

Evaluation of Flash Drought Identification with Machine Learning Techniques.

Part II: Neural Network 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^c

^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: Flash droughts (FDs) describe the rapid onset of drought conditions and can have numerous impacts on agriculture and ecosystems due to rapid desiccation of the landscape. Recent studies into FDs have shown that the surface variables that drive these events include soil moisture, evapotranspiration, and potential evapotranspiration. However, studies that use machine learning (ML) to examine flash drought are still limited, despite the skill they have shown at identifying seasonal-scale droughts. This study is the second part in an investigation of FD representation and prediction using ML methods. Part I investigated traditional ML techniques, while this part investigated the ability of neural network–based methods to identify FD events. Key variables (soil moisture, evaporation, potential evaporation, temperature, and precipitation) were incorporated to train the ML models. Artificial neural networks (ANNs) struggled to identify FDs because they could not learn the surface interactions that characterize FD events. Convolutional U-networks were more capable of learning FD patterns but still overemphasized spatial patterns. Recurrent neural networks (RNNs), using a long short-term memory layer, most frequently had the highest skill score when predicting FD patterns and reduced much of the overprediction of FDs that occurred with the standard ML algorithms. RNNs could also recreate regions where FDs were more frequent as well as the seasonality climatology of FDs better than standard algorithms for some FD identification methods (but struggled more with the seasonality for other identification methods). However, the RNNs still struggled with the precise timing of FD events. Overall, RNNs showed some skill and promise in representing FDs on climatological time scales.

KEYWORDS: Drought; Atmosphere–land interaction; Hydrology; Deep learning; Machine learning; Neural networks

1. Introduction

Climatological dry extremes have received a great deal of attention because they are one of the most impactful forms of natural disasters (Hao et al. 2018; Dikshit et al. 2020; AghaKouchak et al. 2022), with some of the less resilient regions having impacts as large as food crises (Dikshit et al. 2020; Krishnamurthy R et al. 2022; Lesk et al. 2022), and they occur regularly, near a yearly time scale in some locations (Peters-Lidard et al. 2021). In particular, interest has grown in the last decade in a subset of drought events, termed flash drought (FD) for their rapid intensification of drought conditions (on the order of a month; Svoboda et al. 2002; Otkin et al. 2018; Lisonbee et al. 2021; Tyagi et al. 2022). FD events have proved more difficult and complex to identify in real time and predict due to their rapid intensification on the subseasonal time scale (2-week–3-month time scale), seasonality (months in which FD development is more favorable or more likely), and favoring certain regions (often agricultural regions, or

energy or moisture gradients; Otkin et al. 2018; Christian et al. 2019a; Tyagi et al. 2022). Advances in FD studies have found a feedback process that involves increased atmospheric demand via precipitation deficits combined with factors such as increased temperatures, decreased humidity, and increased solar insolation. Subsequently, surface soils and vegetation become desiccated due to increased evapotranspiration (ET), which promotes increased sensible heat flux and increased atmospheric demand (Otkin et al. 2018; Christian et al. 2019a, 2021, 2022, 2023; Osman et al. 2021; Otkin et al. 2021; Tyagi et al. 2022). The key variables that drive the feedback process include soil moisture, evapotranspiration, and potential ET (PET). Following Otkin et al. (2018), studies have introduced methods of identifying and quantifying FD events; however, no single definition has been accepted by the community (Lisonbee et al. 2021; Tyagi et al. 2022). Many of these FD identification methods follow similar principles; i.e., a proxy for FD decreases by a certain threshold within a 3–6-week period (to quantify the rapid intensification), and then, the proxy falls below the 20th percentile (to identify the drought). As such, these methods find similar climatological hotspots (regions where FDs occur more frequently or are more likely to occur) for FD activity (e.g., the U.S. Great Plains; Christian et al. 2019b, 2021; Otkin et al. 2021). Ultimately, the definitions are different via some of the finer details (Osman et al. 2021; Alencar and Paton 2022).

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/AIES-D-24-0069.s1>.

Corresponding author: Stuart Edris, sgedris@ou.edu

Another new feature in drought studies gaining attention over the past decade is the use of machine learning (ML) techniques to find and predict drought events (Mohamadi et al. 2020; Dikshit et al. 2022; Hsieh 2022). These ML techniques fall under the category of statistical models and have been employed in recent studies to predict droughts via climatological and hydrologic metrics that change on the order of months to years (e.g., Ganguli and Reddy 2014; Rhee and Im 2017; Mohamadi et al. 2020; Dikshit and Pradhan 2021). Many ML studies have focused on techniques such as random forests (RFs), support vector machines (SVMs), and boosted trees (Mohamadi et al. 2020; Dikshit et al. 2022). Other studies have explored neural network (NN)-based techniques to estimate drought indices and have utilized artificial neural networks (ANNs) often one layer deep. However, some variations, such as the adaptive neuro-fuzzy interface system (ANFIS), have also become popular (e.g., Mohamadi et al. 2020). Overall, these NN methods frequently outperform other ML techniques. For example, Dikshit and Pradhan (2021) found that a single-layer ANN model reduced the generalization error and better represented the temporal trends in standardized precipitation evaporation index (SPEI) than SVMs and RFs. Likewise, Mohamadi et al. (2020) found that a two-layer ANN and ANFIS both outperformed SVMs (and ANFIS outperformed the ANN), and ANNs better captured the spatial characteristics of soil moisture than RFs and boosted trees in southern Asia (Pradhan et al. 2021). Thus, NNs have generally performed well in predicting drought indices. Other types of NNs have also been used in drought studies. For example, recurrent NNs (RNNs) have occasionally been used to study drought severity (e.g., Dhyani and Pandya 2021), as well as convolutional NNs (CNNs; e.g., Dhyani and Pandya 2021; Liu et al. 2024). The combination of RNNs and CNNs and other deep learning methods have also been used in drought monitoring (Kaur and Sood 2020; Foroumandi et al. 2023). However, these methods are largely unexplored for FD, except for two known examples. As such, this study aims to add to this by examining and testing the feasibility and performance of using such techniques for FD identification and prediction.

Research into FD using ML has thus far been limited, with only two known publications on the topic (Zhang et al. 2022; Foroumandi et al. 2024) despite the potential for ML to overcome many of the challenges in FD prediction (Tyagi et al. 2022). Zhang et al. (2022) predicted rapid intensification from a set of drought events using RFs, RNNs, and multilinear regressions (MLRs), with RFs outperforming the RNNs and having the most consistent results. Foroumandi et al. (2024) predicted standardized soil moisture with a generative adversarial network (GAN), which was then correlated to FD for FD monitoring. The GAN was able to learn complex features in the soil moisture and captured sudden changes in the soil moisture. To expand further on this research, this project, the second in a set of projects investigating ML representation of FDs, seeks to conduct a comprehensive examination of how well different NN techniques, such as ANNs, RNNs, and convolutional U-Nets, can represent FD events directly. Edris et al. (2025, hereafter Part I), focused on the more standard ML

technique, such as RFs, SVMs, and AdaBoosted trees. The contributions of Part II is to investigate the capability of standard NN-based models to represent FD and determine how well, if at all, they can improve on the predictions made by the ML models in Part I and what is further needed to improve FD predictions with ML methods.

2. Datasets

Input and output datasets for the machine learning models were identical to those used in Part I, with the main addition being the application of NN techniques. The reader is referred to Part I for full details and justification of variables. In brief, this project employed key variables from the North American Regional Reanalysis (NARR; Mesinger et al. 2006). The input data include variables known to drive or signal FD events, including temperature T , precipitation P , ET, PET, and the mean 0–40-cm soil moisture (SM). These variables were averaged over pentads to eliminate day-to-day variability and retain key signals. Pentad differences in the critical variables that drive the feedback process for FD development (ET, PET, and SM; Christian et al. 2019b, 2023; Otkin et al. 2021; Osman et al. 2021; Tyagi et al. 2022) were also calculated and filtered through a 30-day (six pentads) leading moving average (referred to as ΔET , ΔPET , ΔSM) to represent the rapid development of drought conditions. Output data were a binary label, indicating whether FD occurred at a specific grid point and pentad (1) or not (0).¹ Because there is no consensus on FD identification, multiple output datasets were made using different FD identification methods, and the NN techniques were trained to each one individually. The FD identification methods used in this study include the methods outlined in Christian et al. (2023; hereafter C23), in Noguera et al. (2020; hereafter N20), in Pendergrass et al. (2020; hereafter P20), in Liu et al. (2020; hereafter L20), and in Otkin et al. (2021; hereafter O21).

Last, Fig. 1 shows the frequency of FD events or the target climatology for the NN models to predict, according to each of the five methods listed above. The general spatial trend in FD was similar for each method (with small variations in the spatial hotspots), but each method had different FD frequencies and each method highlighted some hotspots more strongly than others (some omit hotspot regions altogether), agreeing with the general findings in Osman et al. (2021).

3. ML algorithms and verification

Figure 2 displays the neural network architectures used in this study. They were built and trained using the TensorFlow package and the Keras application programming interface (API) (Abadi et al. 2015; Chollet et al. 2015). A description of neural networks, some of the standard structures, and terminologies can be found in Chase et al. (2023). A small sample of the dataset (10 folds across four rotations) was used

¹ Similar to Part I, there was a hidden third class (2) for sea grid points. Sample weights for this third class were set 0 to avoid impacting training.

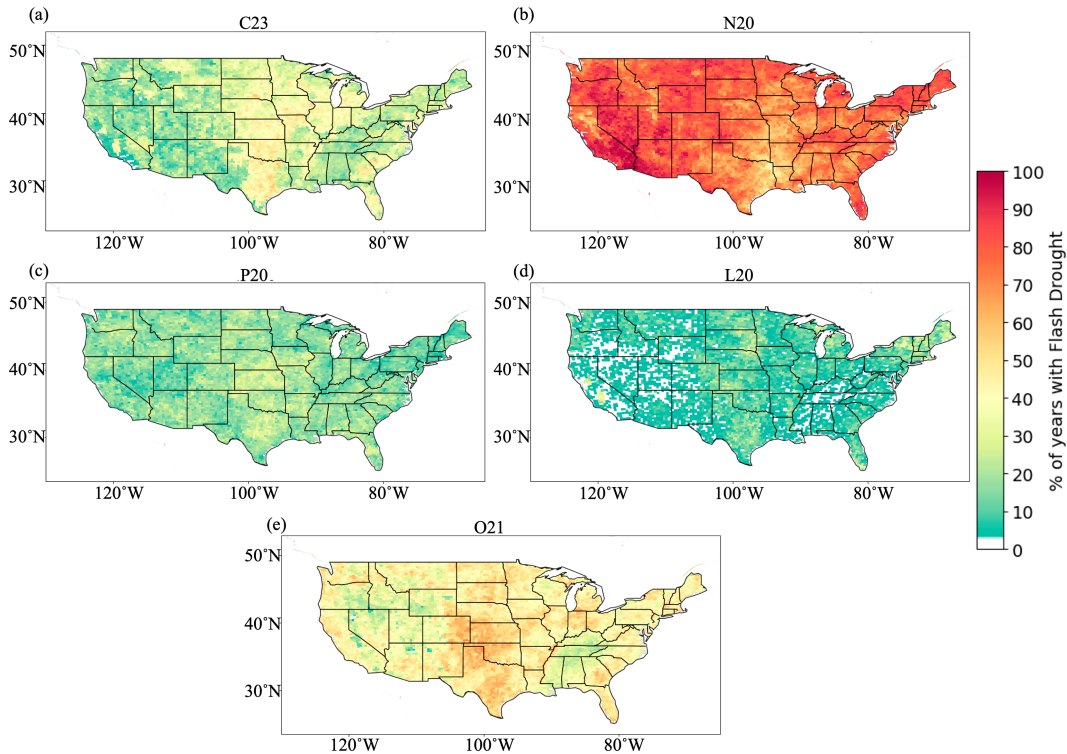


FIG. 1. Percentage of years FD experienced according to (a) the C23 method, (b) the N20 method, (c) the P20 method, (d) the L20 method, and (e) the O21 method (reproduced from Fig. 1 in Part I).

to test different network architectures and parameters, and the ones with the best validation performance and did the best at reconstructing the spatial climatology and temporal time series of FD were chosen for this project. In these tests, the main parameter that affected the performance of the model was not necessarily the architecture of the neural networks but the sample weights applied to account for the class imbalance. Further, since the main difference between each FD identification method is how frequently they identify each FD, the sample weights were modified according to the FD identification method (supplemental Table 1 in the online supplemental material), while the network architecture and other parameters were kept the same. Sample weights are weights that are applied to each grid and pentad when the loss function is calculated, with higher weights placing greater emphasis in the loss function. For this study, grids and pentads without FD were given sample weights of 1, and grids and pentads with FD were given larger sample weights (supplemental Table 1 for precise values) to help emphasize FD in the loss function and account for class imbalance. Regularization is also helpful to prevent overfitting (Marcolongo et al. 2022) and to help account for any spatial autocorrelation in the dataset (Madakumbura et al. 2021). For these reasons, L2 regularization was applied to each network, as well as dropout to help further reduce overfitting (supplemental Table 2). Other general parameters were equally applied for each network and FD identification method (see supplemental Table 2). Training results (such as training time and final loss values) can be found in supplemental Table 3.

For a description of the verification methods, the reader is referred to Part I. In brief, ML algorithms were verified using K -fold cross validation, where the data were split into 43 folds (1 per growing season): 41 folds were used for training, 1 fold was used for validation, and 1 fold was used for testing. ML algorithm skill was determined by the true skill statistic or Pierce's skill score (Woodcock 1976), whose values showed similar trends to other skill scores such as the critical score index (CSI)/threat score and the equitable threat score (ETS); however, difference in performance was found to be clearer with fewer decimal places with the Pierce's skill score. ML predictions were also statistically compared to the true predictions via composite mean differences, with statistical significance determined via the Monte Carlo bootstrapping method (with $N = 5000$, without replacement). Interpretability of the ANNs was determined via Shapley values from the Shapley additive explanations (SHAP) package (Lundberg and Lee 2017) with the same simplifying assumptions as in Part I. Shapley values (called SHAP values in the SHAP package) determine the importance of individual features by comparing model performance when "with" the feature and "without" it. Note that the RNNs and U-Nets could not be interpreted as easily by Shapley values as their outputs had too many dimensions (number of spatial pixels by number of pentads by number of classes for RNNs and number of pentads by x spatial dimension by y spatial dimension by number of classes for U-Nets), yielding ambiguous results when using Shapley values.

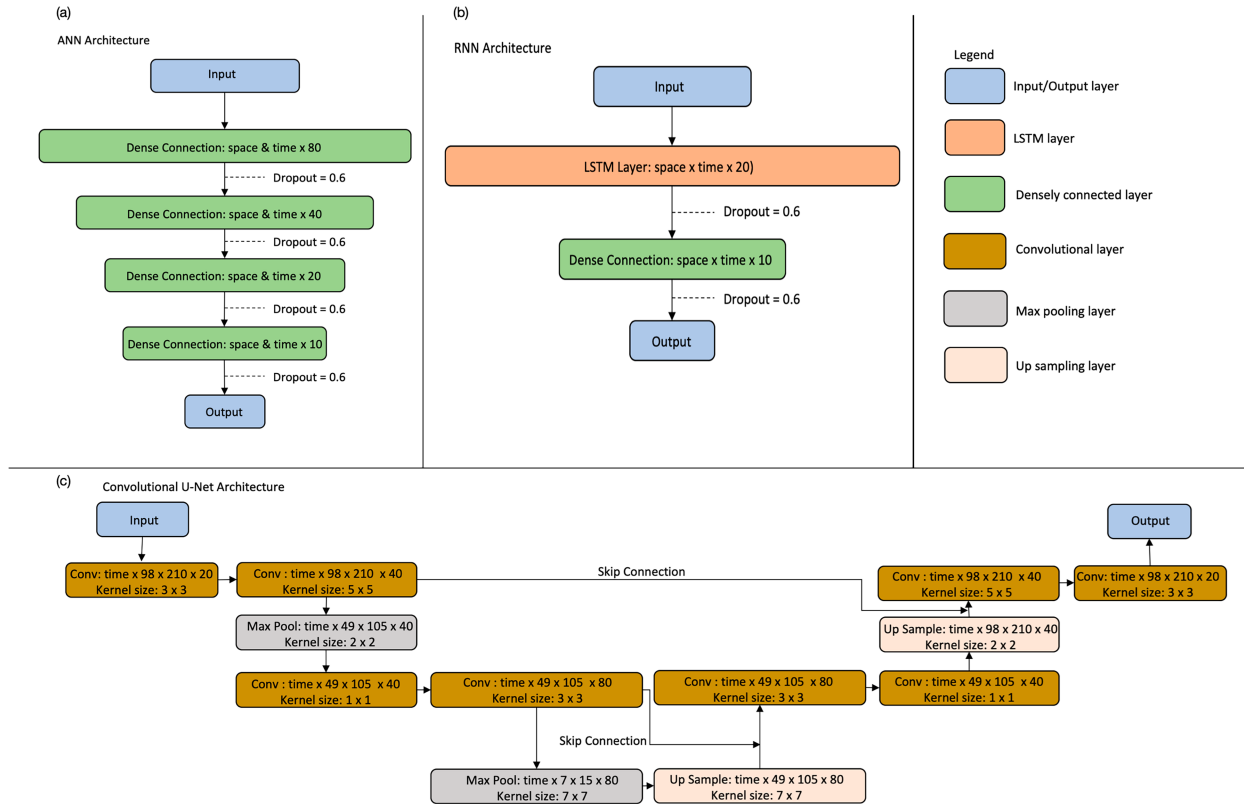


FIG. 2. NN architectures used to identify FD. (a) ANN architecture, (b) RNN architecture, and (c) convolutional U-network. Note that the U-network had a spatial dropout of 0.2 applied after every convolutional layer (not shown). Shapes for the input layers were (a) $N_{\text{space}} \times N_{\text{time}} \times N_{\text{features}}$ for ANNs, (b) $N_{\text{space}} \times N_{\text{time}} \times N_{\text{features}}$ for RNNs, and (c) $N_{\text{time}} \times N_{\text{space}} \times N_{\text{features}}$ for the U-Nets. Shapes for output layers were as in input layers, except N_{labels} was outputted instead of N_{features} . Spatial dimensions (N_{space}) for the NARR grid was 210×98 (or 20580 in the ANNs and RNNs). Time dimension (N_{time}) was 43 pentads per fold. All output layers used a softmax activation function.

a. Artificial/densely connected neural networks

ANNs, sometimes called densely connected NNs (DNNs) or fully connected networks to distinguish them from 1D CNNs, are the standard form of neural networks and the ones most commonly used in drought studies (Mohamadi et al. 2020; Dikshit et al. 2022; Hsieh 2022). These networks take the inputs and connect all of them to each unit in the next layer, the output of which is connected to all the units in the next layer, and so on. Interestingly, many drought studies do not use deep ANNs but rather tend to stick to ANNs with one or two hidden layers (Dikshit and Pradhan 2021; Dikshit et al. 2022; Hsieh 2022). The ANN structure used for this project is a standard structure for ANNs, with a large number of units in the first layer and steadily decreasing number of units for each layer. A four-layer model was chosen, starting with 80 units in the first layer and decreasing to 10 units in the final layer (Fig. 2). When testing on the subset of FD data, the structure of the network did not greatly impact the predictions as much as the number of parameters that are trained in the NN. If there were too many parameters, the network would get stuck in a trivial solution (predicting no FD for all grid points and pentads) and have trouble finding a solution

that predicts FD and learns FD patterns (i.e., it gets stuck in a local minimum in the cross-entropy loss function). Conversely, if there were too few parameters, the network would not be able to sufficiently learn anything. From experimentation, approximately ~ 2000 – 7000 parameters struck a balance where the networks could learn spatial and temporal features of FD without getting trapped in a trivial solution (however, approximately 11%–21% of rotations still got trapped in a trivial solution for the P20, L20, and O21 methods). However, the ANNs were the simplest NNs examined and were by far the

TABLE 1. 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 pentads). Highest skill score for each identification method is bolded.

	AdaBoosting	ANNs	RNNs	U-Nets
C23	0.33 (± 0.019)	0.03 (± 0.006)	0.25 (± 0.034)	0.12 (± 0.015)
N20	0.13 (± 0.007)	0.00 (± 0.002)	0.23 (± 0.046)	0.18 (± 0.025)
P20	0.06 (± 0.008)	0.00 (± 0.000)	0.08 (± 0.025)	0.00 (± 0.002)
L20	0.27 (± 0.016)	0.06 (± 0.023)	0.16 (± 0.027)	0.06 (± 0.023)
O21	0.28 (± 0.016)	0.18 (± 0.050)	0.10 (± 0.016)	0.15 (± 0.011)

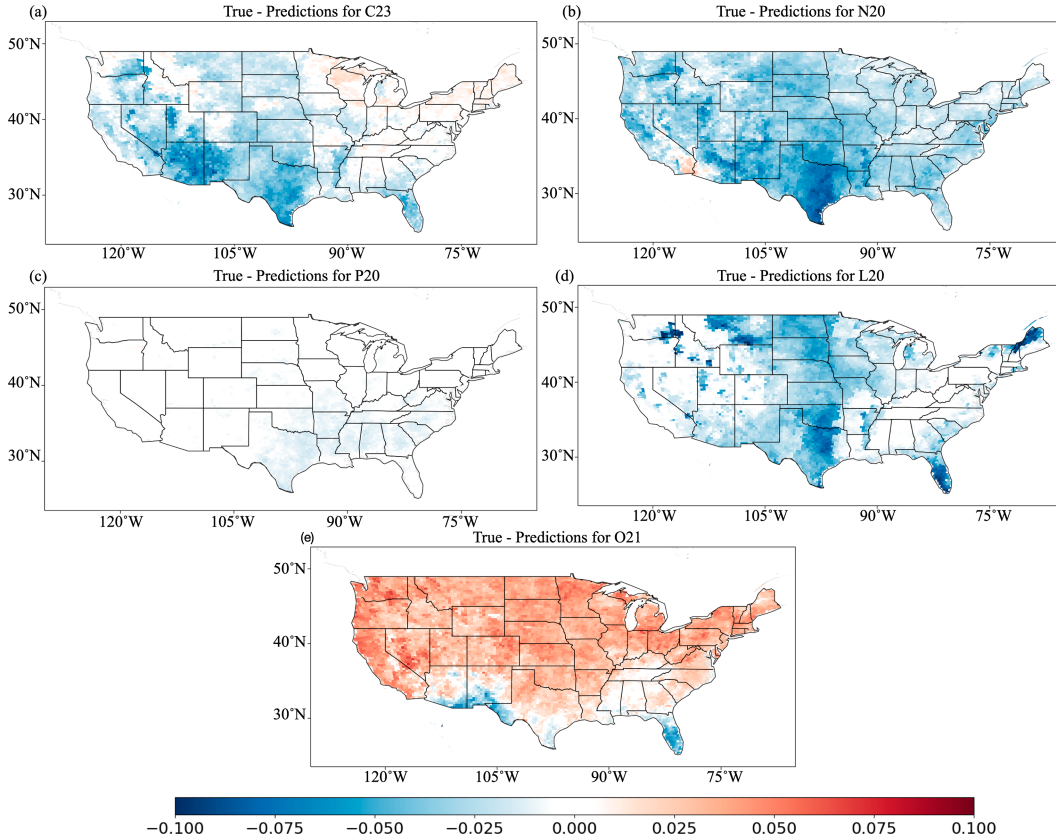


FIG. 3. Composite mean difference (unitless) of true minus predicted FD labels for each FD identification method using RNNs.

fastest to train (supplemental Table 3). The proposed architecture for this study contained 5023 parameters. For the ANNs in this project, each grid and pentad in the dataset was treated as an example, while the input features were densely connected in the hidden layer. This meant that the input layer for the ANNs had a shape of $N_{\text{space}}N_{\text{time}} \times N_{\text{features}}$ or $20580 N_{\text{time}} \times 8$ (where the time dimension was split into different folds and had 1763 pentads for the training set and 43 pentads for the validation and test sets).

b. Recurrent neural networks

RNNs are networks better attuned for learning temporal data based on its recurrent or recursive feature. RNNs take in a time series for each input feature (which are densely connected in an ANN fashion) and, for a given time frame, will use the input features to learn the classification. Additionally, the network will also recursively use information from the previous time frames to learn the temporal features of the data. However, the weighting of the previous time frames becomes smaller the further back the recurrent layer looks, resulting in a standard recursive layer having short-term memory. Additional recurrent modules have been developed so that RNNs can have better long-term memory. The most common modules are the long short-term memory (LSTM) module and the gated recurrent unit (GRU) module

(Hochreiter and Schmidhuber 1997; Sak et al. 2014; Zaremba et al. 2015; Cho et al. 2014). These modules use a series of gates to regulate what gets remembered and what is forgotten to help improve memory, with the parameters of the modules being trained as part of the training process. Drought studies, and atmospheric studies in general, favor LSTMs exclusively whenever RNNs are used, as LSTMs are a more complete form of the GRU modules (Dikshit et al. 2022; Hsieh 2022; Zhang et al. 2022). While both were tested, LSTM layers were shown to slightly improve ML predictions and were used in this study. Thus, this study used an RNN structure with one LSTM layer followed by a densely connected layer after the time series has been learned (Fig. 2). Additional LSTM layers (up to five with over 200 000 parameters) were tested but did not noticeably improve the performance. The final RNN model had fewer parameters than the ANN (2563 parameters total) and did not have as many issues in getting stuck in a trivial solution (though this still occurred with the N20 method for 11 of the rotations). For the RNNs, each grid point was treated as a sample, while the input features were connected in each unit, and the recursive axis was trained along the time axis. This meant that the RNNs had an input shape of $N_{\text{space}} \times N_{\text{time}} \times N_{\text{features}}$ or $20580 \times N_{\text{time}} \times 8$. Note that the time dimension/sequence length was left generic in the model structure since it was different between training and validation/test sets.

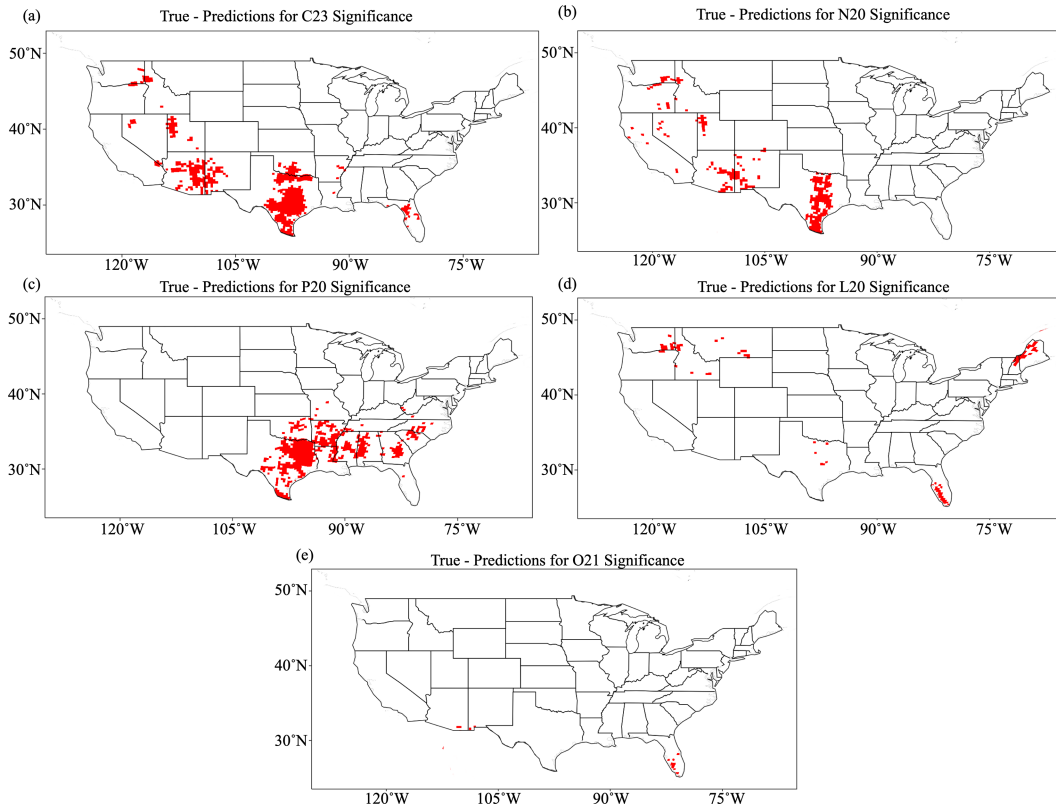


FIG. 4. The 95% statistical significance of composite mean differences of the RNN predictions.

c. Convolutional U-networks

Convolutional neural networks differ from ANNs in that not all of the nodes are connected to all of the subsequent nodes. Rather, only a set of nodes from the previous layer connects to a node in the next layer in a convolutional operation. In a 2D sense, this means that a rectangular set of grids are connected to a single grid in the next layer, while another grid in the next layer uses inputs from another rectangle of grids (the rectangular box is shifted by one or more grids to the right, depending on a parameter termed stride, unless the rectangular box is at the right end of the map, and then, it shifts to the left edge and one or more grids down), and so on until all the grids in the next layer have been filled in. The size of the box used for convolution depends on the kernel size of the convolution layer (e.g., a kernel size of 3×3 means a sliding box of 3×3 grids will be used to derive values in the next layer). In addition, convolutional networks incorporate pooling layers or layers that use an $n \times m$ kernel and either the average or maximum value of grids within that kernel (depending on the type of pooling used) for the pooling layer. For meteorological purposes, zero padding (adding zeros around the edges of the map) is typically used so that convolution layers do not change the grid size (they are set so that the sliding rectangular boxes on slide one grid point at a time, so that the convolution layer has the same grid as the previous layer), while pooling is used to reduce the grid size (a pattern this study follows; Liu et al. 2016; Boulaguiem et al. 2022; Marcolongo et al. 2022;

Foroumandi et al. 2024). An operation, termed upsampling, can be used to go in the opposite direction from pooling (increasing the grid size). Using a sequence of convolutional and pooling layers followed by a sequence of convolutional and upsampling layers forms a network that gives gridded outputs and can generate maps and images, termed a U-network (because the architecture resembles a U).

On the top of the convolutional networks, which are known to effectively capture spatial patterns, U-Nets have been shown to effectively recreate maps of meteorological variables, predict said variables, and learn complex spatial-temporal patterns (e.g., Liu et al. 2016; Boulaguiem et al. 2022; Marcolongo et al. 2022; Foroumandi et al. 2024). Therefore, this study employs a convolutional U-Net to identify FD, with the architecture given in Fig. 2c. This model was deeper than the other networks explored here and contained a larger number of parameters to train (257 875 parameters), but the U-Nets were still able to adeptly identify FD. However, despite the greater number of parameters, the U-Nets took a similar amount of time (for half the number of epochs; supplemental Table 3) and computation resources to train as the RNNs due to the simpler computations for the convolutional layers and the fewer samples for U-Nets to train on. For this experiment, the convolution was performed on maps of FD variables (i.e., along the spatial axes), while the input variables were densely connected to the “filters” (a third dimension in the U-Nets that densely connects the input variables, to identify patterns and features

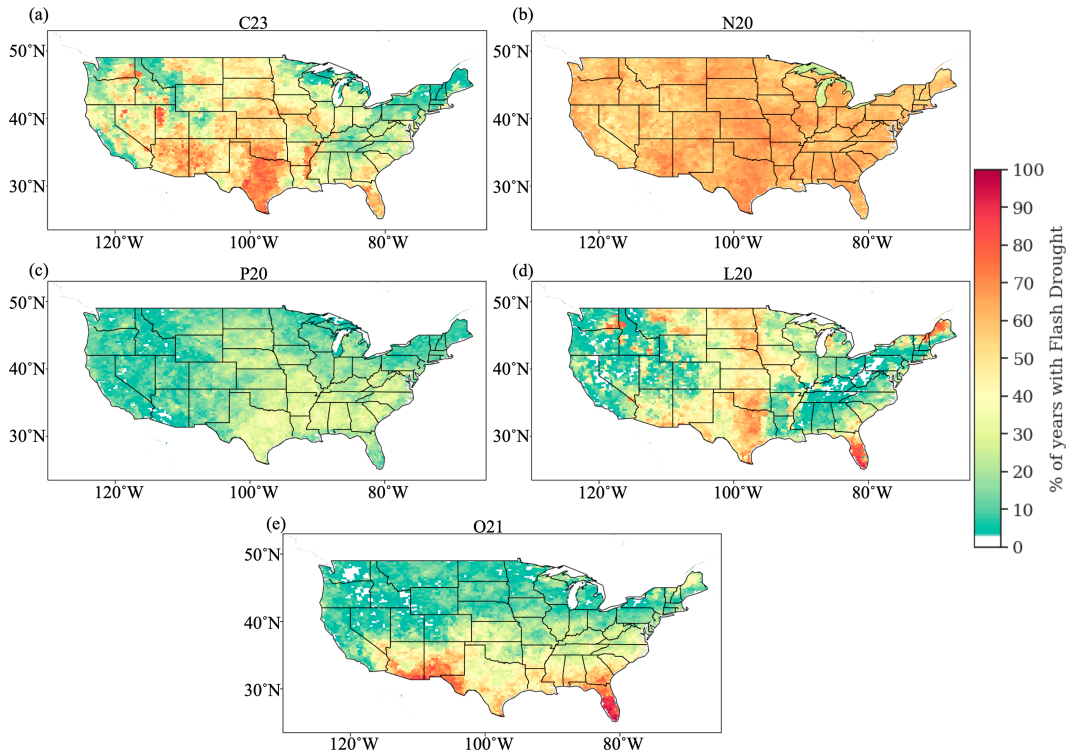


FIG. 5. Predicted frequency climatology of FDs by the RNNs for the (a) C23, (b) N20, (c) P20, (d) L20, and (e) O21 methods.

relevant to the input variables). The U-Nets used the time axis as the number of training examples (giving the U-Nets an input shape of $N_{\text{time}} \times N_{\text{space}} \times N_{\text{features}}$). Note that this means the U-Nets had fewer samples to train on than the other networks (e.g., a training set had 41 folds, with 43 pentads per fold for 1763 samples compared to 20 580 samples for when each grid point was treated as an example).

4. Results and discussion

a. Statistical performance

The true skill statistic for all NN structures are given in Table 1, along with the skill score for the AdaBoosted trees (best-performing ML algorithm in Part I) for reference. In this case, the RNN architecture managed to improve the skill of the ML algorithms (up to a 0.11 increase in the best case) for two FD identification methods (N20 and P20). As the RNNs most frequently showed the highest skill scores of all the NNs, the remaining figures showing ML model performance are for RNN predictions. Similar figures using U-Net predictions can be found in supplemental Figs. 5–12, and ANN results can be found in supplemental Figs. 13–21. In comparison, the standard ANN structure struggled to make FD predictions, and the U-Nets overemphasized spatial patterns (i.e., U-Nets overpredicted FD hotspots, occasionally at the expense of identifying FD in other locations) more than the RNNs. ANN results could, in part, be not only due to the tendency of the ANNs to get trapped in local minima that

are trivial/null solutions (solution that do not predict any FD) more frequently than the other ML algorithms but also because it overemphasized SM above all other variables, similar to SVMs (supplemental Fig. 15; Part I) and generally did not learn the surface interactions involved in FD.

In examining the difference between the true and predicted FD labels for the RNNs (Fig. 3), a widespread improvement can be seen in the climatology predictions compared to the AdaBoosted trees for four of the FD identification methods. In each case, the amount of overprediction of FD in hotspot regions has been notably reduced, with fewer cases of underprediction except for the Great Lakes region and Northeast in the C23 method and southeastern California in the N20 method. The consistent underpredictions of FD in the O21 method suggest that the RNNs focused heavily on the events in the southern United States and, therefore, struggled to identify FD in other regions. Moreover, the composite difference for the P20 method looks quite good, with only small differences compared to the truth labels. However, all of those differences are statistically significant (Fig. 4), and the predictions using the P20 method had a low skill score, suggesting that the RNNs predicted similar spatial climatology but might have struggled with the timing of FD events.

b. Climatological performance

Figure 5 shows the predicted FD frequency climatology by the RNNs. The RNNs showed notable improvement to the predicted climatology compared to the AdaBoosted trees in

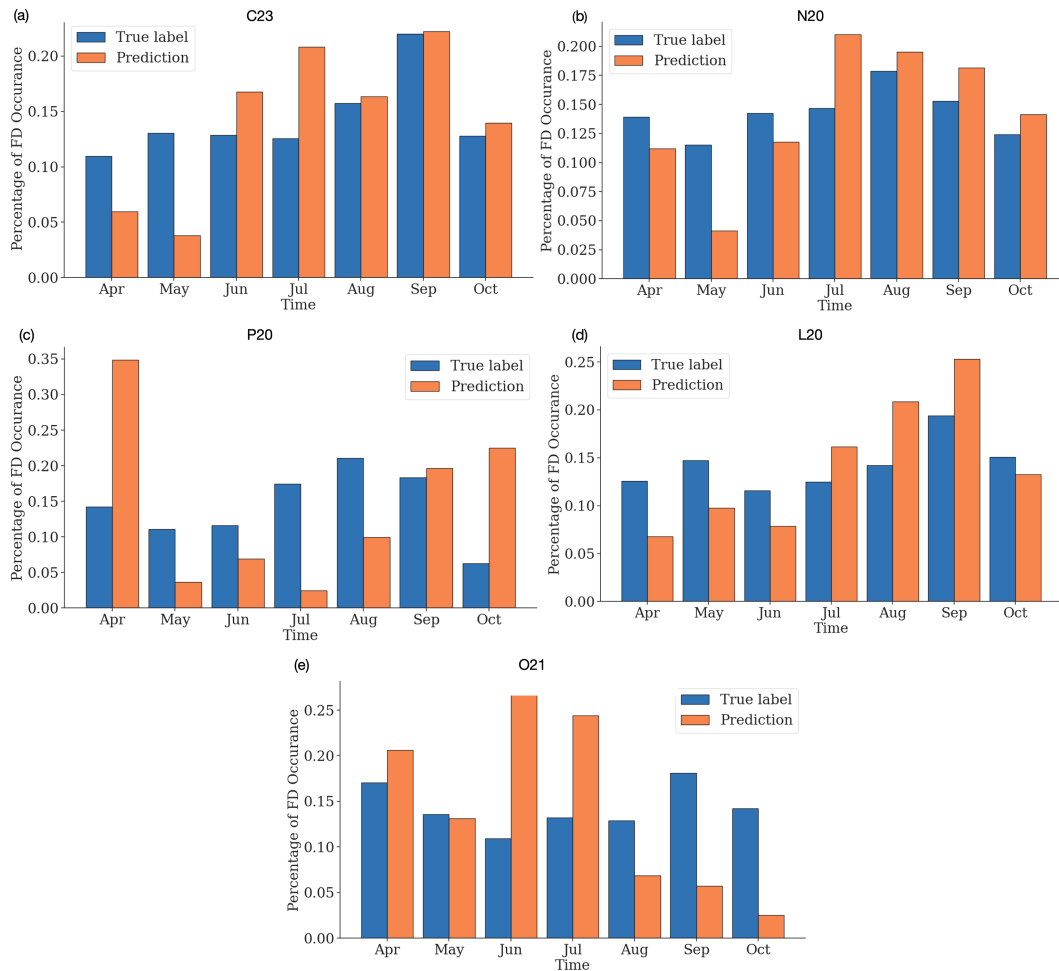


FIG. 6. FDs seasonality (percentage of FD occurrence) for true labels (blue) and predicted labels by the RNNs (orange) for the (a) C23, (b) N20, (c) P20, (d) L20, and (e) O21 methods.

Part I, with the overemphasis on hotspot regions reduced in the RNNs compared to the AdaBoosted trees. This is an interesting note as the RNNs and AdaBoosted trees had comparable skill scores. In general, throughout these results, the RNNs will improve on the climatological predictions of FD but still struggle with the timing of FD events. Since the true skill score compares the predictions pentad to pentad and grid point to grid point, the RNNs and AdaBoosted trees give similar skill scores. The agricultural regions that the AdaBoosted trees identified as hotspots were still present in some RNN predictions (e.g., for the C23 method) and absent in others (P20). In addition, the overemphasis in Florida was also still present for some FD identification methods (i.e., C23; L20; O21). Further, while the RNNs created hotspots in some agricultural regions (e.g., eastern Washington and southern Idaho), the signal there is much more muted compared to what the traditional ML algorithms (RFs, SVMs, and AdaBoosted trees) learned. In exchange, however, the RNNs learned to identify an additional hotspot in Arizona and western New Mexico. This is particularly interesting as while this is not a hotspot for FD in the true labels (Fig. 1), it is a hotspot

for rapid intensification of dry conditions without going into drought (Edris et al. 2023). This tendency toward identifying the American Southwest along with the other successes the RNNs had for the C23 method suggests the models were leaning toward predicting areas of rapid intensification, as much or more so than FD. In contrast, the RNN predictions for the O21 method focused heavily on this hotspot region but struggled to learn FD elsewhere. The emphasis on the southern United States suggests that the RNNs, for the O21 method, may have been over reliant on T and/or PET for FD prediction for that FD identification method, given its similarity to the T and PET climatologies. As the RNNs seemed to overpredict FD in similar locations as the AdaBoosted trees for all methods except O21, they may have relied on similar variables as the AdaBoosted trees to make predictions (SM, P , and PET), but the performance in overpredictions and emphasis on other locations also suggest that the RNNs were more effectively incorporating information from other variables and were not as strongly reliant on a select few as the AdaBoosted trees were.

Also in contrast to the AdaBoosted trees, the RNNs may also have learned the seasonality of FD more effectively. The

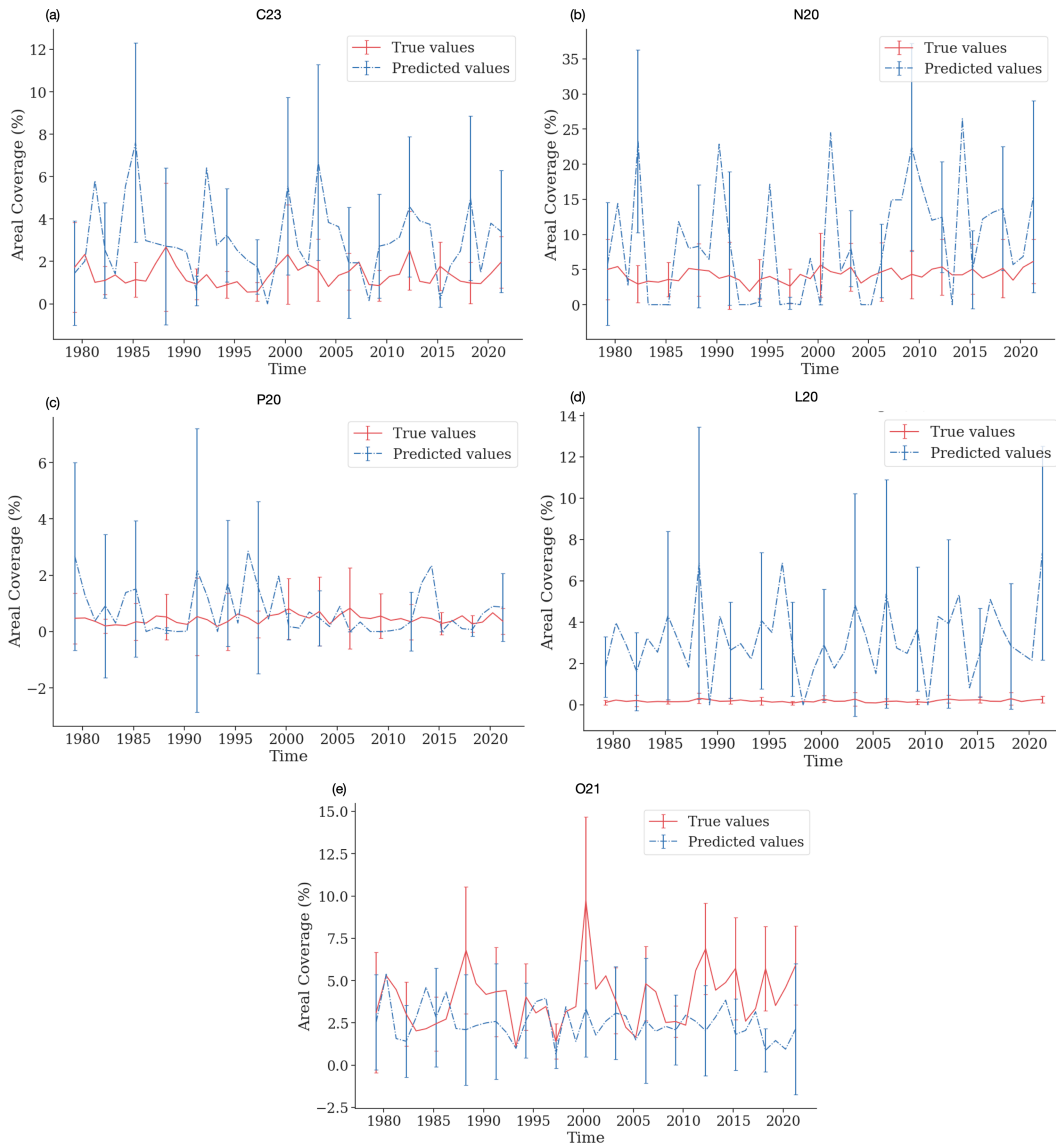


FIG. 7. True (red) and predicted (blue) annual average spatial coverage of FD across the domain for the (a) C23, (b) N20, (c) L20, and (e) O21 methods. Error bars denote one standard deviation in the annual average. Predicted labels were made by the RNNs.

RNNs did not focus on a select few months at the end of the growing season but predicted varying seasonalities depending on the FD identification method (Fig. 6). The RNN seasonalities also matched the true labels more closely in some cases (e.g., C23; N20; L20, though it had trouble finding early season FDs in C23), but that attempt at learning the seasonality also resulted in poorer results in the P20 and O21 methods (Figs. 6c,e). In addition, the resulting time series of FD coverage had more mixed results compared to the AdaBoosted trees. The RNNs in general did not learn as well which years had larger FD events as well (Fig. 7), resulting in some years with much more widespread FD predicted when the true labels show otherwise. Note that years when no FD was predicted for the N20 method represent rotations when the

RNNs predicted trivial solutions (Fig. 7b), further suggesting the timing of particular FD events was a limitation for the RNNs.

c. Case studies

A case study for the FD of 2012 is given here (Fig. 8) as a direct comparison to the predictions from the AdaBoosted trees in Part I. In that case, the boosted trees performed relatively well and captured much of the spatial extent of the FD but struggled with the timing, identifying FD prematurely or months too late. In comparison, the RNNs were also able to capture and even improve on the spatial coverage of FD (e.g., creating a more realistic interpolation of FD in the P20 method); however, it also struggled with the exact timing of the FD onset. Unlike the AdaBoosted trees, which identified FD too early or too late,

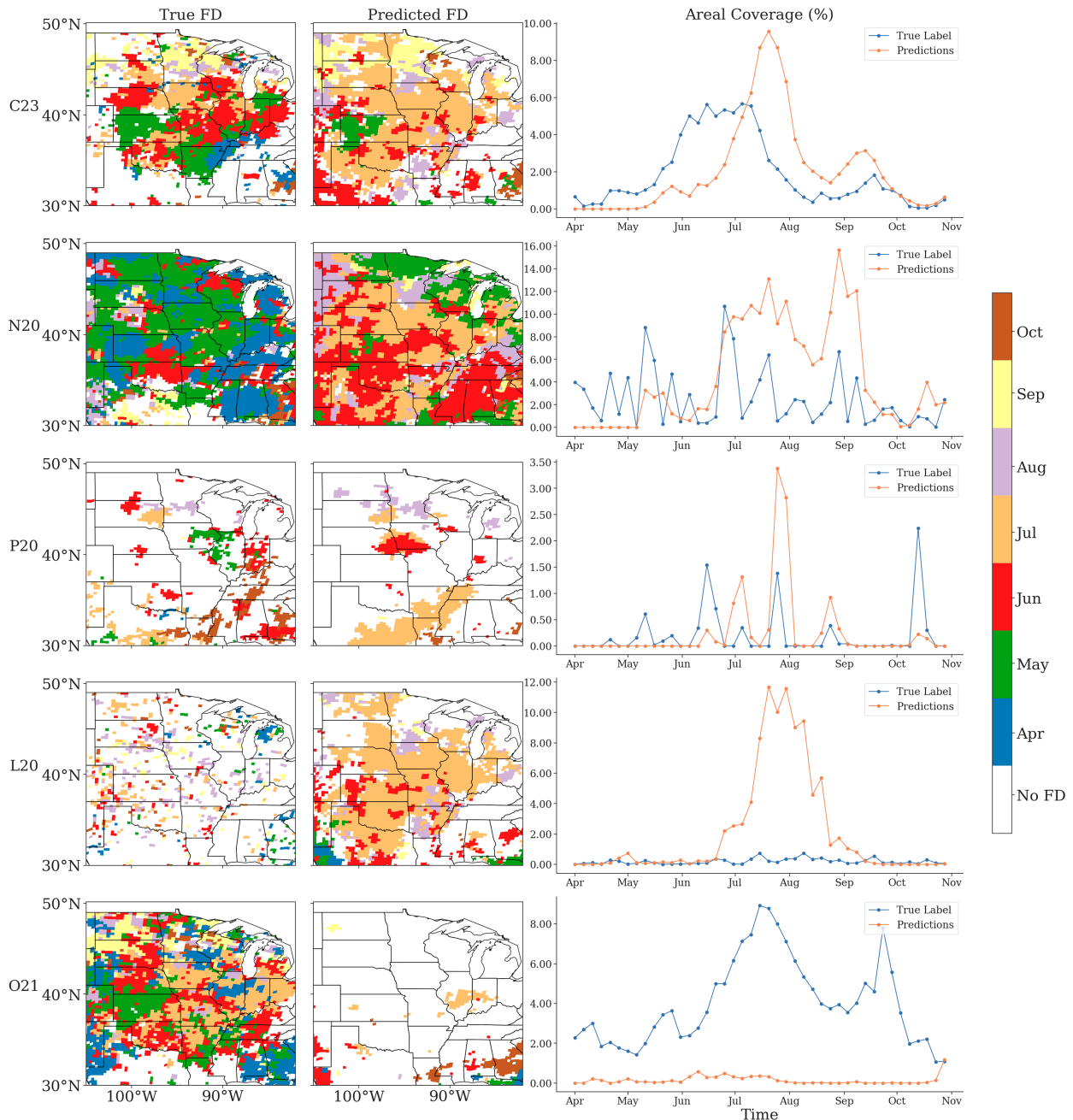


FIG. 8. FD case study for 2012 for each FD identification method. (left) True labels, (center) average predicted month across all rotations for the entire dataset using RNN models, and (right) true and predicted FD coverage over the domain. FD predictions were made by RNN models.

depending on the FD identification method, the RNNs consistently identified FD a month or two late in this case study, shifting the bulk of the identification down by four to eight pentads. In addition, the U-Nets were also able to capture the spatial extent of FD for the **C23** and **O21** methods, though not necessarily the timing (supplemental Fig. 10).

Another case study that is examined here is in 2003, in which FD developed early in the growing season and then

propagated into the central portion of the Great Plains in the summer as SM declined and vegetation desiccation resulted in decreased ET (supplemental Fig. 3) through the rest of the growing season, creating substantial crop damage (Fig. 9; Otkin et al. 2014). This was an interesting case study because the RNNs learned the coverage and the timing of the FD more effectively, having the bulk of the FD identified at the same time as the true labels for the **C23** and **N20** methods.

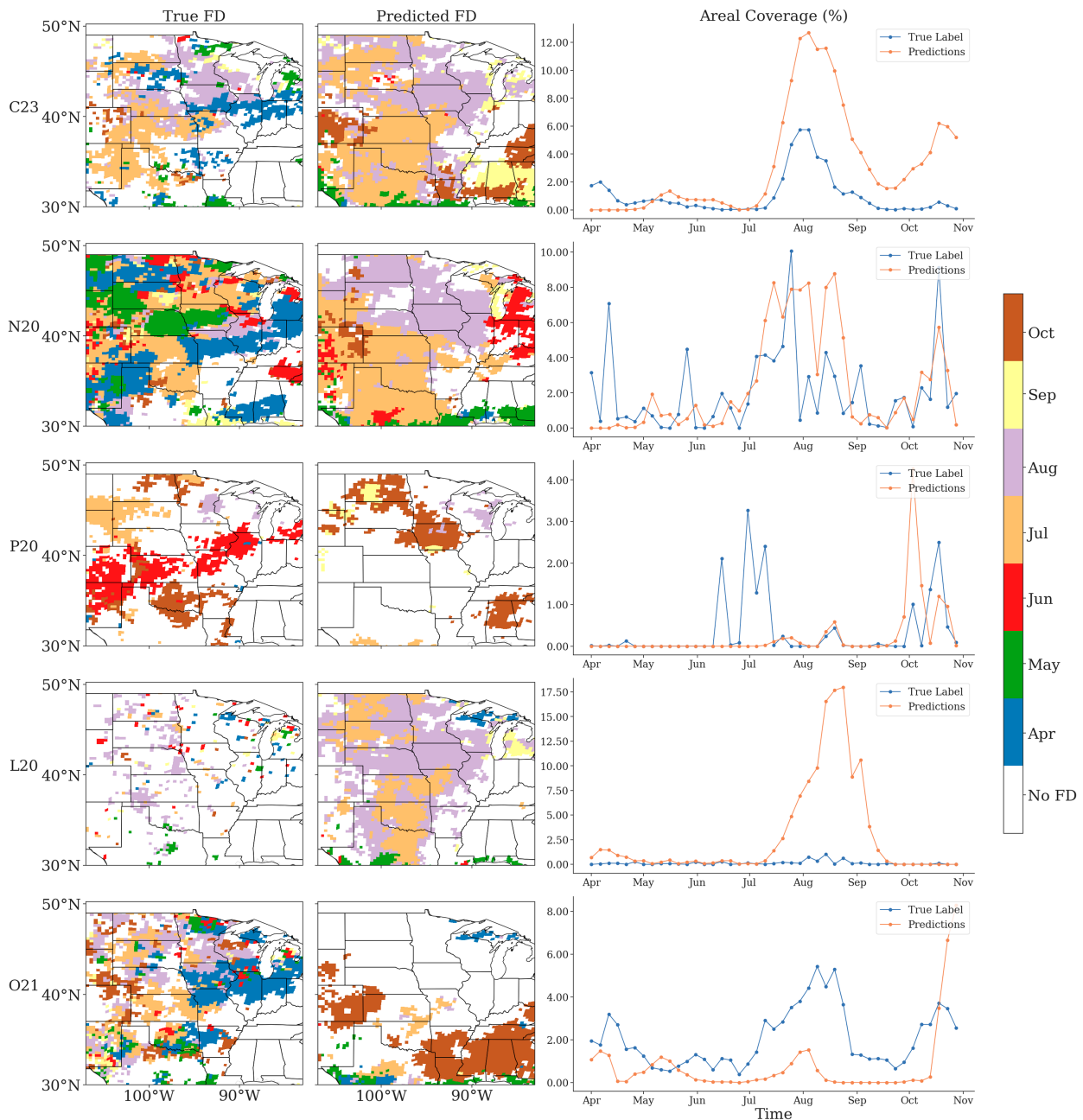


FIG. 9. FD case study for 2003 for each FD identification method. (left) True labels, (center) average predicted month across all rotations for the entire dataset using RNN models, and (right) true and predicted FD coverage over the domain. FD predictions were made by RNN models.

Given that this particular FD was driven more by SM decline (and resulting decline in ET; supplemental Fig. 3) rather than a combination of increasing PET, decreasing SM, and decreasing ET (note most of the FD occurred in June–August, when SM was declining in supplemental Fig. 3, and after PET had started to peak), this case study ultimately proved to be a simpler event that the RNN models were able to capture as compared to 2012 or 2017. However, in the C23 and N20 methods, the RNNs failed to find the regions where the FD initiated

early and thus did not necessarily capture the propagation of FD (Fig. 9, rows 1 and 2). Combined with the good representation of FD coverage for the same three identification methods, the RNNs well outperformed the AdaBoosted trees in this case study (not shown).

The final case study examined was in 2017, in which large parts of eastern Montana and western North and South Dakota fell in severe drought for much of the growing season, resulting in notable agricultural impacts (up to 2.6 billion U.S. dollars in

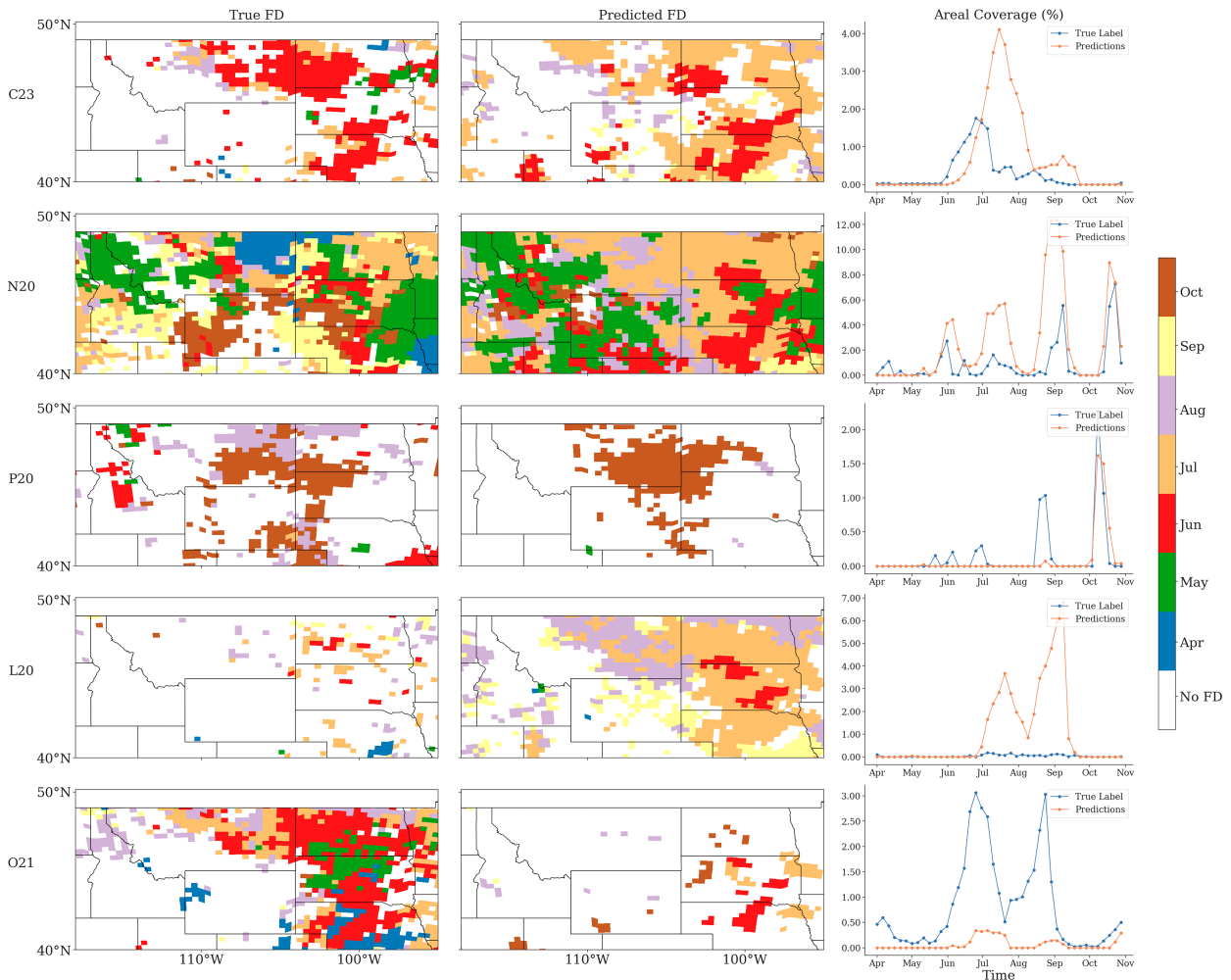


FIG. 10. FD case study for 2017 for each FD identification method. (left) True labels, (center) average predicted month across all rotations for the entire dataset using RNN models, and (right) true and predicted FD coverage over the domain. FD predictions were made by RNN models.

agricultural losses; He et al. 2019; Lowman et al. 2023). Rising temperatures (and evaporative demand) without corresponding precipitation increase in May resulted in early FD identification in the N20 method, while the C23, L20, and O21 methods identified FD in early summer (June) during the rapid decline of SM and ET, and the P20 method identified the FD at the end of the growing season, as evaporative demand rapidly increased (Fig. 10 and supplemental Fig. 4). The RNNs represented FD fairly well for this event, with the spatial extent of the FD fairly well captured by the RNNs. While the RNNs had difficulty getting the timing of FD for the C23 and N20 methods (where FD was predicted about a month late), they were able to capture the timing of the FD better with the N20, P20, and L20 methods.

5. Discussion and conclusions

This paper continued on the work from Part I by predicting FD with NN algorithms. The standard ANN structure had a surprisingly difficult time identifying FD. U-Nets improved on the ANNs but were broadly not able to outperform the

AdaBoosted trees. However, the RNNs showed some improvement at identifying FD over the AdaBoosted trees in Part I, except for the O21 method. Specifically, the RNNs reduced the rate of overprediction of FD in hotspot regions (fewer false positives) by up to 70.83% for the C23 method, 86.15% for the P20 method, and 40.76% for the L20 method (for the N20 method, AdaBoosted trees underpredicted FD, so there was no improvement in overpredictions for that method). RNNs also gave more realistic climatological predictions outside of the hotspot regions. The RNNs also learned the seasonality of FD more effectively for some FD methods than others (i.e., the C23, N20, and L20 methods), and they compared well to the AdaBoosted trees in specific case studies (i.e., 2012) while outperforming them in others (i.e., 2003). However, the RNNs struggled to learn the precise timing of FD events and often identified the initiation of an FD event 1–2 months later than in verification, resulting in low skill scores. Interestingly, ML algorithms predicted the C23 and O21 methods better than others, suggesting they may have an easier time recognizing patterns similar to or correlated with SM patterns and climatologies.

Taken together with the results from Part I, our contributions were an initial look at the capability of ML algorithms to directly represent FD events (i.e., to directly predict if there is or is not FD), contributing a proof of concept of the ability of these standard ML approaches to directly represent FD events. The ML algorithms showed some promise in representing FD events as a whole, with AdaBoosted trees and RNNs (using LSTM layers) showing the most skill. One caveat is that the standard ML algorithms often over-relied on some variables (such as SM, *P*, and PET) to predict FD. RNNs seemed to improve on this but still struggled with the timing of FD events. In addition, the RNNs also had a difficult time learning the patterns and timing of FD for some identification methods (i.e., P20; O21). This is possibly because the RNNs failed to properly synthesize surface variables and learn surface interactions (e.g., the predicted frequency of FDs for the O21 method strongly resembled PET and temperature climatologies, suggesting the RNNs mostly relied on those variables). As a note, this study sought to identify FD directly as a classification problem, which may have different results compared to a regression problem where an FD indicator is predicted (e.g., in Foroumandi et al. 2024). Regression problems that aim to predict one of the driving variables for FD [such as the GAN model in Foroumandi et al. (2024), which was able to represent SM well] may have differing results and offer different perspectives than a classification approach that synthesizes several difficult to represent surface interactions. As such, future steps in this research should examine regression models, as well as investigate more advanced ML models such as transformers and graph neural networks.

Another contribution of these studies was a unique opportunity to examine the climatologies and seasonalities of multiple FD identification methods and how they examine different case study events. And in general, they showed similar, though varying, patterns. For example, C23 focused more heavily on agricultural regions and energy transition regions, where the evaporative stress from vegetation compared to native vegetation is higher. In contrast, the N20 method focused more on the western United States and earlier summer due to its emphasis on meteorological water deficit (i.e., use of SPEI). The P20 also focused more on mid- to early summer FDs due to its emphasis on atmospheric demand [using evaporative demand drought index (EDDI; Hobbins et al. 2016) to identify drought]. Similarly, the L20 and O21 methods had climatological patterns similar to SM, identifying hotspots more closely to the C23 method and identifying FD more in the late growing season (August–October) when SM was lower (Illston et al. 2004) and in early spring. In general, each FD identification method gives its own perspective and flavor, on FD, and examining and comparing their climatologies and how they handle different FD events can provide useful insights and are recommended for future studies. Given the plethora of FD identification methods, each giving its own perspective, a “true” identification of FD or prediction of FD would likely come from an ensemble approach, as recommended in Alencar and Paton (2022), or a convergence of evidence approach combined with expert opinions, similar to what the U.S. Drought Monitor uses (Svoboda et al. 2002).

Future studies should also examine some of the select ML algorithms (viz., the AdaBoosted trees and RNN structures) with multiple explainable artificial intelligence (AI) (XAI) methods, to better understand how the ML algorithms use the input variables and make their predictions, and what is needed to improve them (e.g., if more variables are needed or if more samples are needed). Future works, for example, are recommended to add other variables (such as ENSO, AMO, or other climatological drivers) or breaking down some variables (such as PET) into parts to examine if that forces the ML algorithms to focus on other variables and learn the surface interactions that drive FD events more effectively. Examination into these patterns can help make ML algorithms that are interpretable and improve them, potentially to the point of being able to predict FD on the seasonal to subseasonal time scale (a difficult task for which there are few tools).

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

- Abadi, M., and Coauthors, 2015: TensorFlow: Large-scale machine learning on heterogeneous systems. <https://www.tensorflow.org/>.
- 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>.
- Boulaguiem, Y., J. Zscheischler, E. Vignotto, K. van der Wiel, and S. Engelke, 2022: Modeling and simulating spatial extremes by combining extreme value theory with generative adversarial networks. *Environ. Data Sci.*, **1**, e5, <https://doi.org/10.1017/eds.2022.4>.
- Chase, R. J., D. R. Harrison, G. M. Lackmann, and A. McGovern, 2023: A machine learning tutorial for operational meteorology. Part II: Neural networks and deep learning. *Wea. Forecasting*, **38**, 1271–1293, <https://doi.org/10.1175/WAF-D-22-0187.1>.
- Cho, K., B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, 2014: Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv, 1406.1078v3, <https://doi.org/10.48550/ARXIV.1406.1078>.

- Chollet, F., and Coauthors, 2015: Keras. <https://keras.io>.
- 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. Mesheske, and Y. Zhou, 2022: Flash drought identification from satellite-based land surface water index. *Remote Sens. Appl. Soc. Environ.*, **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>.
- 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>.
- , —, and M. Santosh, 2022: Artificial neural networks in drought prediction in the 21st century—A scientometric analysis. *Appl. Soft Comput.*, **114**, 108080, <https://doi.org/10.1016/j.asoc.2021.108080>.
- Edris, S., A. McGovern, J. Basara, J. Christian, J. Furtado, H. Olayiwola, and X. Xiao, 2025: Evaluation of flash drought identification with machine learning techniques. Part I: Traditional machine learning algorithms. *Artif. Intell. Earth Syst.*, **4**, e240068, <https://doi.org/10.1175/AIES-D-24-0068.1>.
- Edris, S. G., J. B. Basara, J. I. Christian, E. D. Hunt, J. A. Otkin, S. T. Salesky, and B. G. Illston, 2023: Analysis of the critical components of flash drought using the standardized evaporative stress ratio. *Agric. For. Meteorol.*, **330**, 109288, <https://doi.org/10.1016/j.agrformet.2022.109288>.
- 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>.
- 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>.
- He, M., J. S. Kimball, Y. Yi, S. Running, K. Guan, K. Jensco, B. Maxwell, and M. Maneta, 2019: Impacts of the 2017 flash drought in the US Northern Plains informed by satellite-based evapotranspiration and solar-induced fluorescence. *Environ. Res. Lett.*, **14**, 074019, <https://doi.org/10.1088/1748-9326/ab22c3>.
- 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>.
- Hochreiter, S., and J. Schmidhuber, 1997: Long short-term memory. *Neural Comput.*, **9**, 1735–1780, <https://doi.org/10.1162/neco.1997.9.8.1735>.
- 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>.
- Krishnamurthy R. P. K., J. B. Fisher, R. J. Choularton, and P. M. Kareiva, 2022: Anticipating drought-related food security changes. *Nat. Sustainable*, **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>.
- Lisonbee, J., M. Woloszyn, and M. Skumanich, 2021: 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., and Coauthors, 2016: Application of deep convolutional neural networks for detecting extreme weather in climate datasets. arXiv, 1605.01156v1, <https://doi.org/10.48550/ARXIV.1605.01156>.
- , 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>.
- 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. *NIPS'17: Proc. 31st Int. Conf. on Neural Information Processing Systems*, Long Beach, CA, Curran Associates, Inc., 4768–4777, <https://dl.acm.org/doi/10.5555/3295222.3295230>.
- Madakumbura, G. D., C. W. Thackeray, J. Norris, N. Goldenson, and A. Hall, 2021: Anthropogenic influence on extreme precipitation over global land areas seen in multiple observational datasets. *Nat. Commun.*, **12**, 3944, <https://doi.org/10.1038/s41467-021-24262-x>.

- Marcolongo, A., M. Vladymyrov, S. Lienert, N. Peleg, S. Haug, and J. Zscheischler, 2022: Predicting years with extremely low gross primary production from daily weather data using Convolutional Neural Networks. *Environ. Data Sci.*, **1**, e2, <https://doi.org/10.1017/eds.2022.1>.
- 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 nomadic 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, and M. Svoboda, 2014: Examining the relationship between drought development and rapid changes in the evaporative stress index. *J. Hydrometeorol.*, **15**, 938–956, <https://doi.org/10.1175/JHM-D-13-0110.1>.
- , 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>.
- 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>.
- 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>.
- Sak, H., A. Senior, and F. Beaufays, 2014: Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition. arXiv, 1402.1128v1, <https://doi.org/10.48550/ARXIV.1402.1128>.
- 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>.
- 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).
- Zaremba, W., I. Sutskever, and O. Vinyals, 2015: Recurrent neural network regularization. arXiv, 1409.2329v5, <https://doi.org/10.48550/ARXIV.1409.2329>.
- 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>.