algorithm) were trained to identify FD for each method individually. An overview of each FD identification method used can be found in Table 1 and is reviewed below.

The first method for FD identification used was the method outlined in Christian et al. (2019b). This method employed evaporative stress (using SESR, which is the standardized ratio of ET to PET) to determine FD via a set of four criteria. However, Christian et al. (2022) employed an improved version of this method with satellite-derived land surface water index (LSWI), which was later incorporated into SESR (C23). The improved method used the Savitzky–Golay filter to remove high-frequency noise while retaining the signal in the pentads. The removal of this high-frequency noise also allowed for two criteria to be merged together, creating a simpler form of FD criteria. This project employs the improved and simplified version of FD identification in C23. This method required that the rapid intensification event lasted a minimum of 30 days, that the average change in the filtered SESR was below the 25th percentile (quantification of rapid intensification), and that the filtered SESR was below the 20th percentile at the end of the intensification (quantification of drought).

The second method used is outlined in N20. N20 used SPEI on the weekly time scale to represent the demand and supply of atmospheric moisture. Similar to the C23 method, N20 required that a rapid intensification period lasted at least 4 weeks (28 days; 30 days on the pentad time scale), that SPEI decreased by at least two standard deviations in the intensification period

(quantification of rapid intensification), and that SPEI was 1.28 standard deviations below normal at the end of the intensification period (quantification of drought).

The third method is given in P20. P20 used EDDI to identify FD when the EDDI increases by 50th percentiles over 2 weeks (or three pentads; rapid intensification quantification), and when that increased EDDI is maintained for another 2 weeks. P20 also used another method of FD identification which involves how rapidly drought intensifies according to the U.S. Drought Monitor (USDM). The USDM is still relatively new, having started in 2002 (Svoboda et al. 2002), and its methods for categorizing drought have evolved in time, with the allowance of more rapid changes in drought being implemented more in the past decade. Thus, the USDM dataset did not contain enough samples for the purpose of this study and was not used for FD identification.

L20 details the fourth FD method used in this project. L20 based their FD definition on the work done by Ford and Labosier (2017) and used weekly SM percentile data, requiring that, in a 3-week (four pentad) window of rapid intensification, the change in SM percentiles must average -6.5 th percentiles/week (or a 5th percentiles/pentad decrease), or have a minimum of -10 th percentiles/week (or -7.5 th percentiles/pentad; rapid intensification quantification), and that the minimum SM percentile is below the 20th percentile (drought quantification).

The final FD identification method used in this project is from O21. O21 defined the FDII to quantify the occurrence and intensity of FD using SM percentiles at the pentad time scale. The O21 method quantified rapid intensification and drought separately and identified FD by multiplying the components together. The calculation of FDII is more involved

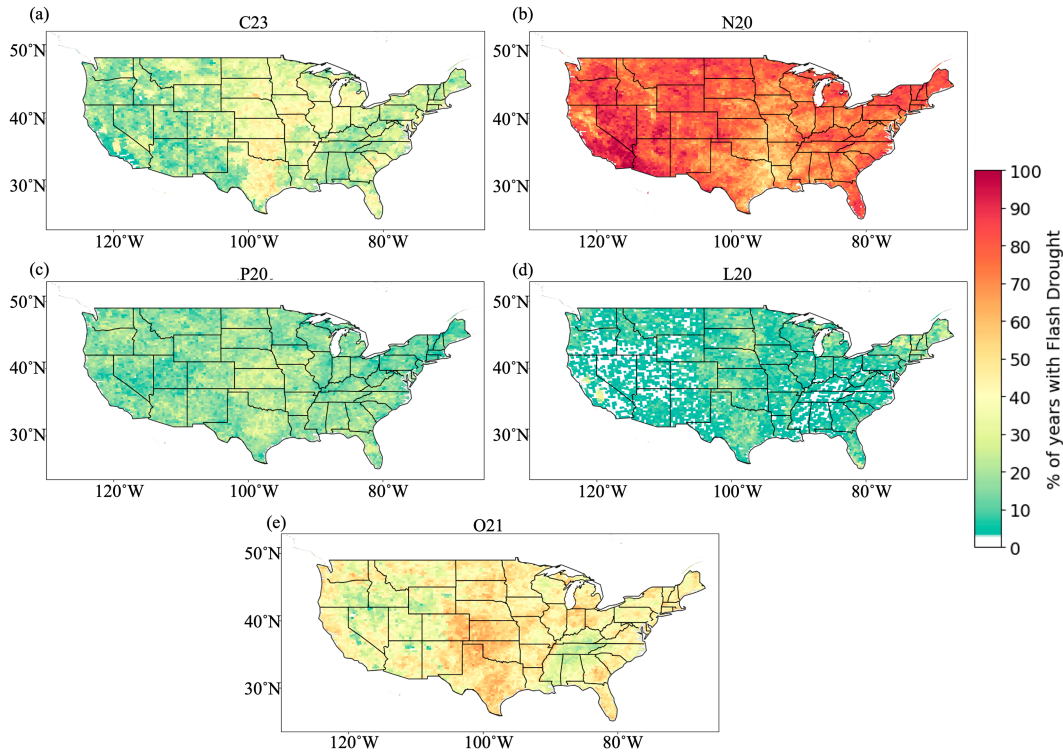


FIG. 2. Percentage of years from 1979 to 2021 in which FD occurred according to the (a) C23 method, (b) N20 method, (c) P20 method, (d) L20 method, and (e) O21 method.

than the previous methods, and the reader is referred to O21 for details. For simplicity, FD is said to occur when $FDII > 0$.

The frequency of FD events according to each method can be seen in Fig. 2. The general spatial trend in FD is similar for each method (with some small variations in the spatial hotspots), but each method has different FD frequencies and highlights some hotspots more strongly than others (some omit hotspot regions altogether), agreeing with the general findings in Osman et al. (2021). These frequency climatologies were one of the targets that the ML algorithms were learning to predict. An additional note is that the relative importance of each variable in describing FD depends on the identification method (Fig. 1). For example, SM is useful to describe FD using the C23 method, while ET, PET, and ΔSM are additional variables useful for the O21 method, and so on. These are also patterns the ML models should learn.

3. ML algorithms and verification

Traditional ML algorithms for this work were created using the sklearn package in Python (Pedregosa et al. 2011; Buitinck et al. 2013). To fine-tune model parameters, a subset of the dataset (tenfolds) and several rotations were used. The parameters with the best validation performance [based on various metrics such as area under the receiver operating characteristic curve (AUC), precision, and accuracy, as well as the ability of the test to recreate FD climatology] were chosen for the full experiments. As discussed above, no two FD identification

methods are the same, and thus each FD identification method needs different parameters. In testing the ML algorithms, the most important parameter was the class weights used to account for class imbalance. This may be expected, considering one important difference between each method is how frequently they identify FD. Thus, during these experiments, most of the parameters were kept the same across each FD identification method, and the class weights were modified to fit each identification method (Table S1 in the online supplemental material). Descriptions of these ML algorithms and the associated terminology can be found in Chase et al. (2022). As a note, XGBoosting, another boosting method, was found to be successful for other drought studies. But when examined here, it produced similar predictions to Ada boosting with lower skill (Table 2 and supplemental Figs. 24–32) and so was omitted from the discussion for brevity.

a. RFs

Decision trees are one of the basic types of ML algorithms. They use a set of yes/no questions (e.g., is the variable above or below a threshold?) that branch out to multiple possibilities leading to a specific classification. An RF is an ensemble of decision trees, each perturbed to slightly different conditions, and the most common classification is taken as the final classification value for the RF. The simplicity of this model, the ease of interpretation, and how readily it provides estimates for feature importance have led to RFs being one of the more common and preferred types of ML algorithms used for drought

TABLE 2. True skill statistic over all grid points and pentads for each ML algorithm and FD identification method. Numbers in the parentheses indicate 95% confidence intervals derived from a one-sample t test (calculated across all folds). Highest skill score for each identification method is bolded.

	RFs	Ada boosting	SVMs	XGBoost
C23	0.20 (± 0.016)	0.33 (± 0.019)	0.13 (± 0.016)	0.18 (± 0.016)
N20	0.17 (± 0.013)	0.13 (± 0.007)	0.22 (± 0.012)	0.22 (± 0.009)
P20	0.11 (± 0.016)	0.06 (± 0.008)	0.05 (± 0.013)	0.01 (± 0.002)
L20	0.15 (± 0.014)	0.27 (± 0.016)	0.07 (± 0.011)	0.02 (± 0.004)
O21	0.14 (± 0.009)	0.28 (± 0.016)	0.08 (± 0.011)	0.06 (± 0.010)

identification and prediction (Rhee and Im 2017; Rahmati et al. 2020; Dikshit and Pradhan 2021; Hsieh 2022) as well as other meteorological phenomena such as soil moisture predictions (e.g., Li et al. 2022) and investigations of mesoscale convection systems and their climatology (Cui et al. 2021). Tests showed that 150 trees, each allowed to grow 15 branches deep, were most effective for this study, similar to findings from other works (e.g., Park et al. 2016; Rhee and Im 2017; Rahmati et al. 2020).

b. Ada boosted decision trees

Boosting algorithms are based on fitting an ML model (in this case, decision trees) to a dataset and then adjusting the weights of the ML model with each iteration. The weight adjustments

are based on the gradient of an error function that is being minimized. Boosting is fairly common, with XGBoosting being the most favored in traditional drought studies (Park et al. 2016; Rahmati et al. 2020; Prodhon et al. 2021; Hsieh 2022). Ada boosting creates additional decision tree classifiers and adjusts the weights with each iteration to improve their performance. For this project, it was found 80 estimators were able to identify FD with a learning rate of 10^{-4} .

c. SVMs

SVMs are similar to perceptrons, with a model that is minimized by a loss similar to a linear model. Various kernels can be used for SVMs, with the radial basis function (RBF) kernel being most common for drought studies for its ability to represent nonlinear data more effectively (Ganguli and Reddy 2014; Mohamadi et al. 2020; Dikshit and Pradhan 2021). However, for this study, RBF kernels did not have a significant improvement over linear kernels, whereas they increased the computation time by an order of magnitude. Thus, linear SVMs were used for this study with the squared hinge loss function minimized, L2 regularization, and a stopping error of 10^{-6} .

d. Verification and interpretations

The skill of the ML algorithms was determined using the true skill statistic or Peirce's skill score (Woodcock 1976).

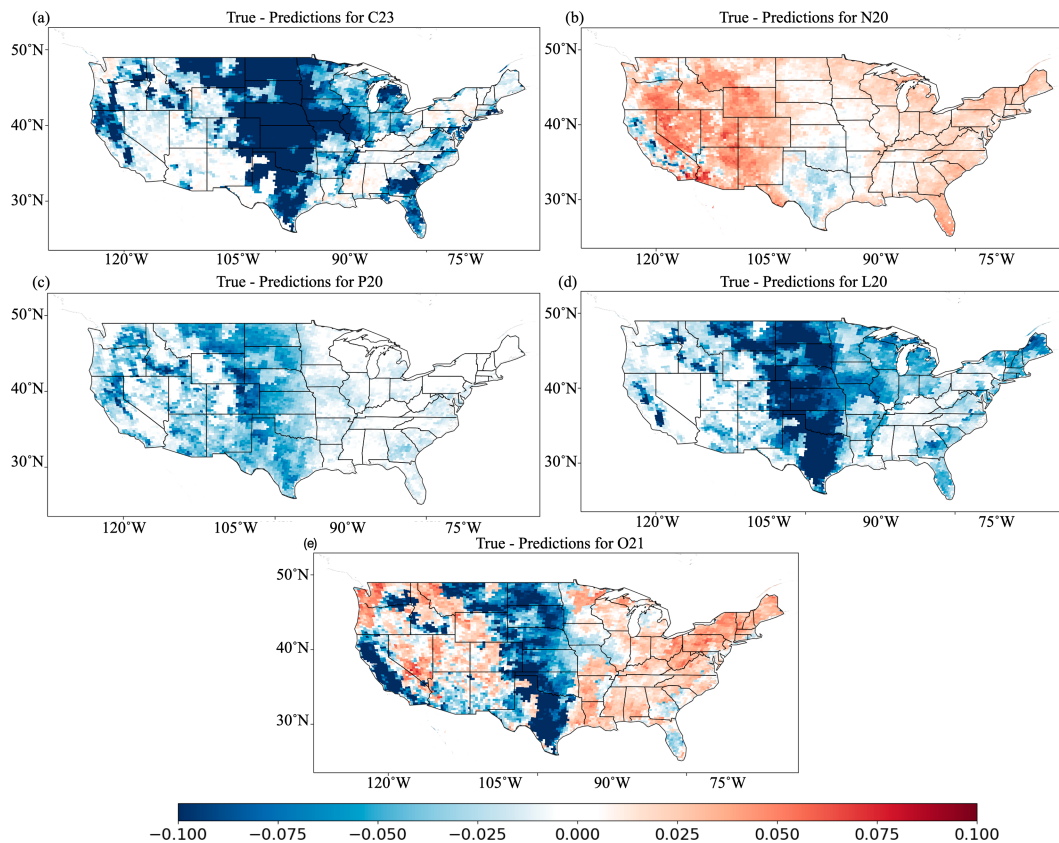


FIG. 3. Composite mean difference (unitless) of true minus predicted FD labels for each FD identification method using Ada boosted tree predictions.

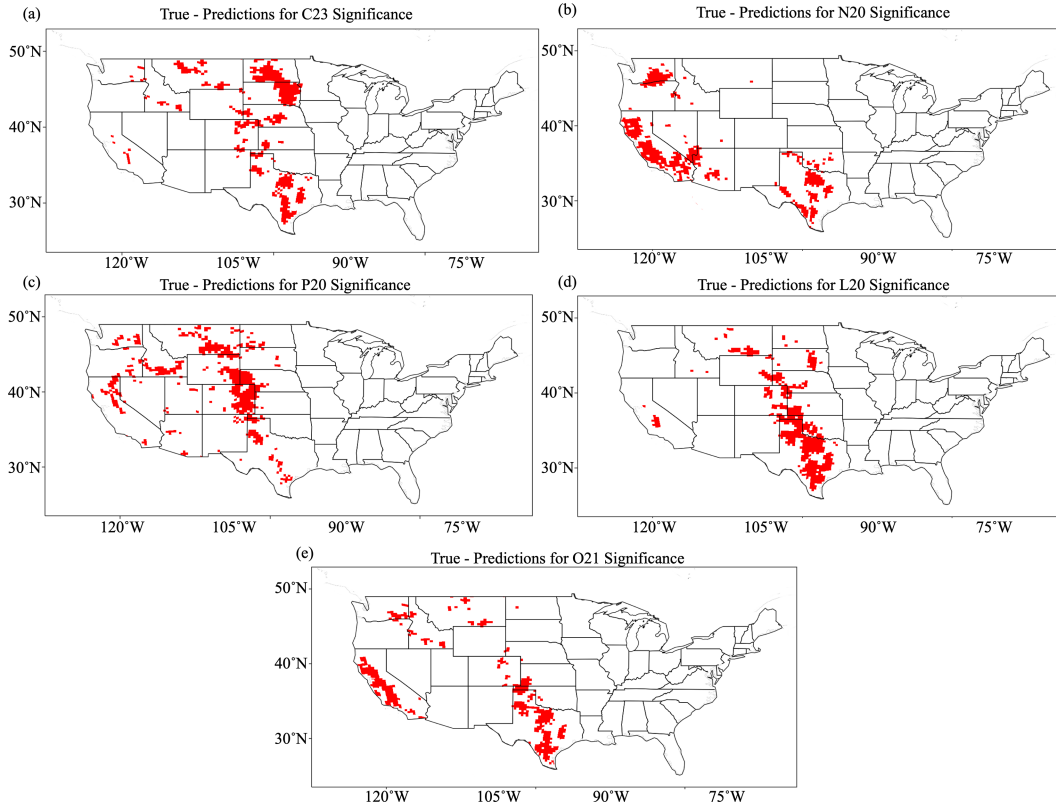


FIG. 4. The 95% statistical significance of composite mean differences between the true FD labels and the AdaBoost predictions. Statistical significance was determined by the Monte Carlo bootstrapping method ($N = 5000$).

The true skill statistic uses all four components of a confusion matrix² to determine the skill score. This score is independent of climatological event frequency, can be expressed similar to the equitable threat score, and can be expressed as the probability of detection (POD) of the event minus the probability of false detection (POFD). The true skill score can be expressed as

$$\begin{aligned} \text{skill} &= \frac{\text{true positives}}{\text{true positives} + \text{false negatives}} \\ &\quad - \frac{\text{false positives}}{\text{false positives} + \text{true negatives}} \\ &= \text{POD} - \text{POFD}. \end{aligned} \quad (2)$$

While other skill metrics such as the critical skill index (CSI) could also be used to determine the skill of the ML algorithms, the general comparison between ML algorithms was the same between multiple metrics (not shown), and the true skill statistic showed the difference in ML algorithm performance at two decimal places, whereas the others often required three or more decimal places. Models were also verified using K -fold cross validation, and skill metrics show the average skill score for test folds across all folds.

The main statistical comparison used to compare true and predicted FD labels was the composite mean difference, which

is the true labels averaged in time minus the predicted labels averaged in time. Statistical significance of the differences was determined from the bootstrap method³ at the $p < 0.05$ level, with a difference of zero used as the null hypothesis. Spatial distributions of other metrics, such as accuracy, precision, recall, and area under the receiver operating characteristic (ROC) curve, were examined with consistent results.

Last, feature importance was obtained via multiple methods to be compared. The first method used was Shapely values, determined from the Shapely additive explanation (SHAP) package (Lundberg and Lee 2017). Shapely values (called SHAP values in the SHAP package) assume that a model output or quality can be represented by a linear sum of contributions, and coefficients of those contributions then make up the feature attribution of a given feature (Lundberg and Lee 2017; Flora et al. 2024). The coefficients are estimated by comparing model performance “with” the feature to the model performance “without” the feature. Feature importance can be determined from attribution by performing a global average of the absolute values of the attributions. A second set of feature importances

³ The Monte Carlo bootstrapping method randomizes the data and repeats the statistics (in this case, the time series was randomized at each grid point and the composite mean difference calculated) N times to obtain a distribution. The actual statistic is compared to the distribution to obtain a p value. In this experiment, $N = 5000$.

² Also called truth tables or contingency tables.

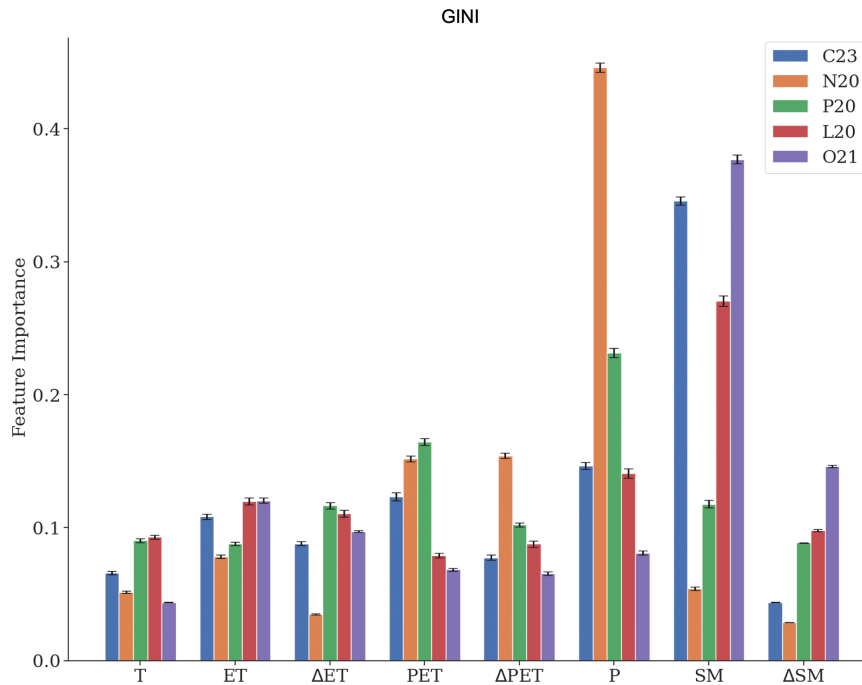


FIG. 5. Average feature importance for the Ada boosted decision trees according to the GINI method. Feature importance was determined for all rotations, and bars show the average importance across each rotation. Error bars indicate the 95% confidence interval from a one-sample t test (average and standard deviations taken across all rotations).

was calculated using permutation importance using the ranked probability skill score (RPSS) as the multiclass metric to evaluate the permuted skill (McGovern et al. 2019; Au et al. 2022; Flora et al. 2024). Permutation permutes (i.e., randomizes) one or more features and examines the decrease in the skill metric (e.g., RPSS) to determine the importance of the feature. The third method for determining feature importance for tree-based models was the GINI importance, obtained as part of the training from the total reduction of the tree criterion caused by certain features (Pedregosa et al. 2011; McGovern et al. 2019). Feature importance calculations are computationally expensive. Because of this, some simplifying assumptions were made. For example, calculations for the Shapely values were performed on the spatially averaged time series (over the domain) for test data for each experiment in the K -fold cross validation, with the assumption that the spatial average approximately represented the data. Additionally, the permutation importance was done with 10% of the test data in calculations.

4. Results and discussion

a. Statistical performance

The skill scores for the ML algorithms are shown in Table 2. Overall, of the ML algorithms explored, the best performing in most cases was the Ada boosted decision trees, with the exception being with the N20 and P20 method for which RFs performed best. To note, since Ada boost trees most frequently showed the most skill, the remaining figures showing ML algorithm performance and skill are for test set predictions made by

Ada boosted trees. For the same figures based on predictions for the RFs, the reader is referred to supplemental Figs. 6–14 and 15–23 for the SVM results. The skill scores for the test set predictions were fairly low (e.g., CSI < 0.35 and the true skill statistic < 0.35 for all ML algorithms). This was likely due to mistiming of the ML algorithms of FD events by a few pentads up to 2 months, as well as an overemphasis of spatial hotspots and climatological seasonality. The overemphasis on spatial hotspots is seen in the composite mean-difference maps (Fig. 3). Regions that generally experienced more frequent FD events, most notably the Great Plains, were overpredicted by the ML algorithms (these overpredictions are also where the difference in true and predicted labels were statistically significant; Fig. 4). In contrast, the ML algorithms performed better outside of climatological hotspots for some FD identification methods (C23; P20; and L20) and underpredicted FD in other FD identification methods (N20 and O21). This pattern of overpredicting the hotspot regions and potentially underpredicting in other regions was found when the true skill score was plotted in space, as well as in standard performance metrics, such as accuracy, precision, recall, and AUC (Fig. S1), though no temporal trends in ML algorithm performance were detected (Fig. S2).

The last feature to note is the interpretability of the ML algorithms (Fig. 5 for Ada boosted trees). For this study, several approaches to estimating feature importance were examined and compared. The ranking of the most important feature importance varied notably from one FD identification method to another. While some had P as the most important variable and PET as the second most important, others ranked SM as

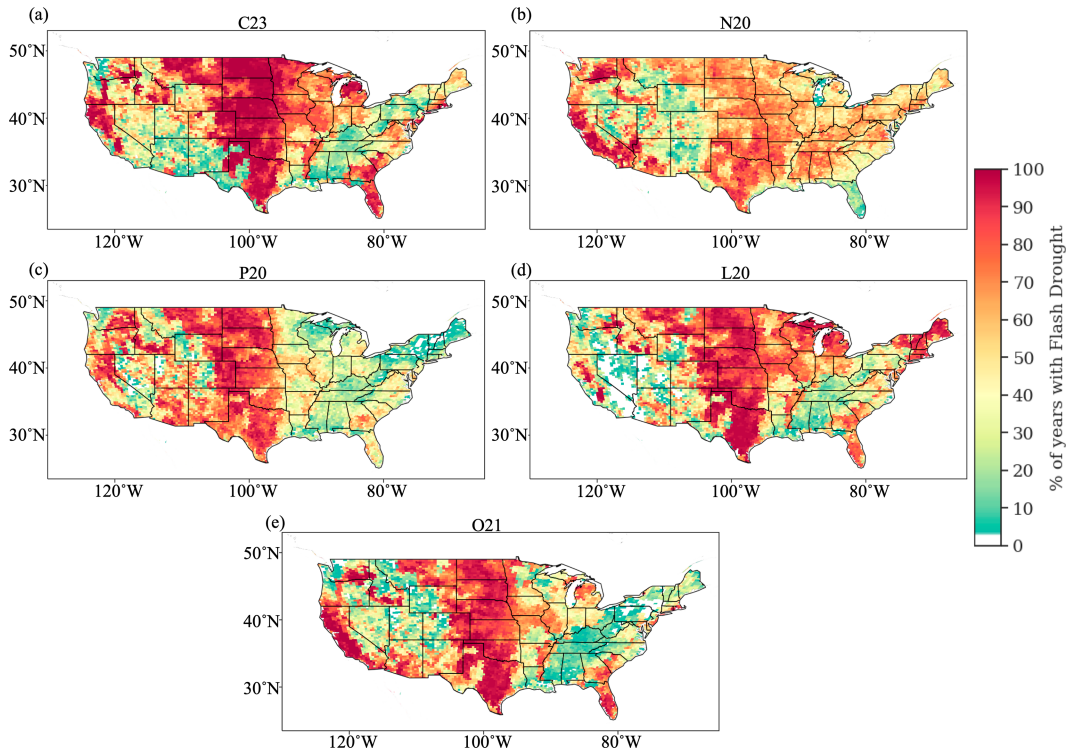


FIG. 6. Predicted frequency climatology of FDs by the Ada boosted trees for the (a) C23, (b) N20, (c) P20, (d) L20, and (e) O21 methods.

the most important. The SHAP feature importance and permutation importance were omitted from the Ada boosted tree analysis as all values were small (10^{-5} for the O21 method, and 10^{-7} and lower for the other methods), and the permutation importance method also struggled with the feature importances as the values were both small (absolute value < 0.02) and statistically identical for each variable. In contrast, the GINI estimations were notably different, with P and SM often the most important variables (as much as twice as important as the second ranking variable). However, the precise ranking of variables and magnitude varied more notably with the GINI method. According to the GINI method, the Ada boosted trees incorporated surface features more readily than the permutation importances would imply.

Furthermore, the ML algorithms learned to predict FD for some identification methods better than others. This finding may be because of the variables the ML algorithms emphasized. Since the L20 and O21 methods used SM to identify FD, these methods had the larger skill score (according to GINI feature importance; Fig. 5). In addition, SESR (for the C23 method) is strongly correlated with SM (via ET and PET used to determine it). This allowed an emphasis on SM for the C23 method to produce a larger skill score. Meanwhile, predictions for the N20 method (which uses SPEI) relied most on P for the predictions. However, the boosted trees were not using PET or T as much, or learning as much of the surface interactions, resulting in a lower skill score. Finally, the P20 method uses EDDI to identify FD, which is not strongly correlated with

variables like P and SM. Focus on these variables for the predictions resulted in more mistiming of FD events and a lower skill score (Table 2). This is more accentuated when compared to the sensitivity analysis in Fig. 1. For example, SM, ET, and PET were among the most important for identifying FD with the C23 method (Fig. 1). Since the Ada boosted trees learned much of, and relied heavily on, SM, ET, PET, and ΔET (Fig. 5), this allowed the boosted trees to obtain the highest skill score for that method (similarly with the O21 method). In contrast, the P20 method can be described by P , SM, PET, and ΔET , whereas the Ada boosted trees incorporated more P , PET, and ΔPET . In short, the Ada boosted trees struggled more noticeably to capture the surface interactions described in the P20 method, resulting in a much lower skill score (similarly for the N20 and L20 methods).

b. Predicted climatology

Apart from performance metrics, we also examined how well the ML algorithms reproduced the statistical patterns of FD in space and time (i.e., the climatology of FD). The predicted frequency climatology (Fig. 6) showed similar patterns to those discussed. Compared to the true labels (Fig. 2), the ML algorithms overpredicted FD events in climatological hotspots. The ML algorithms also highlighted major agricultural regions in the western United States (such as the Central Valley in California and southern Idaho) and overpredicted FD in these locations as well. Given the development of FD via evaporative feedback from the vegetation (i.e., crops

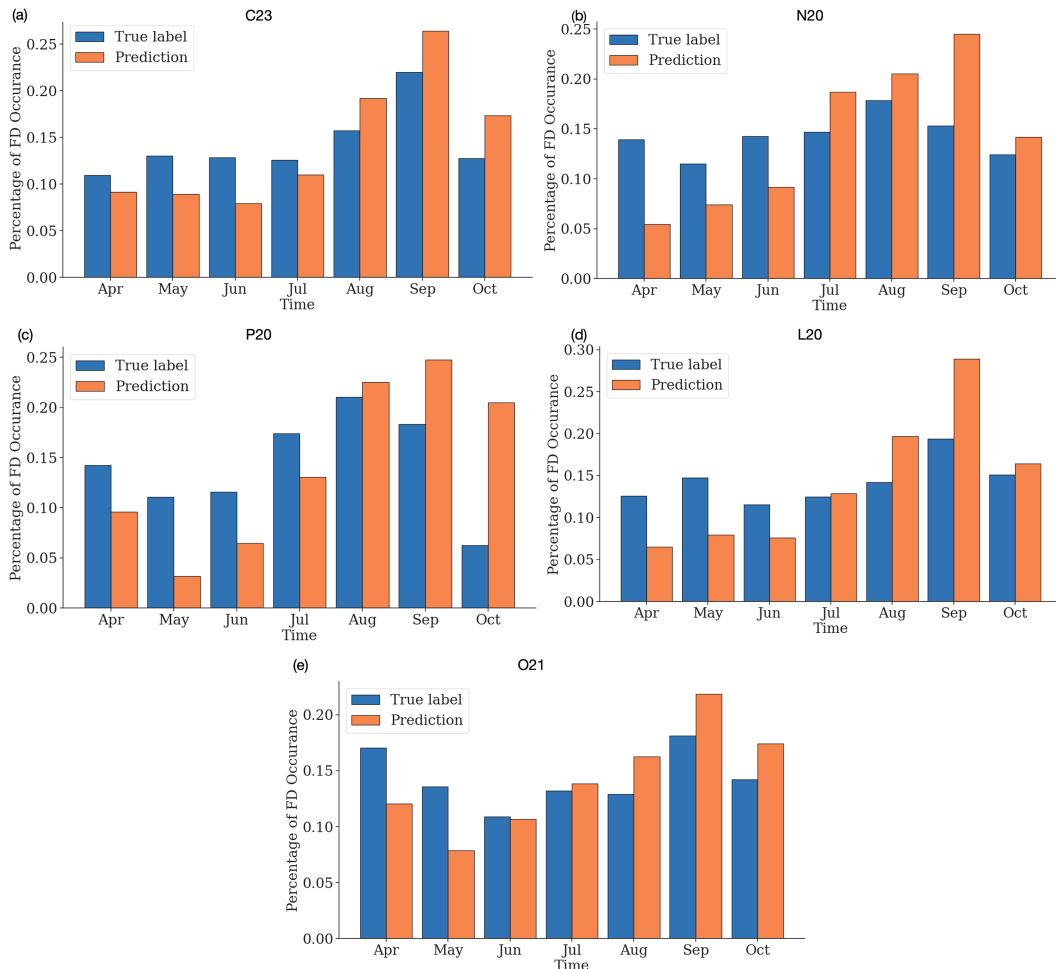


FIG. 7. FD seasonality (percentage of FD occurrence) for true labels (blue) and predicted labels by the Ada boosted trees (orange) for the (a) C23, (b) N20, (c) P20, (d) L20, and (e) O21 methods.

that tend to transpire more than local vegetation), it made sense that these regions would be hotspots for FD, as is the case for other major agricultural regions (e.g., C23; Lowman et al. 2023; Mahto and Mishra 2023; Yuan et al. 2023), but they were not present in the true labels (Fig. 2). However, this suggested that the ML algorithms were able to learn these hotspots (i.e., they learned what spatial regions experience FD more frequently), and they learned to identify FD more frequently for major agricultural regions.

The ML algorithms overemphasized the seasonality of FD events, identifying FD more frequently during favored months and underpredicting FD in months less emphasized in that climatology (Fig. 7). That said, this also shows that ML algorithms were able to learn the seasonality of FD events for some identification methods and overall capture the heterogeneity of these events. The full extent of the ML models' ability to capture FD events, however, is debatable as they tended to focus FD identification during the end of the summer season (August–October) for all methods, rather than just the FD identification methods that have that seasonality. This is a potential issue from the ML algorithms having over relied on one

or two select variables (such as SM, according to the GINI method). For example, the ML algorithms may simply have learned the SM climatology, which decreases in the late growing season (Illston et al. 2004). This would also explain why FD occurrence in some agricultural regions was overpredicted, as these would see larger declines in SM due to intensive agriculture and irrigation than regions dominated by native vegetation. That said, the incorporation of other variables in the C23 and O21 methods, such as ET and ΔET (Fig. 5), allowed them to more accurately identify in the spring and early summer season when those variables naturally increase due to increasing vegetative activity (though they may have over relied on SM, resulting in underpredictions). In terms of the FD coverage in time, the ML algorithms learned the general patterns of FD coverage for the C23, N20, and O21 methods, showing similar spikes and dips as compared to the true labels (i.e., the ML algorithms were able to learn which years contained more extensive FD events and which years were less extensive; Fig. 8). However, learning the magnitude of the FD coverage is a limiting factor in the ML models, though the Ada boosted trees showed some promise with the O21 method.

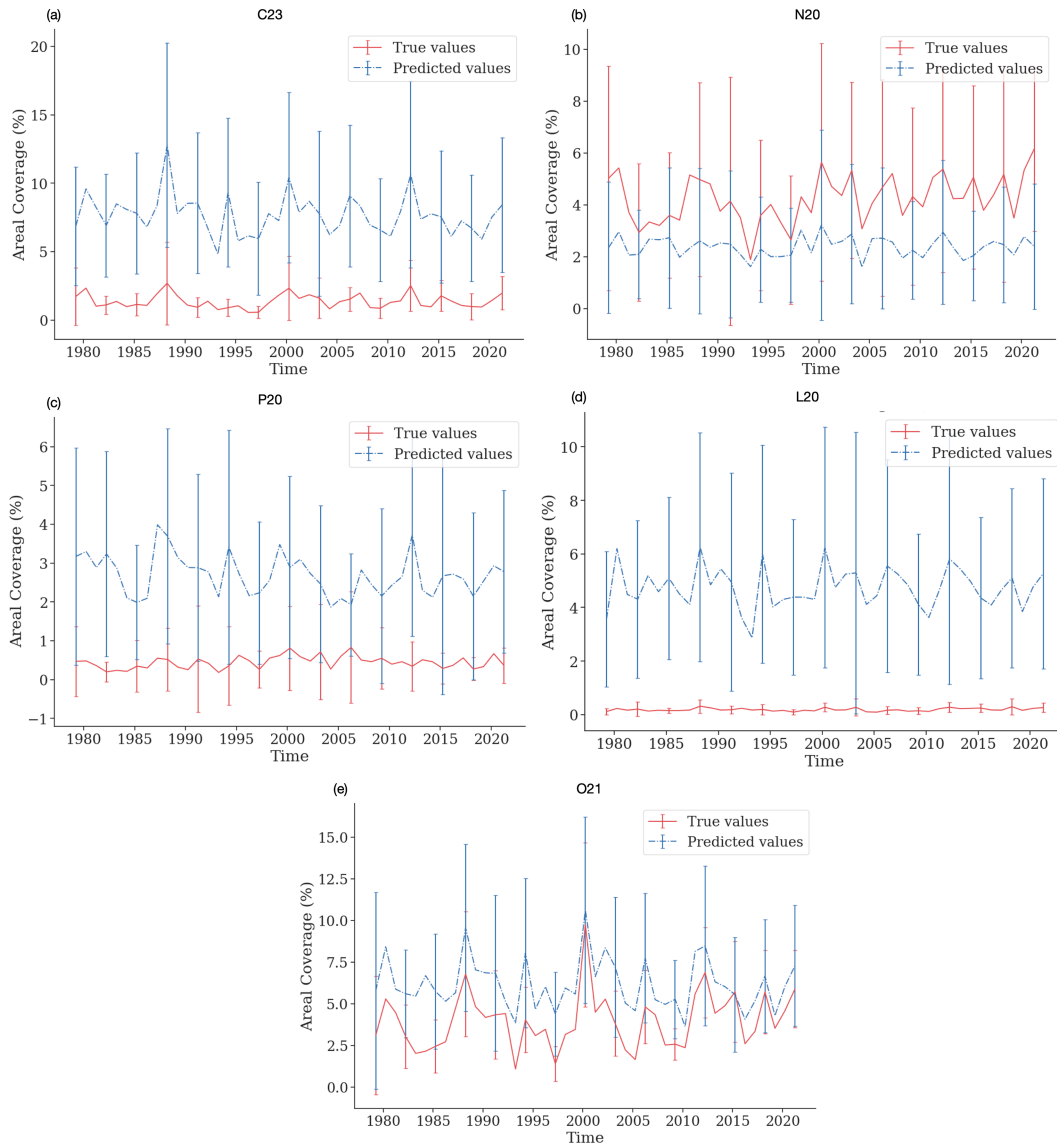


FIG. 8. True (red) and predicted (blue) annual average spatial coverage of FD across the domain for the (a) C23, (b) N20, (c) L20, and (e) O21 methods. Error bars denote one std dev in the annual average. Predicted labels were made by the Ada boosted trees.

c. Case studies

Last, we investigated how well the ML algorithms were able to represent more specific, individual events. To this end, certain FD years, 1988, 2011, and 2012, were chosen due to their high spatial FD coverage and their well-known timing and impacts.

1) 1988

To start, in 1988 prolonged drought conditions initiated in April and May in the Midwest of the United States. Throughout the early growing season, dry conditions propagated further into the Great Plains and Midwest, with some regions undergoing rapid drying as the drought spread into new

regions (Fig. 9). By the end of the summer, the majority of the United States, especially the north-central United States, was in a severe, historic-level drought (Kim and Rhee 2016; Kim et al. 2019). While 1988 is generally viewed as a major drought event, the occurrence of rapid drying with this event has not been discussed much in the literature. FD was recorded for all FD identification methods, though the different ways the FD identification methods examine conditions resulted in different timing and locations for FD, depending on the method. For example, the C23 and N20 methods saw the 1988 drought begin as a FD as a whole (with the end result that most of the domain experienced FD by the end of the growing season), while P20 only recorded FD as the drought propagated through the Great Plains in the early summer.

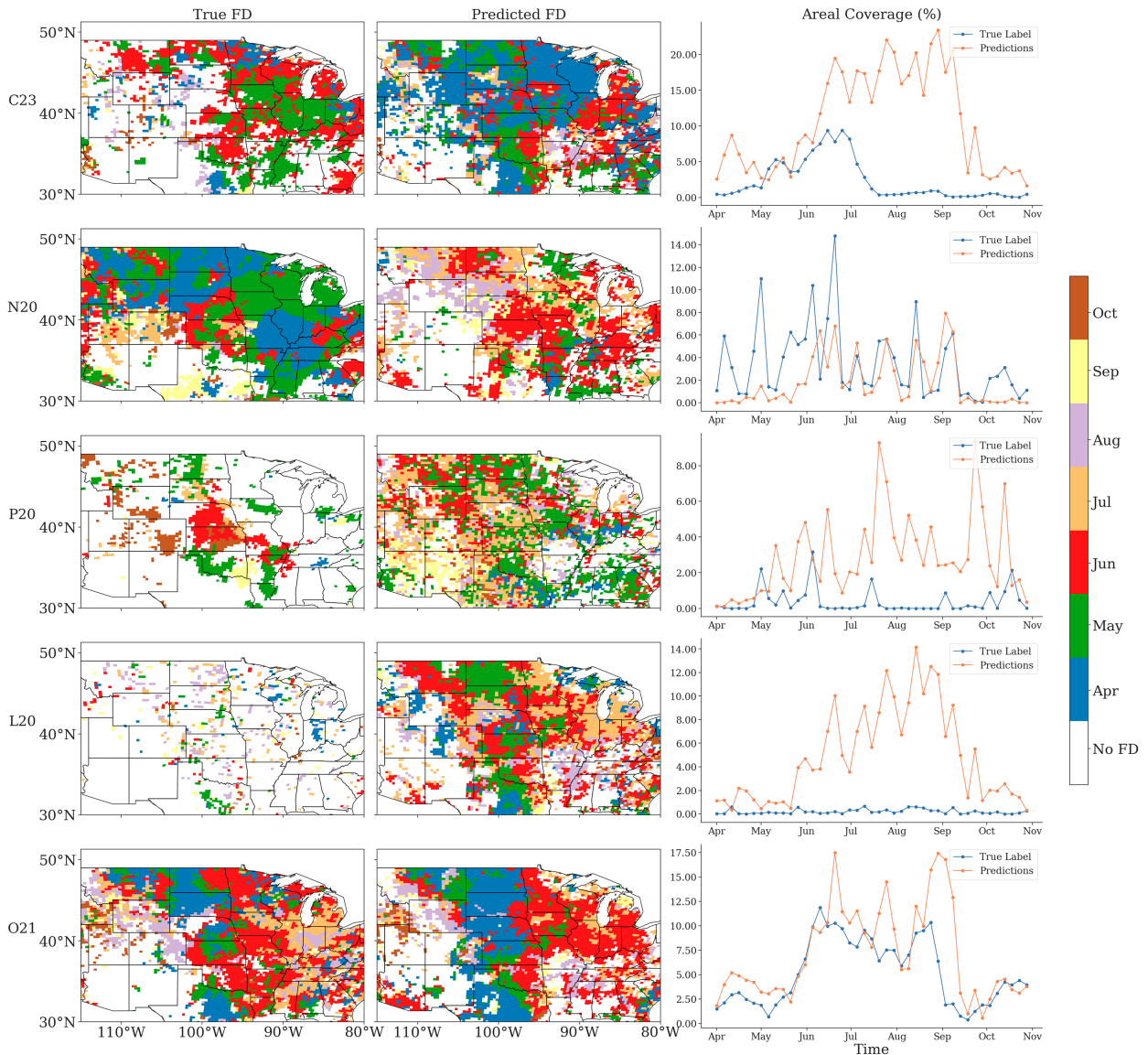


FIG. 9. FD case study for 1988 for each FD identification method. (left) True labels, (center) predicted labels from the test dataset, and (right) true and predicted FD coverage over the domain. FD predictions were made by Ada boosted trees.

Some of the seasonality bias can be seen as the ML algorithms identified FD up to 1–2 months late for some methods (C23 and N20), and prematurely early in another (P20), the bias being toward the late seasonality the ML algorithms predicted as the large decline in SM was not until mid- to late June (Fig. S3). The spatial coverage of the 1988 event was generally well represented by the ML algorithms, with the notable exception that the ML algorithms liked to make continuous patches of FD as compared to discontinuous grids in the L20 method.

2) 2011

During early 2011, abnormally hot and dry conditions started to initiate the development of drought in the Southern Great Plains (Texas and Oklahoma). However, as the spring

and summer continued, the drying increased in rate and spread eastward to much of the southern United States, creating a large FD event that began in April and May and continued into June and July (with much of Texas, Oklahoma, and the southeastern United States settling into drought for the remaining growing season; Fig. 10; Otkin et al. 2013; Ford et al. 2015; McEvoy et al. 2016). For the C23 method, the ML algorithms captured the timing for this event fairly well (since the Ada boosted trees were, but potentially due to the increased PET as well; Fig. 5 and supplemental Fig. 4), but largely overpredicted the coverage of FD initiation in May. For the O21 method, the coverage of the FD event in the Southern Great Plains was well represented, but the FD extended further north than verification (again, a bias toward climatology). Moreover, the ML

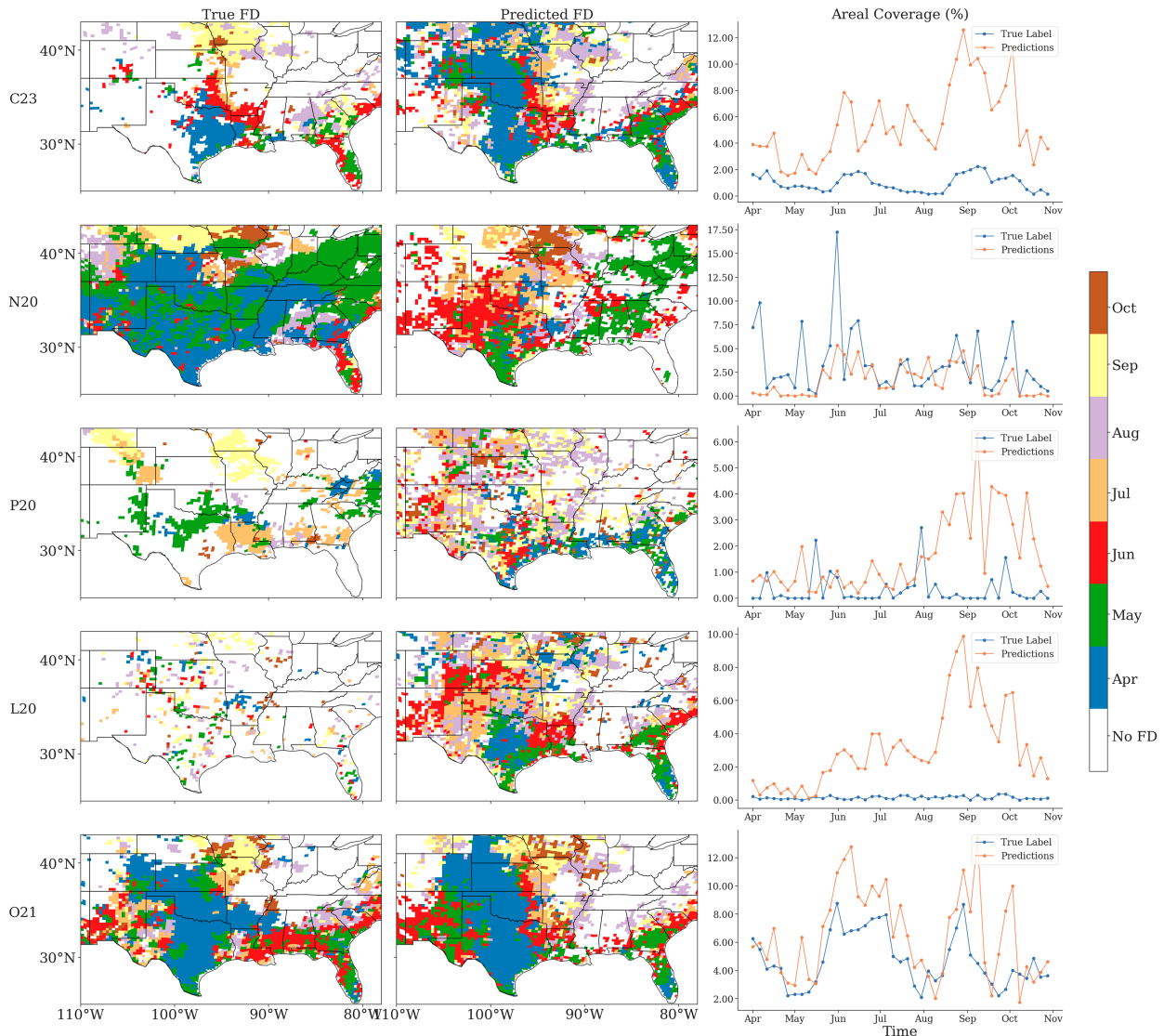


FIG. 10. FD case study for 2011 for each FD identification method. (left) True labels, (center) predicted labels from the test dataset, and (right) true and predicted FD coverage over the domain. FD predictions were made by Ada boosted trees.

algorithms identified FD up to 2 months late in many locations. For the remaining FD identification methods, the ML algorithms either overpredicted (P20 and L20) or underpredicted (N20) FD coverage, and they mistimed FD occurrence by 1–2 months. Given the Ada boosted tree reliance on SM for the L20 method, and P for the P20 and N20 methods (Fig. 5), these late predictions of FD may be due to a combination of a minimum SM in late August and September, and a sharper decrease in P in September than during the rest of the event (for the Ada boosted trees predicting the L20 method, ΔSM was less important than P ; supplemental Figs. 4b,e).

3) 2012

In 2012, parts of eastern Arkansas and southern Missouri dried out at time scales faster than normal. As the spring

continued, the drought propagated north and westward, causing regions that were above-normal moisture conditions and not expecting drought conditions to rapidly deteriorate into drought. The result was widespread desiccation and crop loss (Fig. 11; Ford et al. 2015; Otkin et al. 2016; McEvoy et al. 2016; Basara et al. 2019). The end result saw drought settle into much of the central United States, creating one of the more famous and heavily examined examples for FD. For this case study, the ML algorithms generally identified FD development a month later than truth (in June and July, when the SM first reached its minimum; Fig. S5), rather than May, with the exception being for the C23 method, which identified FD early. The ML algorithms also had a harder time predicting the aerial coverage for this FD event, as they often overpredicted the extent of the FD for the C23, P20, and L20 methods.

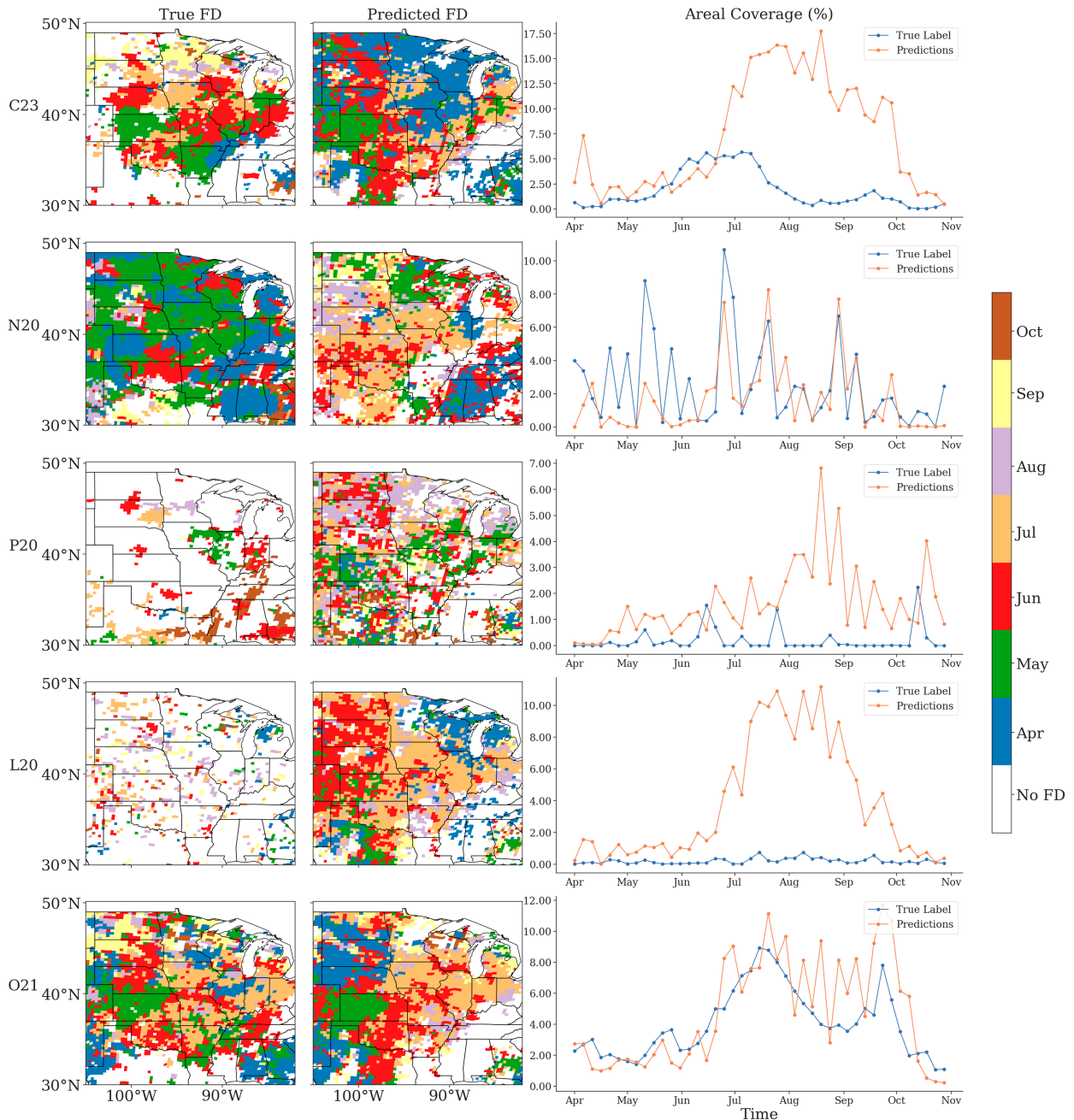


FIG. 11. FD case study for 2012 for each FD identification method. (left) True labels, (center) predicted labels from the test dataset, and (right) true and predicted FD coverage over the domain. FD predictions were made by Ada boosted trees.

However, the ML algorithms were able to represent the coverage of the FD relatively well for the **N20** and **O21** methods.

5. Summary and conclusions

This study contributed an initial investigation of FD identification using three standard ML algorithms: RFs, Ada boosted decision trees, and SVMs. The models were trained on five different FD identification methods individually, using key

variables known to drive FD events. The results found that the most skillful ML algorithm (Ada boosted trees) showed promise in being able to represent FD events across multiple identification methods. The ML algorithms were able to learn which spatial regions experienced the most frequent FD events, as well as some degree of seasonality of the FD events. Case studies showed that the spatial coverage of FD events was often (though not always) well represented. In addition, predictions for the **C23** and **O21** methods consistently showed well-structured spatial

patterns. This is likely due to the fact that the ML algorithms were better able to learn which variables describe those methods, and thence represent surface interactions involved in FD more accurately. However, feature importances showed that the ML algorithms were over relying on select variables (SM, P , and PET), and sometimes the wrong variables when compared to the sensitivity analysis, and the ML algorithms may have learned climatologies and patterns in those variables as opposed to synthesizing the information to identify FD. As a result, the ML algorithms struggled to learn the timing of FD events and favored a specific seasonality (i.e., late summer into fall), due to both negative precipitation and soil moisture anomalies being most prominent at that time. The ML algorithms also often over-emphasized hotspot regions, creating numerous false positives in those areas.

The second part of this study examined more complicated algorithms in neural networks, which better represented FD but are harder to interpret. Given this study focused on standard ML techniques, future studies are recommended to investigate the ability of deep learning models to predict FD, especially given their general improvement over standard ML to represent droughts (Mohamadi et al. 2020; Dikshit and Pradhan 2021; Proadhan et al. 2021). Future studies should also focus on more specific algorithms (such as Ada boosted trees) and methods to improve their performance, such as physics-constrained models or augmenting the input data with climatological variables such as ENSO and the Atlantic multidecadal oscillation, or taking other such approaches so that models learn to incorporate other variables and improve their predictions, as well as apply other explainable AI tools to make the models more interpretable, explainable, and trustworthy.

Acknowledgments. This work was part of the USDA Southern Great Plains Climate Hub, National Science Foundation, and NSF AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography (AI2ES). This material was based upon work supported by the National Science Foundation under Grant OIA-1946093 and Grant ICER-2019758. We are grateful for the constructive and insightful feedback and suggestions from three anonymous reviewers to improve the quality of this article.

Data availability statement. Code work and processed machine learning data for this project can be found on GitHub: https://github.com/Rarell/ML_and_FD. Larger datasets, such as raw data and machine learning results, are available upon request.

REFERENCES

- AghaKouchak, A., and Coauthors, 2022: Status and prospects for drought forecasting: Opportunities in artificial intelligence and hybrid physical–statistical forecasting. *Philos. Trans. Roy. Soc.*, **A380**, 20210288, <https://doi.org/10.1098/rsta.2021.0288>.
- Alencar, P. H. L., and E. N. Paton, 2022: How do we identify flash droughts? A case study in central European croplands. *Hydrol. Res.*, **53**, 1150–1165, <https://doi.org/10.2166/nh.2022.003>.
- Au, Q., J. Herbringer, C. Stachl, B. Bischl, and G. Casalicchio, 2022: Grouped feature importance and combined features effect plot. *Data Min. Knowl. Discovery*, **36**, 1401–1450, <https://doi.org/10.1007/s10618-022-00840-5>.
- Basara, J. B., J. I. Christian, R. A. Wakefield, J. A. Otkin, E. H. Hunt, and D. P. Brown, 2019: The evolution, propagation, and spread of flash drought in the central United States during 2012. *Environ. Res. Lett.*, **14**, 084025, <https://doi.org/10.1088/1748-9326/ab2cc0>.
- Buitinck, L., and Coauthors, 2013: API design for machine learning software: Experiences from the scikit-learn project. arXiv, 1309.0238, <https://doi.org/10.48550/arXiv.1309.0238>.
- Chase, R. J., D. R. Harrison, A. Burke, G. M. Lackmann, and A. McGovern, 2022: A machine learning tutorial for operational meteorology. Part I: Traditional machine learning. *Wea. Forecasting*, **37**, 1509–1529, <https://doi.org/10.1175/WAF-D-22-0070.1>.
- Chen, L. G., J. Gottschalk, A. Hartman, D. Miskus, R. Tinker, and A. Artusa, 2019: Flash drought characteristics based on U.S. drought monitor. *Atmosphere*, **10**, 498, <https://doi.org/10.3390/atmos10090498>.
- Christian, J. I., J. B. Basara, J. A. Otkin, and E. D. Hunt, 2019a: Regional characteristics of flash droughts across the United States. *Environ. Res. Commun.*, **1**, 125004, <https://doi.org/10.1088/2515-7620/ab50ca>.
- , —, —, —, R. A. Wakefield, P. X. Flanagan, and X. Xiao, 2019b: A methodology for flash drought identification: Application of flash drought frequency across the United States. *J. Hydrometeorol.*, **20**, 833–846, <https://doi.org/10.1175/JHM-D-18-0198.1>.
- , —, E. D. Hunt, J. A. Otkin, J. C. Furtado, V. Mishra, X. Xiao, and R. M. Randall, 2021: Global distribution, trends, and drivers of flash drought occurrence. *Nat. Commun.*, **12**, 6330, <https://doi.org/10.1038/s41467-021-26692-z>.
- , —, L. E. L. Lowman, X. Xiao, D. Mescheske, and Y. Zhou, 2022: Flash drought identification from satellite-based land surface water index. *Remote Sens. Appl.*, **26**, 100770, <https://doi.org/10.1016/j.rsase.2022.100770>.
- , and Coauthors, 2023: Global projections of flash drought show increased risk in a warming climate. *Commun. Earth Environ.*, **4**, 165, <https://doi.org/10.1038/s43247-023-00826-1>.
- Cui, W., X. Dong, B. Xi, and Z. Feng, 2021: Climatology of linear mesoscale convective system morphology in the United States based on random forests method. *J. Climate*, **34**, 7257–7276, <https://doi.org/10.1175/JCLI-D-20-0862.1>.
- Dhyani, Y., and R. J. Pandya, 2021: Deep learning oriented satellite remote sensing for drought and prediction in agriculture. *2021 IEEE 18th India Council Int. Conf. (INDICON)*, Guwahati, India, Institute of Electrical and Electronics Engineers, 1–5, <https://doi.org/10.1109/indicon52576.2021.9691608>.
- Dikshit, A., and B. Pradhan, 2021: Interpretable and explainable AI (XAI) model for spatial drought prediction. *Sci. Total Environ.*, **801**, 149797, <https://doi.org/10.1016/j.scitotenv.2021.149797>.
- , —, and A. M. Alamri, 2020: Temporal hydrological drought index forecasting for New South Wales, Australia using machine learning approaches. *Atmosphere*, **11**, 585, <https://doi.org/10.3390/atmos11060585>.
- , —, A. Huete, and H.-J. Park, 2022: Spatial based drought assessment: Where are we heading? A review on the current status and future. *Sci. Total Environ.*, **844**, 157239, <https://doi.org/10.1016/j.scitotenv.2022.157239>.
- Edris, S. G., A. McGovern, J. B. Basara, J. I. Christian, J. C. Furtado, H. Olayiwola, and X. Xiao, 2025: Evaluation of flash drought identification with machine learning techniques. Part II:

- Neural network algorithms. *Artif. Intell. Earth Syst.*, **4**, e240069, <https://doi.org/10.1175/AIES-D-24-0069.1>.
- Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarpley, 2003: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *J. Geophys. Res.*, **108**, 8851, <https://doi.org/10.1029/2002JD003296>.
- Flora, M. L., C. K. Potvin, A. McGovern, and S. Handler, 2024: A machine learning explainability tutorial for atmospheric sciences. *Artif. Intell. Earth Syst.*, **3**, e230018, <https://doi.org/10.1175/AIES-D-23-0018.1>.
- Ford, T. W., and C. F. Labosier, 2017: Meteorological conditions associated with the onset of flash drought in the eastern United States. *Agric. For. Meteorol.*, **247**, 414–423, <https://doi.org/10.1016/j.agrformet.2017.08.031>.
- , D. B. McRoberts, S. M. Quiring, and R. E. Hall, 2015: On the utility of in situ soil moisture observations for flash drought early warning in Oklahoma, USA. *Geophys. Res. Lett.*, **42**, 9790–9798, <https://doi.org/10.1002/2015GL066600>.
- Foroumandi, E., V. Nourani, J. Jeanne Huang, and H. Moradkhani, 2023: Drought monitoring by downscaling GRACE-derived terrestrial water storage anomalies: A deep learning approach. *J. Hydrol.*, **616**, 128838, <https://doi.org/10.1016/j.jhydrol.2022.128838>.
- , K. Gavahi, and H. Moradkhani, 2024: Generative adversarial network for real-time flash drought monitoring: A deep learning study. *Water Resour. Res.*, **60**, e2023WR035600, <https://doi.org/10.1029/2023WR035600>.
- Ganguli, P., and M. J. Reddy, 2014: Ensemble prediction of regional droughts using climate inputs and the SVM-copula approach. *Hydrol. Processes*, **28**, 4989–5009, <https://doi.org/10.1002/hyp.9966>.
- Gavahi, K., P. Abbaszadeh, and H. Moradkhani, 2022: How does precipitation data influence the land surface data assimilation for drought monitoring? *Sci. Total Environ.*, **831**, 154916, <https://doi.org/10.1016/j.scitotenv.2022.154916>.
- Hao, Z., V. P. Singh, and Y. Xia, 2018: Seasonal drought prediction: Advances, challenges, and future prospects. *Rev. Geophys.*, **56**, 108–141, <https://doi.org/10.1002/2016RG000549>.
- Hastie, T., J. H. Friedman, and R. T. Tibshirani, 2009: *The Elements of Statistical Learning*. 2nd ed. Springer, 745 pp.
- Hobbins, M. T., A. Wood, D. J. McEvoy, J. L. Huntington, C. Morton, M. Anderson, and C. Hain, 2016: The evaporative demand drought index. Part I: Linking drought evolution to variations in evaporative demand. *J. Hydrometeorol.*, **17**, 1745–1761, <https://doi.org/10.1175/JHM-D-15-0121.1>.
- Hsieh, W. W., 2022: Evolution of machine learning in environmental science—A perspective. *Environ. Data Sci.*, **1**, e3, <https://doi.org/10.1017/eds.2022.2>.
- Illston, B. G., J. B. Basara, and K. C. Crawford, 2004: Seasonal to interannual variations of soil moisture measured in Oklahoma. *Int. J. Climatol.*, **24**, 1883–1896, <https://doi.org/10.1002/joc.1077>.
- Kaur, A., and S. K. Sood, 2020: Deep learning based drought assessment and prediction framework. *Ecol. Inf.*, **57**, 101067, <https://doi.org/10.1016/j.ecoinf.2020.101067>.
- Kim, D., and J. Rhee, 2016: A drought index based on actual evapotranspiration from the Bouchet hypothesis. *Geophys. Res. Lett.*, **43**, 10 277–10 285, <https://doi.org/10.1002/2016GL070302>.
- , W.-S. Lee, S. T. Kim, and J. A. Chun, 2019: Historical drought assessment over the contiguous United States using the generalized complementary principle of evapotranspiration. *Water Resour. Res.*, **55**, 6244–6267, <https://doi.org/10.1029/2019WR024991>.
- Krishnamurthy R, P. K., J. B. Fisher, R. J. Choularton, and P. M. Kareiva, 2022: Anticipating drought-related food security changes. *Nat. Sustainability*, **5**, 956–964, <https://doi.org/10.1038/s41893-022-00962-0>.
- Lesk, C., W. Anderson, A. Rigden, O. Coast, J. Jägermeyr, S. McDermid, K. F. Davis, and M. Konar, 2022: Compound heat and moisture extreme impacts on global crop yields under climate change. *Nat. Rev. Earth Environ.*, **3**, 872–889, <https://doi.org/10.1038/s43017-022-00368-8>.
- Li, J., Z. Wang, X. Wu, J. Chen, S. Guo, and Z. Zhang, 2020: A new framework for tracking flash drought events in space and time. *CATENA*, **194**, 104763, <https://doi.org/10.1016/j.catena.2020.104763>.
- Li, W., and Coauthors, 2022: Widespread increasing vegetation sensitivity to soil moisture. *Nat. Commun.*, **13**, 3959, <https://doi.org/10.1038/s41467-022-31667-9>.
- Lisonbee, J., M. Woloszyn, and M. Skumanich, 2022: Making sense of flash drought: Definitions, indicators, and where we go from here. *J. Appl. Serv. Climatol.*, **2021**, 1–19, <https://doi.org/10.46275/JOASC.2021.02.001>.
- Liu, W., Y. Huang, and H. Wang, 2024: Effective deep learning seasonal prediction model for summer drought over China. *Earth's Future*, **12**, e2023EF004409, <https://doi.org/10.1029/2023ef004409>.
- Liu, Y., Y. Zhu, L. Zhang, L. Ren, F. Yuan, X. Yang, and S. Jiang, 2020: Flash droughts characterization over China: From a perspective of the rapid intensification rate. *Sci. Total Environ.*, **704**, 135373, <https://doi.org/10.1016/j.scitotenv.2019.135373>.
- Lorenz, D., J. A. Otkin, and B. F. Zaitchik, 2024: Improving sub-seasonal soil moisture and evaporative stress index forecasts through machine learning: The role of initial land state versus dynamical model output. *38th Conf. on Hydrology*, Baltimore, MD, Amer. Meteor. Soc., 540, <https://ams.confex.com/ams/104ANNUAL/meetingapp.cgi/Paper/434746>.
- Lowman, L. E. L., J. I. Christian, and E. D. Hunt, 2023: How land surface characteristics influence the development of flash drought through the drivers of soil moisture and vapor pressure deficit. *J. Hydrometeorol.*, **24**, 1395–1415, <https://doi.org/10.1175/JHM-D-22-0158.1>.
- Lundberg, S. M., and S.-I. Lee, 2017: A unified approach to interpreting model predictions. *Proc. 31st Int. Conf. on Neural Information Processing Systems*, Long Beach, CA, Curran Associates, Inc., 4768–4777, <https://dl.acm.org/doi/pdf/10.5555/3295222.3295230>.
- Mahto, S. S., and V. Mishra, 2023: Increasing risk of simultaneous occurrence of flash drought in major global croplands. *Environ. Res. Lett.*, **18**, 044044, <https://doi.org/10.1088/1748-9326/ac8ed>.
- McEvoy, D. J., J. L. Huntington, M. T. Hobbins, A. Wood, C. Morton, M. Anderson, and C. Hain, 2016: The evaporative demand drought index. Part II: CONUS-wide assessment against common drought indicators. *J. Hydrometeorol.*, **17**, 1763–1779, <https://doi.org/10.1175/JHM-D-15-0122.1>.
- McGovern, A., R. Lagerquist, D. John Gagne II, G. E. Jergensen, K. L. Elmore, C. R. Homeyer, and T. Smith, 2019: Making the black box more transparent: Understanding the physical implications of machine learning. *Bull. Amer. Meteor. Soc.*, **100**, 2175–2199, <https://doi.org/10.1175/BAMS-D-18-0195.1>.
- McKee, T. B., N. J. Doesken, and J. Kleist, 1993: The relationship of drought frequency and duration to time scales. *Proc. Eighth Conf. on Applied Climatology*, Anaheim, CA, Amer. Meteor. Soc., 179–183, <https://www.droughtmanagement.info/>

- literature/AMS_Relationship_Drought_Frequency_Duration_Time_Scales_1993.pdf.
- , —, and —, 1995: Drought monitoring with multiple time scales. *Proc. Ninth Conf. on Applied Climatology*, Dallas, TX, Amer. Meteor. Soc., 233–236, https://iridl.ldeo.columbia.edu/documentation/spi/McKee_et_al_1995.pdf.
- Mesinger, F., and Coauthors, 2006: North American regional reanalysis. *Bull. Amer. Meteor. Soc.*, **87**, 343–360, <https://doi.org/10.1175/BAMS-87-3-343>.
- Mohamadi, S., and Coauthors, 2020: Zoning map for drought prediction using integrated machine learning models with a no-madic people optimization algorithm. *Nat. Hazards*, **104**, 537–579, <https://doi.org/10.1007/s11069-020-04180-9>.
- Noguera, I., F. Domínguez-Castro, and S. M. Vicente-Serrano, 2020: Characteristics and trends of flash droughts in Spain, 1961–2018. *Ann. N. Y. Acad. Sci.*, **1472**, 155–172, <https://doi.org/10.1111/nyas.14365>.
- Osman, M., B. F. Zaitchik, H. S. Badr, J. I. Christian, T. Tadesse, J. A. Otkin, and M. C. Anderson, 2021: Flash drought onset over the contiguous United States: Sensitivity of inventories and trends to quantitative definitions. *Hydrol. Earth Syst. Sci.*, **25**, 565–581, <https://doi.org/10.5194/hess-25-565-2021>.
- Otkin, J. A., M. C. Anderson, C. Hain, I. E. Mladenova, J. B. Basara, and M. Svoboda, 2013: Examining rapid onset drought development using the thermal infrared-based evaporative stress index. *J. Hydrometeorol.*, **14**, 1057–1074, <https://doi.org/10.1175/JHM-D-12-0144.1>.
- , and Coauthors, 2016: Assessing the evolution of soil moisture and vegetation conditions during the 2012 United States flash drought. *Agric. For. Meteorol.*, **218–219**, 230–242, <https://doi.org/10.1016/j.agrformet.2015.12.065>.
- , M. Svoboda, E. D. Hunt, T. W. Ford, M. C. Anderson, C. Hain, and J. B. Basara, 2018: Flash droughts: A review and assessment of the challenges imposed by rapid-onset droughts in the United States. *Bull. Amer. Meteor. Soc.*, **99**, 911–919, <https://doi.org/10.1175/BAMS-D-17-0149.1>.
- , and Coauthors, 2021: Development of a flash drought intensity index. *Atmosphere*, **12**, 741, <https://doi.org/10.3390/atmos12060741>.
- Palmer, W. C., 1965: *Meteorological Drought*. Vol. 30. U.S. Department of Commerce, Weather Bureau, 58 pp., <https://books.google.com/books?id=kyYZgnEk-L8C>.
- Park, S., J. Im, E. Jang, and J. Rhee, 2016: Drought assessment and monitoring through blending of multi-sensor indices using machine learning approaches for different climate regions. *Agric. For. Meteorol.*, **216**, 157–169, <https://doi.org/10.1016/j.agrformet.2015.10.011>.
- Pedregosa, F., and Coauthors, 2011: Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.*, **12**, 2825–2830.
- Pendergrass, A. G., and Coauthors, 2020: Flash droughts present a new challenge for subseasonal-to-seasonal prediction. *Nat. Climate Change*, **10**, 191–199, <https://doi.org/10.1038/s41558-020-0709-0>.
- Peters-Lidard, C. D., D. M. Mocko, L. Su, D. P. Lettenmaier, P. Gentile, and M. Barlage, 2021: Advances in land surface models and indicators for drought monitoring and prediction. *Bull. Amer. Meteor. Soc.*, **102**, E1099–E1122, <https://doi.org/10.1175/BAMS-D-20-0087.1>.
- Prodhan, F. A., and Coauthors, 2021: Deep learning for monitoring agricultural drought in South Asia using remote sensing data. *Remote Sens.*, **13**, 1715, <https://doi.org/10.3390/rs13091715>.
- Qing, Y., S. Wang, B. C. Ancell, and Z.-L. Yang, 2022: Accelerating flash droughts induced by the joint influence of soil moisture depletion and atmospheric aridity. *Nat. Commun.*, **13**, 1139, <https://doi.org/10.1038/s41467-022-28752-4>.
- Rahmati, O., and Coauthors, 2020: Machine learning approaches for spatial modeling of agricultural droughts in the South-East region of Queensland Australia. *Sci. Total Environ.*, **699**, 134230, <https://doi.org/10.1016/j.scitotenv.2019.134230>.
- Rhee, J., and J. Im, 2017: Meteorological drought forecasting for ungauged areas based on machine learning: Using long-range climate forecast and remote sensing data. *Agric. For. Meteorol.*, **237–238**, 105–122, <https://doi.org/10.1016/j.agrformet.2017.02.011>.
- Scott, E. L., R. D. Leeper, and D. A. Coates, 2024: Sensitivity of flash drought prediction based on inter- and intra-annual modes of variability to drought index and region. *38th Conf. on Hydrology*, Baltimore, MD, Amer. Meteor. Soc., 11A.1, <https://ams.confex.com/ams/104ANNUAL/meetingapp.cgi/Paper/439776>.
- Shah, J., and Coauthors, 2022: Increasing footprint of climate warming on flash droughts occurrence in Europe. *Environ. Res. Lett.*, **17**, 064017, <https://doi.org/10.1088/1748-9326/ac6888>.
- Svoboda, M., and Coauthors, 2002: The Drought Monitor. *Bull. Amer. Meteor. Soc.*, **83**, 1181–1190, <https://doi.org/10.1175/1520-0477-83.8.1181>.
- Tyagi, S., X. Zhang, D. Saraswat, S. Sahany, S. K. Mishra, and D. Niyogi, 2022: Flash drought: Review of concept, prediction and the potential for machine learning, deep learning methods. *Earth's Future*, **10**, e2022EF002723, <https://doi.org/10.1029/2022EF002723>.
- Vicente-Serrano, S. M., S. Beguería, and J. I. López-Moreno, 2010: A multiscale drought index sensitive to global warming: The standardized precipitation evapotranspiration index. *J. Climate*, **23**, 1696–1718, <https://doi.org/10.1175/2009JCLI2909.1>.
- Wells, N., S. Goddard, and M. J. Hayes, 2004: A self-calibrating Palmer Drought Severity Index. *J. Climate*, **17**, 2335–2351, [https://doi.org/10.1175/1520-0442\(2004\)017<2335:ASPSI>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<2335:ASPSI>2.0.CO;2).
- Woodcock, F., 1976: The evaluation of yes/no forecasts for scientific and administrative purposes. *Mon. Wea. Rev.*, **104**, 1209–1214, [https://doi.org/10.1175/1520-0493\(1976\)104<1209:TEOYFF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1976)104<1209:TEOYFF>2.0.CO;2).
- Yuan, X., L. Wang, P. Wu, P. Ji, J. Sheffield, and M. Zhang, 2019: Anthropogenic shift towards higher risk of flash drought over China. *Nat. Commun.*, **10**, 4661, <https://doi.org/10.1038/s41467-019-12692-7>.
- , Y. Wang, P. Ji, P. Wu, J. Sheffield, and J. A. Otkin, 2023: A global transition to flash droughts under climate change. *Science*, **380**, 187–191, <https://doi.org/10.1126/science.abn6301>.
- Zhang, L., and Coauthors, 2022: Analysis of flash droughts in China using machine learning. *Hydrol. Earth Syst. Sci.*, **26**, 3241–3261, <https://doi.org/10.5194/hess-26-3241-2022>.